We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Molecular Interactions in Chromatographic Retention: A Tool for QSRR/QSPR/QSAR Studies

Vilma Edite Fonseca Heinzen[1*], Berenice da Silva Junkes[2],
Carlos Alberto Kuhnen[3] and Rosendo Augusto Yunes[1]
*[1]Department of Chemistry, Federal University of Santa Catarina,
University Campus, Trindade, Florianópolis, Santa Catarina,
[2]Federal Institute of Education, Science and Technology of Santa Catarina,
Mauro Ramos Avenue, No. 950, Center, Florianópolis, SC,
[3]Department of Physics, Federal University of Santa Catarina, University Campus,
Trindade, Florianópolis, Santa Catarina,
Brazil*

1. Introduction

Molecular interactions play a fundamental role in the behavior of the chemical and physical properties of any physico-chemical system. In gas chromatography (GC) the chromatographic retention is a very complex process. It involves the interaction of molecules through multiple intermolecular forces, such as dispersion (or London forces), orientation (dipole–dipole or Keesom forces), induction (dipole–induced dipole or Debye forces), and electron donor–acceptor forces including hydrogen–bonding, leading to the partition of the solute between the gas and liquid phases (Kaliszan, 1987; Peng, 2000; Yao et al., 2002). Other factors, such as steric hindrance of substituent groups within the solute molecule, can also affect the chromatographic behavior (Fritz et al., 1979; Peng et al., 1988). It is clear that correlations between gas chromatographic retention indices (RIs) and molecular parameters provide significant information on the molecular structure, retention time and the possible mechanism of absorption and elution (Körtvélyesi et al., 2001). Several topological, geometric, electronic, and quantum chemical descriptors have been used in research on quantitative structure–property and structure–activity relationships (QSPR/QSAR) (Karelson et al., 1996; Katritzky & Gordeeva, 1993; Kier & Hall, 1990). The topological descriptors have shown their efficacy in the prediction of diverse physicochemical and biological properties of various types of compounds (Amboni et al., 2000; Arruda et al., 1993; Estrada, 2001a, 2001b; Heinzen & Yunes, 1993, 1996; Heinzen et al., 1999a; Kier & Hall, 1986; Randic, 1993, 2001; Ren, 2002a, 2002b, 2002c; Sabljic & Trinajstic, 1981). In general, these indices are numbers containing relevant information regarding the structure of molecules. Most of the measured physicochemical properties are steric properties, and therefore they may be reasonably well described by topological indices. However, in some cases, these indices also contain structural information related to the

electronic and/or polar features of molecules (Galvez et al., 1994; Hall et al., 1991). The molecular size, shape, polarity, and ability to participate in hydrogen bonding are among the different factors that can contribute to the physicochemical properties or biological activities of a molecule. It is well known that these factors are related to intermolecular interactions such as van der Waals forces.

The use of graph–theoretical topological indices in QSPR/QSAR studies has sparked great interest in recent years. The topological indices have become a powerful tool for predicting numerous physicochemical properties and/or biological activities of compounds, as well as for molecular design. One of the most important properties that have been extensively studied is the chromatographic retention (Estrada & Gutierrez, 1999; Ivanciuc, O. et al., 2000; Ivanciuc T. & Ivanciuc, O., 2002; Katritzky et al., 1994; Katritzky et al., 2000; Pompe & Novic, 1999; Ren, 1999, 2002a). Quantitative structure–chromatographic retention relationship (QSRR) studies have been widely investigated by gas chromatography (GC) and high–performance liquid chromatography (HPLC) (Markuszenwski &. Kaliszan, 2002). Topological indices (TI) are obtained *via* mathematical operations from the corresponding molecular graphs of compounds (Ivanciuc, O. et al., 2002; Kier & Hall, 1976; Liu, S.-S. et al. 2002; Marino et al., 2002; Rios–Santamarina et al., 2002; Toropov & Toropova, 2002) in contrast to the physicochemical characterization used by traditional QSAR (García–Domenech et al., 2002). One of the main advantages of TI is that they can be easily and rapidly computed for any constitutional formula yielding good correlation abilities. However, important disadvantages should be noted including the difficulties encountered in encoding stereo–chemical information, for example, to distinguish between *cis*– and *trans*–isomers, and their lack of physical meaning. Many topological indices have been proposed since the pioneering studies by Wiener (Wiener, 1947) and by Kier on the use of QSAR (Kier & Hall, 1976). The TI developed for QSAR/QSRR studies can be illustrated by Estrada's approach to edge weights using quantum chemical parameters (Estrada, 2002) and by Ren's atom–type AI topological indices derived from the topological distance sums and vertex degree (Ren, 2002d).

Based on a chromatographic behavior hypothesis, our group developed a topological index called the semi–empirical topological index (I_{ET}). This index was initially developed to predict the chromatographic retention of linear and branched alkanes and linear alkenes, with the objective of differentiating their *cis*– and *trans*– isomers and obtaining QSRR models (Heinzen et al., 1999b). The excellent results achieved stimulated our group to extend the new topological descriptor to other classes of compounds (Amboni et al., 2002a, 2002b; Arruda et al., 2008; Junkes et al., 2002a, 2002b, 2003a, 2003b, 2004; Junkes et al., 2005; Porto et al., 2008). The equation obtained to calculate the I_{ET} was generated from the molecular graph and the values of the carbon atoms, and the functional groups were attributed observing the experimental chromatographic behavior and supported by theoretical considerations. This was carried out due to the difficulty in obtaining a complete theoretical description of the interaction between the stationary phase and the solute. Based only on theoretical equations or hypotheses it is not possible, for example, to estimate how the molecular conformation of the solute affects the intermolecular forces. In view of this, it seems reasonable to assume that from the experimental behavior we can obtain insights regarding these factors in order to apply them to other processes involved in QSPR studies. Thus, it can be noted that the semi–empirical topological index has a clear physical meaning.

The semi–empirical topological index (I_{ET}) allowed the creation of a new descriptor, the electrotopological index, I_{SET}, which was recently developed by our group and applied to QSPR studies to predict the chromatographic retention index for a large number of organic compounds, including aliphatic hydrocarbons, alkanes and alkenes, aldehydes, ketones, esters and alcohols (Souza et al., 2008, 2009a, 2009b, 2010). The new descriptor for the above series of molecules can be quickly calculated from the semi-empirical, quantum-chemical, AM1 method and correlated with the approximate numerical values attributed by the semi-empirical topological index to the primary, secondary, tertiary and quaternary carbon atoms. Thus, unifying the quantum-chemical with the topological method provided a three-dimensional picture of the atoms in the molecule. It is important to note that the AM1 method portrays more reliable semi-empirical charges, dipoles and bond lengths than those obtained from time-consuming, low-quality, *ab initio* methods, that is, when employing a minimal basis set in *ab initio* calculations. Despite the fact that the calculated partial atomic charges may be less reliable than other molecular properties, and that different semi-empirical methods give values for the net charges with poor numerical agreement, it is important to recognize that their calculation is easy and that the values at least indicate the trends of the charge density distributions in the molecules. Since many chemical reactions or physico-chemical properties are strongly dependent on local electron densities, net atomic charges and other charge-based descriptors are currently used as chemical reactivity indices.

For alkanes and alkenes, this correlation allowed the creation of a new semi-empirical electrotopological index (I_{SET}) for QSRR models based on the fact that the interactions between the solute and the stationary phase are due to electrostatic and dispersive forces. This new index, I_{SET}, is able to distinguish between the *cis-* and *trans-*isomers directly from the values for the net atomic charges of the carbon atoms that are obtained from quantum-chemical calculations (Souza et al., 2008). For polar molecules like aldehydes, ketones, esters and alcohols, the presence of heteroatoms like oxygen changes considerably the charge distribution of the corresponding hydrocarbons, leading to a small increase in the interactions between the solute and the stationary phase (Souza et al., 2009a, 2009b, 2010). An appropriate way to calculate the I_{SET} was developed, taking into account the dipole moment exhibited by these molecules and the atomic charges of the heteroatoms and the carbon atoms attached to them. By considering the stationary phase as non-polar material the interaction between these molecules and the stationary phase becomes electrostatic with the contribution of dispersive forces. These interactions were slowly increased relative to the corresponding hydrocarbons. Hence, the interactions between the molecules and the stationary phase were slowly increased as a result of the charge redistribution that occurred in presence of the heteroatom. This charge redistribution accounted for the dipole moment of the molecules. Clearly the main outcomes in terms of the charge distribution due the presence of the (oxygen) heteroatoms occur in the neighborhood, and the excess charge of these atoms leads to electrostatic interactions that are stronger relative to the weak dispersive dipolar interactions.

2. Semi-empirical topological index (I_{ET})

Three important factors led us to develop the semi-empiric topological index: (i) no topological index alone was able to differentiate between the *cis-* and *trans-* isomeric structures of alkenes; (ii) if all the carbon atoms have a value of 100 as indicated by Kovàts,

from the experimental results it is not possible to determine a constant value for each of the different carbon atoms (secondary, tertiary and quaternary) of alkanes; (iii) when the Kovàts indices of retention for very branched hydrocarbons (alkanes) are correlated with the number of carbon atoms an unacceptable linearization is observed. It is known that the chromatographic process of separation results from the forces that operate between solute molecules and the molecules of the stationary phase. The retention of alkanes and alkenes is due to the number of carbon atoms and the interaction of each specific carbon atom with the stationary phase. The interaction of the stationary phase with the carbon atoms is determined by its electrical properties and by the steric hindrance to this interaction by other carbon atoms attached to it. The values attributed to the carbon atoms were based on the results of the experimental chromatographic behavior of the molecules that measure the real electrical and steric characteristics of the carbons. For this reason the index is denominated semi-empirical. The representation of the molecules was based on the molecular graph theory, where the carbon atoms are considered as the vertexes of the graph and the hydrogens are suppressed (Hansen & Jurs, 1988). Thus, it is called a topological index.

