

# Topological Quantum Similarity Indices Based on Fitted Densities: Theoretical Background and QSPR Application

Anna Gallegos<sup>†</sup> and Xavier Gironés<sup>\*,‡</sup>

Institute of Computational Chemistry, Department of Chemistry, University of Girona, Campus Montilivi, 17071 Girona, Spain, and Medicinal Chemistry Department, AstraZeneca, R&D Mölndal, Pepparedsleden 1, S-43183 Mölndal, Sweden

Received September 15, 2004

A novel methodology to derive adjacency matrices for 3-dimensional molecules is presented in this work. These matrices, which are derived from atomic quantum similarity calculations, allow redefining several adjacency matrix-based topological indices, such as Randić, Zagreb or Chi's. The present derivation is built upon a previous work, where a simpler class of density function was used to describe the atoms composing the molecule, and the present proposal suggests using fitted atomic densities from the Atomic Shell Approximation procedure, which has been proved to appropriately reproduce, in its application space, atomic ab initio densities. The construction of such matrices as well as the derived indices are presented, along with some QSPR, where these topological indices are used as molecular descriptors for heat of formation, inhibition, and toxicity with promising results.

## INTRODUCTION

The recent years have witnessed the growth of both the foundation of new concepts based on graph theory<sup>1</sup> as well as the formulation of new topological descriptors and their application in different fields of chemistry in general and in QSPR in particular.<sup>2–6</sup> Starting from the work of Wiener on the boiling points of paraffins,<sup>7</sup> where molecules were described by path counts on the basis of the molecular graph representation, namely the Wiener Index, a constantly growing list of about 400 different Topological Indices (TI) have been formulated, which describe different molecular features such as shape (Kier and Hall kappa index<sup>8</sup>), ramification (Randić's index<sup>9</sup>), cyclization (Balaban's index<sup>10</sup>), and electronic features (Kier and Hall E-State<sup>11</sup> and Gálvez charge distribution indices<sup>12</sup>) to quote a few ones.

Many of these initial topological indices are based on the 2-dimensional (2D) molecular graph theory, from which the known adjacency and topological distance matrices are derived, and in many cases this simple representation suffices to obtain a reasonable molecular model; in other cases, a more realistic, namely 3D representation, is needed. To deal with this problem, some indices have been modified to account for a 3D molecular graph by using the real 3D interatomic distance matrix instead of the topological distance, arising a new breed of 3D-TI;<sup>1,13,14</sup> particularly, some indices were defined to introduce quantum chemical characteristics,<sup>15–17</sup> thus taking into account the molecular electronic distribution.

Parallel to these developments, Lobato et al.<sup>18</sup> introduced the concept of Topological Quantum Similarity Index (TQSI), which boards the 3D problem not only by looking

at the distance matrix but also by focusing on the adjacencies. In this case, instead of providing discrete values, depending on if a pair of atoms is actually bonded, a matrix built up from scaled interatomic quantum similarity values was used. However, this approach lacked a proper clear definition of the atomic densities used to describe the individual atoms, and hence we suggest the use of fitted atomic densities from the extensively validated Atomic Shell Approximation (ASA)<sup>19–23</sup> frame, to provide a robust, yet computationally affordable layer to compute novel TI.

## COMPUTATIONAL METHODS

**Quantum Similarity Measures.** The basic idea behind Quantum Similarity Measures (QSM) is that comparison between different objects, namely molecules, can be done from their respective electron distributions; so that given two prealigned molecules, a QSM is defined as the integral:

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

However, if QSM involve just pairs of atoms instead of full molecules, the previous equation still holds but allows to obtain pairwise atomic similarity measures, where the only parameter, besides the required atomic densities, is the distance between them, which can be easily extracted from available 3D molecular coordinates.

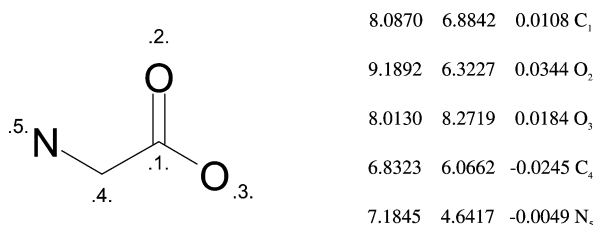
**Atomic Shell Approximation.** This breed of fitted densities arose from the cost of calculating eq 1 using ab initio densities. The idea behind Atomic Shell Approximation (ASA)<sup>19–23</sup> is to express electron density as a discrete contribution of 1S atomic densities, which minimize the error between the original density and the fitted one. Mathematically, the atomic ASA density is expressed as

$$\rho_a(\mathbf{r}) = \sum_{i \in a} w_i \exp(-\alpha_i |\mathbf{r} - \mathbf{R}_i|^2) \quad (2)$$

\* Corresponding author phone: +460 31 7064011; fax: +46 31 7763700; e-mail: Xavier.Girones@AstraZeneca.com.

