Available online at www.sciencedirect.com

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 12 (2004) 6351–6369

Bioorganic & Medicinal Chemistry

Total and local (atom and atom type) molecular quadratic indices: significance interpretation, comparison to other molecular descriptors, and QSPR/QSAR applications

Yovani Marrero Ponce*

Department of Pharmacy, Faculty of Chemical Pharmacy and Department of Drug Design, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba

Received 19 August 2004; revised 20 September 2004; accepted 21 September 2004
Available online 12 October 2004

Abstract—This paper describes the significance interpretation, comparison to other molecular descriptors, and QSPR/QSAR applications of a new set of molecular descriptors: atom, atom type, and total molecular quadratic indices. The features of the *k*th total and local quadratic indices are illustrated by examples of various types of molecular structures, including chain lengthening, branching, heteroatoms content, and multiple bonds. The linear independence of the local (atom type) quadratic indices to others 0D, 1D, 2D, and 3D molecular descriptors is demonstrated by using principal component analysis for 42 heterogeneous molecules. It is concluded that the local quadratic indices are independent indices containing important structural information to be used in QSPR/QSAR and drug design studies. In this sense, molecular quadratic indices were used to the description and prediction of the boiling point of 28 alkyl alcohols and to the modeling of the partition coefficient (log *P*), specific rate constant (log *k*), and antibacterial activity of 2-furylethylene derivatives. These models were statistically significant and showed very good stability to data variation in leave-one-out (LOO) cross-validation experiment. The comparison with the other approaches also revealed good behaviors of our method in this QSAR study.

1. Introduction

The basis for various quantitative structure–property/activity relationship (QSPR/QSAR) methods is the 'description' of the molecular structures by means of numbers. At present, there are a great number of molecular descriptors that can be used in QSAR/QSPR studies.[1–5] For instance, computer programs such as Dragon[6] compute up to 1800 descriptors, which may have very different complexity but can be classified according to their 'dimensionality' in: zero-dimensional (0D), 1D, 2D, and 3D molecular descriptors.

The so-called topological indices (TIs) are among the most useful molecular descriptors known nowadays.[7–11] These descriptors have become a powerful tool for predicting physical–chemical properties and biological activities of organic compounds as well as for 'rational' drug design.[7–11] TIs are molecular descriptors derived from graph-theoretical invariants and accounts for structural information contained in two-dimensional representation of molecules.[1] These can be classified as 'global' and 'local' according to the way in which they characterize the molecular structure.[7,12] However, most of TIs known today can be considered as global molecular descriptors and most also largely limit their field of application for lack of information on multiple bonds and/or heteroatoms in molecular graphs. One exception in this sense is the index of electrotopological state (E-state).[11,13,14] This approach has been used successfully in a variety of QSPR/QSAR studies of complex molecules.[11,13–15] However, the development of the atomic-chemical level topological indices is not well advanced. This should be the primary driving force to find new atomic level topological indices to describe different physical properties and activities.

In this context, the present author has recently introduced the novel computer-aided molecular design scheme TOMOCOMD (acronym of TOpological MOlecular COMputer Design). It calculates several new

Keywords: TOMOCOMD-CARDD software; Total and local (atom and atom type) quadratic indices; Factor analysis; QSAR/QSAR studies.

* Tel.: +53 42 281192/281473; fax: +53 42 281130/281455; e-mail addresses: yovanimp@qf.uclv.edu.cu; ymarrero77@yahoo.es

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.09.034

2D/3D families of total and local (atom and atom type) topologic and stochastic molecular descriptors, such as quadratic and linear indices; defined by analogy with the quadratic and linear mathematical maps.[16,17] This point of view was very recently successfully applied to the prediction of physical properties and Caco-2 permeability of organic compounds and drugs, respectively.[16,18,19] The method is very flexible and makes possible the study of small molecules as well as macromolecules such as nucleic acids.[20] Interestingly, molecular quadratic indices can be generalized to allow the codification of 3D-structural features.[21] These earlier communications stressed that the *k*th total and local (atom and atom type) quadratic index has useful applications in chemistry, which is yet to be investigated thoroughly.

In this sense, the main aims of this paper are the following ones: (1) to indicate the most important characteristic for these novel indices by means of several structure changes in organic molecules, including chain lengthening, branching, heteroatoms content, and multiple bonds, (2) to check if, the information contained in the total and local (atom and atom type) quadratic indices are different from that of others 0D, 1D, 2D, and 3D molecular descriptors presently in use in QSPR/QSAR and drug design practice, and (3) to compare the total and local (atom and atom type) quadratic indices with total and local spectral moments, with 2D/3D connectivity indices (vertex and edge ones) and with quantum chemical descriptors in modeling of the boiling point and partition coefficient ($\log P$), specific rate constant ($\log k$) as well as antibacterial activity of alkyl alcohols and 2-furylethylene derivatives, respectively.

2. Molecular quadratic indices

Implemented in the subprogram CARDD (acronym of Computer-Aided 'Rational' Drug Design) of the TOMOCOMD software, the quadratic indices can be calculated for the 'molecular pseudograph's atom adjacent matrix' of small-to-medium sized organic compounds. The general principles of the quadratic indices have been explained in some detail elsewhere.[16–21] However, an overview of this approach will be given.

For a given molecule composed of *n* atoms, the 'molecular vector' (X) is constructed and the *k*th total quadratic indices, $\boldsymbol{q}_k(x)$ are calculated as quadratic forms as shown in Eq. 1

$$\boldsymbol{q}_k(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} {}^{k}a_{ij}x_ix_j \tag{1}$$

where *n* is the number of atoms of the molecule and $x_1,\ldots,x_n$ are the coordinates or components of the 'molecular vector' (X) in a system of canonical basis vectors of $\Re^n$. The components of the 'molecular' vector are numeric values, which can be considered as weights (atom labels) for the vertices of the pseudograph. Certain atomic properties (electronegativity, density, atomic radii, etc.) can be used with this purpose. In this work Pauling electronegativities are selected as atom weights.[22]

The coefficients ${}^{k}a_{ij}$ are the elements of the *k*th power of the symmetric square matrix $\mathbf{M}(G)$ of the molecular pseudograph (G) and are defined as follows:

$$\begin{aligned} a_{ij} &= P_{ij} \quad \text{if } i \neq j \text{ and } \exists\ e_k \in E(G) \\ &= L_{ii} \quad \text{if } i = j \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{2}$$

where $E(G)$ represents the set of edges of G. P_{ij} is the number of edges (bonds) between vertices (atoms) v_i *y* v_j and L_{ii} is the number of loops in v_i.

Eq. 1 for $\boldsymbol{q}_k(x)$ can be written as the single matrix equation:

$$\boldsymbol{q}_k(x) = \mathbf{X}^{\mathrm{t}}\mathbf{M}^{k}\mathbf{X} \tag{3}$$

where $\mathbf{X}$ is a column vector (a $n \times 1$ matrix), $\mathbf{X}^{\mathrm{t}}$ the transpose of $\mathbf{X}$ (a $1 \times n$ matrix) and $\mathbf{M}^k$ the *k*th power of the matrix $\mathbf{M}$ of the molecular pseudograph G (mathematical quadratic form's matrix).

In addition to total quadratic indices, computed for the whole molecule, local-fragment (atom and atom type) formalisms can be developed. These descriptors are termed local quadratic indices, $\boldsymbol{q}_{kL}(x)$.[16–21] The definition of these descriptors is as follows:

$$\boldsymbol{q}_{kL}(x) = \sum_{i=1}^{m} \sum_{j=1}^{m} {}^{k}a_{ijL}x_ix_j \tag{4}$$

where *m* is the number of atoms of the fragment of interest and ${}^{k}a_{ijL}$ is the element of the row '*i*' and column '*j*' of the matrix $\mathbf{M}_L^k$. This matrix is extracted from the $\mathbf{M}^k$ matrix and contains the information referred to the vertices of the specific molecular fragments and also of the molecular environment.

The matrix $\mathbf{M}_L^k = [{}^{k}a_{ijL}]$ with elements ${}^{k}a_{ijL}$ is defined as follows:

$$\begin{aligned} {}^{k}a_{ijL} &= {}^{k}a_{ij} \quad \text{if both } v_i(x_i) \text{ and } v_j(x_j) \text{ are atoms} \\ &\qquad\qquad \text{contained within the molecular fragment} \\ &= \frac{1}{2}{}^{k}a_{ij} \quad \text{if } v_i(x_i) \text{or } v_j(x_j) \text{ is an atom contained} \\ &\qquad\qquad \text{within the molecular fragment} \\ &\qquad\qquad \text{but not both} \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{5}$$

These local analogues can also be expressed in matrix form by the expression:

$$\boldsymbol{q}_{kL}(x) = \mathbf{X}^{\mathrm{t}}\mathbf{M}_L^{k}\mathbf{X} \tag{6}$$

Note that the above scheme follows the spirit of a Mulliken population analysis.[23] Also note that for every partitioning of a molecule into *Z* molecular fragment there will be *Z* local molecular fragment matrices. In this case, if a molecule is partitioned into *Z* molecular fragments, the matrix $\mathbf{M}^k$ can be partitioned into *Z* local matrices

$\mathbf{M}_L^k$, $L = 1,\ldots,Z$, and the kth power of matrix $\mathbf{M}$ is exactly the sum of the kth power of the local Z matrices. In this way, the total quadratic indices are the sum of the quadratic indices of the Z molecular fragments:

$$\boldsymbol{q}_k(x) = \sum_{L=1}^{Z} q_{kL}(x) \tag{7}$$

Atom and atom-type quadratic indices are specific cases of local quadratic indices. In this sense, the kth atom-type quadratic indices are calculated by summing the kth atom quadratic indices of all atoms of the same atom type in the molecule.

In the atom-type quadratic indices formalism, each atom in the molecule is classified into an atom type (fragment), such as heteroatoms, H bonding to heteroatoms (O, N, and S), halogens, aliphatic carbon chain, aromatic atoms (aromatic rings), and so on. For all data sets, including those with a common molecular scaffold as well as those with very diverse structure, the kth atom-type quadratic indices provide important information.

As mentioned above, the present approach codifies 3D information by the introduction of a local *trigonometric 3D-chirality correction factor* in molecular vector, X.[21] In these sense, a chirality molecular vector is obtained (*X), where the components of X (for instance, Pauling electronegativity (x_A)[22] of the atom A) are substituted by the following term [$x_A + \sin((\omega_A + 4\Delta)\pi/2)$]. The trigonometric 3D-chirality correction factor use a dummy variable, ω_A[25–27] and an integer parameter, Δ:[21]

$$\begin{aligned}\omega_A &= 1 \text{ and } \Delta \text{ is an odd number when } A \text{ has (rectus), } E \text{ (entgegen), or } a \text{ (axial)notation according to Cahn–Ingold–Prelog rules}\\ &= 0 \text{ and } \Delta \text{ is an even number, if } A \text{ does not have specific environment}\\ &= -1 \text{ and } \Delta \text{ is an odd number when } A \text{ has (sinister), } Z \text{ (zusammen), or } e \text{ (ecuatorial)notation according to Cahn–Ingold–Prelog rules}\end{aligned} \tag{8}$$

Thus, this 3D-chirality factor $\sin((\omega_A + 4\Delta)\pi/2)$ takes different values in order to codify specific stereochemical information such as chirality, Z/E isomerism, and so on. This factor therefore takes values in the following order $1 > 0 > -1$ for atoms that have specific 3D environments. The chemical idea here is not that the attraction of electrons by an atom depends on their chirality, due to experience shows that chirality does not change the electronegativities of atoms in the molecule in an isotropic environment in an observable way.[24,27] This correction has principally a mathematical means and must not be source of any misunderstanding. The present trigonometric 3D-chiral correction factor is invariant with respect to the selection of other chirality scales[27] and gets ever the values 1, 0, and −1 for $R = E = a$, non-chiral = no Z/E isomerism involved = no a/e substitution, and $S = Z = e$ atoms, respectively.[21]

A very interesting point is that the present 3D-chiral descriptor reduces to simples (2D) quadratic indices ones for molecules without specific 3D characteristics because $\sin(0 + 4\Delta)\pi/2 = 0$, being Δ zero or any even number. That is, when all the atoms in the molecule are not chiral, the TOMOCOMD-CARDD molecular descriptors do not change upon the introduction of this factor. This means that $^*\text{X} = \text{X}$ and thus, $^*\boldsymbol{q}_k(x) = \boldsymbol{q}_k(x)$.[21]

3. Results and discussion

3.1. Interpretation and influence of structure change on total and local (atom) quadratic indices

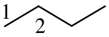
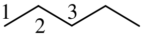
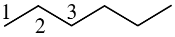
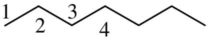
The influence of structure on the kth quadratic indices may be revealed by examining several sets of calculations in which features are systematically varied.[14] In this sense, some effect of structure on kth total and local quadratic indices are illustrated in several ways in the following examples: (a) effect of chain length, (b) effect due to branching, (c) effect across multiple bonds, and (d) effect due to heteroatom change. The influences of these structural features on our molecular descriptors are shown in Tables 1–4, respectively.