2.1 Calculation of I_{ET} for alkanes and alkenes

Values were attributed to the carbon atoms (vertex of the molecular graphs) according to the following considerations. (i) According to the Kovàts convention, the correlation between the retention index and number of carbon atoms is linear for the alkanes (Kovàts, 1968). However, branched alkanes do not present this linear relationship with the Kovàts index, since the retention of the tertiary and quaternary carbon atoms is decreased by the steric effects of their neighboring groups. It is evident that secondary, tertiary and quaternary carbon atoms have values of less than 100 u.i., as previously attributed by Kovàts. (ii) Observing the experimental chromatographic behavior, approximate numerical values were attributed: 100 u.i. for the carbon atom in the methyl group in agreement with Kovàts, 90 u.i. for the secondary carbon atoms, 80 u.i. for the tertiary and 70 u.i. for the quaternary. All values were divided by 100 to make them consistent with the common topological values. (iii) The contribution of these carbon atoms to the chromatographic retention is also dependent on the neighboring substituent groups due to steric effects. In order to estimate the steric effects, it was observed that the values for the experimental RI decreased as the branch increased, showing a log trend. Therefore, it was necessary to add the value of the logarithm of each adjacent carbon atom. Thus, the new semi-empirical topological index (I_{ET}) is expressed as:

$$I_{ET} = \sum_i (C_i + \delta_i) \tag{1}$$

$$\delta_i = \sum_{j \sim i} log C_j$$

where C_i is the value attributed to each carbon atom in the molecule and δ_i is the sum of the logarithm of the value for each adjacent carbon atom (C_1, C_2, C_3 and C_4) and ~ means 'adjacent to'. (iv) For alkenes, the main interaction force between the solute and stationary phase is the dispersive force, which is reduced by neighboring steric effects, however, the electrostatic force is also involved. The influence of conformational effects on the intermolecular forces makes it

very difficult to predict these effects based only on theoretical considerations. For this reason, the values attributed to the carbon atom of the double bond for alkenes were calculated by numerical approximation based on the experimental retention indices, as described in our previous publication (Heinzen et al., 1999b; Junkes et al., 2002a).

2.2 Calculation of I_{ET} for compounds with oxygen-containing functional groups

The values attributed to the carbon atoms and functional groups (vertex of the molecular graphs) were based on the following considerations: (i) For this group of compounds, the main intermolecular forces that contribute to their chromatographic behavior on low polarity stationary phases are dispersive and inductive forces. The values attributed to functional groups are also based on the experimental retention index. (ii) The –COO- (ester), C=O (ketone or aldehyde) and C-OH (alcohol) groups were considered as a single vertex of the molecular graph of the compounds studied. This was carried out due to the difficulty and the inconsistency associated with calculating the individual values of the carbon atoms and the oxygen atoms of these groups. Thus, better numerical approximations were obtained, capable of reflecting the experimental chromatographic behavior of these compounds, when these groups were treated as a single vertex. (iii) The same considerations that were taken into account during the development of the semi-empirical topological method for the prediction of retention indices of alkanes and alkenes (Heinzen et al., 1999b; Junkes et al., 2002a) were employed to develop the I_{ET} for oxo-compounds. (iv) The contribution of the carbon atoms and functional groups to the chromatographic retention was represented by a single symbol, C_i, as indicated in Equation 1. The semi-empirical topological index can be expressed by a general Equation, for the entire set of compounds included in this work, where: C_i = value attributed to the –COO- (ester), C=O (ketone or aldehyde), C–OH (alcohol) groups and/or to each carbon atom, i, in the molecule. δ_i = the sum of the logarithm of the values of each adjacent carbon atom (C_1, C_2, C_3, and C_4) and/or the logarithm of the value of the –COO- (ester), C=O (ketone or aldehyde), C-OH (alcohol) groups, and ~ means 'adjacent to'. In a first step, an approximate I_{ET} (I_{Eta}) was calculated for each compound. This was achieved using the equation previously obtained for linear alkanes containing from 3 to 10 carbon atoms and Kovàts experimental retention indices of compounds (Heinzen et al., 1999b). (v) Subsequently, the values of C_i for primary and secondary carbon atoms, previously attributed to alkanes (Heinzen et al., 1999b), and the approximate I_{ET}, calculated above, were used in Equation 1 in order to calculate the values of –COO-, C=O and C-OH groups of linear compounds. Thus, values were attributed to each class of functional group according to the position of the group in the carbon chain. (vi) One of the fundamental factors taken into consideration for the development of this topological index was the importance of the steric and other mutual intramolecular interactions between the functional group and nearby atoms. Therefore, for branched molecules, different values were attributed to carbon atoms in the α, β, and γ position with respect to the functional groups compared to those previously attributed to alkanes (Heinzen et al., 1999b) as described in the literature (Amboni et al., 2002a, 2002b; Junkes et al., 2003b, 2004).

The values of C_i for the carbon atoms and the values attributed to the functional groups of esters, aldehydes, ketones and alcohols are listed in Table 1 of Junkes et al. (Junkes et al., 2004,).

2.3 Calculation of I_{ET} for alkylbenzene compounds

The same considerations employed in the generation of the semi-empirical topological index, I_{ET}, for linear and branched alkanes and alkenes (Heinzen et al., 1999b; Junkes et al., 2002a) were applied to this group of compounds (alkylbenzenes). Firstly, the molecules were represented by hydrogen-suppressed molecular graphs based on chemical graph theory (Hansen & Jurs, 1988) where the carbon atoms were considered as vertexes of the molecular graph of these compounds. The contribution of each carbon atom to the chromatographic retention is represented by a single symbol, C_i, as can be observed from Eq. (1) where Ci is the value attributed to (=C<) fragments and/or each carbon atom i in the molecule; and δ_i is the sum of the logarithm of the values for each adjacent carbon atom (C_1, C_2, C_3 and C_4). The values of Ci for the carbon atoms of linear, branched, *ortho*, *meta* and *para* substituted, tri-substituted and tetra-substituted alkyl benzenes can be seen in Porto et al. (Porto et al., 2008).

2.4 Calculation of I_{ET} for halogenated aliphatic compounds

The present approach is based on the representation of molecules by hydrogen-suppressed molecular graphs which, in turn, are based on chemical graph theory, where the carbon atoms (Ci) are the graph vertexes. As with the carbon atoms the C–X and X–C–X fragments (where X = chlorine, bromine, or iodine atom) are considered a vertex of the molecular graph of these compounds, as previously considered for the functional groups (Heinzen et al., 1999b). The I_{ET} is expressed as equation (1) where Ci is the value attributed to each carbon atom i and/or to C–X or X–C–X fragments in the molecule; and δi is the sum of the logarithm of the values for each adjacent carbon atom (C1, C2, C3, and C4) and/or the logarithm of the values of the adjacent C–X and X–C–X fragment. The values to be attributed to the carbon atoms, and to the functional group (Ci) for halogenated hydrocarbons, are calculated by numerical approximation based on the experimental retention index (RIExp) values and supported by theoretical considerations. The values of Ci for the carbon atoms of linear and branched halogenated aliphatic compounds can be obtained in Arruda et al. (Arruda et al., 2008).

2.5 Development of QSRR models using the I_{ET}

As the starting point, the I_{ET} was developed for alkanes on a low polarity stationary phase. These are the simplest compounds and their properties are almost completely dependent on topological features. Subsequently, this novel topological descriptor was extended to different classes of organic compounds with more complex structural features. A summary of the best simple linear regression models (RI = b + a I_{ET}) and the statistical data for each data set of compounds, obtained in previous QSRR studies, is given in Table 1.

3. The semi-empirical electrotopological index, I_{SET}

The semi–empirical topological index (I_{ET}) discussed in the previous section allows the creation of a new descriptor, the electrotopological index, I_{SET}, which was developed and applied to QSPR studies to predict the retention index, boiling points and octanol/water partition coefficient (Log P), for a large amount of organic compounds, including aliphatic hydrocarbons alkanes and alkenes, aldehydes, ketones, esters and alcohols (Souza et al.,

No	Data Set	a	b	N	r	SD	Ref.
01	Alkanes	116.8	-19.05	157	0.9901	26.20	Heinzen et al., 1999b
02	Cis-/trans-linear alkenes	122.8446	-41.7054	79	0.99996	2.35	Heinzen et al., 1999b
03	Branched alkenes	120.4671	-29.0457	59	0.9985	5.76	Junkes et al., 2002a
04	Methyl-branched alkanes	123.1610	-39.5251	178	0.99998	4.31	Junkes et al., 2002b
05	Esters	123.79	-48.14	81	0.9995	5.79	Amboni et al., 2002a
06	Aldehydes and ketones	123.4951	-45.6553	54	0.9999	5.01	Junkes et al., 2004
07	Alcohols	124.1239	-51.3739	44	0.9991	5.70	Junkes et al., 2003b
08	Alkylbenzene	123.0824	-39.7381	122	0.9998	5.5	Porto et al., 2008
09	Halogenated compounds	124.7788	-56.8944	141	0.9997	8.0	Arruda et al., 2008

Table 1. Summary of the best simple linear regressions ($RI_{Calc} = a + b\ I_{ET}$) found for different data set on low polarity stationary phases.

