Generalized Topological Indices. Modeling Gas-Phase Rate Coefficients of Atmospheric Relevance[†]

Ernesto Estrada[*,‡] and Adelio R. Matamala[§]

Complex Systems Research Group, X-Rays Unit, RIAIDT, Edificio CACTUS,
University of Santiago de Compostela, Santiago de Compostela 15782, Spain, and
Facultad de Ciencias Químicas, Universidad de Concepción, Casilla 160-C, Concepción, Chile

We develop the idea that the use of ad hoc molecular descriptors in QSAR/QSPR studies is not an optimal solution. Instead, we propose to optimize these descriptors for the specific properties under study. In the case of topological indices (TIs) we propose the use of the *generalized topological indices* (GTIs), which account for several of the classical TIs in one single graph invariant. GTIs represent points in a six-dimensional space of topological parameters, which can be optimized for describing a specific property. The situation shows some resemblance with the geometry optimization procedures used to minimize molecular energy. Here we study the reaction rate coefficients for the gas-phase radical reactions between alkanes and cycloalkanes with OH, Cl, and NO_3 radicals. These reaction rate coefficients were studied recently by using some "classical" TIs, such as Balaban and Randić indices. Despite the fact that the Randić index produces acceptable QSPR models for some of these reactions, the models developed are still far from being optimal. Using the GTI approach we have improved these QSPRs by reducing the standard deviation by almost 50%. In addition, the current approach permits the illustration of the similarities and differences among the different descriptors studied, indicating possible directions for searching new optimal molecular descriptors.

INTRODUCTION

Topological indices (TIs) play an unquestionable role in research fields related to the prediction of physical, chemical, and biological properties of molecules. The use of these molecular descriptors have permitted the prediction of many properties of chemical, pharmaceutical, toxicological, and environmental relevance.[1−3] Beyond that "traditional" applications of TIs in quantitative structure−property (QSPR)[4] and quantitative structure−activity (QSAR) relationships[5] they have been successful in drug design,[6−8] chemical genomics,[9] proteomics,[10−12] and the study of biological interaction networks.[13,14] In all these studies the TIs used can be considered as "classical" molecular descriptors, where the term "classic" is used here to distinguish the indices that have been defined in an ad hoc way in the literature, and which are not necessarily optimal for modeling or predicting a specific property or activity. Some of these "classic" TIs were created in contexts which are very different from the QSAR/QSPR research field. This is the case of the Zagreb indices which were introduced in the context of molecular orbitals in acyclic polyenes.[15,16] Thus, there is no guarantee that the use of these "classical" TIs can predict effectively a property/activity for an arbitrary set of compounds, even for the same series of compounds originally used to define the TI. For instance, the Randić index was designed to account for molecular branching, and it was tested by correlating with the boiling points of alkanes.[17]

Basically, there are two approaches used to try to amend the lack of correlation of one particular TI with an experimental property. One of these approaches is to try with other TI with the hope that this index will give a better correlation. Another is to try by combining several TIs into a multilinear regression model. The first approach has the disadvantage that we would test a very large number of indices before obtaining a good QSAR/QSPR model. The second approach can fail when the problems of overfitting arise. Instead of these nonoptimal approaches we have proposed the use of an optimization approach for TIs. Our approach is based on the use of the generalized topological indices (GTI) scheme in combination with a simplex optimization algorithm, which permits the obtaining of optimal TIs for describing an experimental property.[18−25] An analogous situation is that of the molecular orbital methods which necessarily involves the optimization of molecular geometry before the calculation of molecular properties. There have been previous attempts to optimize TIs individually through the use of the so-called "variable TI", such as variable connectivity indices[26−28] or variable Zagreb indices.[29] However, our approach encompasses most of the "classical" TIs into one single graph invariant based on a vector-matrix-vector multiplication procedure using the generalized graph matrix. Using this approach we are able to express Zagreb indices, Randić index, Balaban index, Wiener index, Harary indices, Schultz MTI, Gutman index, and hyper-Wiener index using only one graph invariant.[18−25] Here, we will explain in detail this approach, and we will use it for optimizing models to predict gas-phase reaction coefficients of alkanes and cycloalkanes in reaction with OH, Cl and NO_3 radicals. These reactions are of relevance for tropospheric chemistry, and the predic-

---

[†] Dedicated to Professor Nenad Trinajstić on the occasion of his 70th birthday.

[*] Corresponding author fax: +34 981 547 077; e-mail: estrada66@yahoo.com.

[‡] University of Santiago de Compostela.

[§] Universidad de Concepción.

GENERALIZED TOPOLOGICAL INDICES

J. Chem. Inf. Model., Vol. 47, No. 3, 2007 **795**

tion of these rate coefficients have been considered recently in the "atmospheric chemistry" literature.

We start by considering a brief review about a recent work that suggests the usefulness of TIs in the study of free radical abstractions reactions of alkanes. Then we will continue with the main philosophy of our approach as well as the technical details of the Simplex-GTI method. Finally, we will develop QSPR models for predicting the experimental rate constants of saturated hydrocarbons, and we will compare them with the models using nonoptimal TIs.

"CLASSICAL" TOPOLOGICAL INDICES AND GAS-PHASE FREE RADICAL ABSTRACTION REACTIONS OF ALKANES

Alkanes and cycloalkanes are among the volatile organic compounds (VOC) that undergo atmospheric transformation.[30−32] In a recent paper McGillen et al.[33] asked the following question: "Can topological indices be used to predict gas-phase rate coefficients of importance to tropospheric chemistry?". The answer given by the authors to the above question was "Rate coefficients appear to share a strong linear relationship with the degree of branching of organic moieties.".[33] The authors used two "classical" TIs for correlating the rate coefficients of free radical abstraction reactions of a series of alkanes and cycloalkanes for radical moieties OH, Cl, and NO_3. Accordingly, the Randić index was found to correlate better than the ionization potential (IP) calculated from quantum-chemical semiempirical method PM3 for each of the radicals studied. The acceptable result obtained by Randić index is due the well-known power of this TI to encode the molecular branching and molecular size. So, these authors open a new way to the use of TIs: "The data presented here suggests that topological indices are an effective and accessible means of estimating the rate coefficients of alkanes with respect to OH, Cl, and NO_3 radicals.".

After the promissory result found by McGillen et al.,[33] the natural point now is how to improve the application of TIs to model the rate coefficients. Obviously, the problem is not solved by searching into the vast pool of TIs defined in the literature. All of these TIs have been defined in an ad hoc way, and none of them is necessarily optimal for predicting the free radical abstraction reaction of alkanes and cycloalkanes. Thus, the solution to this problem is simply to "build" a TI which is optimal to describe this property. This building procedure is not ad hoc but employs an optimization procedure which minimizes the error in predicting the experimental property by starting from a nonoptimal TI.

THE OPTIMIZATION PHILOSOPHY

Basically, all properties of molecules are a direct or an indirect consequence of the spatial distribution of its electron density. Both ab initio and semiempirical electronic structure methods provide that kind of information after a molecular geometry optimization procedure. The main idea is that the molecular energy of a particular electronic state is a function of geometric conformation adopted by the nuclei in the space. So, the total molecular energy for an electronic state is a function of the geometric parameters that define the conformation of a molecule, such as distances, bond angles, and torsion angles. Thus, it is possible to "modify" the total molecular energy by systematic change of these geometric

parameters. There is a very large number of possible conformations which depends on the values taken by these parameters, but only one or a few of them will represent stable molecular conformations. The objective of the molecular optimization procedures is to find the molecular wave function for those conformations of minimal energy by modifying systematically the geometric parameters of the molecule under study and, in a subsequent step, from this molecular wave function calculate the expected property.

Graph theory offers a different approach to calculate molecular properties and/or molar properties. Although the underlying idea still continues being the same: the distribution of molecular electron density is the responsibility of the molecular properties, now that kind of information is encrypted by the so-called "molecular topology". For instance, TIs modeling molecular branching are nothing else than invariants that encode the inhomogeneous distribution of electron density due to the atom−atom connectivity within the molecule.