Firstly, note that the mathematical quadratic maps' matrices, $\mathbf{M}^k$, are graph-theoretic electronic-structure models, like an 'extended MO Hückel' model. The $\mathbf{M}^1$ matrix considers all valence-bond electrons (σ- and π-networks) in one step and their power ($k = 0, 1, 2, 3, \ldots$) can be considering as an interacting-electron chemical-network model in k step. This model can be seen as an intermediate between the quantitative quantum-mechanical Schrödinger equation and classical chemical bonding ideas.[28] Recently, Gonzalez et al. have developed a new method based on the Markov chain theory, which has been successfully employed in QSPR and QSAR studies.[29–32] This approach also describes changes in the electron (stochastic) distribution and vibrational decay with time throughout the molecular backbone using Markov chain formalism.

The present approach is based on a simple model for the intramolecular movement of all valence-bond electrons. Let us consider a hypothetical situation in which a set of atoms is free in space at an arbitrary initial time (t_0). In this time, the electrons are distributed around atom nucleus. Alternatively, these electrons can be distributed around cores in discrete intervals of time t_k. In this sense, the electron in an arbitrary atom i can move to other atoms at different discrete time periods t_k ($k = 0, 1, 2, 3, \ldots$) throughout the chemical-bonding network.[21]

For this reason, each kth total and local quadratic indices encode particular information of the molecular structure. For instance, atom-type quadratic index of zero order has information on the molecular size of the fragment and it depends on the number and type of atoms that are contained in the fragment under study.[16] In this connection, all the carbon-atom's

Table 1. The changes in *k*th total and local quadratic indices due to chain lengthening in alkanes

Atom (*i*)	$\boldsymbol{q}_0(x_i)$	$\boldsymbol{q}_1(x_i)$	$\boldsymbol{q}_2(x_i)$	$\boldsymbol{q}_3(x_i)$	$\boldsymbol{q}_4(x_i)$	$\boldsymbol{q}_5(x_i)$	$\boldsymbol{q}_6(x_i)$	$\boldsymbol{q}_7(x_i)$
								
C_1	6.5025	6.5025	13.005	19.5075	32.5125	52.02	84.5325	136.5525
C_2	6.5025	13.005	19.5075	32.5125	52.02	84.5325	136.5525	221.085
				...				
Total	26.01	39.015	65.025	104.04	169.065	273.105	442.17	715.275
								
C_1	6.5025	6.5025	13.005	19.5075	39.015	58.5225	117.045	175.5675
C_2	6.5025	13.005	19.5075	39.015	58.5225	117.045	175.5675	351.135
C_3	6.5025	13.005	26.01	39.015	78.03	117.045	234.09	351.135
				...				
Total	32.5125	52.02	91.035	156.06	273.105	468.18	819.315	1404.54
								
C_1	6.5025	6.5025	13.005	19.5075	39.015	65.025	123.5475	214.5825
C_2	6.5025	13.005	19.5075	39.015	65.025	123.5475	214.5825	396.6525
C_3	6.5025	13.005	26.01	45.5175	84.5325	149.5575	273.105	487.6875
				...				
Total	39.015	65.025	117.045	208.08	377.145	676.26	1222.47	2197.845
								
C_1	6.5025	6.5025	13.005	19.5075	39.015	65.025	130.05	221.085
C_2	6.5025	13.005	19.5075	39.015	65.025	130.05	221.085	442.17
C_3	6.5025	13.005	26.01	45.5175	91.035	156.06	312.12	533.205
C_4	6.5025	13.005	26.01	52.02	91.035	182.07	312.12	624.24
				...				
Total	45.5175	78.03	143.055	260.1	481.185	884.34	1638.63	3017.16

quadratic indices [$\boldsymbol{q}_0(x_i)$] have the same value (see Tables 1–4): 6.5025 in Pauling electronegativity scale.[22]

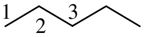
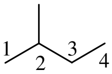
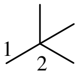
In a similar way, total quadratic indices of this same order [$\boldsymbol{q}_0(x)$] encode information about size and heteroatom content, which can be easily observed in Tables 1 and 4, respectively. In this sense, with lengthening of the chain (from butane to heptane) the $\boldsymbol{q}_0(x)$ value progressively increases. Therefore, for a homologous series this descriptor increases 6.5025 (squared carbon's Pauling electronegativity) for the addition of each methylene group. In addition, the value of this index for butane, for propylamine, propan-1-ol and for 1-fluoro-propane is increased in this same order ($\boldsymbol{q}_0(x)$ of 26.01, 28.7491, 31.3411, and 35.3479, respectively), in correspondence to the value of electronegativity of the following atoms: C, N, O, F.

Finally, total (and local) quadratic indices of zero order can be classified according to their 'dimensionality' as one-dimensional descriptors (1D). This index does not take into consideration the effect of branching and unsaturated atoms. Some examples of these are shown in Tables 2 and 3, respectively. This is a logical result, because in this initial time (t_0) the set of atoms is free in space and the electrons are distributed around atom nucleus.

On the other hand, total quadratic indices of the first order, $\boldsymbol{q}_1(x)$ is capable of discriminating between saturated and unsaturated (double and triple bonds) isomers (see Table 3), even though it cannot be considered as a unique descriptor and so some isomers have identical values. For instance, it not discriminates the 1-butene (52.02) and 1-butyne (65.025) from their isomers 2-butene (52.02) and 3-butyne (65.025), respectively. Further, $\boldsymbol{q}_1(x)$ is unable to differentiate between ramified isomers (see Table 2) and their value systematically varied due to chain lengthening in linear alkanes (see Table 1).

In another way, local (atom) quadratic indices of first order, $\boldsymbol{q}_1(x_i)$ are very influenced by the effect of branching in alkanes. Several examples illustrate this effect in Table 2. Methyl groups at a branch point have $\boldsymbol{q}_1(x_i)$

Table 2. Changes in *k*th total and local quadratic indices due to branching in the pentanes' skeleton

Atom (*i*)	$\boldsymbol{q}_0(x_i)$	$\boldsymbol{q}_1(x_i)$	$\boldsymbol{q}_2(x_i)$	$\boldsymbol{q}_3(x_i)$	$\boldsymbol{q}_4(x_i)$	$\boldsymbol{q}_5(x_i)$	$\boldsymbol{q}_6(x_i)$	$\boldsymbol{q}_7(x_i)$
								
C_1	6.5025	6.5025	13.005	19.5075	39.015	58.5225	117.045	175.5675
C_2	6.5025	13.005	19.5075	39.015	58.5225	117.045	175.5675	351.135
C_3	6.5025	13.005	26.01	39.015	78.03	117.045	234.09	351.135
				...				
Total	32.5125	52.02	91.035	156.06	273.105	468.18	819.315	1404.54
								
C_1	6.5025	6.5025	19.5075	26.01	65.025	91.035	221.085	312.12
C_2	6.5025	19.5075	26.01	65.025	91.035	221.085	312.12	754.29
C_3	6.5025	13.005	26.01	39.015	91.035	130.05	312.12	442.17
C_4	6.5025	6.5025	13.005	26.01	39.015	91.035	130.05	312.12
				...				
Total	32.5125	52.02	104.04	182.07	351.135	624.24	1196.46	2132.82
								
C_1	6.5025	6.5025	26.01	26.01	104.04	104.04	416.16	416.16
C_2	6.5025	26.01	26.01	104.04	104.04	416.16	416.16	1664.64
				...				
Total	32.5125	52.02	130.05	208.08	520.2	832.32	2080.8	3329.28

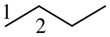
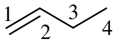
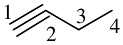
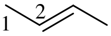
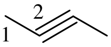
values of 6.5025. A significant increase in $\boldsymbol{q}_1(x_i)$ value is found on the carbon at the branch point (13.005, 19.5075, and 26.01 for the carbon atom with two, three, and four adjacent carbon atom, respectively). The unsaturated atoms exhibit the same behavior (see Table 3). This calculated effect mirrors the inductive effect and also the reduction in topological freedom or a raise in steric crowding. Additionally, the introduction of a heteroatom into an alkane molecule produces an effect on the $\boldsymbol{q}_1(x_i)$ of the adjacent carbon atoms, which is proportionate with Pauling electronegativity[22] value of the heteroatom. For example, $\boldsymbol{q}_1(x_i)$ of the adjacent carbon atom to the nitrogen, oxygen, and fluorine atom are 14.2545, 15.2745, and 16.6515, respectively (see Table 4).

Conversely, the total and local quadratic indices of second order can be considered as branching and multiple bonds molecular descriptors. For example, with branching of the chain, the atom quadratic indices of second-order value of the methyl group connected with the *n*-, *i*-, and *tert*-butyl group, steadily increase: $\boldsymbol{q}_2(x_{\mathrm{C1}})$ of 13.005, 19.5075, and 26.01, respectively (see Table 2). This table depicts that total quadratic index of second order, $\boldsymbol{q}_2(x)$ is able to discriminate among the pentane's branching isomers. In this case, $\boldsymbol{q}_2(x)$ values are also increased due to branching in the skeleton: $\boldsymbol{q}_2(x)$ of 91.035, 104.04, and 130.05 for pentane, for 2-methyl-butane, and for 2,2-dimethyl-propane, respectively (see Table 2). The unsaturated atoms exhibit the same behavior (see Table 3). For instance, this atoms (sp^2 and sp) have elevated $\boldsymbol{q}_2(x_i)$ values, which depend of the nature and topology of the atoms involved. Furthermore, terminal unsaturation result in lower $\boldsymbol{q}_2(x_i)$ values for this atoms relative to the unsaturated atoms in midchain. This behavior also mirrors the inductive effect and the reduction in topological freedom or a raise in steric crowding. That is, this local term represents the accessible from outside to atoms in a path of length 2 in the molecule. Thus, the second order total and local quadratic indices are interpreted as a component of the 'molecular accessibility' coming from contributions of paths of length 2 in the molecule.

In a similar way, we can interpret the effect of structure on 'higher' order of total and local quadratic indices in a molecule, starting from contributions of 'subgraphs' of different lengths 3, 4, 5, etc. The physical meaning of this approach is based on the fact the valence-shell electrons in an arbitrary atom *i* can move to other atoms at different discrete time periods t_k ($k = 0, 1, 2, 3, \ldots$) throughout the chemical-bonding network.

In any case, if a complete series of indices is considered, a specific characterization of the chemical structure is obtained (whole structure or fragment), which is not repeated in any other molecule. The generalization of the descriptors to 'superior analogs' is necessary for the evaluation of situations where only one descriptor is unable to bring a good structural characterization.[33] In this sense, the *k*th atom, atom type, and total

Table 3. The influences of unsaturation on the *k*th total and local quadratic indices in hydrocarbons

Atom (*i*)	$\mathbf{q}_0(x_i)$	$\mathbf{q}_1(x_i)$	$\mathbf{q}_2(x_i)$	$\mathbf{q}_3(x_i)$	$\mathbf{q}_4(x_i)$	$\mathbf{q}_5(x_i)$	$\mathbf{q}_6(x_i)$	$\mathbf{q}_7(x_i)$
								
C_1	6.5025	6.5025	13.005	19.5075	32.5125	52.02	84.5325	136.5525
C_2	6.5025	13.005	19.5075	32.5125	52.02	84.5325	136.5525	221.085
				. . .				
Total	26.01	39.015	65.025	104.04	169.065	273.105	442.17	715.275
								
C_1	6.5025	13.005	39.015	78.03	208.08	416.16	1092.42	2184.84
C_2	6.5025	19.5075	39.015	104.04	208.08	546.21	1092.42	2861.1
C_3	6.5025	13.005	26.01	52.02	130.05	260.1	676.26	1352.52
C_4	6.5025	6.5025	13.005	26.01	52.02	130.05	260.1	676.26
Total	26.01	52.02	117.045	260.1	598.23	1352.52	3121.2	7074.72
								
C_1	6.5025	19.5075	78.03	214.5825	799.8075	2184.84	8095.6125	22,101.998
C_2	6.5025	26.01	71.5275	266.6025	728.28	2698.5375	7367.3325	27,284.49
C_3	6.5025	13.005	32.5125	84.5325	299.115	812.8125	2997.6525	8180.145
C_4	6.5025	6.5025	13.005	32.5125	84.5325	299.115	812.8125	2997.6525
Total	26.01	65.025	195.075	598.23	1911.735	5995.305	19,273.41	60,564.285
								
C_1	6.5025	6.5025	19.5075	45.5175	110.5425	266.6025	643.7475	1554.0975
C_2	6.5025	19.5075	45.5175	110.5425	266.6025	643.7475	1554.0975	3751.9425
				. . .				
Total	26.01	52.02	130.05	312.12	754.29	1820.7	4395.69	10,612.08
								
C_1	6.5025	6.5025	26.01	84.5325	279.6075	923.355	3049.6725	10,072.373
C_2	6.5025	26.01	84.5325	279.6075	923.355	3049.6725	10,072.373	33,266.79
				. . .				
Total	26.01	65.025	221.085	728.28	2405.925	7946.055	26,244.09	86,678.325

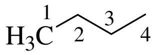
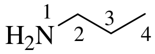
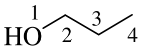
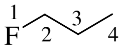
quadratic indices can be used as variables in QSAR/QSPR and 'rational' drug design studies.