2008, 2009a, 2009b, 2010). This new descriptor for this series of molecules can be quickly calculated from atomic charges obtained through the semi-empirical quantum-chemical, AM1 method (Bredow & Jug, 2005; Smith, 1996), since it was found that atomic charges correlated with the approximate numerical values attributed by the semi-empirical topological index to the primary, secondary, tertiary and quaternary carbons atoms.

3.1 Calculation of I_{SET} for alkanes and alkenes

For alkanes and alkenes, the above-mentioned correlation allowed the creation of a new semi-empirical electrotopological index (I_{SET}) for QSRR models based on the fact that the interactions between the solute and the stationary phase are due to electrostatic and dispersive forces (Souza et al., 2008). This new index, I_{SET}, is able to distinguish between the *cis*- and *trans*-isomers directly from the values of the net atomic charges of the carbon atoms that are obtained from quantum-chemical calculations. More precisely, this new semi-empirical electrotopological index, I_{SET}, was developed based on the refinement of the previous semi-empirical topological index, I_{ET}. The values for the Ci fragments that were firstly attributed from the experimental chromatographic retention and theoretical deductions have an excellent relationship with the net atomic charge of the carbon atoms. Thus, the values attributed to the vertices in the hydrogen-suppressed graph of carbon atoms (Ci) are calculated from the correlation between the net atomic charge in each carbon atom, which is obtained from quantum chemical semi-empirical calculations, and the Ci fragments for primary, secondary, tertiary and quaternary carbon atoms (1.0, 0.9, 0.8 and 0.7, respectively) obtained from the experimental values. This shows that it is possible to

calculate a new index, I_{SET} (the semi-empirical electrotopological index) through the net atomic charge values obtained from a Mulliken population analysis using the semi-empirical AM1 method and their correlation with the values attributed to the different types of carbon atoms. This demonstrates that the I_{SET} encodes information on the charge distribution of the solute which drives the dispersive and electrostatic interactions between the solute (alkanes and alkenes) and the stationary phase (Souza et al., 2008).

Since the interactions between the solute and the stationary phase are dispersive for alkanes and electrostatic for alkenes, the chromatographic retention is strongly dependent on the electronic charge distribution of each carbon atom of these molecules. A simple linear regression equation was obtained between the values of the carbon atoms, SETi values, based on experimental gas chromatography retention (for primary (1.0), secondary (0.9), tertiary (0.8) and quaternary (0.7) carbon atoms) and the net atomic charges (δi) of these atoms, as given in Equation (2).

$$SET_i = -1.77125\delta_i + 0.62417 \tag{2}$$

This indicates that the physical reality encoded by the semi-empirical topological index (I_{ET}) developed in our laboratory is completely related to net atomic charges which, as is well known, are important forces in intermolecular interactions. It is clear that the interactions between the non-polar stationary phases and the different compounds were determined predominantly through the electronic charge distribution of the molecular structures of the compounds analyzed by gas chromatography. From Equation (1) it is clear that knowledge of the net atomic charges is sufficient to calculate the SETi value for all kinds of carbon atoms and not only the values given by the carbon models (that is 1.0, 0.9, 0.8 and 0.7) or in specific tables. Hence, the above method of calculating the SETi values of the carbon atoms allows a new index to be created, denominated the semi-empirical electrotopological index, I_{SET}. Considering the steric effects of the neighboring carbon atoms, as was observed in the calculation of I_{ET}, this new index can be calculated according to Equation (3).

$$I_{SET} = \sum_{i,j}(SET_i + logSET_j) \tag{3}$$

In the above expression the *i* sum is over all the atoms of the molecule (excluding the H atoms) and the j is an inner sum of atoms attached to the *i* atom. The *cis*-2-pentene and *trans*-2-pentene molecules represented in the graph below are taken as an example of the I_{SET} calculation.

| *cis*-2-pentene | *trans*-2-pentene |
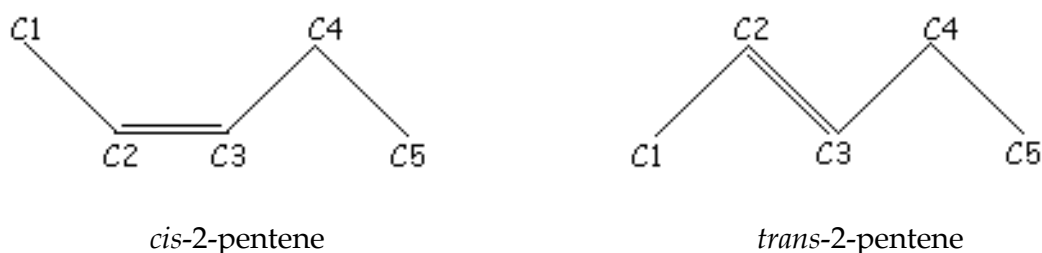
The net atomic charges and SET$_i$ values for the above molecules are given in Table 2 below.

Carbon atom	*cis*-penten-2-eno		*trans*-penten-2-eno	
	δ_j	SET_j	δ_j	SET_j
C1	-0.187	0.9555	-0.185	0.9520
C2	-0.173	0.9307	-0.170	0.9254
C3	-0.165	0.9165	-0.167	0.9200
C4	-0.133	0.8598	-0.132	0.8580
C5	-0.209	0.9945	-0.207	0.9910

Table 2. The net atomic charge (δi) and the SET_i values for each carbon atom of *cis*-2-pentene and *trans*-2-pentene molecules.

The I_{SET} calculation now follows:

I_{SETC1} = 0.9555 + log 0.9307 I_{SETC1} = 0.9520 + log 0.9254

I_{SETC2} = 0.9307 + log 0.9555 + log 0.9165 I_{SETC2} = 0.9254 + log 0.9520 + log 0.9200

I_{SETC3} = 0.9165 + log 0.9307 + log 0.8598 I_{SETC3} = 0.9200 + log 0.9254 + log 0.8580

I_{SETC4} = 0.8598 + log 0.9165 + log 0.9945 I_{SETC4} = 0.8580 + log 0.9200 + log 0.9910

I_{SETC5} = 0.9945 + log 0.8598 I_{SETC5} = 0.9910 + log 0.8580

I_{SET} = 4.3653 I_{SET} = 4.3481

As expected on physical-chemical grounds, the AM1 calculation reveals that the optimized structures of the *cis*- and *trans*- isomers have slightly different charge distributions. As can be seen from the above results, the Mulliken population analysis gives the net atomic charges of the carbon atoms for each isomer, which implies that the difference in the values for the SETi fragments is sufficient to give different I_{SET} values.

3.2 Calculation of I_{SET} for compounds with oxygen-containing functional groups

3.2.1 Ketones and aldehydes

For polar molecules like aldehydes, ketones, esters and alcohols, the presence of heteroatoms like oxygen changes considerably the charge distribution of the corresponding hydrocarbons giving a small increase in the interactions between the solute and the stationary phase. An appropriate way to calculate the I_{SET} was developed that takes into account the dipole moment exhibited by these molecules and the atomic charges of the heteroatoms and the carbon atoms attached to them (Souza et al., 2009a). By considering the stationary phase as non-polar material, the interactions are slowly increased relative to the corresponding hydrocarbons due to the charge redistribution that occurs in presence of the heteroatom. This charge redistribution accounts for the dipole moment of the molecules. Thus, the dipolar charge distribution in such molecules leads to a small increase in the interactions of the solute with the stationary phase relative to hydrocarbons where the dipole moment is zero, or almost zero. Clearly the major effects on the charge distribution due the presence of the (oxygen) heteroatoms occur in the neighborhood and the excess charge of these atoms leads to electrostatic interactions that are stronger relative to the weak dispersive dipolar interactions (Christian, 1990).

In relation to the chromatographic retention it can be observed, for instance, that the molecules 2-hexanone, 3-hexanone and hexanal have experimental retention indices of 767, 764 and 776, respectively, and for the corresponding hydrocarbon molecule in the absence of the heteroatom, that is, the heptane, the retention index is 700. Due to the presence of the heteroatom (oxygen) there is an increase in the retention index of around 10%. Hence, the interactions between the molecules and the stationary phase are slowly increased and clearly this is due to the charge redistribution that occurs in the presence of the heteroatom. This charge redistribution accounts for the dipole moment of molecules like aldehydes and ketones. The dispersive force between these kinds of molecules and the stationary phase includes the charge-dipole interactions and dipole–induced dipole interactions which are weak relative to the electrostatic interactions. Thus, the dipolar charge distribution in such molecules leads to a small increase in the interactions of the solute with the stationary phase relative to hydrocarbons where the dipole moment is zero. Initially, it appears that the above-mentioned factors mean that the retention index can be calculated as in equation 3, and the same applies to the heteroatoms, but including subtle alterations that incorporate the effects of the dispersive dipolar interactions.