In general, molecular/molar properties are not related to electron density in a simple way. "Classical" TIs lack the necessary degree of freedom to model complex functions between molecular topology information and the property under study. So, the existence of parameters to be optimized in the definition of topological indices is highly desired. As a matter of example we will consider the Randić index. According to Randić et al.[34] we can write down this index in the following general form:

$$\chi = \sum_{\text{bonds}} [(\delta_i + m)(\delta_j + n)]^k$$

For the sake of simplicity we will consider, as in Randić et al.[34] paper, that $m = n$. In the "classical" definition of the Randić index $m = 0$ and $k = -1/2$. This is the index used by McGillen et al.[31] to correlate the experimental rate coefficients of alkanes in their reactions with atmospheric radicals. But, how do we know whether these parameters ($m = 0$ and $k = -1/2$) are optimal for describing such a property? The Randić index calculated from these parameters represents a single "conformation" in the space of all possible indices that can be generated by using different values of m and k. Randić et al.[34] explored some of these parameters for correlating the boiling point of small alkanes. They found that the change of the exponent from $k = -1/2$ to $k = -1/3$ decreases the standard deviation from 6.746 to 5.003. However, this second set of topological parameters represents another "conformation" of the index, which is not necessarily the optimal one. Consequently, the solution to this problem consists in carrying out an optimization process in which we explore the whole space of these parameters in order to minimize the error in predicting the property under study. In this case the error represents the analogous of the energy in the molecular geometry optimization procedure as can be seen in the illustrative scheme in Figure 1.
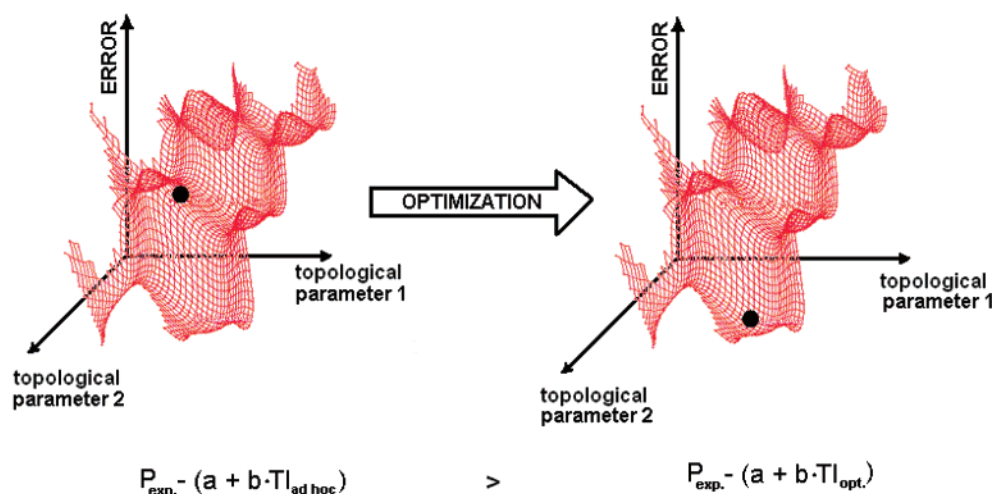
TOPOLOGICAL PARAMETER SPACE AND GENERALIZED TIS

There is a substantial difference between the molecular geometry optimization procedures and the optimization of TIs. In the first we always optimize the same set of geometric parameters, i.e., distances, bond angles, and torsion angles.
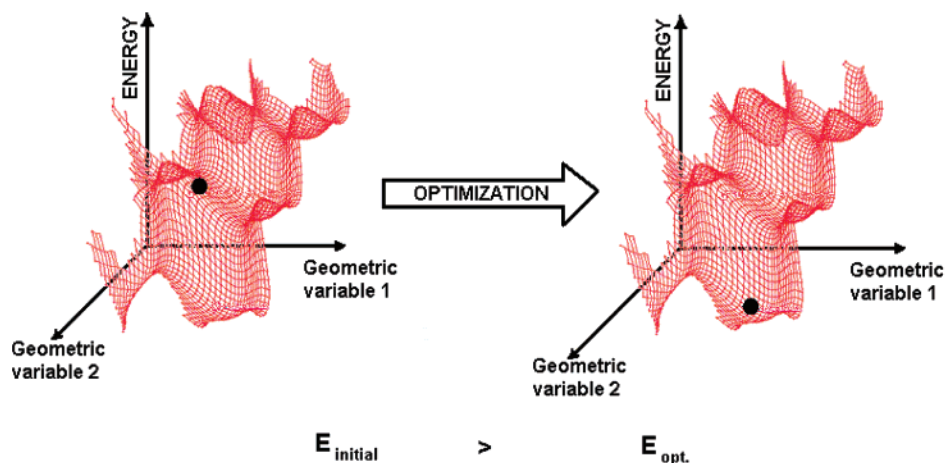
MOLECULAR GEOMETRY OPTIMIZATION

Figure 1. Schematic representation of the similarities between the molecular geometry optimization procedure and the optimization of TIs using the Simplex-GTI approach.

However, for optimizing TIs we need to select a different set of topological parameters for every index. For instance, Zagreb M_2 and Randić indices depend on the product of vertex degrees summed over all bonds, and the only parameter to be optimized is the exponent in the graph invariant. However, in the M_1 Zagreb index the sum is over atoms instead of over bonds. On the other hand, Wiener, Balaban, and Harary indices are not based on degrees but on topological distances, and they use very different invariants in their definition. There are other TIs, such as MTI, which depends on both degrees and distances. Thus, the question here is whether it is possible to find a "universal" topological parameter space in which all these TIs can be represented. If such a topological space exists, then we can consider every single TI as a point in this space, and we can move from one index to another simply by changing the topological parameters. More importantly, we can find the optimal index for correlating an experimental property by means of an optimization procedure that involves these topological parameters. We have found such a topological parameter space by means of the introduction of a generalized graph matrix and the generalized TIs approach, which is explained below.

Let $G(V,E)$ be a molecular graph with $|V| = n$ vertices and $|E| = m$ edges. Let d_{ij} be the entries of the $n \times n$ topological distance matrix of the graph $G(V,E)$. The GTI associated with the graph $G(V,E)$ is defined by the following vector−matrix−vector formula

$$\text{GTI}[G] = C\mathbf{u}^{\text{T}}(G;y,\bar{w},q)[\sum_{i=1} \gamma_i'\Gamma(G;x_i,p_i)]\mathbf{v}(G;z,\bar{s},r) \quad (1)$$

where γ_i' is a coefficient modifying the generalized graph matrices. Thus, for a given set of parameters y, z, q, r, and components of $\bar{w} = (w_1,w_2,...,w_n)$, and $\bar{s} = (s_1,s_2,...,s_n)$ vectors, we obtain

$$\text{GTI}[G] = \sum_{i=1} \gamma_i(\mathbf{u}^{\text{T}}(G;y,\bar{w},q)\Gamma(G;x_i,p_i)\,\mathbf{v}(G;z,\bar{s},r)) \quad (2)$$

where γ_i coefficients contain the information of both C and γ_i'. Eq 2 can be written in the equivalent form

$$\text{GTI}[G] = \sum_{i=1} \gamma_i\text{GTI}_i[G] \quad (3)$$

where

GENERALIZED TOPOLOGICAL INDICES

J. Chem. Inf. Model., Vol. 47, No. 3, 2007 **797**

$$GTI[G] = \begin{pmatrix} p & q & r \\ x & y & z \end{pmatrix} \begin{vmatrix} \bar{s} \\ \bar{w} \end{vmatrix}_G =$$

$$\mathbf{u}^T(G;y,\bar{w},q)\Gamma(G;x,p)\mathbf{v}(G;z,\bar{s},r) \quad (4)$$

where **u** and **v** are column vectors whose components are given by

$$u_i(G;y,q) = [w_i + \sum_{j=1}^{n} g_{ij}(y,1)]^q \quad \text{and}$$

$$v_i(G;z,r) = [s_i + \sum_{j=1}^{n} g_{ij}(z,1)]^r \quad (5)$$

The Γ matrix is the so-called *generalized molecular graph matrix* whose $n \times n$ entries are expressed in terms of the topological distance through

$$g_{ij}(G;x,p) = \begin{cases} 1 & \text{, if } d_{ij} = 1, \\ (d_{ij}x^{d_{ij}-1})^p & \text{, if } i \neq j\,;\, d_{ij} \neq 1, \\ 0 & \text{, otherwise} \end{cases} \quad (6)$$

The vectors $\bar{w} = (w_1, w_2, ..., w_n)$ and $\bar{s} = (s_1, s_2, ..., s_n)$ contain the weighting parameters for differentiating heteroatoms in certain TIs, such as the valence connectivity index. In general, for TIs $\bar{w} = \bar{s} = (0,0,...,0) = \bar{0}$.