3.2. Significance and comparison of the quadratic indices with other molecular descriptors

One of the main aims of this paper is the comparison of the information content of the *k*th total and local quadratic indices with that of the other descriptors used in QSPR/QSAR practice. The existence of linear independence has been claimed by Randic[33] as one of the desirable attributes for novel topological indices. To check the existence or not of linear independence between the total and local quadratic indices and others 229 0D–3D molecular descriptors calculated with DRAGON[6] software I carried out a factor analysis. The comparison was based on a set of 42 chemicals (see Table 5). Even though the number of chemicals is limited, the generality of the comparison was assured by the presence of diverse chemical functionalities and substructures. The symbols and definitions of some of these molecular indices are given in Table 6. The complete list of 0D–3D molecular descriptors used in this study as well as their description is given as Supplementary data. This list of molecular indices is representative of the three 'dimension' of molecular descriptors reported in the literature.

The results of the factor analysis are summarized in Table 7, and the 10 principal factors explain approximately 85.91% of the variance. The first factor explains 33.28% of the variance in the molecular indices studied. The addition of the second factor increases to 47.89% the variance explained, and the addition of the third factor allows 58.29% of the index variance to be accounted for. The other factors explain (% cumulative variance) the 5.91 (64.20), 5.58 (69.78), 4.30 (74.08), 3.71

Table 4. The influence of heteroatoms on the *k*th total and local quadratic indices values

Atom (*i*)	$f_0(x_i)$	$f_1(x_i)$	$f_2(x_i)$	$f_3(x_i)$	$f_4(x_i)$	$f_5(x_i)$	$f_6(x_i)$	$f_7(x_i)$
								
$C_1(CH_3)$	6.5025	6.5025	13.005	19.5075	32.5125	52.02	84.5325	136.5525
C_2	6.5025	13.005	19.5075	32.5125	52.02	84.5325	136.5525	221.085
C_3	6.5025	13.005	19.5075	32.5125	52.02	84.5325	136.5525	221.085
C_4	6.5025	6.5025	13.005	19.5075	32.5125	52.02	84.5325	136.5525
Total	26.01	39.015	65.025	104.04	169.065	273.105	442.17	715.275
								
$N_1(NH_2)$	9.2416	7.752	16.9936	23.256	41.7392	62.016	108.224	162.792
C_2	6.5025	14.2545	19.5075	35.0115	52.02	90.78	136.5525	237.3285
C_3	6.5025	13.005	20.757	32.5125	55.7685	84.5325	146.5485	221.085
C_4	6.5025	6.5025	13.005	20.757	32.5125	55.7685	84.5325	146.5485
Total	28.7491	41.514	70.2631	111.537	182.0402	293.097	475.8575	767.754
								
$O_1(OH)$	11.8336	8.772	20.6056	26.316	49.9832	70.176	129.344	184.212
C_2	6.5025	15.2745	19.5075	37.0515	52.02	95.88	136.5525	250.5885
C_3	6.5025	13.005	21.777	32.5125	58.8285	84.5325	154.7085	221.085
C_4	6.5025	6.5025	13.005	21.777	32.5125	58.8285	84.5325	154.7085
Total	31.3411	43.554	74.8951	117.657	193.3442	309.417	505.1375	810.594
								
F_1	15.8404	10.149	25.9894	30.447	62.1278	81.192	160.394	213.129
C_2	6.5025	16.6515	19.5075	39.8055	52.02	102.765	136.5525	268.4895
C_3	6.5025	13.005	23.154	32.5125	62.9595	84.5325	165.7245	221.085
C_4	6.5025	6.5025	13.005	23.154	32.5125	62.9595	84.5325	165.7245
Total	35.3479	46.308	81.6559	125.919	209.6198	331.449	547.2035	868.428

(77.78), 3.15 (80.93), 2.92 (83.85), and 2.06% (85.91) of the variance in the molecular descriptors studied (see Table 7). Factor loadings from the principal component analysis, after a Varimax normalized rotation of the factors, are shown in Table 8. This table depicts a single portion of the obtained results, and the complete list of molecular descriptors' loadings is given also as Supplementary data.

All *k*th total quadratic indices [$\boldsymbol{q}_k(x)$] are strongly loaded in factor 1 (F_1). Most of the constitutional (e.g., MW, Se, and nSK), empirical (Ui and ARR), molecular properties (MR), topological (first-, second-, and some third-generation; such as: ISIZ, ZM2V, Ram, TI1, D/D, X0-5v, X2sol, XMOD, PW4, PW5, AECC, VAR, TIC4, STN, Eig1p, VRZ1, MPC07, piPC05, piID, D/Dr06, D/Dr10, BEHe6, BELe4, and GGI4), 2D-autocorrelations, (ATS7e), Randic molecular profiles (DP01 and DP10), geometrical (i.e., AGDD, TIE, and G2), RDF (RDF045-55u), 3D-MoRSE (Mor07u and Mor20e), WHIM (L1u), and GETAWAY (ITH) are also robustly loaded (loadings > 0.70) in this factor. Thus, total quadratic indices and the others 'F_1-indices' produce much redundancy and overlapping among them and their relations are very complex. The third factor (F_3) is almost exclusively an atom type (H atoms bonding to heteroatoms) quadratic indices [$\boldsymbol{q}_{kL}(x_{\text{H–E}})$], an hydrophilic factor (Hy), a fragment-based polar surface area (PSA) and a radial distribution function 2.0/unweighted (RDF020u) dimension. This result showed that $\boldsymbol{q}_{kL}(x_{\text{H–E}})$, Hy, PSA, and RDF020u have a strongly parallel relation. In addition, these molecular descriptors and Moriguchi octanol–water partition coefficient (MLOGP) are also connected by the factor 3, as can be seen in Table 8. In this case, there is a weakly opposite relation between these molecular descriptors. This is a logical result, because these indices encode the hydrogen-bonding capabilities in opposite ways. The eighth factor (F_8) appears to be most significant for the atom-type quadratic indices of heteroatoms [$\boldsymbol{q}_{kL}(x_{\text{E}})$]. As subsequently stated (see experimental section), the indices with a high loading in the same factor are interrelated, while no correlation exists between indices having non-zero loadings only in different factors.[34,35] Consequently, it is clear that the atom-type quadratic indices are orthogonal to most of the 0D–3D molecular descriptors. Thus, I can say that the atom-type (heteroatoms) quadratic indices contain structural information not contained in any other 0D–3D molecular descriptors.

3.3. Applications in QSPR/QSAR studies

3.3.1. Describing boiling points of 28 alkyl alcohols. The objective will be to show, in as direct a manner as

Table 5. Chemicals data set for analysis of principal components

No.	Chemical
01	Methane
02	Ethane
03	*n*-Propane
04	*n*-Butane
05	*n*-Pentane
06	*n*-Hexane
07	Isobutane
08	Neopentane
09	2-Methylpentane
10	*cis*-2-Butene
11	*trans*-2-Butene
12	2-Butine
13	Cyclopropane
14	Cyclobutane
15	Cyclopentane
16	Cyclohexane
17	Cyclohexanone
18	Benzene
19	Toluene
20	Phenol
21	Benzoic acid
22	Aniline
23	Nitrobenzene
24	Fluorobenzene
25	Chlorobenzene
26	Bromobenzene
27	Iodobenzene
28	Benzamide
29	Naphthalene
30	Anthracene
31	Pyrrole
32	Furan
33	Thiophen
34	Purine
35	Dibenzofuran
36	Ethanol
37	Trifluoroethanol
38	2-Aminoethanol
39	Propanol
40	2-Propanone
41	2-Propanol
42	2-Propylamine

possible, that the total and local quadratic indices delineated in a previous section yield predictive molecular properties/activities in a QSPR/QSAR analysis. The first data set that will be studied here is composed by 28 alkyl alcohols, 14 are primary, 6 secondary, and 8 tertiary, for which the boiling point (bp) has been reported previously.[11,12] The best linear regression model obtained to describe the bp of these compounds using total quadratic indices is given below:

$$\text{bp }(^\circ\text{C}) = -154.575(\pm 8.33) + 6.53(\pm 0.26)\,\boldsymbol{q}_1^{\text{H}}(x) - 0.763(\pm 0.03)\,\boldsymbol{q}_3^{\text{H}}(x) \qquad (9)$$

$N = 28$, R=0.995, $R^2 = 0.990$, $s = 2.910$, $q_2 = 0.989$, $s_{\text{CV}} = 2.969$, $F(2,25) = 1268.9$, where $\boldsymbol{q}_1^{\text{H}}(x)$ and $\boldsymbol{q}_3^{\text{H}}(x)$ are the total quadratic indices (first and third order, respectively) calculated considering H atoms in molecular pseudograph, N is the number of compounds, R is the regression coefficient, R^2 is the determination coefficient, s is the standard deviation of the regression, F is the Fisher ratio, and q^2 and s_{CV} is the squared regression coefficient and the standard deviation of the LOO cross-validation experiment, respectively.

In Table 9 are depicted the values of experimental and calculated values of the bp for data set, and in Figure 1 is illustrated the linear relationships between them. This model (Eq. 9) explains the 99% of the variance of the experimental bp values.

Quite similar equations were reported by Estrada and Molina[12] and Kier and Hall[11] using spectral moment and E-state/biomolecular encounter parameters as molecular descriptors, respectively. These models explain more than 98% and 92% of the variance of the experimental bp values. Table 9 also depicts the values of experimental and calculated values of the bp of the 28 studied alkyl alcohols using these approaches.[11,12]

Predictability and stability of the obtained model using total quadratic indices (Eq. 9) to data variation is carried out here by means of LOO cross-validation procedure. This model (Eq. 9) showed a q^2 and s_{CV} of 0.989 and 2.969 °C, respectively. Unfortunately, these authors[11,12] do not report the result of the LOO cross validation. It is remarkable that Eq. 9 uses three variables less than the model obtained by Estrada and Molina[12] and one variable less than the model obtained by Kier and Hall.[11] However, Eq. 9 explains a greater percent of the variance of the experimental bp values than that the previously developed models do (a decrease in the standard error of 68.29% and 49.83% with regard to the results archived previously).[11,12] Table 10 summarizes the statistical parameters archived by all these approaches.

3.3.2. Modeling partition coefficients (log *P*) and specific rate constants (log *k*) of 34 2-furylethylenes derivatives. It has been clear from structure–activity relationship studies that the lipophilicity and the nucleophilic addition of thiol groups of some enzymes to the exocyclic double bond of 2-furylethylene derivatives are critical for the development of their antibacterial activity.[36] The partition coefficient *n*-octanol/water (log *P*) and the specific rate constant (log *k*) of nucleophilic addition of the mercaptoacetic acid to the exocyclic double bond has an important role in the understanding of the biological behavior of these 2-furylethylene derivatives.[36] Consequently, I will study these parameters to compare the possibilities of molecular quadratic indices in QSPR and to compare these results to those obtained by Estrada and Molina[12,37] using topological (total and local spectral moment and 2D connectivity indices), topographic and quantum chemical descriptors. The molecular structures of such compounds are depicted in Table 11 and the overall result of the obtained models by these authors[12,37] are summarized in Table 10.