All of these factors can be included in the calculation of the retention index through a small increase in the SET_i values for the heteroatoms and the carbon atoms attached to them. This was carried out by multiplying the SET_i values of these atoms by a function A_μ which is dependent on the dipole moment of the molecule and the net charge at the oxygen and carbon atoms (to include both the electrostatic and dispersive interactions). Since we must have $A_\mu = 1$, when the dipole moment is zero or almost zero (as in the case of alkanes and alkenes) in a first attempt to achieve this function a linear dependence on the molecular dipole moment μ is considered, that is, $A_\mu = 1 + (\mu/\mu_F)$, where μ_F is a local function (in the units of the dipole moment) in the sense that it is dependent on the net charge of oxygen and carbon atoms. On the one hand this definition of A_μ works only if $\mu/\mu_F > 1$, since A_μ must reflect the small increase in the interactions due to dipolar dispersive forces. On the other hand good choices for the definition of μ_F for ketones and aldehydes (as we shall see below) means that the ratio μ/μ_F can be much greater than unity showing clearly that it is not possible to apply the above definition to A_μ. Considering that $\mu/\mu_F > 1$ then A_μ can not be a polynomial function of μ/μ_F. Thus, A_μ must have a weaker dependence on the dipole moment than the linear one and this weak dependence can be achieved through a logarithmic function since it is clear that the function $f(x) = x$ increases much faster than the function $f(x) = \log(1+x)$. Taking these factors into account it is possible to achieve a definition of A_μ that differs slightly from unity and is logarithmically dependent on the dipole moment of the molecule, as seen in equation 4

$$A_\mu = 1 + \log(1 + \frac{\mu}{\mu_F}),\qquad\qquad(4)$$

where μ is the calculated molecular dipole moment and μ_F is a local function which is dependent on the charges of the atoms belonging to the C=O bond. Clearly, μ_F must be directly related to the net charge of the oxygen atoms since it must reflect some contribution to the electrostatic interaction between these molecules and the stationary phase. In this regard, μ_F may also be related to the atomic charge of the carbon atom of the functional group C=O or related to the difference between the atomic charges of these atoms. Hence, μ_F

can be defined in different ways and some definitions of μ_F can be used in preliminary calculations. As expected, after some preliminary calculations, the best choice was for ketones $\mu_F = d|Q_C - Q_O|$ where d is the calculated C=O bond length and $|Q_C - Q_O|$ is the absolute value of the difference between the atomic charges at the carbon and oxygen atoms. This definition of μ_F is an attempt to take into account the contribution of the atomic charge of the oxygen atoms and the respective bonded carbon atom to the electrostatic interactions. For aldehydes the terminal carbon atom of the C=O bond is attached to a hydrogen and thus it is necessary to consider the net positive charge in this polar region of the molecule as the sum of the atomic charges of the carbon and hydrogen atoms. This means that for aldehydes the best choice for μ_F was $\mu_F = d|Q_C + Q_H - Q_O|$. Therefore, equation 4 indicates that there is an increase in the interaction between the molecules and the stationary phase due to the presence of the dipole moment and that this contribution may be screened by the charge located on the heteroatoms (oxygen atoms) if $\mu/\mu_F < 1$, or may be increased if $\mu/\mu_F > 1$. In the case of ketones and aldehydes the local function μ_F is less than the dipole moment showing that A_μ receives an appreciable contribution from the atomic charges of these atoms. This reveals the contribution of oxygen to the electrostatic interaction between the solute and the stationary phase. Therefore, to include the dispersive dipolar interactions in the calculation of the retention index we multiply the SET_i values for the heteroatoms (oxygen) and the carbon atoms attached to them by the dipolar function A_μ given in equation 4. That is, in this model the I_{SET} is calculated as in equation 5

$$I_{SET} = \sum_{i,j} (A_\mu SET_i + \log A_\mu SET_j),\qquad(5)$$

where the SET_i values are obtained using equation 2. As in equation 3, in the above expression the i sum is over the all the atoms of the molecule (excluding the H atoms) and the j is an inner sum of the atoms attached to the i atom. In the above expression, for the I_{SET} the dipolar function A_μ is taken as unity for the remaining carbon atoms of the molecules. Equation 4 reduces to equation 2 when the dipole moment of the molecule is zero or almost zero, as is the case for alkanes and alkenes since $A_\mu = 1$ for $\mu = 0$.
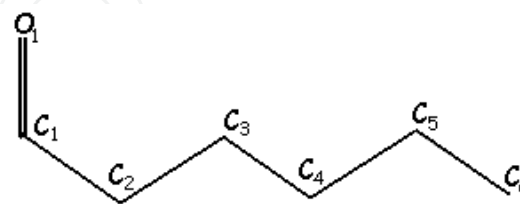
The 3-hexanone and hexanal molecules represented in the graph below are taken as an example of the I_{SET} calculation.
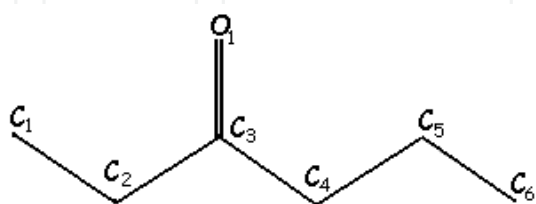
3-hexanone

$\mu_F = d|Q_C - Q_O|$

$\mu_F = 1.2342\,[0.224 - [-0.288]] = 0.6319$

hexanal $Q_H = +0.086$

$\mu_F = d|Q_C + Q_H - Q_O|$

$\mu_F = 1.2314[+0.183 + 0.086 - [-0.289]] = 0.6871$

$$A_\mu = 1 + \log(1 + \mu/\mu_F)$$

$A_\mu = 1 + \log(1 + (2.6790/0.63191)) = 1.7193$ $A_\mu = 1 + \log(1 + (2.7640/0.68712)) = 1.7009$

The net atomic charges and SET_i values are given in Table 3 below.

Atoms	3-hexanone			hexanal		
	δi	SET_i	$AuSETi$	δi	SET_i	$AuSETi$
O_1	-0.288	1.1346	1.9507	-0.289	1.1363	1.9328
C_1	-0.206	0.9892	-	+0.183	0.2995	0.5094
C_2	-0.212	0.9998	-	-0.233	1.0371	-
C_3	-0.224	0.2268	0.3899	-0.155	0.8988	-
C_4	-0.212	0.9998	-	-0.158	0.9041	-
C_5	-0.155	0.8988	-	-0.159	0.9059	-
C_6	-0.212	0.9998	-	-0.211	0.9980	-

Table 3. The net atomic charge (δi) and the SET_i values for each carbon and oxygen atom of 3-hexanone and hexanal molecules.

The I_{SET} calculation now follows:

$$I_{SET} = \sum_{i,j} (A_\mu SET_i + \log A_\mu SET_j)$$

3-hexanone

$SET_{O1} = 1.9507 + \log 0.3899 = 1.5416$

$SET_{C1} = 0.9892 + \log 0.9998 = 0.9891$

$SET_{C2} = 0.9998 + \log 0.9892 + \log 0.3899 = 0.5860$

$SET_{C3} = 0.3899 + \log 0.9998 + \log 1.9507 + \log 0.9998 = 0.6799$

$SET_{C4} = 0.9998 + \log 0.3899 + \log 0.8988 = 0.5444$

$SET_{C5} = 0.8988 + \log 0.9998 + \log 0.9998 = 0.8986$

$SET_{C6} = 0.9998 + \log 0.8988 = 0.9535$

$I_{SET} = 6.1931$

hexanal

$SET_{O1} = 1.9328 + \log 0.5094 = 1.6398$

$SET_{C1} = 0.5094 + \log 1.9328 + \log 1.0371 = 0.8114$

$SET_{C2} = 1.0371 + \log 0.5094 + \log 0.8988 = 0.6978$

$SET_{C3} = 0.8988 + \log 1.0371 + \log 0.9041 = 0.8708$

$SET_{C4} = 0.9041 + \log 0.8988 + \log 0.9059 = 0.8148$

$SET_{C5} = 0.9059 + \log 0.9041 + \log 0.9980 = 0.8612$

$SET_{C6} = 0.9980 + \log 0.9059 = 0.9550$

$I_{SET} = 6.6508$

3.2.2 Esters

For esters the major effects related to the charge distribution are due to the presence of the two oxygen atoms and they occur on these atoms and in their neighborhood (their adjacent carbon atoms). The excess charge of these atoms leads to electrostatic interactions that are stronger than the weak dispersive dipolar interactions. For esters, all these factors were included in the calculation of the retention index through a small increase in the SET_i values for heteroatoms and the carbon atoms attached to them (Souza et al., 2009b). As in the case of ketones and aldehydes, it was verified that the introduction of the dipole moment of the molecule is not sufficient to explain the chromatographic behavior of these molecules. Thus, it was necessary to introduce an equivalent local dipole moment of the (-COOC-) group that contributes to the increase in the retention value. This was carried out by multiplying the SET_i values of the atoms belonging to the O=C-O-C group by the function $\mathbf{A_\mu}$ which is dependent on the dipole moment of the molecule and the net charge of the oxygen and carbon atoms (to include both the electrostatic and dispersive interactions). The same approach used for ketones and aldehydes was applied to esters, that is, considering that $\mathbf{A_\mu}$ has a weaker dependence on the dipole moment than the linear one, as given in equation 4. For esters $\mathbf{\mu_F}$ is an equivalent local dipole moment (in the units of dipole moment) which is dependent on the charges of the atoms belonging to the O=C-O-C group. Clearly $\mathbf{\mu_F}$ must be directly related to the net charge of the oxygen atoms since it must reflect some contribution to the electrostatic interaction between these atoms and the stationary phase. In this regard, $\mathbf{\mu_F}$ may also be related to the atomic charge of the carbon atom of the functional group C=O or related to the difference between the atomic charges of these atoms. Hence, $\mathbf{\mu_F}$ can be defined in different ways and some definitions of $\mathbf{\mu_F}$ can be used in preliminary calculations.