For the sake of simplicity we will use the following compact symbol when $\bar{w} = (0,0,...,0) = \bar{0}$ and $\bar{s} = (0,0,...,0) = \bar{0}$:

$$\begin{pmatrix} p & q & r \\ x & y & z \end{pmatrix} \begin{vmatrix} \bar{0} \\ \bar{0} \end{vmatrix}_G = \begin{pmatrix} p & q & r \\ x & y & z \end{pmatrix}_G \quad (7)$$

The "classical" TIs are expressed in terms of the GTI formula instead of the very different invariants normally used in their definition as can be illustrated in Table 1.

An equivalent form in which GTIs (4) can be written is

$$GTI[G] = C \sum_{k=1}^{diam(G)} c_k \eta^{(k)}[G] \quad (8)$$

where $diam(G) = \max (k)$ is the diameter of the graph G

$$c_k = c_k(x,p) = k^p\, x^{p(k-1)} \quad (9)$$

and

$$\eta^{(k)}[G] = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \langle i,j \rangle_G \Delta_{ij}^{(k)} \quad (10)$$

The kth-order geodesic (shortest path) matrix $\Delta^{(k)}$ of the graph G is defined by the following entries:

$$\Delta_{ij}^{(k)}[G] = \begin{cases} 0, & \text{if } d_{ij} \neq k \text{ in graph } G, \\ 1, & \text{if } d_{ij} = k \text{ in graph } G \end{cases} \quad (11)$$

And, each geodesic-bracket $\langle i,j \rangle_G$ is defined by

$$\langle i,j \rangle_G = \frac{1}{2} (u_i v_j + v_i u_j) \quad (12)$$

Functions $u_i(G;y,q)$ and $v_i(G;z,r)$ adopt the form

$$u_i(G;y,q,\bar{w}) = [w_i[G] + \delta_i[G] + \sum_{k=2}^{diam(G)} k\, y^{k-1} N_i^{(k)}[G]]^q \quad (13)$$

$$v_i(G;z,r,\bar{s}) = [s_i[G] + \delta_i[G] + \sum_{k=2}^{diam(G)} k\, z^{k-1} N_i^{(k)}[G]]^r \quad (14)$$

where $\delta_i[G]$ is the classic degree of the ith vertex in the graph G, whereas the quantity $N_i^{(k)}[G]$ is the number of vertices at distance k from the ith vertex:

$$N_i^{(k)}[G] = \sum_{j=1}^{n} \Delta_{ij}^{(k)}[G] \quad (15)$$

Equation 8 shows that any GTI can be separated in terms of the contributions of the pair of vertices at the same topological distance in the graph. Each $\eta^{(k)}$ term defines the contribution to GTI of all *interactions* between pairs of vertices separated at distance k in the graph. These contributions are scaled by two real parameters through the c_k coefficients. The diameter of the graph is a global descriptor, and its presence in definition (8) reveals the dependence of any GTI on the 'size' of the molecular graph. On the other hand, the molecular-connectivity relationship among atoms in the molecule defines the 'shape' of the molecular graph. This 'shape' is coded by the so-called *geodesic-brackets* (see eq 12). Equations 13 and 14 show that the functions u and v are the generalization of the 'classic' vertex degree notion. Through these functions u and v and by settling the x and p parameters, a pair of weights is assigned to each vertex in the graph. On each vertex these weights code the topological environment around it. From the previous analysis, it is clear that the codification of the topological complexity of any molecular graph relies on the $N_i^{(k)}[G]$ quantities, defined by eq 15.

GTI AND TROPOSPHERIC REACTIONS OF ALKANES/ CYCLOALKANES

Our main objective here is to find quantitative correlations between GTIs and rate parameters for H atom abstraction reactions considered by McGillen et al.[33] Thus, we use the three data sets compiled in the work of McGillen et al.[33] for the reactions between alkanes and cycloalkanes with OH, Cl, and NO_3 radicals. The first of these reactions is generally considered the most important radical reaction in daytime troposphere, and this data set is the most complete of the three data sets studied here. When considering the data set for reaction with Cl radicals we have left out two data points, cyclopropane and 2,2-dimethylhexane. These compounds were detected by McGillen et al.[33] as statistical outliers departing from the general trend followed by the rest of compounds in this data set. The third data set is the least complete of all, reflecting the difficulties with the experimental determination of this slow reaction. It is formed by 19 data points instead of 24 for Cl radical and 32 for the reaction with OH radicals. All data are illustrated in Table 2.

We have used the so-called downhill Simplex optimization procedure for minimizing the error in predicting the reaction rate coefficients under study. We are using only the six-

Table 1. "Classical" Topological Indices, Their Original Definitions, and Compact Representation Using the GTI Approach

name	"classical" definition	GTI formula	refs
Zagreb M_1 index	$M_1(G) = \sum_i (\delta_i)^2$	$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}_G$	15, 16
Zagreb M_2 index	$M_2(G) = \sum_{bonds} (\delta_i\delta_j)$	$\frac{1}{2}\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}_G$	15, 16
Randić χ index	$\chi(G) = \sum_{bonds} (\delta_i\delta_j)^{-0.5}$	$\frac{1}{2}\begin{pmatrix} 1 & -1/2 & -1/2 \\ 0 & 0 & 0 \end{pmatrix}_G$	17
valence connectivity index	$\chi^v(G) = \sum_{bonds} (\delta_i^v\delta_j^v)^{-0.5}$	$\frac{1}{2}\begin{pmatrix} 1 & -1/2 & -1/2 \\ 0 & 0 & 0 \end{pmatrix} \| \frac{\bar{N}}{\bar{N}}_G$ $\hat{N} = (N_1, N_2, ..., N_n) = $ number of π- and p-electrons	5
Wiener W index	$W(G) = \frac{1}{2}\sum_{i,j} d_{ij}$	$\frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}_G$	35
Balaban J index	$J(G) = \frac{m}{C+1}\sum_{bonds} (s_is_j)^{-0.5}$	$\frac{m}{2(C+1)}\begin{pmatrix} 1 & -1/2 & -1/2 \\ 0 & 1 & 1 \end{pmatrix}_G$	36
Harary H indices	$H_k(G) = \frac{1}{2}\sum_{i,j} (d_{ij})^{-k}$	$H_k[G] = \frac{1}{2}\begin{pmatrix} k & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}_G$	37, 38
Gutman index	$S_G(G) = \sum_i \sum_j \delta_i\delta_j d$	$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}_G$	39
Schultz MTI	$MTI(G) = \sum_i [(\mathbf{A}+\mathbf{D})\mathbf{v}]_i$	$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}_G + \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}_G$	40
Reverse MTI	$RMTI(G) = \sum_i [(\mathbf{A}+\mathbf{RD})\mathbf{v}]_i$	$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}_G + \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}_G$	41
Hyper-Wiener index	$WW(G) = \frac{1}{4}\sum_i \sum_j (d_{ij}^2 + d_{ij})$	$\frac{1}{4}\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}_G + \frac{1}{4}\begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}_G$	42