The molecular descriptors included in these equations clearly pointed to the identification of the reaction centers involved in the studied chemical interaction.[12] That is to say, the *k*th local spectral moments calculated for

Table 6. Description of some of the 0D–3D DRAGON's molecular descriptors used in the present study

Symbol	Definition	Class
MW	Molecular weight	Constitutional
Se	Sum of atomic Sanderson electronegativities (scaled on carbon atom)	Constitutional
Ss	Sum of Kier–Hall electrotopological states	Constitutional
Ui	Unsaturation index	Empirical
Hy	Hydrophilic factor	Empirical
ARR	Aromatic ratio	Empirical
MR	Ghose–Crippen molar refractivity	Properties
PSA	Fragment-based polar surface area	Properties
MLOGP	Moriguchi octanol–water partition coeff. (log *P*)	Properties
ZM1V	First Zagreb index by valence vertex degrees	Topological
HNar	Narumi harmonic topological index	Topological
TI2	Second Mohar index TI2	Topological
Rww	Reciprocal hyper-detour index	Topological
J	Balaban J index	Topological
JhetZ	Balaban-type index from *Z* weighted distance matrix (Barysz matrix)	Topological
X2v	Valence connectivity index chi-2	Topological
S1K	1-Path Kier alpha-modified shape index	Topological
PHI	Kier flexibility index	Topological
PW2	Path/walk 2—Randic shape index	Topological
PJI2	2D Petitjean shape index	Topological
SIC2	Structural information content (neighborhood symmetry of 2-order)	Topological
SEigZ	Eigenvalue sum from *Z* weighted distance matrix (Barysz matrix)	Topological
SRW05	Self-returning walk count of order 05	Mol. walk counts
BEHe6	Highest eigenvalue n. 6 of Burden matrix/weighted by atomic Sand. elect.	BCUT
GGI1	Topological charge index of order 1	Galvez charge ind.
ATS2e	Broto-Moreau autocorrelation of a topological structure—lag 2/Sand. elect.	2D autocorrelations
MATS1e	Moran autocorrelation—lag 1/weighted by atomic Sand. electronegativitie	2D autocorrelations
GATS3e	Geary autocorrelation—lag 3/weighted by atomic Sand. electronegativitie	2D autocorrelations
HOMA	Harmonic Oscillator Model of Aromaticity index	Aromat. indices
RCI	Jug RC index	Aromat. indices
AROM	Aromaticity (trial)	Aromat. indices
DP10	Molecular profile no. 10	Randic mol profiles
SHP2	Average shape profile index of order 2	Randic mol profiles
J3D	3D-Balaban index	Geometrical
TIE	E-state topological parameter	Geometrical
FDI	Folding degree index	Geometrical
PJI3	3D Petijean shape index	Geometrical
RDF015u	Radial Distribution Function—1.5/unweighted	RDF
RDF010e	Radial Distribution Function—1.0/weighted by atomic Sand. Electroneg.	RDF
Mor02u	3D-MoRSE—signal 02/unweighted	3D-MoRSE
Mor20e	3D-MoRSE—signal 20/weighted by atomic Sanderson electronegativities	3D-MoRSE
E1e	First component accessibility directional WHIM index /atomic Sand. elect.	WHIM
Gu	G total symmetry index/unweighted	WHIM
HGM	Geometric mean on the leverage magnitude	GETAWAY
H0e	H autocorrelation of lag 0/weighted by atomic Sand. electronegativities	GETAWAY
RCON	Randic-type R matrix connectivity	GETAWAY
R2e	R autocorrelation of lag 2/weighted by atomic Sand. electronegativities	GETAWAY

Table 7. Results of the factor analysis by using the principal component method for 229 0D–3D molecular descriptors as well as the 33 total and local (atom type) quadratic indices for 42 very heterogeneous chemicals

Factors	Eigenvalue	% Total variance	Cumulative eigenvalue	% Cumulative variance
F_1	87.20	33.28	87.20	33.28
F_2	38.26	14.60	125.46	47.89
F_3	27.26	10.40	152.72	58.29
F_4	15.49	5.91	168.20	64.20
F_5	14.61	5.58	182.82	69.78
F_6	11.26	4.30	194.08	74.08
F_7	9.71	3.71	203.79	77.78
F_8	8.25	3.15	212.04	80.93
F_9	7.64	2.92	219.68	83.85
F_{10}	5.40	2.06	225.08	85.91

the atoms 2, 6, and 7 or for the bonds defined by these atoms (C_2–C_6 and C_6–C_7) were selected as the most significant.[12] These atoms are those involved in the exocyclic double bond of the 2-furylethylene and these are the 'target' of the nucleophilic attack by thiol (mercapto) group.

Taking into account this logical result, I calculated the *k*th local quadratic indices for these atoms (bonds C_2–C_6 and C_6–C_7). The best obtained models, using these atom type [$\boldsymbol{q}_k^{\mathrm{H}}(x_{\mathrm{C6-C7}})$ and $\boldsymbol{q}_k^{\mathrm{H}}(x_{\mathrm{C2-C6}})$] and total [$\boldsymbol{q}_k^{\mathrm{H}}(x)$ and $\boldsymbol{q}_k(x)$; using H atoms suppressed- and available-pseudograph, respectively] quadratic indices as molecular descriptors, together with its statistical parameters is given:

Table 8. Factor loadings (varimax normalized rotation) for some of the 229 0D–3D molecular descriptors as well as some of the 33 total and local (atom type) quadratic indices for 42 very heterogeneous chemicals

Index	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
MW	*0.73*	−0.08	−0.03	0.05	0.43	−0.10	0.22	0.11	−0.02	0.19
nSK	*0.90*	0.02	0.05	0.06	0.33	−0.04	0.16	0.18	−0.02	0.08
Ui	*0.71*	−0.42	0.05	−0.01	0.42	−0.21	−0.05	0.22	−0.03	0.13
Hy	−0.19	−0.16	*0.81*	−0.04	−0.20	0.02	0.22	−0.17	0.03	−0.06
MR	*0.88*	0.09	−0.03	0.12	0.38	−0.09	−0.05	0.05	0.02	0.12
PSA	0.06	−0.18	*0.72*	0.02	0.02	0.06	0.37	0.33	0.05	0.04
MLOGP	0.39	0.41	−0.58	0.05	0.18	−0.17	−0.21	−0.23	−0.10	0.15
ISIZ	*0.70*	0.63	−0.04	0.22	0.16	−0.01	−0.05	−0.05	0.07	0.12
IAC	*0.70*	0.23	0.27	0.14	0.33	−0.07	0.37	0.25	0.03	0.16
ZM2V	*0.84*	−0.23	0.10	−0.03	0.19	0.00	0.27	0.34	−0.03	0.01
HNar	0.48	0.05	0.03	−0.15	*0.73*	0.21	−0.02	0.17	−0.33	−0.07
Ram	*0.86*	0.01	0.05	−0.10	0.02	−0.02	0.29	0.18	0.32	0.00
TI1	*0.89*	−0.16	−0.04	−0.21	−0.01	0.20	−0.15	0.02	−0.08	−0.15
TI2	0.12	0.26	0.09	*0.89*	0.02	−0.03	0.24	−0.07	0.12	−0.04
D/D	*0.79*	0.11	0.06	0.26	0.18	−0.17	0.36	0.16	0.08	0.17
JhetZ	0.15	−0.23	0.04	0.05	*0.75*	−0.05	0.39	0.14	0.25	0.05
Jhetv	0.23	0.05	−0.11	0.12	*0.79*	−0.23	−0.11	0.02	0.24	0.04
X0v	*0.79*	0.24	−0.09	0.17	0.45	−0.04	0.05	0.06	0.11	0.12
X5v	*0.90*	0.08	−0.09	−0.05	0.13	−0.09	−0.01	0.00	−0.16	0.02
X2sol	*0.85*	0.01	0.02	−0.10	0.40	−0.06	0.00	0.17	0.12	0.18
S3K	−0.15	0.15	−0.03	*0.84*	−0.08	−0.01	0.24	−0.14	−0.12	−0.09
PHI	−0.08	0.38	0.01	*0.85*	0.01	0.01	0.10	−0.20	0.01	0.09
PW2	0.35	0.21	0.11	−0.11	*0.71*	0.05	0.33	0.13	0.21	0.01
PW5	*0.82*	−0.06	0.02	−0.04	0.28	−0.26	0.00	0.20	−0.17	0.14
AECC	*0.79*	0.12	0.05	0.36	0.40	−0.14	0.16	0.10	−0.09	0.08
VAR	*0.89*	0.03	0.05	0.24	0.05	−0.06	0.27	0.13	0.11	0.11
CIC0	0.37	*0.71*	−0.11	0.24	0.40	0.04	−0.22	−0.14	0.10	−0.13
TIC4	*0.72*	0.18	0.25	0.38	0.23	0.02	0.16	0.14	0.05	0.27
LP1	0.50	0.17	0.08	−0.07	*0.76*	0.11	0.24	0.18	0.00	−0.06
STN	*0.85*	−0.13	0.00	−0.24	0.29	0.10	−0.09	0.19	−0.22	0.02
Eig1p	*0.81*	0.05	0.08	0.13	0.09	0.05	0.44	0.28	−0.02	0.06
SEigZ	0.03	−0.42	0.09	−0.04	0.24	−0.05	*0.71*	0.21	−0.01	0.16
SEigp	−0.04	0.16	−0.23	−0.01	0.12	0.00	*−0.84*	−0.26	0.00	0.12
VRZ1	*0.96*	−0.02	0.03	0.01	0.19	0.02	0.10	0.15	−0.02	0.03
MPC07	*0.91*	−0.03	−0.05	0.00	−0.10	0.29	−0.05	−0.02	0.05	−0.15
piPC05	*0.97*	−0.11	−0.01	−0.03	0.05	0.01	−0.04	−0.04	0.01	−0.10
D/Dr10	*0.74*	0.00	−0.07	−0.05	−0.06	−0.06	−0.07	−0.37	0.02	−0.26
SRW07	0.15	−0.14	0.00	−0.25	0.09	*0.78*	−0.13	0.38	−0.15	0.01
SRW09	0.30	−0.12	0.02	−0.16	0.05	*0.75*	−0.12	0.45	−0.08	0.03
BEHe7	*0.75*	0.35	0.12	0.21	0.16	−0.12	0.00	0.02	0.20	−0.17
BELe4	*0.81*	0.29	0.02	0.11	−0.06	0.10	−0.05	0.12	0.24	0.10
BELe5	*0.89*	0.02	−0.12	0.01	−0.08	0.25	−0.08	−0.21	0.02	−0.12
GGI3	*0.83*	−0.02	0.03	0.15	0.06	0.12	0.09	0.29	0.00	0.20
GGI5	*0.70*	−0.04	0.08	0.11	−0.06	−0.01	0.23	0.18	0.02	0.05
ATS2e	0.09	−0.27	0.19	−0.04	0.08	0.10	*0.81*	0.21	−0.03	−0.09
ATS7e	*0.77*	0.10	0.07	0.24	−0.08	0.12	0.04	−0.10	−0.03	0.12
MATS1e	0.27	0.15	−0.28	−0.02	*0.76*	0.09	0.24	0.01	−0.04	−0.11
MATS3e	0.13	−0.10	0.14	0.09	−0.16	0.13	0.04	0.08	*−0.80*	0.01
GATS3e	−0.01	0.15	0.00	−0.04	0.33	−0.04	0.09	−0.09	*0.76*	−0.07
DP01	*0.72*	0.16	0.07	0.21	0.56	−0.06	0.22	0.17	0.03	0.09
AGDD	*0.82*	0.46	−0.04	0.27	0.15	0.02	−0.05	−0.05	0.06	0.10
MAXDN	0.06	−0.14	0.23	0.00	−0.04	−0.16	*0.90*	0.12	0.14	−0.03
TIE	*0.79*	0.18	0.13	0.10	0.19	−0.10	0.14	0.17	0.03	0.15
G2	*0.86*	−0.09	0.04	0.00	0.36	−0.02	0.19	0.22	−0.04	0.11
SPAM	0.01	*−0.80*	0.14	0.39	0.16	0.06	0.19	0.11	−0.06	0.11
ASP	0.15	−0.18	−0.04	*0.80*	0.23	−0.07	−0.02	−0.11	−0.11	−0.02
FDI	0.42	−0.06	−0.02	0.10	*0.82*	0.01	0.14	0.20	−0.11	−0.03
RDF020u	0.21	−0.06	*0.73*	0.24	−0.03	0.22	0.22	0.01	0.03	−0.10
RDF025u	0.48	*0.72*	−0.11	0.15	0.14	−0.13	−0.04	−0.14	−0.10	0.23
RDF045u	*0.76*	0.20	0.22	0.35	0.03	0.01	−0.11	0.19	0.01	0.05
RDF055u	*0.73*	0.13	−0.05	0.27	−0.09	0.42	−0.07	0.17	0.03	0.12
Mor02u	0.17	*0.95*	−0.02	0.07	0.05	0.12	−0.03	−0.14	0.04	0.03
Mor07u	*0.77*	−0.20	0.07	−0.13	0.40	−0.04	0.05	0.12	−0.31	0.05
Mor19u	−0.23	*0.76*	−0.38	0.24	−0.01	−0.08	−0.18	−0.16	0.01	0.07
Mor20e	*0.77*	−0.39	0.12	−0.06	0.14	−0.23	−0.01	−0.08	0.07	0.00

Table 8 (*continued*)