Esters have two oxygen atoms and thus it is possible to define two local functions, one being dependent on the charges and bond length of the $C=O_1$ bond and another on the charges and bond length of the $C-O_2$ bond. Therefore, it was necessary to perform some calculations with different definitions for the equivalent local dipole moment. After the preliminary calculations it was found that for esters the charge difference, $Q_O - Q_C$, does not give reasonable results because the charges of the oxygen atoms mask the charge of the carbonyl carbon. As expected, our best choice was for the esters $\mathbf{\mu_{F1}} = \mathbf{d_1}|\mathbf{Q_{O1}}|$ and $\mathbf{\mu_{F2}} = \mathbf{d_2}|\mathbf{Q_{O2}}|$, where $\mathbf{d_1}$ and $\mathbf{d_2}$ are the calculated $C_1=O_1$ and C_1-O_2 bond lengths and $|\mathbf{Q_{O1}}|$ and $|\mathbf{Q_{O2}}|$ are the absolute values of the atomic charges of the oxygen atoms (O_1 and O_2). The equivalent local dipole moment is then calculated as the magnitude of the vectorial sum of two dipole moments, that is, $\mathbf{\mu_{F1}} = (\mathbf{\mu^2_{F1}} + \mathbf{\mu^2_{F2}} + 2\,\mathbf{\mu_{F1}}\,\mathbf{\mu_{F2}}\,\mathbf{cos\theta})^{1/2}$, where $\mathbf{\theta}$ is the angle between the $C=O_1$ and $C-O_2$ bonds. For formates, a specific charge distribution occurs in the polar region of the molecules and the best mathematical model for the local moment was that which takes into account the contribution to the electrostatic interactions that originate from the atomic charges of the oxygen atoms, the carbon atoms and the H atom belonging to the $C_1O_1O_2C_{Al}$ group of the formate molecules (C_{Al} represents the carbon on the alcoholic side). Thus, the equivalent dipoles were built from the net charges of the HC_1O_1, HC_1O_2 and O_2C_{Al} groups of atoms. The equivalent dipoles associated with these net charges are:

$\mathbf{\mu_{F1}} = \mathbf{d_1}|\mathbf{Q_H} + \mathbf{Q_{C1}} - \mathbf{Q_{O1}}|$, $\mathbf{\mu_{F2}} = \mathbf{d_2}|\mathbf{Q_H} + \mathbf{Q_{C2}} - \mathbf{Q_{O2}}|$ and $\mathbf{\mu_{F3}} = \mathbf{d_3}|\mathbf{Q_{CA1}} - \mathbf{Q_{O2}}|$ where $\mathbf{d_1}$ and $\mathbf{d_2}$ are the calculated $C_1=O_1$ and C_1-O_2 bond lengths and $\mathbf{d_3}$ is the calculated $C_{Al}-O_2$ bond length. In a first approach, the local moments $\mathbf{\mu_{F2}}$ and $\mathbf{\mu_{F3}}$ are considered to be

collinear and another equivalent dipole is obtained from the difference between μ_{F2} and μ_{F3}, that is, $\mu_{F4} = \mu_{F2 \text{-}} \mu_{F3}$ and the final equivalent local moment is calculated as above, that is, $\mu_F = (\mu^2_{F1} + \mu^2_{F4} + 2 \mu_{F1} \mu_{F4} \cos\theta)^{1/2}$ where θ is the angle between the $C=O_1$ and $C\text{-}O_2$ bonds. Hence, for formates the charge of the hydrogen atom attached to the carbon atom of the COO functional group is also considered, as in the case of aldehydes, because the charge of the H atom contributes explicitly to the positive charge of the local polar region of the molecule. The above-mentioned best definitions for μ_F imply that the present approach to calculating the retention index considers important polar features of the organic functions, such as ketones, aldehydes and esters, through the information carried by the local moment μ_F. In other words, according to Equation (4) there is an increase in the interaction between the molecules and the stationary phase due to the presence of a dipole moment and this contribution may be screened by the charge located on the heteroatoms and the carbon atom of the C=O group if , $\mu_F > \mu$ or may be increased if $\mu_F < \mu$. In the case of esters, the local function μ_F is less than the dipole moment showing that A_μ has an appreciable contribution from the atomic charges of those atoms. This verifies, for esters, the contribution of the oxygen atom to the electrostatic interaction between the solute and the stationary phase.

Therefore, in the case of esters the I_{SET} value is here calculated as in Equation (5), where the SET_i values are obtained using Equation (2) through AM1 calculations of the net atomic charges. As mentioned above, Equation (5) is calculated by multiplying the SET_i values of the atoms belonging to the $C_1O_1O_2C_{Al}$ group by the dipolar function A_μ which is taken as unity for the remaining carbon atoms of the molecules. Hence, Equation (5) is a general definition for the electrotopological index that can be applied to different organic functions, which are specified through appropriate definitions of the equivalent local moment μ_F. The preliminary applications of I_{SET} as given by Equation (5) showed that this expression overestimates the calculated retention index for branched esters and underestimates the results for methyl esters. This finding reveals the need to consider other definitions for the local moment μ_F for branched esters and methyl esters. However, another easy choice is to take into account the steric effects for the branched esters and methyl esters. The simplest way to do this is to consider the steric hindrance of the C_{Al} carbon atom of the $C_1O_1O_2C_{Al}$ group and the carbon atom attached to the acid side of the COO functional group (here named the C_{Ac} carbon). As seen in Equation (2), the log SET_j factor gives, precisely, the steric effect of atom j. Thus, to include a steric correction (sc) in Equation (5) for branched esters the term **sc = n logSET(C_{AC}) + n logSET(C_{Al})** was added, where **n** is the number of branches of the ester. On the other hand, for methyl esters the C_{Al} carbon is bound to three H atoms and it is necessary to remove the overestimated steric effects of the **logA$_\mu$SET$_j$** terms in Equation (5). For methyl esters this is easily achieved by including a second steric correction (ssc) by adding the term **ssc = -log SET(C_{Al})** to equation (5). Very good results were obtained using this approach, which reveals that in this model the complex steric effects in branched esters can be included simply by considering the steric hindrance using the net charge (through the SET_i values) of the two carbon atoms bound to the alcoholic and acid sides of the COO functional group. The calculation of I_{SET} for a large amount of molecules is easily carried out by means of a FORTRAN code developed in our lab that calculates I_{SET} by reading the output data (calculated net charges, dipole moment and atomic positions) from AM1 semi-empirical calculations.

3.2.3 Alcohols

As observed for the preceding compounds, for alcohols the major effects on the charge distribution are due the presence of the oxygen atom and they occur at the site of and close to their neighbors (adjacent carbon atoms). The excess charge at these atoms leads to electrostatic interactions that are stronger than the weak dispersive dipolar interactions. Thus, it is clear that it is necessary to introduce an equivalent local dipole moment for each of the organic functions that participate in increasing the retention value. For alcohols, as in the case of ketones, aldehydes and esters, this was achieved by multiplying the SET_i values of the atoms belonging to the C-OH group by a function $\mathbf{A_\mu}$ as defined by equation 4, with $\mathbf{\mu_F}$ being the equivalent local dipole moment which is dependent on the charges of the atoms belonging to the C-OH group (Souza et al., 2010). Clearly, $\mathbf{\mu_F}$ is directly related to the net atomic charge of the oxygen atoms since it must reflect some contribution to the electrostatic interaction between these atoms and the stationary phase. Thus, $\mathbf{\mu_F}$ may also be related to the atomic charge of the carbon atom of the functional group C-OH or to the difference between the atomic charges of these atoms. Hence, as with the other organic functions, $\mathbf{\mu_F}$ can be defined in different ways and some of these definitions can be used in the preliminary calculations. For primary alcohols the terminal carbon atom of the C-O bond is attached to two hydrogen atoms and thus it is necessary to consider the net positive charge in this polar region of the molecule as the sum of the atomic charges of the carbon and hydrogen atoms. Thus, for primary alcohols we found that the best definition of the local moment is related to the charges of all atoms at the polar head of the molecules, that is, $\mathbf{\mu_F} = \mathbf{d|Q_C +(Q_{H1} +Q_{H2})/2 - Q_O -Q_{HO}|}$ where $\mathbf{d}$ is the calculated C-O bond length and $\mathbf{|Q_C +(Q_{H1} +Q_{H2})/2 - Q_O -Q_{HO}|}$ is the absolute value of the difference between the net atomic charge at the carbon (Q_C) plus the average charge of the hydrogen atoms attached to it $\mathbf{(Q_{H1} + Q_{H2})/2}$ and the charges of the oxygen atom (Q_O) and the hydrogen attached to it (Q_{Ho}). For secondary, tertiary and quaternary alcohols the best choice for the local moment is related to the net atomic charge of the C and O atoms only, that is, $\mathbf{\mu_F = d|Q_C - Q_O|}$, where $\mathbf{d}$ is the length of the C-O bond and $\mathbf{|Q_C - Q_O|}$ is the absolute value of the difference between the charge of the carbon atom and oxygen atom attached to it. These definitions of $\mathbf{\mu_F}$ attempt to take into account the contribution to the electrostatic interactions originating from the polar region of the molecules. Therefore, this shows again that Equation 4 represents a dipolar contribution to the interactions between the molecules and the stationary phase (which originates from the presence of a molecular dipole moment) and this contribution is decreased by the charge of the heteroatoms (oxygen atoms) when $\mathbf{\mu_F} > \mu$, or increased when $\mathbf{\mu_F} < \mu$. This reveals the contribution of oxygen to the electrostatic interaction between the solute and stationary phases.