dimensional parameter subspace formed by all 6-tuples (x,y,z,p,q,r). Let consider that $\Theta = \{\theta_1, \theta_2, ..., \theta_g\}$ is the set of reaction rate coefficients for the alkanes and cycloalkanes studied represented by the set $\hat{G} = \{G_1, G_2, ..., G_g\}$ of molecular graphs. We are assuming the existence of a 6-tuple $(x_0, y_0, z_0, p_0, q_0, r_0)$ that minimizes the six-dimensional scalar function Q

$$Q_{xyz}^{pqr} = 1 - |R(\text{GTI}[G] \leftrightarrow \Theta)| \quad (16)$$

where the symbol $|\cdot|$ means absolute value, and $R(P \leftrightarrow Q)$ denotes the linear correlation coefficient between the data P and Q. Now, to find the minimum (local or global) of the function (16), we use the downhill simplex method of optimization, which has been implemented in a computer program termed GTI-Simplex where each simplex involves seven 6-tuples of the form (x,y,z,p,q,r). Each 6-tuple defines a GTI response point and GTI-Simplex explores the GTI-space to find a minimum of function (16).

Using the GTI-simplex approach we have obtained the results illustrated in Table 3, where we give the coefficients of the linear regression models using GTI as well as Wiener and Randić indices used for comparison. The linear models obtained are of the following form

$$\log(k_x \, \text{cm}^3 \, \text{molecule}^{-1} \, \text{s}^{-1}) = A + B \cdot TI(G) \quad (17)$$

where X = OH, Cl, and NO_3. We call GTI_1, GTI_2, and GTI_3 the optimal GTIs obtained for the reactions with OH, Cl, and NO_3 radicals, respectively. As can be seen in Table 3 GTI clearly improves the results previously obtained by McGillen et al.[33] using the Randić index.[17] We recall that

the Randić index was found by McGillen et al.[33] as the best descriptor in correlating the reaction rates for the compounds studied. We have improved the correlation coefficients from 0.75 to 0.93 for the reaction with OH radicals, from 0.86 to 0.96 for the reaction with Cl radicals, and from 0.63 to 0.82 for the reaction with NO_3 radicals. The last results continue to be the least robust, but there is an appreciable improvement respect to the previous models. The standard deviation of the regressions for the reactions with OH and Cl radicals were reduced by 45% with respect to those obtained by using the Randić index, while this reduction was 26% for NO_3. In all cases we have carried out a leave-one-out cross-validation of the models and obtained the regression coefficient (CVR) and standard deviation (CVS) of the cross-validated models. As can be seen in Table 2 the values of CVR and CVS prove that these models are stable to the inclusion/exclusion of datapoints.

Despite the dramatic reduction in the standard deviations for the three reaction rates analyzed the errors in predicting these properties are still very far from the experimental errors. These experimental errors are, of course, different for different compounds. For the sake of comparison we selected the three isomers of pentane: pentane, 2-methylbutane, and 2,2-dimethylpropane. The experimental errors for the reaction rates of these three compounds in the reaction with OH radicals are respectively 0.03, 0.03, and 0.10, which are ten times lower than the standard deviation of the best model using GTI_1. On the other hand, the experimental errors for the reactions with Cl radicals for these three compounds are respectively 0.02, 0.015, and 0.05, which are below the standard deviation of the best regression model.

Table 2. Recommended Rate Parameters for the Gas-Phase Reactions Considered in This Study and the Optimal GTIs for Describing Each of Them

	molecule	$\log (k_{OH})$	$\log (k_{Cl})$	$\log (k_{NO_3})$	GTI$_1$	GTI$_2$	GTI$_3$
1	ethane	−12.61	−10.23	−17.70	1.0000	1.0000	1.0000
2	propane	−11.96	−9.98	−17.66	0.7866	0.8929	0.7420
3	cyclopropane	−13.09	−12.95[a]		1.1679		
4	n-butane	−11.63	−9.65	−16.33	0.5009	0.7491	0.2321
5	2-methylpropane	−11.67	−9.86	−16.01	0.7743	0.8806	0.4083
6	cyclobutane	−11.69	−9.98		0.5938	0.8384	
7	n-pentane	−11.42	−9.60	−16.09	0.3911	0.6701	0.1147
8	2-methylbutane	−11.44	−9.71	−15.81	0.5099	0.7536	0.1459
9	2,2-dimethylpropane	−12.08	−9.95		0.7775	0.8868	
10	cyclopentane	−11.30	−9.64		0.5399	0.7738	
11	n-hexane	−11.28	−9.51	−16.97	0.3226	0.6187	0.0750
12	2-methylpentane	−11.28	−9.59	−15.77	0.3810	0.6663	0.0777
13	3-methylpentane	−11.28	−9.61	−15.69	0.4344	0.7002	0.0778
14	2,2-dymethylbutane	−11.65			0.5445		
15	2,3-dimethylbutane	−11.24	−9.70	−15.39	0.4815	0.7316	0.0815
16	cyclohexane	−11.16	−9.51	−15.87	0.3555	0.6627	0.0210
17	n-heptane	−11.17	−9.44	−15.86	0.2903	0.5860	0.0585
18	2-methylhexane		−9.51			0.6099	
19	2,4-dimethylpentane	−11.32	−9.59	−15.84	0.3607	0.6414	0.0467
20	2,2,3-trimethylbutane	−11.42	−9.77	−15.65	0.4952	0.7469	0.0585
21	cycloheptane	−10.91			0.3227		
22	methylcyclohexane	−11.03	−9.46		0.3612	0.6689	
23	n-octane	−11.09	−9.39	−15.74	0.2741	0.5662	0.0522
24	2,2,4-trimethylpentane	−11.48	−9.65	−16.12	0.3641	0.6515	0.0355
25	n-nonane	−11.01	−9.37	−15.71	0.2710	0.5560	0.0515
26	3,3-diethylpentane	−11.32			0.4300		
27	n-decane	−10.96	−9.31	−15.59	0.2762	0.5535	0.0551
28	2,2,3,3-tetramethylbutane	−12.01	−9.81	−16.31	0.4922	0.7482	0.0401
29	2,2-dimethylhexane		−11.52[a]				
30	2,3,4-trimethylpentane	−11.19			0.3954		
31	3,4-diethylhexane	−11.16			0.3436		
32	isopropylcyclopropane	−11.58			0.4832		
33	n-butyl cyclohexane	−10.83			0.3114		
34	cyclooctane	−10.88			0.2479		

[a] Compounds not included in the correlations in a similar way as in ref 33.

Table 3. Statistical Parameters of the Linear Regression Models Relating the Experimental Rate Parameters with TIs According to a Linear Model of the Form: $\log (k_X \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}) = A + B \cdot TI(G)$

|	X = OH			X = Cl			X = NO$_3$		
	$W(G)$	$\chi(G)$	GTI$_1(G)$	$W(G)$	$\chi(G)$	GTI$_2(G)$	$W(G)$	$\chi(G)$	GTI$_3(G)$
A	−11.768	−12.528	−10.427	−9.825	−10.251	−8.364	−16.429	−17.315	−15.752
B	0.0069	0.362	−2.121	0.004	0.211	−1.809	0.007	0.421	−2.020
R	0.581	0.754	−0.930	0.690	0.857	−0.958	0.448	0.627	−0.818
S	0.405	0.327	0.183	0.165	0.118	0.065	0.599	0.522	0.386
CVR	0.504	0.702	0.919	0.612	0.824	0.950	0.264	0.467	0.773
CVS	0.261	0.267	0.179	0.141	0.111	0.067	0.349	0.370	0.350

In Figure 2 we illustrate the intercorrelation among the different GTIs as well as the Wiener and Randić indices. As can be seen GTI$_1$ and GTI$_2$ are linearly related, and they also show a modest correlation with GTI$_3$. The first two GTIs are not well correlated to the "classical" TIs, while GTI$_3$ appears to have some nonlinear relation with the Wiener and Randić indices.