Index	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
L1u	*0.78*	0.09	0.00	0.58	0.10	0.10	−0.05	0.01	0.06	0.00
P1u	0.28	−0.15	0.13	*0.87*	0.07	0.16	0.04	0.03	0.08	−0.16
Ku	0.27	−0.23	0.08	*0.82*	0.34	0.09	0.02	0.01	−0.04	−0.09
HIC	0.54	*0.72*	0.01	0.16	0.36	−0.03	0.01	−0.07	0.08	0.09
HGM	−0.38	−0.26	−0.06	−0.28	*−0.81*	−0.04	−0.08	−0.01	−0.07	0.12
HTu	0.48	*0.81*	−0.03	0.15	0.17	−0.04	−0.07	0.00	0.14	0.10
HATS0u	−0.46	−0.19	0.02	−0.01	*−0.74*	0.01	−0.04	0.16	0.02	−0.27
HATS5u	0.07	0.09	−0.08	*0.75*	0.08	−0.27	−0.25	0.20	0.01	−0.21
H4e	0.16	*0.78*	−0.09	0.33	−0.09	0.00	−0.01	−0.16	0.22	0.21
HATS2e	−0.39	−0.16	−0.11	−0.07	*−0.83*	−0.02	0.14	0.00	−0.03	−0.15
RCON	0.17	*0.93*	0.08	0.09	0.19	0.08	−0.02	0.08	0.01	−0.12
R3u	−0.22	*0.87*	0.06	−0.17	0.07	0.13	0.17	−0.14	−0.13	−0.12
R5u	0.37	0.32	−0.04	*0.73*	0.06	−0.18	−0.18	0.12	−0.09	0.03
RTu	−0.03	*0.96*	0.03	0.11	−0.05	0.03	−0.03	0.01	0.18	−0.07
R2u+	−0.41	−0.25	−0.08	−0.01	*−0.81*	−0.02	−0.11	0.01	0.00	−0.07
R5u+	0.27	0.15	0.15	*0.71*	0.08	−0.26	−0.17	0.19	−0.02	−0.08
R2e+	−0.39	−0.30	−0.07	−0.04	*−0.72*	0.04	0.26	0.03	−0.03	−0.12
$\boldsymbol{q}_0(x)$	*0.81*	−0.05	0.11	0.05	0.27	−0.05	0.41	0.24	−0.02	0.05
$\boldsymbol{q}_1(x)$	*0.88*	−0.16	0.06	0.00	0.32	−0.04	0.12	0.25	−0.05	0.05
$\boldsymbol{q}_5(x)$	*0.92*	−0.26	0.04	0.01	0.20	−0.01	−0.01	0.21	0.00	0.00
$\boldsymbol{q}_{10}(x)$	*0.95*	−0.19	0.01	0.03	0.09	0.07	−0.04	0.14	0.03	−0.08
$\boldsymbol{q}_{0L}(x_E)$	0.07	−0.26	0.33	−0.01	0.01	−0.01	*0.80*	0.40	0.00	−0.03
$\boldsymbol{q}_{1L}(x_E)$	0.15	−0.19	0.25	−0.02	0.05	0.03	0.47	*0.74*	0.00	−0.02
$\boldsymbol{q}_{2L}(x_E)$	0.18	−0.19	0.25	−0.02	0.03	0.01	0.54	*0.73*	0.01	−0.02
$\boldsymbol{q}_{9L}(x_E)$	0.31	−0.13	0.21	0.00	0.06	0.10	0.10	*0.86*	−0.01	0.02
$\boldsymbol{q}_{0L}(x_{E\text{–}H})$	−0.05	−0.08	*0.95*	0.03	−0.05	−0.04	0.21	−0.08	0.04	−0.03
$\boldsymbol{q}_{1L}(x_{E\text{–}H})$	−0.06	−0.08	*0.94*	0.03	−0.05	−0.04	0.23	−0.09	0.04	−0.03
$\boldsymbol{q}_{4L}(x_{E\text{–}H})$	−0.01	−0.08	*0.97*	−0.02	0.01	−0.03	0.07	0.09	0.02	−0.01
$\boldsymbol{q}_{10L}(x_{E\text{–}H})$	0.11	−0.11	*0.78*	−0.03	0.11	−0.07	−0.11	0.39	−0.04	0.07

$$\begin{aligned}\log P = {} & 12.117(\pm 2.343) + 0.027(\pm 0.001)\boldsymbol{q}_0^{\mathrm{H}}(x) \\ & + 0.31(\pm 0.021)\boldsymbol{q}_0^{\mathrm{H}}(x_{\mathrm{E}}) \\ & - 0.053(\pm 0.0033)\boldsymbol{q}_2^{\mathrm{H}}(x_{\mathrm{E}}) \\ & - 0.211(\pm 0.05)\boldsymbol{q}_1^{\mathrm{H}}(x_{\mathrm{E\text{–}H}}) \\ & + 0.019(\pm 0.007)\boldsymbol{q}_3^{\mathrm{H}}(x_{\mathrm{E\text{–}H}}) \\ & - 0.267(\pm 0.048)\boldsymbol{q}_1^{\mathrm{H}}(x_{\mathrm{C6\text{–}C7}}) + 2.418 \\ & \times 10^{-9}(\pm 2.16 \times 10^{-10})\boldsymbol{q}_{15}^{\mathrm{H}}(x_{\mathrm{C6\text{–}C7}}) \qquad (10)\end{aligned}$$

$N = 34$, $R = 0.985$, $R^2 = 0.969$, $s = 0.142$, $q^2 = 0.951$, $s_{\mathrm{CV}} = 0.156$, $F(7,26) = 116.76$.

$$\begin{aligned}\log k = {} & 13.496(\pm 0.725) - 0.138(\pm 0.022)\boldsymbol{q}_0(x) \\ & + 7.04 \times 10^{-9}(\pm 1.96 \times 10^{-9})\boldsymbol{q}_{13}(x) \\ & + 0.98 \times 10^{-3}(\pm 0.21 \times 10^{-3})\boldsymbol{q}_4^{\mathrm{H}}(x) \\ & + 0.049(\pm 0.006)\boldsymbol{q}_1(x_{\mathrm{E}}) \\ & + 0.446(\pm 0.056)\boldsymbol{q}_2^{\mathrm{H}}(x_{\mathrm{C6\text{–}C7}}) \\ & - 0.017(\pm 0.036)\boldsymbol{q}_4^{\mathrm{H}}(x_{\mathrm{C6\text{–}C7}}) \\ & - 0.029(\pm 0.005)\boldsymbol{q}_4^{\mathrm{H}}(x_{\mathrm{C2\text{–}C6}}) \qquad (11)\end{aligned}$$

$N = 34$, $R = 0.984$, $R^2 = 0.969$, $s = 0.285$, $q^2 = 0.922$, $s_{\mathrm{CV}} = 0.298$, $F(7,26) = 115.14$. These equations explained 96.9% of the variance of both $\log P$ and $\log k$. These statistics are slightly better than those obtained previously (see Table 10).[12,37] The experimental and calculated values of $\log P$ and $\log k$ obtained with 2D and 3D connectivity indices, quantum chemical descriptors, total and local spectral moments as well as molecular quadratic indices are shown in Tables 12 and 13, respectively. Plots of observed versus calculated $\log P$ and $\log k$ according to the Eqs. 10 and 11 are illustrated in Figures 2 and 3, respectively.

Finally, LOO cross-validation procedure was used in order to assess the predictive ability of developed models. Using this approach, the models 10 and 11 had a LOO q^2 of 0.951 and 0.922, respectively. These values of q^2 ($q^2 > 0.5$) can be considered as a proof of the high predictive ability of the models.[38] In this sense, the equations obtained with the vertex and edge connectivity indices, with the topographic descriptors, and with the quantum chemical descriptors (Eqs. 10, 11, and 13 in Ref. 34) showed a smaller predictive abilities (s_{CV} of 0.247, 0.176, and 0.370, respectively) that Eq. 10 ($s_{\mathrm{CV}} = 0.156$), achieved with the total and local quadratic indices. Unfortunately, the authors[12] do not report the result of the LOO cross-validation experiment for $\log k$ (see Table 10).

3.3.3. Classification of 34 2-furylethylene derivatives as antibacterial. Linear discriminant analysis (LDA) will be used here to obtain a classification model of 2-furylethylene compounds according to their antibacterial activity. The classification model obtained is given below together with the statistical parameters of LDA:

$$\begin{aligned}\mathrm{Act} = {} & 4.465 - 0.025\boldsymbol{q}_2^{\mathrm{H}}(x) + 0.681\boldsymbol{q}_{L1}^{\mathrm{H}}(x_{\mathrm{E}}) \\ & - 0.716\boldsymbol{q}_{L2}^{\mathrm{H}}(x_{\mathrm{E}}) + 0.125\boldsymbol{q}_{L4}^{\mathrm{H}}(x_{\mathrm{E}}) \\ & - 2.023 \times 10^{-3}\boldsymbol{q}_L^{\mathrm{H}}7(x_{\mathrm{E}}) \qquad (12)\end{aligned}$$

Table 9. Experimental and predicted values of the boiling point of alcohols R–OH used in this study

Alcohol–R	Exp. bp[a] (°C)	Pred. bp[b] (°C)	Res.[c]	Res-CV[d]	Pred. bp[e] (°C)	Pred. bp[f] (°C)
$(CH_3)_2CH$–	82.3	83.4	−1.1	−1.3	82.9	91.1
$CH_3CH_2CH_2$–	97.2	96.4	0.8	1.0	96	97.4
$CH_3(CH_2)_3$–	117.7	116.2	1.5	1.7	115.2	113.6
$CH_3CH(CH_3)CH_2$–	107.8	108.7	−0.9	−1.0	108	109
$CH_3CH_2C(CH_3)_2$–	102.4	101.6	0.8	0.9	105.4	112.4
$CH_3CH_2CH_2CH(CH_3)$–	119.3	122.3	−3.0	−3.2	114.4	120.3
$CH_3CH(CH_3)CH_2CH_2$–	131.1	128.9	2.2	2.3	134.5	127.4
$CH_3CH_2CH(CH_3)CH_2$–	128	128.3	−0.3	−0.3	127.3	125.2
$CH_3(CH_2)_4$–	137.9	136.0	1.9	2.1	134.3	131.8
$CH_3C(CH_3)_2CH(CH_3)$–	120.4	119.6	0.8	0.9	129.3	123
$CH_3(CH_2)_2C(CH_3)_2$–	121.1	121.4	−0.3	−0.3	124.9	128.9
$(CH_3CH_2)_2C(CH_3)$–	122.4	120.5	1.9	2.1	121.9	126.3
$CH_3CH_2C(CH_3)_2CH_2$–	136.5	133.3	3.2	3.3	142.5	138.4
$CH_3CH(CH_3)CH_2CH(CH_3)$–	131.6	135.1	−3.5	−3.6	133.9	133.4
$CH_3CH(CH_3)CH(CH_3CH_2)$–	126.5	133.7	−7.2	−7.5	121.9	128.7
$CH_3CH(CH_3)CH(CH_3)CH_2$–	144.5	140.8	3.7	3.8	146.7	138.3
$CH_3CH_2CH_2CH(CH_3)CH_2$–	149	148.1	0.9	1.0	146.4	143.4
$CH_3(CH_2)_5$–	157.6	155.8	1.8	2.0	153.4	169.8
$(CH_3CH(CH_3))_2CH$–	138.7	145.8	−7.1	−7.6	136.4	139
$CH_3CH(CH_3)CH_2CH(CH_3)CH_2$–	159	160.8	−1.8	−1.9	165.5	157.7
$(CH_3CH_2)_3C$–	142	139.5	2.5	2.8	138.6	138.5
$CH_3(CH_2)_6$–	176.4	175.6	0.8	1.0	172.5	172.2
$(CH_3CH_2CH_2)_2(CH_3)C$–	161	160.1	0.9	1.0	160.9	161.3
$(CH_3(CH_2)_3)(CH_3CH_2)(CH_3)C$–	163	160.1	2.9	3.3	160.5	162.7
$CH_3CH(CH_3)CH_2(CH_2)_4$–	188	188.3	−0.3	−0.4	191.6	188.3
$CH_3(CH_2)_7$–	195.1	195.4	−0.3	−0.3	191.6	193
$CH_3(CH_2)_5C(CH_3)_2$–	178	181.7	−3.7	−4.2	182.2	188.4
$(CH_3CH_2CH_2)_2(CH_3CH_2)C$–	182	179.1	2.9	3.6	177.6	177

[a] Experimental values taken of the Ref. 12.

[b] Predicted values using total quadratic indices (Eq. 9).

[c] Residual values: bp(Exp) − bp(Pred).

[d] Residual values of the LOO cross-validation experiment (deleted residual).