For alcohols, the I_{SET} values are calculated as in Equation (5), where the SET_i values are obtained using Equation (2) through AM1 calculations of the net atomic charges.

3.3 Development of QSPR models using the I_{SET}

The molecular descriptor I_{SET} was developed first for alkanes and alkenes on a low polarity stationary phase and then extended to oxo-compounds through the inclusion of the molecular dipole moment and a local dipole moment in its definition. The models for the best simple linear regression between the retention index and the molecular descriptor (RI = b + a ISET) and the statistical data for each class of compounds, obtained in previous QSRR

studies, are given in Table 4. For esters and alcohols good correlations between the retention index and the I_{SET} were obtained also for stationary phases with different polarities (not included in Table 4). The good statistical results achieved (Table 4) employing I_{SET} are better or equivalent to those obtained using multiple linear regression employing many molecular descriptors.

	Compounds	Phase	N	a	b	r^2	r^2_{CV}	SD	Ref.
01	Alkanes and Alkenes	SQ	179	120.8	-36.8	0.9980	0.9980	10.7	Souza et al., 2008
02	Aldehydes and ketones	HP-1	42	123.9	-13.2	0.9993	0.9993	11.7	Souza et al., 2009a
03	Esters	SE-30	100	115.0	-74.7	0.9981	0.9980	7.6	Souza et al., 2009b
04	Alcohols	SE-30	31	126.0	-186.6	0.9990	0.9977	9.3	Souza et al., 2010

Table 4. Summary of the best simple linear regressions ($RI_{Calc} = b + a\, I_{SET}$) found for different classes of compounds on low-polarity stationary phases.

As can be seen from Table 4, the QSRR models for 179 representative linear and branched alkanes and alkenes, obtained with the I_{SET} using the net atomic charge to calculate more precisely the Ci fragment values of I_{ET}, were of good quality for the statistical parameters obtained. This new descriptor I_{SET} contains information on the 3D features of molecules, and discriminates between geometrical isomers, such as *cis-* and *trans-*alkenes, and between conformers, and the elution sequence is correct for the majority of the compounds.

The results obtained for aldehydes and ketones are similar to those reported by Ren (Ren, 2003) in multiple linear regression models for 33 aldehydes and ketones using Xu and AI topological indices and by Héberger and co-workers using quantum-chemical descriptors (SW and μ) (Héberger et al., 2001) and physico-chemical properties (T_{Bp}, M_W, log P) (Héberger et al., 2000) for 31 and 35 compounds, respectively. For esters the results obtained by single linear regression using the I_{SET} are better than those reported by Lu et al. (Lu et al., 2006). For SE-30 and OV-7 stationary phases the results are also better than those found by Liu et al. (Liu et al., 2007) and for more polar stationary phases the statistical parameters differ only slightly. Both of these studies use multiple linear regression (MLR) between RI and the topological indices for 90 saturated esters on stationary phases with different polarities.

Several authors have developed QSRR models, based on MLRs, to predict the RI values for saturated alcohols. For example, Guo et al. (Guo et al., 2000), using the MLR analysis and artificial neural networks technique, obtained the statistical parameters r^2=0.9982, SD=8.21, N=19 for an SE-30 stationary phase. In a previous study, the best statistical parameters of the MLR models obtained by Farkas and Héberger (Farkas & Héberger, 2005), employing four molecular descriptors, were r=0.9804, SD=14.22, r^2_{CV}=0.9801 and N=44 for an OV-1 stationary phase. Therefore, our prediction results, on low polarity stationary phases, using the I_{SET} as a single descriptor, showed statistical quality comparable to similar studies reported by the above authors. Furthermore, the statistical parameters of the present approach have a good agreement with those obtained for alkanes and alkenes, aldehydes

and ketones, for saturated esters and for alcohols using the semi-empirical topological index, I_{ET}, previously developed. These results show clearly that I_{SET} is a molecular descriptor that embodies in an appropriated manner the net atomic charges and charge distribution of molecules since the retention index embodies the intermolecular interactions between the stationary phase and the molecules.

The fact that properties that are determined by intermolecular forces can be adequately modeled by the I_{SET} descriptor can be easily seen in its relationship with the boiling point (BP). For alcohols a good correlation was obtained through a simple linear model (BP = b + a I_{SET}), as can be seen in Table 5. The QSPR model obtained for the experimental BP of 134 compounds showed high values for the coefficient of determination and cross validation coefficient showing the good predictive capacity of the model.

	Alcohols	N	a	b	r^2	r^2_{CV}	SD	Ref.
01	Boiling Point	134	22.0	-32.9	0.9818	0.9811	5.4	Souza et al., 2010

Table 5. The coefficients and statistical parameters for linear regression between experimental boiling point and I_{SET}.

This model can explain 98.20% of the variances in the experimental values and most of these compounds (N=101) are not included in the initial model used to build the I_{SET}, showing the external stability of the model. These results are similar to those obtained using I_{ET} for 146 aliphatic alcohols and can be compared with those obtained by Ren (Ren, 2002b), but using MLR models, for 138 compounds with five descriptors.

The octanol-water partition coefficient (log P) of compounds, which is a measure of hydrophobicity, is widely used in numerous Quantitative Structure-Activity Relationship (QSAR) models for predicting the pharmaceutical properties of molecules. The partition coefficient is a property that is determined by intermolecular forces and thus it is expected that it can be described by a molecular descriptor such as I_{SET}. The results obtained in the statistical analysis of the single linear regression between experimental log P values and I_{SET} are shown in Table 6 for each class of compounds.

Compound	N	a	b	r^2	r^2_{CV}	SD
Hydrocarbon	23	3.90×10^{-3}	0.9997	0.9971	0.9964	0.104
Alcohols	60	3,24	0,6394	0.9876	0.9870	0.183
Aldehydes	9	1.60×10^{-3}	1.0014	0.9972	0.9961	0.058
Ketones	19	-2.7182	0.6693	0.9864	0.9831	0.158
Esters	14	-3.1575	0.6587	0.9903	0.9838	0.118

Table 6. The coefficients and statistical parameters for linear regression between experimental log P values and I_{SET}.

The results in Table 6 indicate that the theoretical partition coefficients calculated using the I_{SET} method give good agreement with the experimental partition coefficients. The QSPR models obtained with I_{SET} showed high values for the correlation coefficient (r > 0.99), and the leave-one-out cross-validation demonstrates that the final models are statistically significant and reliable (r^2_{cv} > 0.98). As can be observed, this model explains more than 99%

of the variance in the experimental values for this set of compounds. Among the various classes of compounds the best results obtained with the I_{SET} method are for hydrocarbons (Table 6), which is related to the fact that the present model was developed initially for this class of organic compounds. As can be seen in Table 6, the lowest standard deviation was obtained for the correlation of aldehydes and for alcohols the correlation was stronger. The range of standard deviations obtained verifies the applicability of the present approach to different classes of organic compounds. For alcohols, the earlier approach of Duchowicz et al. (Duchowicz et al., 2004), based on the concept of flexible topological descriptors and on the optimization of correlation weights of local graphic invariants, is applied to model the octanol/water partition coefficient of a representative set of 62 alcohols, resulting in a satisfactory prediction with a standard deviation of 0.22. Recently, Liu et al. (Liu et al. 2009) carried out a QSPR study to predict the log P for 58 aliphatic alcohols using novel molecular indices based on graph theory, by dividing the molecular structure into substructures obtaining models with good stability and robustness, and values predicted using the multiple linear regression method are close to the experimental values (r = 0.9959 and SD = 0.15). The above results show the reliability of the present model calculation based on the semi-empirical calculation of atomic charges and local dipole moments using only one descriptor, I_{SET}. This new approach to polar molecules, with the introduction of the remodeled I_{SET} index including the contribution of the dipole moment of the molecule and an effective local dipole moment associated with the net charges of the atoms of the carbonyl group, opens new possibilities for studies on the chromatographic and other properties of different organic functions.

4. Conclusions

It is known that the chromatographic process of separation results from the forces that operate between solute molecules and the molecules of the stationary phase. These forces are called van-der-Waals forces since van der Waals recognized them as the reason for the non-ideal behavior of the real gases. Intermolecular forces are usually classified according to two distinct categories: i) the first category corresponds to the directional, induction and dispersion forces which are non-specific; and ii) the second group corresponds to hydrogen bonding forces and the forces of charge transfer or electron-pair donor-acceptor forces which are specific.

In the development of the semi-empirical topological index (I_{ET}) it was considered that the retention of alkanes is due to the number and interaction of each specific carbon atom with the stationary phase, considered as non-polar, which is determined by its electrical characteristic and by the steric hindrance by other carbon atoms attached to it. In this case only dispersion forces due to the continuous electronic movement, at any instant, result in a small dipole moment which can fluctuate and polarize the electron system of the neighboring atoms or molecules. For the alkenes, some carbon atoms with greater electronegativity give the molecules a dipole moment and for this reason besides the dispersion forces, electrostatics forces play an important role. However, in this method the behavior of this kind of carbon atom is determined from the experimental data and indicated in specific tables. As the values were obtained from the experimental data they encode the real physical interaction force. In the case of oxo-compounds, the presence of atoms with different carbon atom electronegativity introduces a dipole moment in the

functional group and a change in the dipole moment of the whole molecule. These factors were considered in order to obtain the different values for the functional groups and they were able to encode the physical force involved in the chromatographic separation.