## ON THE NUMBER OF OPTIMIZED PARAMETERS

Here we are using GTIs in a QSAR/QSPR model developed by optimizing a series of GTI parameters that produces the best fit with the experimental property. Then, apparently, the comparison of the GTIs with the "classical" TIs is not fair because the QSAR/QSPR models using GTIs have 6 + 2 optimized parameters, while the classical indices have only 2 adjusted parameters. The first six optimized parameters in the GTIs correspond to the *x, y, z, p, q, r*

parameters, and the last two correspond to the intercept and slope of the regression equation. However, this interpretation is not correct as we have previously shown that these "classical" TIs are also obtained from the GTI approach. We explain this point in the following paragraphs.

First, we consider a six-dimensional space of the *x, y, z, p, q, r* parameters represented by mean of a radial plot. Let consider that we define the optimization space by restricting the parameters to the range between −3 and 3, e.g., −3 ≤ *x* ≤ 3. In the radial plot this optimization region is represented by a hexagonal shadowed area as illustrated in Figure 3A. Then, a particular set of *x, y, z, p, q, r* points represents a TI, which corresponds to a particular hexagon in the radial plot (see Figure 3A). Now, let start by fixing the values of the GTI parameters to a value of zero, i.e., *x = y = z = p = q = r* = 0. This initial situation is represented by a dotted hexagon in Figure 3B.
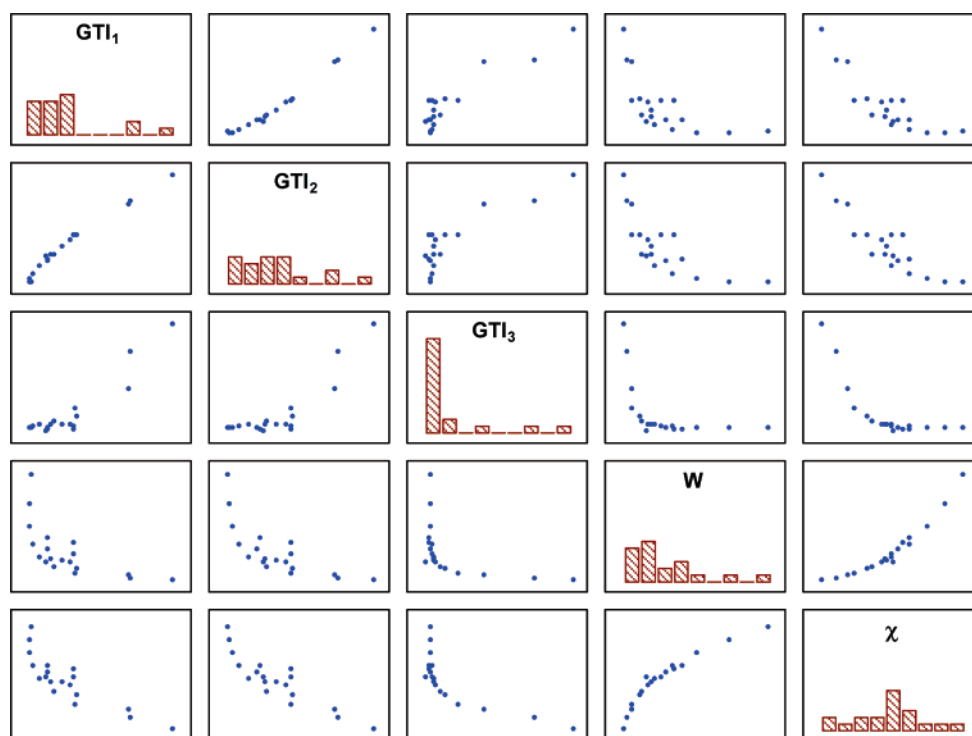
Figure 2. Intercorrelation among the three GTIs studied here as well as two "classical" TIs: Wiener and Randić indices.
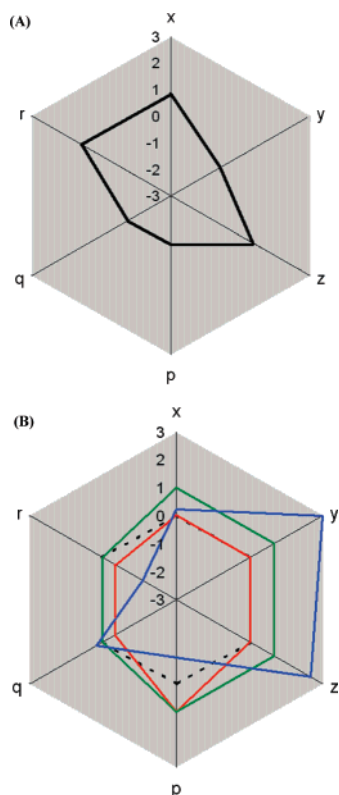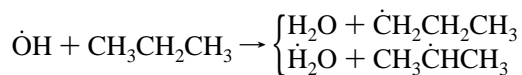

Figure 3. (A) Representation of the six dimensional space of GTIs in the form of a radial plot. The shadowed area represents a hypothetical region of optimization and the hexagon represented in black corresponds to a particular GTI. (B) Representation of the Randić (red hexagon), Wiener (green hexagon), and optimized GTI$_1$ (blue hexagon) indices. The hypothetical starting configuration for the optimization search is represented by a black dotted hexagon.

The optimization process consists in changing systematically the values of the GTI parameters in order to improve the correlation coefficient with the experimental property

under study by using a linear regression model. When the GTI parameters take the values $x = y = z = 0$, $p = 1$, $q = r = -0.5$, the hexagon represented in the radial plot (red in Figure 3B) corresponds to the Randić index. Another set of GTI parameters represents another TI. For instance, $x = y = z = p = 1$, $q = r = 0$ correspond to the Wiener index (green in Figure 3B). The situation represented in Figure 3b by a blue hexagon corresponds to the GTI which is optimal for describing the reaction rate of alkanes and cycloalkanes against OH radicals. By using this analysis it is obvious that the number of optimizable parameters in all the descriptors analyzed here are exactly the same, i.e., $6 + 2$. The cases of "classical" TIs can be considered as "intermediate states", which are not necessarily optimal for describing a given property. However, if the Randić index is optimal for describing a given property, by using the optimization process described here we will obtain the following GTI values $x = y = z = 0$, $p = 1$, $q = r = -0.5$, which obviously correspond to the mentioned index. In this case, these values of GTI parameters are optimal for describing such a property, and the Randić index can be considered as an optimized GTI descriptor to account for such property.