[e] Predicted values using spectral moments.[12]

[f] Predicted values using E-state.[11]

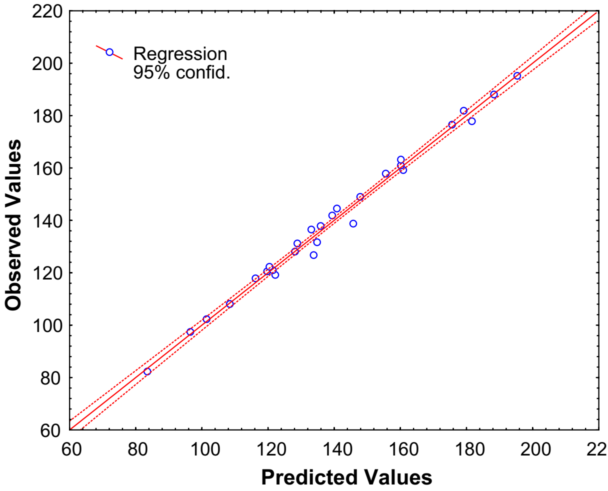


Figure 1. Correlation between experimental and calculated boiling point by Eq. 9.

$N = 34, \quad \lambda = 0.259, \quad D^2 = 11.78, \quad F(2,28) = 15.975$ $p < 0.0000$, where λ is Wilk's statistic, D^2 is the squares of Mahalanobis distances, and F is the Fisher ratio. The statistical analysis showed that exist appropriate discriminatory power for differentiating between the two respective groups. The calculation of percentages of good classification in the data set and external prediction set permitted us to carry out the assessment of the models.

Model 12 classified correctly 97.06% of the compounds in the training data set (100.0% and 95.0% of good classification in active and inactive training data set, respectively), misclassifying only 1 compound of a total of 34. The percentage of false actives in this data set was only 2.94%, that is 1 inactive compound was classified as active from 34 cases. Conversely, no compound from the group of actives was misclassified as inactive ones (0.00% of misclassification). However, compound 13 was consider as not-classified (NC) by this model, because the Eq. 12 give an a posteriori probability of active that does not differ from that of inactive by more than 5%.

The statistical analysis of three models obtained previously using 2D and 3D connectivity and quantum chemical descriptors showed quite similar results. In this case, the overall accuracy of the three models was 91.2%, 94.1%, and 88.2%, respectively.[37]

The classification of all compounds in the complete training data set provides some assessment of the goodness of fit of the model, but it does not provide a thorough criterion of how the model can predict the biological proper-

Table 10. Statistical parameters of the QSPR/QSAR models obtained using different molecular descriptors

Index	n	R^2	s	q^2	s_{CV}	F
Boiling point of 28 alkyl alcohols						
Quadratic indices	2	0.990	2.910	0.989	2.969	1268.9
Local spectral moments[12]	5	0.982	4.2	[a]	[a]	23.8
E-state/encounter parameters[11]	3	0.926	5.8	[a]	[a]	204
Partition coefficient n-octanol/water (log P) of 34 2-furylethylenes						
Quadratic indices	7	0.969	0.142	0.951	0.156	116.76
Vertex and edge conn. indices[37]	7	0.939	0.199	[a]	0.247	56.9
Topographic descriptors[37]	7	0.964	0.155	[a]	0.176	84.6
Quantum chemical descriptors[37]	Used the Rogers and Cammarata approach[34]	0.875	0.319	[a]	0.370	45.5
Reactivity (log k) of 34 2-furylethylenes						
Quadratic indices	7	0.969	0.285	0.922	0.298	115.14
Connectivity indices	7	0.821	0.681	[a]	[a]	17.1
Global spectral moments[12]	7	0.843	0.655	[a]	[a]	18.8
Local spectral moments[12]	7	0.964	0.320	[a]	[a]	70.4
Quantum chemical descriptors[12]	7	0.968	0.288	[a]	[a]	112.2
		λ	D^2	Good class. training	Good class. test	
Classification of 34 2-furylethylene derivatives as antibacterial						
Quadratic indices	5	0.259	11.78	97.06%	100%	15.975
Vertex and edge conn. indices[37]	5	0.43	5.7	91.2%	100%	7.7
Topographic descriptors[37]	5	0.38	6.7	94.1%	100%	9.1
Quantum chemical descriptors[37]	5	0.44	5.2	88.2%	100%	7.1

[a] Values are not reported in the literature.

ties of new compounds. To assess such predictive power, the use of an external test set is essential.[38–40] In this sense, the activity of the compounds in such set was predicted with the obtained discrimination function. The overall accuracy for this group was 100.0%, with one chemical not-classified (NC). Using this same external test set of nine new 2-furylethylenes, the QSAR models obtained with 2D and 3D connectivity and quantum chemical descriptors have also 100.0% of global good classification, including one NC compound.[37] The results of global classification of compounds in both, training and external prediction sets archived with all these approaches are shown in Table 14.

Finally, the improvement in the statistical parameters of our model (Eq. 12) compared to that using 2D and 3D connectivity indices as well as quantum chemical descriptors is easily detected by the decrease in the Wilk's λ parameter and an increase in the Mahalanobis square distance (see Table 10).

4. Conclusion

I have proved that the total and local (atom and atom type) quadratic indices produce the best results for describing the bp and the log P, the log k as well as the antibacterial activity for the 28 alkyl alcohols and 34 furylethylenes studied, respectively. The observation that the quadratic indices approach performs comparably to E-state/biomolecular encounter parameters, spectral moments, 2D and 3D connectivity as well as chemical quantum descriptors for several properties is very interesting. That is, in all examples I have shown that quadratic indices produced somewhat better linear regressions than the other sets of descriptors (see Table 10). This result shows that the molecular quadratic indices are promising total and local-level molecular descriptors for QSPR/QSAR and drug design studies.

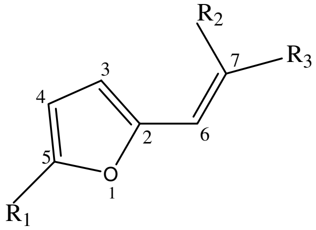
In addition, evidence of significant information is presented in this paper in several ways. The variation of the kth total and local quadratic indices values with alkyl-chain lengthening, branching, heteroatoms-content, and multiple bonds agrees with usual organic intuition. The principal component analysis presented in this work indicated that the information carried by local quadratic indices is markedly different from that codified in various 0D, 1D, 2D, and 3D molecular descriptors presently in QSPR/QSAR and drug design practice. On the contrary, much redundancy and overlapping was found among total quadratic indices and the most of the other structural indices using in this study.

5. Experimental

5.1. Data sets

5.1.1. Chemicals data set for analysis of principal components. The 42 chemicals used for this study were selected from DRAGON's example data[6] and are listed in Table 5. As evident, the sample is relatively small, but very heterogeneous, thus allowing for a general characterization of total and local quadratic indices information independently from individual chemical sets.

Table 11. Chemical structures and numbering of atoms in the 2-furylethylene compounds used in this study



No.	R_1	R_2	R_3
1	H	NO_2	$COOCH_3$
2	CH_3	NO_2	$COOCH_3$
3	Br	NO_2	$COOCH_3$
4	I	NO_2	$COOCH_3$
5	$COOCH_3$	NO_2	$COOCH_3$
6	NO_2	NO_2	$COOCH_3$
7	NO_2	$COOC_2H_5$	$COOC_2H_5$
8	NO_2	H	NO_2
9	H	H	NO_2
10	NO_2	H	$CONH_2$
11	NO_2	H	$CONHCH_3$
12	NO_2	H	$CON(CH_3)_2$
13	NO_2	H	$CONHC_2H_5$
14	NO_2	H	$CONH(CH_2)_2CH_3$
15	NO_2	H	$CONHCH(CH_3)_2$
16	NO_2	H	$CONH(CH_2)_3CH_3$
17	NO_2	H	$CONHCH_2CH(CH_3)_2$
18	NO_2	H	$CONHCH(CH_3)C_2H_5$
19	NO_2	H	$CONHC(CH_3)_3$
20	NO_2	H	$CONHCH_2C(CH_3)_3$
21	NO_2	H	$COOCH_3$
22	NO_2	H	$COOC_2H_5$
23	NO_2	H	$COO(CH_2)_2CH_3$
24	NO_2	H	$COOCH(CH_3)_2$
25	NO_2	H	$COO(CH_2)_3CH_3$
26	NO_2	H	$COOCH_2CH(CH_3)_2$
27	NO_2	H	$COOCH(CH_3)C_2H_5$
28	NO_2	H	$COOC(CH_3)_3$
29	NO_2	H	$COO(CH_2)_4CH_3$
30	NO_2	H	Br
31	NO_2	H	CN
32	NO_2	H	OCH_3
33	NO_2	H	H
34	NO_2	CN	$COOCH_3$
Novel R_1, R_2-substituted 2-furylethylenes ($R_3 = NO_2$) used as *external test set* to assess the predictive power of the classification model for antibacterial activity			
1	Br	Br	NO_2
2	I	I	NO_2
3	Br	H	NO_2
4	H	Br	NO_2
5	I	H	NO_2
6	H	I	NO_2
7	H	CH_3	NO_2
8	Br	CH_3	NO_2
9	I	CH_3	NO_2
	H	I	

5.1.2. Chemicals data sets for QSPR/QSAR studies. In order to illustrate the possibilities of the total and local (atom and atom type) quadratic indices in the QSPR/QSAR studies, I have selected the following two series to be investigated: (1) boiling point of 28 alkyl alcohols (see Table 9) firstly studied by Kier and Hall using E-state/biomolecular encounter parameters[11] and recently by Estrada and Molina[12] using the local spectral moments of the edge adjacency matrix, and (2) a set of 34 2-furylethylene derivatives previously studied using total and local spectral moments, 2D/3D connectivity indices (vertex and edge ones) and to quantum chemical descriptors to model their partition coefficient (log *P*), specific rate constant (log *k*) and antibacterial activity. These chemicals have different substituents at position 5 of the furan ring as well as at the β position of the exocyclic double bond.[36,37] The structures of these 34 furylethylene derivatives are given in Table 11.

The 2-furylethylene compounds have been well known as antimicrobials, antitumoral, and cytotoxic during many years.[41–43] The values of the *n*-octanol/water log *P* and log *k* (for nucleophilic addition of the mercaptoacetic acid) of these compounds have been experimentally determined and reported in the literature.[36] Tables 12 and 13 depict these values, respectively. The antibacterial activity of these compounds was determined as the inverse of the concentration *C* that produces 50% of growth inhibition in *E. coli* at six different times and reported as log(1/*C*).[36] This antibacterial activity was used to classify furylethylenes in two groups by Estrada and Molina.[37] The group of active compounds is composed of those compounds having values of log(1/*C*) < 3, while the rest formed group of inactive compounds. Table 14 illustrates the classification of 2-furylethylene derivatives as antibacterial according to this experimental cutoff value. This table also depicts the antibacterial activity of a series of nine new 2-furylethylenes using by Estrada and Molina[37] like external prediction (test) set. These compounds have a NO_2 group at position R_3 and a Br or I at positions R_1 and/or R_2. All these compounds were shown to have antibacterial activity in different assays.[43,44] The structures of these compounds are given at the bottom of Table 11.

5.2. Computational methods

5.2.1. TOMOCOMD-CARDD approach. TOMOCOMD is an interactive program for molecular design and bioinformatics research.[45] The program is composed by four subprograms, each one of them dealing with drawing structures (drawing mode) and calculating 2D and 3D molecular descriptors (calculation mode). The modules are named CARDD (computed-aided 'rational' drug design), CAMPS (computed-aided modeling in protein science), CANAR (computed-aided nucleic acid research) and CABPD (computed-aided bio-polymers docking). In this paper, I outline salient features concerned with only one of these subprograms: CARDD. This subprogram was developed based on a user-friendly philosophy. That is to say, this computer graphics software shows a great efficiency of interaction with user, without *prior* knowledge of programming skills (e.g., practicing pharmaceutic and organic chemist, teacher, university student, and so on).