The new semi-empirical electrotopologiocal index (I_{SET}) demonstrated that the values for the carbon atoms that are not tetrahedral and functional groups (considering the new local dipole created by the heteroatom) can be calculated from the net atomic charges that are obtained from quantum-chemical calculations. In the case of esters, the major effects are due to the presence of the two oxygen atoms and their adjacent carbon atoms. As in the case of aldehydes and ketones it was verified that the introduction of the dipole moment of the molecules is not sufficient to explain the chromatographic behavior. Thus, it was necessary to introduce an equivalent local dipole moment of the ester group that contributes to the increase in the retention value. In the case of esters two local functions must be considered according to the charges and the bond lengths of the C=O and C-O bonds. Thus, the semi-empirical electrotopological index was developed based on the refinement of the previously developed semi-empirical topological index, unifying the quantum-quantum chemical with the topological method to provide a three-dimensional picture of the atoms in the molecule.

The I_{ET} and I_{SET} were generated to predict the chromatographic retention indices and other physical-chemical properties and to obtain the quantitative structure-retention relationship (QSRR/QSPR). The efficiency and the applicability of these descriptors were demonstrated through the good statistical quality and high internal stability obtained for the different classes of compounds studied.

5. Acknowledgement

6. References

Amboni, D. M.; Junkes, B. S.; Yunes, R. A &, Heinzen, V. E. F. (2000). *Journal of Agricultural Food Chemistry*, Vol.48, pp. 3517–3521.

Amboni, R. D. M. C; Junkes, B. S.; Yunes, R. A. & Heinzen, V. E. F. (2002a). Semi-empirical topological method for prediction of the chromatographic retention of esters. *Journal of Molecular Structure (Theochem)*, Vol.579, pp. 53–62.

Amboni, R. D. M. C; Junkes, B. S.; Yunes, R. A. & Heinzen, V. E. F. (2002b). Quantitative structure-property relationship study of chromatographic retention indices and normal boiling points for oxo compounds using the semi-empirical topological method. *Journal of Molecular Structure (Theochem)*, Vol.586, pp. 71–80.

Arruda, A. C. S.; Souza, E. S.; Junke, B. S.; Yunes, R. A. & Heinzen, V. E. F. (2008). Semi-empirical topological index to predict properties of halogenated aliphatic compounds. *Journal of Chemometrics*, Vol.22, pp. 186-194.

Arruda, A. C.; Heinzen V. E. F. & Yunes, R. A. (1993). *Jornal of Chromatography A*, Vol.630, pp. 251–256.

Bredow, T. & Jug, K. (2005). Theory and range of modern semiempirical molecular orbital methods. *Theorical Chemistry Accounts,* Vol.113, pp. 1–14.

Christian, R. (1990). Solvents and Solvent effects in Organic Chemistry, 2nd ed. VCH , Germany.

Duchowicz, P. R.; Castro, E. A.; Toropov, A. A.; Nesterova, A. I. & Nabiev, O. M. (2004). QSPR Modeling of the octanol/water partition coefficient of alcohols by means of optimization of correlation weights of local graph invariants. *Journal of the Argentine Chemical Society,* Vol.92, pp. 29-42.

Estrada, E. & Gutierrez, Y. (1999). Modeling chromatographic parameters by novel graph theoretical sub–structural approach, *Journal of Chromatography A*, Vol.858, pp. 187–199.

Estrada, E. (2001). Generalization of topological indices. *Chemical Physics Letters*, Vol.336, pp. 248–252.

Estrada, E. (2001). Recent Advances on the Role of Topological Indices in Drug Discovery Research. *Current Medicinal Chemistry*, Vol.8, pp. 1573–1588.

Estrada, E. (2002). The Structural Interpretation of the Randić Index. *Internet Electronic Journal of Molecular Design,* Vol.*1*, pp. 360–366, http://www.biochempress.com.

Farkas, O. & Héberger, K. (2005). Comparison of ridge regression, partial least-squares, pairwise correlation, forward-and best subset selection methods for prediction of retention indices for aliphatic alcohols. *Journal of Chemical Information and Modeling,* Vol.45, pp. 339–346.

Fritz, D. F.; Sahil, A. & Kováts, E. (1979). Study of the adsorption effects on the surface of poly–(ethylene glycol)–coated column packings. *Journal of Chromatography*, Vol.*186*, pp. 63–80.

Galvez, J.; Garcia, R.; Salabert, M. T. & Soler, R. (1994). *Journal of Chemistry Information Computational and Science*, Vol.34, pp. 520–525.

García–Domenech, R. ; Catalá–Gregori, A.; Calabuig, C.; Antón–Fos, G.; del Castillo, L. & Gálvez, J. (2002). Predicting Antifungal Activity: A Computational Screening Using Topological Descriptors. *Internet Electronic Journal of Molecular Design,* Vol.1, pp. 339–350, http://www.biochempress.com.

Guo, W.; Lu, Y. & Zheng, X. M. (2000). The predicting study for chromatographic retention index of saturated alcohols by MLR and ANN. *Talanta,* Vol.51, pp. 479–488

Hall, L. H.; Mohney, B. & Kier, L. B. (1991). *Journal of Chemistry Information Computational and Science*, Vol.31, pp. 76–82.

Hansen, P. J. & Jurs, P. C. (1988). Chemical applications of graph theory. Part I. Fundamental and topological indices. *Journal of Chemical Education,* Vol.65, pp. 574-580.

Heinzen, V. E. F. & Yunes, R. A. (1993). *Jornal of Chromatography A*, Vol.654, pp. 83–89.

Heinzen, V. E. F. & Yunes, R. A. (1996). *Jornal of Chromatography A*, Vol.719, pp. 462–467.

Heinzen, V. E. F.; Cechinel Filho V. & Yunes, R. A. (1999a). *IL Farmaco*, Vol.54, pp. 125–129.

Heinzen, V. E. F.; Soares, M. F. & Yunes, R.A. (1999b). Semi-empirical topological method for prediction of the chromatographic retention of cis- and trans- alkene isomers and alkanes. *Journal of Chromatography* A, Vol.849, pp. 495–506.

Héberger, K.; Görgényi, M. & Sjörström, M. (2000). Partial least squares modeling of retention data of oxo compounds in gas chromatography. *Chromatographia,* Vol.51, pp. 595–600.

Ivanciuc, O.; Ivanciuc, T., Carbol–Bass, D. & Balaban, A. T. (2000). Comparison of Weighting schemes for molecular graph descriptor: Application in quantitative structure–retention relationship models for alkylphenols in gas–liquid chromatography, *Journal of Chemistry Information Computational and Science*, Vol.40, pp. 732–743.

Ivanciuc T. & Ivanciuc, O. (2002). Quantitative Structure–Retention Relationship Study of Gas Chromatographic Retention Indices for Halogenated Compounds, *Internet Electronic Journal of Molecular Design,* Vol.1, pp. 94–107, http://www.biochempress.com.

Ivanciuc, O.; Ivanciuc, T.; Cabrol–Bass, D. & Balaban, A. T. (2002). Optimum Structural Descriptors Derived from the Ivanciuc–Balaban Operator. *Internet Electronic Journal of Molecular Design, Vol.1*, pp. 319–331, http://www.biochempress.com.

Junkes, B. S.; Amboni, R. D. M. C; Yunes, R. A. & Heinzen, V. E .F. (2002a). Use of a semiempirical topological method to predict the chromatographic retention of branched alkenes. *Chromatographia,* Vol.55, pp. 75–81.

Junkes, B. S.; Amboni, R. D. M. C; Yunes, R. A. & Heinzen, V. E .F. (2002b). Quantitative structure-retention relationships (QSRR), using the optimum semiempirical topological index, for methylbranched alkanes produced by insects. *Chromatographia,* Vol.55, pp. 707–715.

Junkes, B. S.; Amboni, R. D. M. C; Yunes, R. A. & Heinzen, V. E .F. (2003a). Semi-empirical topological index: a novel molecular descriptor for quantitative structure-retention relationship studies. *Internet Electronic Journal of Molecular Design*, Vol.2, pp. 33–49.

Junkes, B. S.; Amboni, R. D. M. C; Yunes, R. A. & Heinzen, V. E .F. (2003b). Prediction of chromatographic retention of saturated alcohols on stationary phases of different polarity applying the novel semi-empirical topological index. *Analytica Chimica Acta,* Vol.477, pp. 29–39.

Junkes, B. S.; Amboni, R. D. M. C; Yunes, R. A. & Heinzen, V. E .F. (2004). Application of the semiempirical topological index in the QSRR of aliphatic ketones and aldehydes on stationary phases of different polarity. *Journal of Brazilian Chemical Society*, Vol.15, pp. 183–189.

Junkes, B. S.; Arruda, A. C. S.; Yunes, R. A.; Porto, L. C. & Heinzen, V. E. F. (2005). Semiempirical topological index: application to QSPR/QSAR modeling. *Journal of Molecular Modeling,* Vol.11, pp. 128–134.

Kaliszan, R. (1987). *Quantitative structure–chromatographic retention relationships*, Wiley–Interscience, New York, USA.

Karelson, M.; Lobanov, V. S. & Katritzky, A. R. (1996). Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chemical Reviews*, Vol.96, pp. 1027–1044.

Katritzky, A. R. & Gordeeva, E. V. (1993). Traditional topological indices VS electronic, geometrical and combined molecular descriptors in QSAR/QSPR research. *Journal of Chemistry Information Computational and Science*, Vol.33, pp. 835–857.