THE PROBLEM OF INTERPRETATION

One of the hot points of discussion about the use of topological indices in QSPR/QSAR is the one of interpretation. Here we will discuss only the problems that arise when we try to obtain a structural interpretation of the QSPR models obtained for modeling the hydrogen abstraction reactions in hydrocarbons. For the sake of brevity we will concentrate our analysis on the Randić and GTI$_1$ indices and the reaction with OH radicals. Let us take, for instance, the reaction of an OH radical with propane, whose initial step is given by

GENERALIZED TOPOLOGICAL INDICES

J. Chem. Inf. Model., Vol. 47, No. 3, 2007 **801**

$$\dot{O}H + CH_3CH_2CH_3 \rightarrow \begin{cases} H_2O + \dot{C}H_2CH_2CH_3 \\ \dot{H}_2O + CH_3\dot{C}HCH_3 \end{cases}$$

Thus, the first difficulty concerning the interpretation of a QSPR using the Randić or GTI_1 indices is the following. While the H abstraction reaction takes place on atomic centers, i.e., carbon atoms, the TIs involved in the QSPR models give bond contributions to the property under study. In order to explain this point we will use the Randić index and the QSPR model obtained by McGillen et al.[33] The model in question is given below:

$$\log(k_X \, \text{cm}^3 \, \text{molecule}^{-1} \, \text{s}^{-1}) = -12.528 + 0.362\chi$$

Now, let express the Randić index in terms of bond contributions

$$\chi = |CH_3 - CH_3| + 0.707|CH_3 - CH_2| + 0.577|CH_3 - CH| + 0.5|CH_3 - C| + 0.5|CH_2 - CH_2| + \cdots$$

where the symbol $|\cdots|$ is used to designate the number of bonds of every type. Then, by simple substitution of this linear combination into the QSPR model we can find the contribution of every bond to the reaction of H abstraction in alkanes and cycloalkanes. This substitution gives, for instance, a contribution of 0.256 for every CH_3-CH_2 bond and 0.209 for every CH_3-CH bond. But we know that the reaction is taking place on the carbon atoms and not on the carbon−carbon bonds. Thus, using this approach we cannot know what is the specific contribution of one CH_3 group directly bonded to a CH_2 or to a CH. The same situation occurs to the GTI_1 and to many other TIs. In closing, we cannot interpret these QSPR models in terms of the finer possible structural contributions, such as the contributions of chemical groups obtained by Neeb.[43]

Another choice is to interpret the QSPR models obtained in terms of the global effects of the molecular structure on the reactivity. As we have previously seen the Randić index and the GTI_1 are inversely related. While the Randić index increases with size and decreases with branching, GTI_1 decreases with size and increases with branching. Then, they have positive and negative correlations with the reactivity of hydrocarbons, respectively. Consequently, these indices account for two important effects affecting the reactivity of hydrocarbons with OH radicals: the increase of reactivity with the increase of size and the increase of reactivity with the increase of branching. We can think that the rate coefficient values increase with the number of carbon atoms as a consequence of the increment of the number of sites where the hydrogen atoms can be abstracted. On the other hand, it is well-known[44] that the reactivity trend of C−H groups toward a radical is primary < secondary < tertiary. It is obvious that a quaternary carbon will not display this reaction, because it does not have any hydrogen atom attached. Both descriptors, the Randić index and the optimized GTI_1, account very well for these two effects. Then, where is the advantage of using the optimized descriptor? From a pragmatic point of view the answer is obvious: Because the optimized descriptor improves very much the quality of the QSPR model. These improvements are a consequence of the better fit of the size and branching contributions to the reactivity of these compounds with OH radicals. In addition
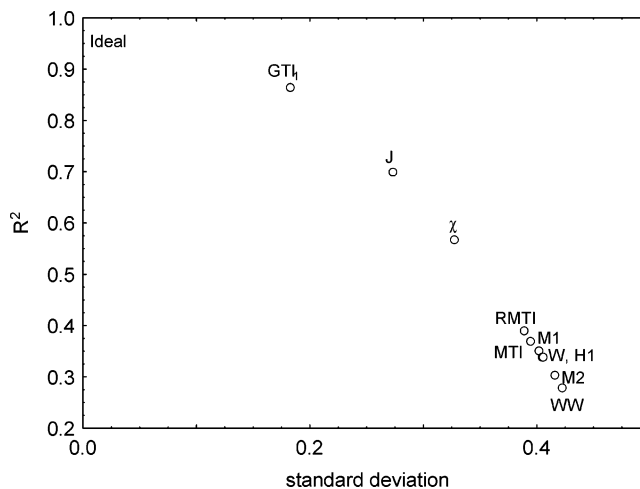
Figure 4. Plot of the squared correlation coefficient and standard deviation of the regression for the correlation of TIs and the reaction rate of hydrocarbons and OH radicals.

and as a consequence of this better fit, GTI_1 gives a better discrimination of certain pairs of molecules not well-accounted by the Randić index. As a matter of example take cyclobutane and 2,2-dimethylpropane. According to the Randić index these two molecules should display the same reactivity toward OH radicals as they both have $\chi = 2.000$. However, GTI_1 predicts that cyclobutane is more reactive than 2,2-dimethylpropane as their GTIs are 0.5938 and 0.7775, respectively. In fact, the values of $\log k$ for these two compounds are −11.69 and −12.08, which indicates that cyclobutane is 10 times more reactive than 2,2-dimethylpropane. This is an obvious additional value of the optimized descriptors with respect to the ad hoc ones.

THE PROBLEM OF SELECTING TIS

Currently there is a vast amount of molecular descriptors including several hundreds of TIs. Thus an obvious problem for developing QSPR/QSAR models is how to select the best molecular descriptor(s) from this pool. The most used approach is simply to take all of them and using appropriate statistical techniques to select the "best" one for fitting the property under study. However, as we have seen in this work no one of the existing TIs is necessarily an optimal descriptor for fitting a specific property because they have been defined in an ad hoc way. Then, we can consider the Zagreb indices, Randić index, Balaban index, Wiener index, Harary indices, Schultz MTI, Gutman index, and hyper-Wiener index as our pool of TIs to describe the reactivity of hydrocarbons toward OH radicals. Note that McGillen et al.[33] uses a definition of the Balaban index which differs from the original one. In Balaban's original definition, as in fact which is known as the Balaban index, the sum of bond weights (see Table 1) is multiplied by the term $m/(m - n + 2)$. However, McGillen et al.[33] multiplied the sum by the number of bonds m, which give a catastrophic result for the QSPR model. Here we use the original definition of the Balaban index.[36]

We obtain the results illustrated in Figure 4 for the correlation coefficient and the standard deviation of the regression model. As can be seen the optimized GTI gives the best results compared to the rest of TIs. GTI_1 is the closest index to the ideal case of correlation coefficient one and standard deviation zero. Obviously, there are several other

Table 4. Optimal GTI Parameters Obtained from the Simplex Optimization Procedure for the Three Reactions Studied in This Work

	x	y	z	p	q	r
X = OH	0.234	2.965	2.495	−0.949	0.254	−1.615
X = Cl	0.249	3.081	1.275	−0.057	0.291	−1.711
X = NO$_3$	1.308	4.169	1.291	−2.551	1.213	−4.367

TIs not included in this analysis. If we include these other indices, then we need to compare them with their optimized versions, which are not accounted by the GTI approach at its current state. Thus, such a comparison is not appropriate in the spirit of the current work.

Another advantage of the GTI approach is that it permits the analysis of the similarities and differences among the different TIs. This kind of comparison permits the analysis of what are the structural topological parameters that describe better any experimental property. If we compare, for instance, the optimal GTI for describing the reaction with OH radicals with the Randić index, then we can understand why this "classical" TI fails in describing this experimental property effectively. As can be seen in the compact bracket representation of these two indices given below the Randić index does not consider the "interactions" between nonbonded atoms, which are accounted for by the x,y,z parameters:

$$\chi = \frac{1}{2}\begin{pmatrix} 1 & -1/2 & -1/2 \\ 0 & 0 & 0 \end{pmatrix}_G$$

$$GTI_1 = \frac{1}{2}\begin{pmatrix} -0.949 & 0.254 & -1.615 \\ 0.224 & 2.965 & 2.495 \end{pmatrix}_G$$

These kinds of "interactions" taken into account by the x,y,z parameters, which takes values different from zero in GTI, appear to play an important role in explaining the reaction rates of alkanes and cycloalkanes not only versus OH radicals but also versus Cl and NO$_3$ radicals (see the values of these parameters in Table 4).