The calculation of total and local (atom and atom type) quadratic indices for any organic molecule (or any drug-

Table 12. Experimental and calculated values of the partition coefficient *n*-octanol/water (log *P*) for the furylethylenes studied

No.	Obsd.[a]	Pred.[b]	Res.[c]	Res-CV[d]	Topol.[e]	Topog.[f]	QC[g]
1	1.879	1.971	−0.092	−0.109	1.894	1.955	1.836
2	2.439	2.446	−0.007	−0.008	2.482	2.398	2.239
3	2.739	2.859	−0.120	−0.164	2.753	2.748	2.405
4	2.999	2.566	0.433	0.529	2.905	2.898	2.510
5	1.869	1.823	0.046	0.061	1.763	1.930	1.976
6	1.599	1.597	0.002	0.004	1.619	1.550	1.679
7	2.504	2.772	−0.268	−0.336	2.703	2.640	2.706
8	1.303	1.411	−0.108	−0.144	1.191	1.338	1.456
9	1.583	1.791	−0.208	−0.281	1.453	1.783	1.583
10	0.649	0.645	0.004	0.009	0.433	0.300	0.180
11	0.984	0.965	0.019	0.024	0.999	1.091	1.076
12	0.819	0.724	0.095	0.193	1.160	0.870	2.149
13	1.386	1.388	−0.002	−0.002	1.583	1.423	1.482
14	1.860	1.841	0.019	0.022	2.311	1.941	1.858
15	1.803	1.813	−0.010	−0.011	1.966	2.084	1.906
16	2.356	2.289	0.067	0.076	2.168	2.332	2.240
17	2.225	2.294	−0.069	−0.078	2.493	2.526	2.241
18	2.284	2.266	0.018	0.020	2.384	2.383	2.277
19	2.333	2.240	0.093	0.107	2.316	2.316	2.346
20	2.605	2.748	−0.143	−0.172	2.382	2.575	2.618
21	1.652	1.723	−0.071	−0.082	1.347	1.585	1.830
22	2.098	2.121	−0.023	−0.026	1.984	1.947	2.126
23	2.673	2.573	0.100	0.109	2.733	2.459	2.504
24	2.641	2.521	0.120	0.131	2.484	2.666	2.592
25	2.827	3.021	−0.194	−0.220	2.726	2.837	2.902
26	3.135	3.026	0.109	0.124	3.052	3.034	2.902
27	3.091	2.973	0.118	0.133	3.018	2.952	2.943
28	3.060	2.923	0.137	0.155	2.994	3.002	3.029
29	3.404	3.466	−0.062	−0.077	3.227	3.252	3.266
30	2.447	2.390	0.057	0.080	2.510	2.469	2.132
31	1.050	0.975	0.075	0.130	1.365	1.258	1.344
32	1.591	1.596	−0.005	−0.008	1.510	1.500	1.711
33	1.611	1.688	−0.077	−0.118	1.738	1.515	1.590
34	1.488	1.540	−0.052	−0.128	1.309	1.424	1.504

[a] Experimental values taken of the Ref. 37.

[b] Predicted values using total and local (atom and atom-type quadratic indices (Eq. 10)).

[c] Residual values: log *P*(Obsd) − log *P*(Pred).

[d] Residual values of the LOO cross-validation experiment (deleted residual).

[e] Predicted values using topological indices (vertex and edge connectivity ndices).[37]

[f] Predicted values using topographic descriptors.[37]

[g] Predicted values using quantum chemical molecular descriptors.[37]

like compounds) was implemented in the TOMOCOMD-CARDD software.[45] The main steps for the application of this method in QSAR/QSPR and drug design can be briefly resumed as follows:

1. Draw the molecular pseudographs for each molecule of the data set, using the software drawing mode. This procedure is carried out by a selection of the active atom symbol belonging to different groups of the periodic table.
2. Use appropriated atom weights in order to differentiate the molecular atoms. In this work, I used as atomic property the Pauling electronegativity[22] for each kind of atom.
3. Compute the total and local quadratic indices of the molecular pseudograph's atom adjacency matrix. They can be carried out in the software calculation mode, where you can select the atomic properties and the family descriptor previously to calculate the molecular indices. This software generates a table in which the rows correspond to the compounds and columns correspond to the total and local quadratic indices or any others family molecular descriptors implemented in this program.
4. Find a QSPR/QSAR equation by using mathematical techniques, such as multilinear regression analysis (MRA), neural networks (NN), linear discrimination analysis (LDA), and so on. That is to say, I can find a quantitative relation between a property ***P*** and the quadratic indices having, for instance, the following appearance,

$$\boldsymbol{P} = a_0\boldsymbol{q}_0(x) + a_1\boldsymbol{q}_1(x) + a_2\boldsymbol{q}_2(x) + \cdots + a_k\boldsymbol{q}_k(x) + c \tag{13}$$

   where ***P*** is the measurement of the property, $\boldsymbol{q}_k(x)$ is the *k*th total quadratic indices, and the $\boldsymbol{a}_k$'s are the coefficients obtained by the linear regression analysis.
5. Test the robustness and predictive power of the QSPR/QSAR equation by using internal and external cross-validation techniques.

Table 13. Experimental and calculated values of the specific rate constant for the reaction of nucleophilic addition of thiols (log*k*) to the exocyclic double bond of the studied 2-furylethylenes

No.	Obsd.[a]	Pred.[b]	Res.[c]	Res-CV[d]	TIs[e]	Total moments[f]	QC[g]	Local moments[h]
1	6.591	7.159	−0.568	−0.771	7.532	6.281	6.78	6.679
2	6.518	6.772	−0.254	−0.422	6.983	5.583	6.646	6.563
3	6.914	6.391	0.523	0.717	6.352	6.893	7.021	6.979
4	6.982	6.640	0.342	0.446	6.997	6.554	6.935	7.069
5	7.176	6.658	0.518	0.719	6.75	7.39	7.137	7.291
6	7.602	8.017	−0.415	−0.880	6.685	7.41	7.466	7.572
7	5.255	5.444	−0.189	−0.257	5.533	5.002	5.212	4.934
8	6.763	6.385	0.378	0.732	4.965	5.962	6.737	6.662
9	5.623	5.576	0.047	0.079	4.997	4.839	5.454	5.599
10	3.813	3.929	−0.116	−0.128	4.831	4.455	3.695	4.318
11	3.840	3.847	−0.007	−0.008	3.636	4.876	3.616	4.034
12	3.874	3.876	−0.002	−0.003	4.162	2.911	3.537	3.448
13	3.825	3.679	0.146	0.159	3.916	4.255	3.562	3.875
14	3.623	3.542	0.081	0.087	3.798	4.032	3.55	3.841
15	3.751	3.612	0.139	0.149	3.528	3.255	3.522	3.721
16	3.784	3.398	0.386	0.421	3.114	3.437	3.566	3.842
17	3.697	3.505	0.192	0.213	4.175	3.468	3.57	3.826
18	3.705	3.486	0.219	0.241	3.934	3.157	3.541	3.711
19	3.697	3.650	0.047	0.054	3.264	3.686	3.482	3.798
20	3.650	3.569	0.081	0.107	3.526	4.269	3.581	3.851
21	4.000	4.199	−0.199	−0.228	3.693	5.184	3.985	3.735
22	3.920	4.007	−0.087	−0.097	3.974	4.202	3.92	3.617
23	3.790	3.874	−0.084	−0.091	3.856	4.409	3.926	3.596
24	3.763	3.917	−0.154	−0.166	3.869	3.438	3.849	3.552
25	3.623	3.731	−0.108	−0.117	3.172	3.893	3.921	3.6
26	3.650	3.841	−0.191	−0.207	4.233	4.178	3.917	3.61
27	3.592	3.795	−0.203	−0.221	4.193	3.613	3.873	3.566
28	3.584	3.933	−0.349	−0.383	4.059	3.596	3.819	3.881
29	3.590	3.585	0.005	0.005	3.478	3.204	3.918	3.6
30	2.987	3.048	−0.061	−0.103	3.463	3.534	3.281	2.785
31	3.273	3.383	−0.110	−0.207	3.07	3.779	4.089	3.651
32	2.140	2.380	−0.240	−0.984	3.07	2.794	2.287	2.751
33	3.553	3.359	0.194	0.373	3.78	2.716	3.338	3.485
34	5.557	5.520	0.037	0.128	5.114	5.474	4.969	5.524

[a] Experimental values taken of the Ref. 12.
[b] Predicted values using total and local (atom and atom-type quadratic indices (Eq. 11)).
[c] Residual values: log*k*(Obsd) − log*k*(Pred).
[d] Residual values of the LOO cross-validation experiment (deleted residual).
[e] Predicted values using topological indices (connectivity indices).[12]
[f] Predicted values using total spectral moments.[12]
[g] Predicted values using quantum chemical molecular descriptors.[12]
[h] Predicted values using local spectral moments.[12]

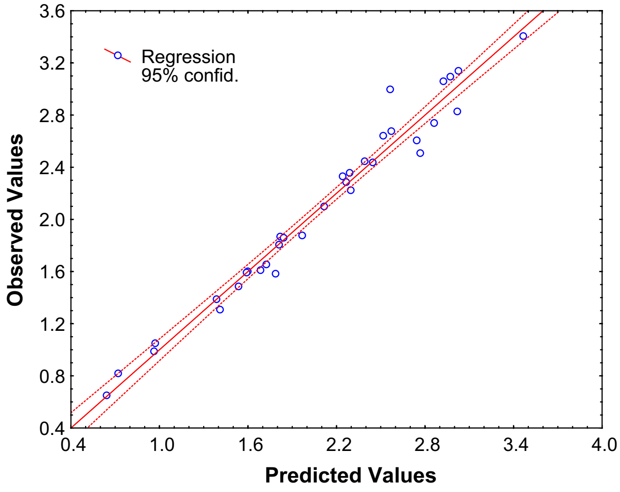


Figure 2. Linear correlations of observed versus calculated log*P* according to the model obtained from molecular quadratic indices.

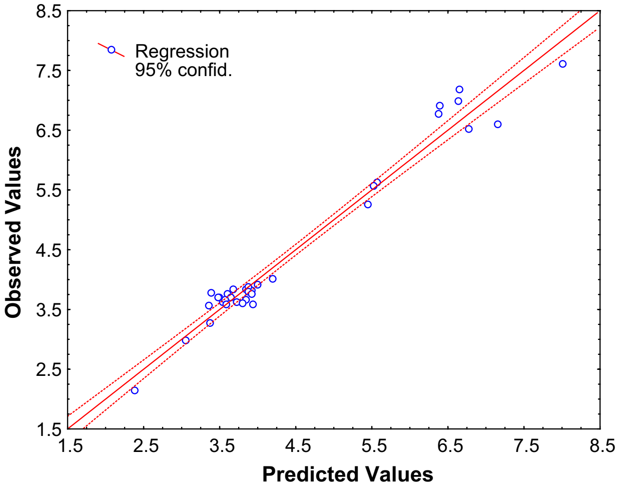


Figure 3. Observed versus predicted log*k* (Eq. 11) for the reaction of nucleophilic addition of thiols to the exocyclic double bond of the 2-furylethylene derivatives.

Table 14. Classification of 2-furylethylene derivatives as antibacterial according to the four models obtained with molecular quadratic indices, 2D and 3D connectivity as well as quantum chemical descriptors

Compd.	Obsd.[37]	Quadratic indices		2D conn.[37]		3D conn.[37]		Quantum[37]	
		Class.	Prob.	Class.	Prob.	Class.	Prob.	Class.	Prob.
1	+	+	99.99	+	95.43	+	99.49	+	99.72
2	+	+	99.02	+	91.67	+	95.83	+	99.86
3	+	+	100.00	+	84.95	+	96.22	+	98.31
4	+	+	100.00	+	79.65	+	95.78	+	97.67
5	+	+	99.93	+	99.72	+	99.63	+	99.66
6	+	+	100.00	+	99.85	+	99.98	+	98.91
7	+	+	99.99	+	94.29	+	91.77	+	98.57
8	+	+	99.87	+	74.81	+	57.81	+	92.50
9	+	+	94.32	–	9.86	–	8.11	+	77.08
10	+	+	81.05	+	99.13	+	99.28	–	32.29
11	+	+	98.30	+	88.24	+	57.28	–	9.46
12	+	+	99.80	+	66.00	+	86.94	–	4.26
13	+	NC	48.89	+	57.89	+	71.56	–	12.79
14	–	–	23.16	–	6.25	–	46.43	–	12.96
15	–	–	1.22	–	28.14	–	36.15	–	8.72
16	–	–	1.18	–	0.92	–	1.10	–	11.78
17	–	–	8.53	–	2.35	–	6.19	–	11.05
18	–	–	0.38	–	37.62	–	4.56	–	9.96
19	–	–	0.01	–	8.96	–	2.97	–	9.96
20	–	–	2.75	–	1.14	–	0.08	–	9.75
21	–	+	70.78	+	55.73	+	88.95	–	8.59
22	–	–	6.12	–	22.77	–	18.96	–	7.60
23	–	–	2.70	–	1.36	–	8.23	–	7.66
24	–	–	0.13	–	7.01	–	4.83	–	6.44
25	–	–	0.11	–	0.19	–	0.09	–	8.56
26	–	–	1.14	–	0.33	–	0.58	–	8.49
27	–	–	0.06	–	7.59	–	0.43	–	7.94
28	–	–	0.00	–	1.41	–	0.27	–	7.02
29	–	–	0.01	–	0.02	–	0.04	–	7.21
30	–	–	19.48	–	4.65	–	7.56	–	0.32
31	–	–	2.49	–	29.58	–	37.61	–	3.04
32	–	–	2.75	–	23.67	–	14.96	–	5.30
33	–	–	2.85	+	58.87	–	14.08	–	0.42
34	+	+	97.88	+	97.13	+	97.50	+	62.36
Test set									
1	+	+	96.53	+	88.53	+	95.81	+	87.18
2	+	+	96.66	+	86.87	+	94.59	+	85.53
3	+	+	97.03	+	59.01	+	65.00	+	54.95
4	+	+	93.82	+	96.35	+	99.59	+	96.12
5	+	+	97.16	NC	50.07	+	53.00	NC	50.01
6	+	+	93.58	+	96.72	+	99.51	+	97.68
7	+	NC	48.09	+	95.92	+	99.62	+	94.76
8	+	+	62.39	+	86.38	+	96.06	+	84.27
9	+	+	63.69	+	81.75	+	95.74	+	79.29

6. Develop a structural interpretation of obtained QSAR/QSPR model using total and local (atom and atom type) quadratic indices as molecular descriptors.