Katritzky, A. R.; Ignatchenko, E. S.; Barcock, R. A.; Lobanov, V. S. & Karelson, M. (1994). Prediction of gas chromatographic retention times and response factors using a general quantitative structure–property relationship treatment. *Analytical Chemistry*, Vol.66, pp. 1799–1807.

Katritzky, A. R.; Chen, K.; Maran, U. & Carlson, D. A. (2000). QSRR correlation and predictions of GC retention indexes for methyl–branched hydrocarbons produced by insects, *Analytical Chemistry*, Vol.72, pp. 101–109.

Kier L. B. & Hall, L H. (1976). Molecular connectivity chemistry and drug research, Academic Press, New York, USA.

Kier, L. B. & Hall, L. H. (1986). Molecular connectivity in structure activity studies. Research Studies Press, Letchworth, UK.

Kier, L. B. & Hall, L. H. (1990). An Electrotopological State Index for Atoms in Molecules. *Pharmaceutical Research*, Vol.7, pp. 801–807.

Körtvélyesi, T., Görgényi, M. & Hérberger, K. (2001). Correlation between retention indices and quantum–chemical descriptor of ketones and aldehydes on stationary phases of different polarity. *Analytica Chimica Acta*, Vol.*428*, pp. 73–82.

Kováts, E. (1968). Zu fragen der polarität. Die method der linearkombination der wechselwirkungskräfte (LKWW). *Chimie*, Vol.22, pp. 459-462.

Liu, S.-S.; Liu, H.-L. ; Shi, Y.-Y. & Wang, L.-S. (2002). QSAR of Cyclooxygenase–2 (COX–2) Inhibition by 2,3-Diarylcyclopentenones Based on MEDV–13. *Internet Electronic Journal of Molecular Design*, Vol.*1*, pp. 310–318, http://www.biochempress.com.

Liu, F.; Liang, Y.; Cao, C. & Zhou, N. (2007). QSPR study of GC retention indices for saturated esters on seven stationary phases based on novel topological indices, *Talanta*, Vol.72, pp. 1307–1315.

Liu, F.; Cao, C. & Cheng, B. (2011). A quantitative structure-property relationship (QSPR) study of aliphatic alcohols by the method of dividing the molecular structure into substructure. *International Journal of Molecular Sciences,* Vol.12, pp. 2448-2462.

Lu, C.; Guo, W. & Yin, C. (2006). Quantitative structure-retention relationship study of gas chromatographic retention indices of saturated esters on different stationary phases using novel topological indices, *Analytica Chimica Acta,* Vol.561, pp. 96–102.

Marino, D. J. G.; Peruzzo, P. J.; Castro, E. A. & Toropov, A. A. (2002). QSAR Carcinogenic Study of Methylated Polycyclic Aromatic Hydrocarbons Based on Topological Descriptors Derived from Distance Matrices and Correlation Weights of Local Graph Invariants. *Internet Electronic Journal of Molecular Design, Vol.1*, pp. 115–133, http://www.biochempress.com.

Markuszenwski, M. &. Kaliszan, R. (2002). Quantitative structure–retention relationships in affinity high–performance liquid chromatography, *Journal of Chromatography B*, Vol.768, pp. 55–66.

Peng, C. T.; Ding, S. F.; Hua, R. L. & Yang, Z. C. (1988). Prediction of retention indexes I. Structure–retention index relationship on apolar columns. *Journal of Chromatography*, Vol.*436*, pp. 137–172.

Peng, C. T. (2000). Prediction of retention indices V. Influence of electronic effects and column polarity on retention index. *Journal of Chromatography A*, Vol.*903*, pp. 117–143.

Pompe, M. & Novic, M. (1999). Prediction of gas–chromatographic retention indices using topological descriptors, *Journal of Chemistry Information Computational and Science*, Vol.*39*, pp. 59–67.

Porto, L. C.; Souza, E. S.; Junkes, B. S.; Yunes, R. A. & Heinzen ,V. E. F. (2008). Semi-Empirical Topological Index: development of QSPR/QSRR and optimization for alkylbenzenes. *Talanta,* (Oxford). Vol.76, pp. 407-412.

Randic, M. (1993). Comparative structure-property studies: Regressions using a single descriptor. *Croatica Chemica Acta* (CCACAA), Vol.66, pp. 289–312.

Randic, M. (2001). The connectivity index 25 years after. *Journal of Molecular Graphic and Modeling,* Vol.20, pp. 19–35.

Ren, B. (1999). A new topological index for QSRR of alkanes. *Journal of Chemistry Information Computational and Science*, Vol.*39*, pp. 139–143.

Ren, B. (2002a). Application of novel atom–type AI topological indices to QSRR studies of alkanes, *Computer & Chemistry*, Vol.*26*, pp. 357–369.

Ren, B. (2002b). Novel Atomic-Level-Based AI Topological Descriptors: Application to QSPR/QSAR Modeling. *Journal of Chemical Information and Computer Sciences*, Vol.42, pp. 858–868.

Ren, B. (2002c). Novel atom-type AI indices for QSPR studies of alcohols. *Computer & Chemistry,* Vol.26, pp. 223–235.

Ren, B. (2002d). Application of novel atom-type AI topological indices in the structure-property correlations. *Journal of Molecular Structure,* (Theochem), Vol.586, pp. 137–148.

Ren, B. (2003). Atom-type-based AI topological descriptors for quantitative structure–retention index correlations of aldehydes and ketones. *Chemometrics Intelligent Laboratory Systems,* Vol.66. pp. 29–39.

Rios–Santamarina, I.; García–Domenech, R.; Cortijo, J.; Santamaria, P.; Morcillo, E. J. & Gálvez, J. (2002). Natural Compounds with Bronchodilator Activity Selected by Molecular Topology. *Internet Electronic Journal of Molecular Design,* Vol.*1*, pp. 70–79, http://www.biochempress.com.

Sabljic, A. & Trinajstic, N. (1981). Quantitative structure activity relationships: The role of topological indices. *Acta Pharmaceutica Jugoslavica,* Vol.31, pp. 189–214.

Smith, W. B. Introduction to Theoretical Organic Chemistry and Molecular Modeling. VCW: New York, 1996; 142–143.

Souza, E. S.; Junkes, B. S.; Kuhnen, C. A.; Yunes, R. A. & Heinzen, V. E. F. (2008). On a new semiempirical electrotopological index for QSRR models. *Journal of Chemometrics,* Vol.22, pp. 378-384.

Souza, E. S.; Kuhnen, C. A.; Junkes, B. S.; Yunes, R. A. & Heinzen, V. E. F. (2009a). Modelling the semiempirical electrotopological index in QSPR studies for aldehydes and ketones. *Journal of Chemometrics, Vol.*23, pp. 229-235.

Souza, E. S.; Kuhnen, C. A.; Junkes, B. S.; Yunes, R. A. & Heinzen, V. E. F. (2009b). Quantitative structure retention relationship modelling of esters on stationary phases of different polarity. *Journal of Molecular Graphics and Modelling,* Vol.28, pp. 20-27.

Souza, E. S.; Kuhnen, C. A.; Junkes, B. S.; Yunes, R. A. & Heinzen, V. E. F. (2010). Development of semi-empirical electrotopological índex using the atomic charge in QSPR/QSRR models for alcohols. *Journal of Chemometrics,* Vol.24, pp. 149-157.

Toropov, A. A. & Toropova, A. P. (2002). QSAR Modeling of Mutagenicity Based on Graphs of Atomic Orbitals. *Internet Electronic Journal of Molecular Design*, Vol.*1*, pp. 108–114, http://www.biochempress.com.

Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American Chemical Society,* Vol.69, pp. 17–20.

Yao, X.; Zhang, X.; Zhang, R.; Liu, M.; Hu, Z. & Fan, B. (2002). Prediction of gas chromatographic retention indices by the use of radial basis function neural networks. *Talanta*, Vol.*57*, pp. 297–306.
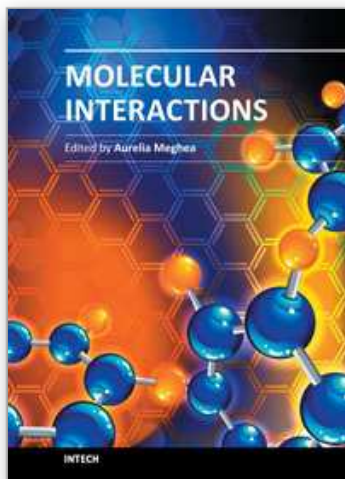
Molecular Interactions

Edited by Prof. Aurelia Meghea

ISBN 978-953-51-0079-9

Hard cover, 400 pages

Publisher InTech

Published online 29, February, 2012

Published in print edition February, 2012

In a classical approach materials science is mainly dealing with interatomic interactions within molecules, without paying much interest on weak intermolecular interactions. However, the variety of structures actually is the result of weak ordering because of noncovalent interactions. Indeed, for self-assembly to be possible in soft materials, it is evident that forces between molecules must be much weaker than covalent bonds between the atoms of a molecule. The weak intermolecular interactions responsible for molecular ordering in soft materials include hydrogen bonds, coordination bonds in ligands and complexes, ionic and dipolar interactions, van der Waals forces, and hydrophobic interactions. Recent evolutions in nanosciences and nanotechnologies provide strong arguments to support the opportunity and importance of the topics approached in this book, the fundamental and applicative aspects related to molecular interactions being of large interest in both research and innovative environments. We expect this book to have a strong impact at various education and research training levels, for young and experienced researchers from both academia and industry.

INTECH

open science | open minds