In the GTI approach every TI represents a single point in a 6-dimensional space formed by the (x,y,z,p,q,r) parameters. In order to visualize the representation of these indices we use the radial graphic with six axes previously used. Every index is then represented as a hexagon as illustrated in Figure 5A (see the parameters (x,y,z,p,q,r) for the three GTIs in Table 4). We can use the Euclidean distance as a measure of the dissimilarity among the different indices. If $p_i(I_1)$ is any of the six parameters (x,y,z,p,q,r) corresponding to the index I_1, then the dissimilarity between the indices I_1 and I_2 is obtained as follows:

$$d(I_1,I_2) = \sqrt{\sum_i [p_i(I_1) - p_i(I_2)]^2}$$

In Figure 5B we illustrate the dendogram representing the dissimilarities among 5 "classical" TI (Zagreb M_1 and M_2,[15,16] Randić χ index,[17] Wiener W index,[35] and Balaban J index[36]) and the three GTIs obtained in this work. As can be seen Zagreb M_1 and M_2 indices are the most similar pair of descriptors among these eight TIs. Wiener W index and Balaban J index form a second cluster indicating their similarity produced by the fact that both are based on topological distances. Curiously, Randić χ index occupies an intermediate place between Zagreb indices and the cluster
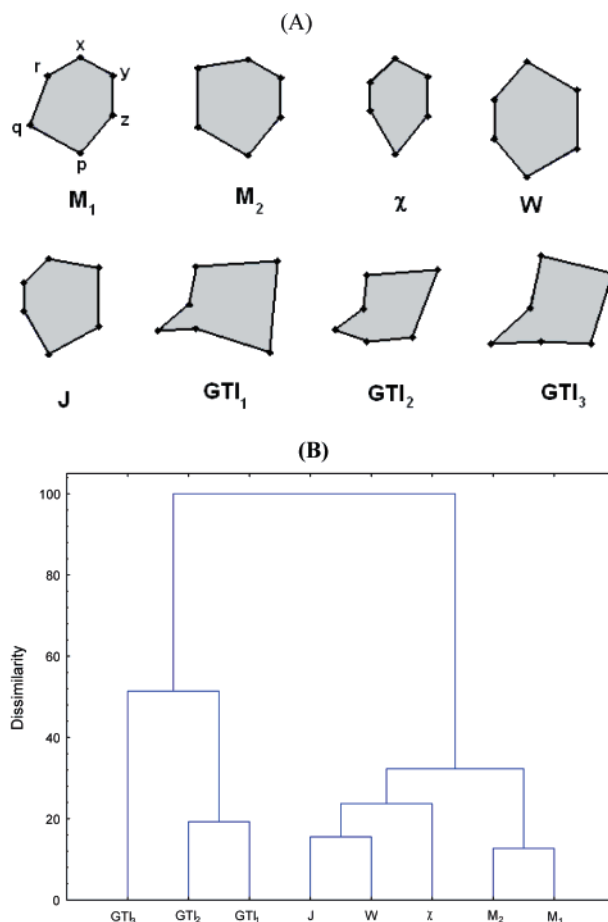
Figure 5. (A) Representation of some "classical" TIs as well as the three optimized GTIs obtained in this work. This representation is based on considering a radial system consisting on the six topological parameters of the GTI approach and then taking the hexagonal figures formed for every index. (B) Dendogram showing the percentages of dissimilarities among the different TIs and GTIs studied in this work.

formed by W and J. This index is based on vertex degrees as the Zagreb indices but uses the same invariant as the Balaban J index. The most interesting question in this analysis is that the three optimal GTIs are not similar to any of the "classical" TIs.

THE PROBLEM OF MULTILINEAR CORRELATIONS

Multilinear correlation is the most used approach in QSPR/QSAR studies. It consists of obtaining linear fits between the property under study and two or more molecular descriptors. One approach is to use a set of related descriptors defined in a similar way which conform a family of topological indices, such as the molecular connectivity indices. Another choice is to use a large pool of different descriptors and to find the best multilinear correlation for the property under study. We will consider only this second approach here. The reason is that the GTI approach has been defined to account for a series of different TIs, such as the ones defined in Table 1. In order to compare GTI with a series of ad hoc descriptors forming a family, such as the molecular connectivity, we would need to define a way of optimizing every single index in this family, which is obviously out the scope of the current work. Thus, we will develop multilinear QSPR models using the descriptors defined in Table 1.

GENERALIZED TOPOLOGICAL INDICES

J. Chem. Inf. Model., Vol. 47, No. 3, 2007 **803**

Using linear regression analysis we have derived the best multilinear fits using two, three, and four TIs. No one significant model was obtained by using only two of the TIs described in Table 1. The best three-variables model is the one obtained by combining the Zagreb M1 index, the Randić index, and the Wiener index

$$\log k = -12.181 + 0.788M1 - 1.167\chi - 0.031W$$

which has a square correlation coefficient of 0.868 and a standard deviation of the regression of 0.187. In other words, this model using three TIs accounts for the same variance in the experimental property than the QSPR model obtained using one single, but optimized, GTI.

In a previous work one of the current authors proposed an Occam's razor for the QSPR/QSAR model which can be stated in the following way:[45] ***QSPR/QSAR Occam's Razor*—** Given two QSPR/QSAR models with the same predictive error, the simpler one should be preferred because simplicity is desirable in itself.

Assuming this paradigm for selecting QSPR/QSAR models it is obvious that the model using only one single GTI is very much preferred to that using three ad hoc molecular descriptors. In addition, four molecular descriptors, Zagreb M1 index, the Randić index, the Wiener index, and the second-order Harary index, are necessary to account for more variance in the experimental property than the QSPR model obtained using the GTI$_1$. This model has a squared correlation coefficient of 0.894 and a standard deviation of the regression of 0.160. The question about whether this four-variables model is preferred to the one using one single optimized index is open to debate and will depend very much on the necessity for simplicity or accuracy for every specific problem.

Another open question is that here we have considered the optimization of TIs one by one. That is, in order to produce the best QSPR model we have optimized the six GTI parameters to produce a single optimal index. However, the current approach is not limited to this specific objective, and it is possible to optimize this set of parameters to produce two or more optimal GTIs, which will be used in multilinear QSPR/QSAR models in a similar way as the families of indices are used. In any case it is obvious that the optimization of the molecular descriptors is a better route to obtain good QSPR/QSAR models than the use of ad hoc descriptors.

CONCLUDING REMARKS

We have shown here how QSPR models can be improved by using optimized TIs instead of the ad hoc descriptors originally defined. We have reached improvements of about 50% in the linear correlations between the rate coefficients for the gas-phase reaction of alkanes and cycloalkanes with radicals of atmospheric relevance. These improvements have been obtained by considering the optimization of the molecular descriptors using the Simplex-GTI approach. This approach represents a condensed way of expressing TIs because we can define several of the "classical" TIs using the same graph invariant. Among these indices we have been able to represent the Wiener index, Balaban *J* index, Randić χ index, molecular connectivity index, Zagreb indices, Harary indices, Schultz MTI, Gutman S_G index, and hyper-Wiener index. However, more mathematical work is necessary to

account for several other molecular descriptors using the same invariant based on the generalized graph matrix. From a practical point of view the current approach permits the optimal representation of an experimental property in terms of topological molecular characteristics. This simply means that we can obtain optimized TIs that effectively describe a physical, chemical, or biological property. Even more importantly we can obtain insights about the similarities among the optimal TIs and the "classical" ones indicating in which directions we need to search for optimal descriptors. There is still a long road to travel for establishing this approach in the context of QSAR/QSPR studies. We need more studies applying the GTI approach to solving real-world problems. However, we think that the current work is a clear indication of the feasibility of this approach for solving QSAR/QSPR problems of relevant importance.

REFERENCES AND NOTES

(1) *Topological Indices and Related descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon & Breach: Amsterdam, 1999.

(2) Estrada, E.; Uriarte, E. Recent Advances on the Role of Topological Indices in Drug Discovery Research. *Curr. Med. Chem.* **2001**, *8*, 1573−1588.

(3) Votano, J. R. Recent uses of topological indices in the development of in silico ADMET models. *Curr. Opin. Drug Discovery Dev.* **2005**, *8*, 32−37.

(4) Mihalić, Z.; Trinajstić, N. A Graph Theoretical Approach to Structure Property Relationships. *J. Chem. Educ.* **1992**, *69*, 701−712.