5.2.2. DRAGON software. DRAGON[6] is a very sophisticated program for the calculation of molecular descriptors. Among them, 0D (D = 'dimensionality') or constitutional, 1D (e.g., empirical descriptors and molecular properties), 2D (such as: 2D autocorrelations, topological indices, BCUT descriptors, Galvez topological charges indices, molecular walk counts) and 3D (aromaticity indices, Randic molecular profiles, charge-, geometrical-, RDF-, 3D-MoRSE-, GATEWAY-, and WHIM-descriptors) molecular descriptors. In order to compare the information carried by the total and local quadratic indices with 0D–3D QSPR/QSAR descriptors, all these structural indices were calculated by selecting all possible descriptors options from the 'description selection' DRAGON menu.[6] The calculation was based on the set of 42 chemicals (Table 5) depicted in DRAGON's example data. However, I selected 'pair correlation checkbox' in order to exclude from output file the descriptors containing redundant information (correlation coefficient equal or greater than 0.95). In addition, constant descriptors were also eliminated. Subsequently, only the molecular indices 'unweighted' or using Sanderson electronegativities like atomic weighs were saved. This procedure takes into account the fact of that total and local quadratic indices were computed selecting only one atomic property (Pauling electronegativity[22]) in the TOMOCOMD-CARDD's properties menu,[45] and in order to reduce the descriptor's number for the later chemometric analysis. Finally, to compare the total and local quadratic indices to other QSPR/QSAR descriptors, I have obtained a total of 229 (0D–3D) molecular descriptors. The symbols and definitions of some of these molecular indices are given in Table 6 (also see Supplementary data).

5.3. Chemometric methods

5.3.1. Analysis of principal components. To conduct the comparison of the molecular descriptors computed in this work, I performed a factor analysis by using the principal components method. The theoretical aspects of this statistical technique have been extensively exposed in the literature including many chemical applications.[46–52] The main applications of factor analytic techniques are (1) to *reduce* the number of variables, and (2) to *detect structure* in the relationships between variables, that is to *classify* variables.[53] In this approach, factor loadings are obtained from original variables. Thus, these factors capture most of the 'essence' of these molecular descriptors because they are a linear combination of the original items. Factor loadings from the principal component analysis are shown in Table 8 (also see Supplementary data). The results of the factor analysis performed here are summarized in Table 7.

The factor analysis was carried out using 'varimax normalized' as rotational strategy to obtain the factor loadings from the principal component analysis. This strategy makes the structure of factors pattern as simple as possible, permitting a clearer interpretation of the factors without loss of orthogonality between them.[46–52] Finally, some of the most important conclusions that can be drawn from a factor analysis that will be of large usefulness in the present paper are the following:[46–52] (1) variables with a high loading in the same factor are interrelated and will be the more so the higher the loadings, (2) no correlation exists between variables having non-zero loadings only in different factors. These are the principal ideas that permit the interpreting the *factor structure* obtained using the factor analysis as a classification method, and (3) only variables with high loadings in different factors may be combined in a regression equation in order to eliminate collinearities.

5.3.2. Generation of QSPR and classification models. In describing bp, $\log P$, and $\log k$ the multiple linear regression analysis was used as statistical method. This experiment was performed with STATISTICA software package.[53] The tolerance parameter (proportion of variance that is unique to the respective variable) used was the default value for minimum acceptable tolerance, which is 0.01. Forward stepwise was fixed as the strategy for variable selection. The principle of parsimony (Occam's razor) was taken into account as strategy for model selection. In this connection, we select the model with higher statistical signification but having as few parameters (a_k) as possible. The quality of the models was determined examining the regression's statistic parameters and of the cross-validation procedures.[38–40] In this sense, the quality of models was determined by examining the regression coefficients (R), determination coefficients or squared regression coefficient (R^2), Fisher ratio's p level [$p(F)$], standard deviations of the regression (s) and the leave-*one*-out (LOO) press statistics (q^2, s_{CV}).

On the other hand, linear discriminant analysis (LDA) was used to the classification of 34 2-furylethylene derivatives as antibacterial. This statistical analysis was performed using also STATISTICA software.[53] The use of LDA in 'rational' drug design has been extensively used by different authors.[19–21,25,27,32,52,54–56]

In order to test the quality of the discriminant function derived, I used Wilks' λ (U-statistic) and the Mahalanobis distance (D^2). Wilks' λ statistical is helpful to value the total discrimination and can take values between 0 (perfect discrimination) and 1 (no discrimination). The D^2 indicates the separation of the respective groups. The statistical robustness and predictive power of the obtained model was assessed using an external prediction (test) set. In developing classification models the values of 1 and -1 were assigned to active and inactive compounds, respectively. To make the classification of compounds in both groups we preferred the use of the a posteriori probabilities instead of cutoff values. This is the probability that the respective case belongs to a particular group (active or inactive) and it is proportional to the Mahalanobis distance from that group centroid. In closing, the posterior probability is the probability, based on our knowledge of the values of others variables, that the respective case belongs to a particular group. When the probability of being active does not differ more than 5% from the probability of being inactive, the case is considered as not classified (NC). An external test set of nine new compounds was used in order to assess the predictive ability of the obtained LDA model.

Supplementary data

The complete list of the 229 0D–3D DRAGON's molecular descriptors[6] used in this study, as well as their description and loadings (from factor analysis) is available free of charge. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2004.09.034.

References and notes

1. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Germany, 2000.
2. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; John Wiley & Sons: New York, 2000.
3. *QSPR/QSAR Studies by Molecular Descriptors*; Diudea, M. V., Ed.; Nova Science: Huntington, NY, 2001.

4. *From Chemical Graphs to Three-Dimensional Geometry*; Balaban, A. T., Ed.; Plenum: New York, 1997.
5. Balaban, A. *SAR QSAR Environ. Res.* **1998**, *8*, 1.
6. Todeschini, R.; Consonni, V.; Pavan, M. Dragon. Software version 2.1. 2002. http://www.disat.unimib.it/chm/Dragon.htm.
7. Estrada, E.; Uriarte, E. *Curr. Med. Chem.* **2001**, *8*, 1699.
8. Devillers, J.; Balaban, A. T. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon and Breach: Amsterdam, The Netherlands, 1999.
9. Katritzky, A. R.; Gordeeva, E. V. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835.
10. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: London, 1976.
11. Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic: New York, 1999.
12. Estrada, E.; Molina, E. *J. Mol. Graphics Mod.* **2001**, *20*, 54.
13. Kier, L. B.; Hall, L. H. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, 1999; pp 491–562.
14. Kier, L. B.; Hall, L. H. *Pharm. Res.* **1990**, *7*, 801.
15. Boulamwini, J. K.; Raghavan, K.; Fesen, M. R.; Pommier, Y.; Kohn, K. W.; Weinstein, J. N. *Pharm. Res.* **1996**, *13*, 1892.
16. Marrero-Ponce, Y. *Molecules* **2003**, *8*, 687.
17. Marrero-Ponce, Y. *J. Chem. Inf. Comput. Sci.*, in press.
18. Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; Ofori, E.; Montero, L. A. *Int. J. Mol. Sci.* **2003**, *4*, 512.
19. Marrero-Ponce, Y.; Cabrera, M. A.; Romero, V.; González, D. H.; Torrens, F. A. *J. Pharm. Pharm. Sci.* **2004**, *7*, 186.
20. Marrero-Ponce, Y.; Nodarse, D.; González-Díaz, H.; Ramos de Armas, R.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. *CPS: physchem/0401004*.
21. Marrero-Ponce, Y.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2004**, *12*, 5331.
22. Pauling, L. *The Nature of Chemical Bond*; Cornell University Press: New York, 1939.
23. Walker, P. D.; Mezey, P. G. *J. Am. Chem. Soc.* **1993**, *115*, 12423.
24. Eliel, E.; Wilen, S.; Mander, L. *Stereochemistry of Organic Compounds*; John Wiley & Sons: New York, 1994.
25. de Julián-Ortiz, J. V.; Alapont, C. de G.; Ríos-Santamarina, I.; García-Doménech, R.; Gálvez, J. *J. Mol. Graphics Mod.* **1998**, *16*, 14.
26. Golbraikh, A.; Bonchev, D.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 147.
27. González-Díaz, H.; Hernández-Sánchez, I.; Uriarte, E.; Santana, L. *Comput. Biol. Chem.* **2003**, *27*, 217.
28. Klein, D. J. *Internet Electron. J. Mol. Des.* **2003**, *2*, 814.
29. Ramos de, A. R.; González Díaz, H.; Molina, R.; González, M. P.; Uriarte, E. *Bioorg. Med. Chem.* **2004**, *12*, 4815.
30. González-Díaz, H.; Olazábal, E.; Castañedo, N.; Hernádez, S. I.; Morales, A.; Serrano, H. S.; González, J.; Ramos de, A. R. *J. Mol. Mod.* **2002**, *8*, 237.
31. González-Díaz, H.; Gia, O.; Uriarte, E.; Hernádez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J. A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L. A.; Cabrera, M. A. *J. Mol. Mod.* **2003**, *9*, 395.
32. González-Díaz, H.; Bastida, I.; Castañedo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H. S.; Ramos de, A. R. *Bull. Math. Biol.* **2004**, *66*, 1285.
33. Randic, M. *J. Math. Chem.* **1991**, *7*, 155.
34. Malinowski, E. R.; Howery, D. G. *Factor Analysis in Chemistry*; Wiley-Interscience: New York, 1980.
35. Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam, 1984; pp 188–197.
36. Balaz, S.; Sturdik, E.; Rosenberg, M.; Augustin, J.; Skara, B. *J. Theor. Biol.* **1988**, *131*, 115.
37. Estrada, E.; Molina, E. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 791.
38. Wold, S.; Erikson, L. Statistical Validation of QSAR Results. Validation Tools. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: New York, 1995; pp 309–318.
39. Golbraikh, A.; Tropsha, A. *J. Mol. Graphics Mod.* **2002**, *20*, 269.
40. Belsey, D. A.; Kuh, E.; Welsch, R. E. *Regression Diagnostics*; Wiley: New York, 1980.
41. Dore, J. Ch.; Viel, C. *Farmaco* **1975**, *30*, 81.
42. Sturdik, E.; Drobnica, L.; Balaz, S. *Coll. Czch. Chem. Commun.* **1985**, *50*, 470.
43. Blondeau, J. M.; Castañedo, N.; Gonzalez, O.; Medina, R.; Silveira, E. *Antimicrob. Agents Chemother.* **1999**, *11*, 16663.
44. Castañedo, N.; Goizueta, R.; Perez, J.; Gonzalez, J.; Silveira, E. Cuesta, M.; Martinez, A.; Lugo, E.; Estrada, E.; Carta, A.; Navia, O.; Delgado, M. Cuban patent 22446, 1994; European patent, application number 955000567; Canadian patent, application number 2,147,594; Japan patent, application number 222002; U.S. patent, application number 60008011.
45. Marrero-Ponce, Y.; Romero, V. TOMOCOMD software. Central University of Las Villas. 2002. TOMOCOMD (TOpological MOlecular COMputer Design) for Windows, version 1.0 is a preliminary experimental version; in future a professional version will be obtained upon request to Y. Marrero: yovanimp@qf.uclv.edu.cu; ymarrero77@yahoo.es.
46. Cramer, R. D., III. *J. Am. Chem. Soc.* **1980**, *102*, 1849.
47. Cramer, R. D., III. *J. Am. Chem. Soc.* **1980**, *102*, 1837.
48. Needham, D. E.; Wei, I. C.; Seybold, P. G. *J. Am. Chem. Soc.* **1988**, *110*, 4186.
49. Estrada, E.; Gonzáles, H. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 75.
50. Estrada, E.; Rodríguez, L. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1037.
51. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1042.
52. González-Díaz, H.; Marrero-Ponce, Y.; Hernández, I.; Bastida, I.; Tenorio, E.; Nasco, O.; Uriarte, U.; Castañedo, N.; Cabrera, M. A.; Aguila, E.; Marrero, O.; Morales, A.; Pérez, M. *Chem. Res. Toxicol.* **2003**, *16*, 1318.
53. STATISTICA ver. 5.5, Statsoft, 1999.
54. Estrada, E.; Peña, A.; García-Domenech, R. *J. Comput.-Aided Mol. Design.* **1998**, *12*, 583.
55. Estrada, E.; Peña, A. *Bioorg. Med. Chem.* **2000**, *8*, 2755.
56. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. *J. Med. Chem.* **2000**, *43*, 1975.