(5) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; John Wiley Publ.: London, 1986.

(6) Grassy, G.; Calas, B.; Yasri, A.; Lahana, A.; Woo, J.; Iyer, S.; Kaczorek, M.; Floc'h, R.; Buelow, R. Computer-assisted rational design of immunosuppressive compounds. *Nature Biotech.* **1998**, *16*, 748−752.

(7) Mahmoudi, N.; de Julián-Ortiz, J.-V.; Ciceron, L.; Gálvez, J.; Mazier, D.; Danis, M.; Derouin, F.; García-Domenech, R. Identification of new antimalarial drugs by linear discriminant analysis and topological virtual screening. *J. Antimicrob. Chemother.* **2006**, *57*, 489−497.

(8) Duart, M.; García-Domenech, R.; Gálvez, J.; Alemán, P. A.; Martín-Algarra, R. V.; Antón-Fos, G. M. Application of a mathematical topological pattern of antihistaminic activity for theselection of new drug candidates and pharmacology assays. *J. Med. Chem.* **2006**, *49*, 3667−3673.

(9) Haggarty, S. J.; Clemons, P. A.; Schreiber, S. L. Chemical genomics profiling of biological networks using graph theory and combinations of small molecule perturbations. *J. Am. Chem. Soc.* **2003**, *125*, 10543−10545.

(10) Randić, M. On graphical and numerical characterization of proteomics maps. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1330−1138.

(11) Randić, M.; Nović, M.; Vracko, M. Novel characterization of proteomics maps by sequential neighborhoods of protein spots. *J. Chem. Inf. Model.* **2005**, *45*, 1205−1213.

(12) Randić, M. Quantitative characterizations of proteome: dependence on the number of proteins considered. *J. Proteome Res.* **2006**, *5*, 1575−9.

(13) Estrada, E. Virtual identification of essential proteins withing the protein interaction network of yeast. *Proteomics* **2006**, *6*, 35−40.

(14) Estrada, E. Protein bipartivity and essentiality in the yeast protein-protein interaction network. *J. Proteome Res.* **2006**, *5*, 2177−2184.

(15) Gutman, I.; Trinajstić, N. Graph theory and molecular orbitals. Total -electron energy of alternant hydrocarbons. *Chem. Phys. Lett.* **1972**, *17*, 535−537.

(16) Gutman, I.; Ruscić, B.; Trinajstić, N.; Wilcox, C. F. Graph theory and molecular orbitals. XII. Acyclic polyenes. *J. Chem. Phys.* **1975**, *62*, 3399−3405.

(17) Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.

(18) Estrada, E. Generalization of topological indices. *Chem. Phys. Lett*. **2001**, *336*, 248−252.

(19) Estrada, E.; Gutierrez, Y. The Balaban J index in the multidimensional space of generalized topological indices. Generalizations and QSPR improvements. *MATCH: Comm. Math. Comput. Chem*. **2001**, *44*, 155−167.

(20) Estrada, E. The Wiener number in the context of generalized topological indices. In *Topology in Chemistry. Discrete Mathematics of Molecules*; Rouvray, D. H., King, R. B., Eds.; Horwood Pub.: Chichester, 2002.

(21) Matamala, A. R.; Estrada, E. Generalized topological indices: Optimization methodology and physico-chemical interpretation. *Chem. Phys. Lett*. **2005**, *410*, 343−347.

(22) Matamala A. R.; Estrada, E. Simplex optimization of generalized topological index (GTI-Simplex): A unified approach to optimize QSPR models. *J. Phys. Chem. A* **2005**, *109*, 9890−9895.

(23) Estrada, E. Generalized graph matrix, graph geometry, quantum chemistry, and optimal description of physicochemical properties. *J. Phys. Chem. A* **2003**, *107*, 7482−7489.

(24) Estrada, E. Three-dimensional generalized graph matrix, Harary descriptors, and a generalized interatomic Lennard-Jones potential. *J. Phys. Chem. A* **2004**, *108*, 5468−5473.

(25) Estrada, E.; Matamala, A. R. GTI-Space: The space of generalized topological indices. *J. Math. Chem*. **2007**, in press.

(26) Randić, M. Novel graph-theoretical approach to heteroatoms in QSAR. *Chemom. Intell. Lab. Syst*. **1991**, 213−227.

(27) Randić, M.; Pompe, M. The variable connectivity index 1chi(f) versus the traditional molecular descriptors: a comparative study of 1chi(f) against descriptors of CODESSA. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 631−638.

(28) Soskić, M.; Plavsić, D. Modeling the octanol-water partition coefficients by an optimized molecular connectivity index. *J. Chem. Inf. Model.* **2005**, *45*, 930−938.

(29) Milicević, A.; Nikolić, S.; Trinajstić, N. On reformulated Zagreb indices. *Mol. Diversity* **2004**, *8*, 393−399.

(30) Seinfeld, J. H. Global atmospheric chemistry of reactive hydrocarbons. In *Reactive Hydrocarbons in the Atmosphere*; Hewitt, C. N., Ed.; Academic Press: London, 1999.

(31) Atkinson, R.; Arey, J.; Hoover, S.; Preston, K.; Atmospheric chemistry of gasoline-related emissions: formation of pollutants of potential concern. *Atmospheric Chemistry Overview*; *EPA*. http://www.oehha-.org/air/pdf/atmosChemGas.pdf (accessed 2006).

(32) Rivett, A. C.; Martin, D.; Gray, J.; Price, C. S.; Kickless, G.; Simmonds, P. G.; O' Doherty, S. J.; Greally, R.; Knights, A.; Shallcross, D. E. The role of volatile organic compounds in the polluted

(33) urban environment of Bristol, England. *Atmos. Chem. Phys*. **2003**, *3*, 1165−1176.

McGillen, M. R.; Percival, C. J.; Raventos-Duran, T.; Sánchez-Reyna, G.; Shallcross, D. E. Can topological indices be used to predict gas-phase rate coefficients of importance to tropospheric chemistry? Free radical abstraction reactions of alkanes. *Atmos. Environ*. **2006**, *40*, 2488−2500.

(34) Randić, M.; Hansen, P. J.; Jurs, P. C. Search for useful graph theoretical invariants of molecular structure. *J. Chem. Inf. Comput. Sci*. **1988**, *28*, 60.

(35) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc*. **1947**, *69*, 17−20.

(36) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett*. **1982**, *89*, 399−404.

(37) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary index for the characterization of chemical graphs. *J. Math. Chem*. **1993**, *12*, 235−250.

(38) Ivanciuc, O.; Balaban, T. S.; Balaban, A. T. Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices. *J. Math. Chem*. **1993**, *12*, 309−318.

(39) Gutman, I. Selected properties of the Schultz molecular topological index. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 1087−1089.

(40) Schultz, H. P. Topological organic chemistry. 1. Graph theory and topological indices of alkanes. *J. Chem. Inf. Comput. Sci*. **1989**, *29*, 227−228.

(41) Schultz, H. P.; Schultz, T. P. Topological organic chemistry. 11. Graph theory and reciprocal Schultz-type molecular topological indices of alkanes and cycloalkanes. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 853−857.

(42) Randić, M. Novel molecular descriptor for structure-property studies. *Chem. Phys. Lett*. **1993**, *211*, 478−483.

(43) Neeb, P. Structure-reactivity based estimation of the rate constant for hydroxyl radical reaction with hydrocarbons. *J. Atmos. Chem*. **2000**, *35*, 295−315.

(44) Carey, A F. A.; Sunberg, R. J. *Advanced Organic Chemistry. Part A: Structure and Mechanisms*; Plenum Press: New York, 1977; Chapter 12, p 524.

(45) Estrada, E.; Patlewicz, G. On the usefulness of graph-theoretic descriptors in predicting theoretical parameters. Phototoxicity of polycyclic aromatic hydrocarbons (PAHs). *Croat. Chem. Acta* **2004**, *77*, 203−211.