<sup>†</sup> University of Girona.

<sup>‡</sup> AstraZeneca.



**Figure 1.** 2D, hydrogen suppressed, representation and 3D coordinates of glycine.

where the  $\{w_i\}$  are coefficients of the linear expansion,  $\{\alpha_i\}$  is the exponents of the 1S functions,<sup>21</sup> and  $\mathbf{R}$  is the position vector. Different ASA fittings can be browsed and downloaded from the WWW.<sup>23</sup>

**Topological Indices based on Quantum Similarity.** As stated in the Introduction section, it is known that many topological indices use the adjacency matrix. Therefore, our proposal, similar to the previous study,<sup>18</sup> is to substitute this matrix by a Quantum Similarity-based one, where the elements of such a matrix are chosen to be intramolecular interatomic quantum similarity measures between the composing atoms. To obtain comparable values in the similarity matrix, the following scaling is performed:

-similarly to classic topology, the diagonal elements are set to 0.

-atomic densities are scaled by the square root of the connectivity, therefore counting hydrogens of the involved atom (e.g. an oxygen in an ether function is scaled by  $\sqrt{2}$ , but by 1 in a carbonyl group). That is, the expression of interatomic similarity is finally expressed as:

$$Z_{ab} = \sqrt{V_a} \sqrt{V_b} \int \rho_a(\mathbf{r}_1) \rho_b(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2$$

$$= \sqrt{V_a} \sqrt{V_b} \sum_i \sum_j w_i w_j \exp \left( - \frac{\alpha_{i,a} \alpha_{j,b}}{\alpha_{i,a} + \alpha_{j,b}} |\mathbf{r}_i - \mathbf{r}_j|^2 \right) \times \left( \frac{\pi}{\alpha_{i,a} + \alpha_{j,b}} \right)^{3/2} \quad (3)$$

As an illustrative example, a comparison between the two matrices is done for the molecule of glycine. The atomic numbering and coordinates are shown in Figure 1, and the resulting matrices as well as the integer topological and float 3D distances in Table 1.

As seen in Table 1, the actual figures in both matrices, even quantitatively different, follow the same tendency, so that bonded atoms provide a higher value than unbounded pairs.

**Formulation of TQSI.** Topological indices within the Quantum Similarity framework can be defined in the same way as classical ones but substituting the required matrices as appropriate:

-Sums over edges are taken identically, as values from Z matrix are never 0, except for the diagonal terms.

-Atomic valences are taken as the sum of the elements of each row.

-Integer topological distances are substituted by the 3D float distances.

As an illustrative example, classical and Quantum Similarity (QS)-based Randić indices are calculated step-by-step for glycine. (Note that the topological valences used to calculate classical Randić index also include bonding patterns [1,3] and not just the adjacency matrix.)

$$\text{Classical } \chi = \frac{1}{\sqrt{4 \times 2}} + \frac{1}{\sqrt{4 \times 1}} \frac{1}{\sqrt{4 \times 2}} \frac{1}{\sqrt{2 \times 1}} = 1.914 \quad (4)$$

$$\text{QS } \chi = \frac{2.3213}{\sqrt{7.3507 \times 2.9778}} + \frac{2.5302}{\sqrt{7.3507 \times 3.0859}} + \frac{2.1032}{\sqrt{7.3507 \times 5.4431}} + \frac{2.7186}{\sqrt{5.4431 \times 3.2431}} = 2.007 \quad (5)$$

So, any topological index definition, which is expressed as contributions of connectivities, can be translated from the classical to the QS formulation by including the interatomic QSM and the 3D distances as appropriate. So far, the indices that have been considered are compiled in Table 2.

As seen in Table 2, no index purely based on topological distances, e.g. the Wiener Index, is explored, as the adjacency matrix does not play any role.

**3D Dependence.** As many 3D molecular descriptors, TQSI are ultimately dependent on the supplied 3D coordinates. Cases, which normally do not affect classical indices, where variability arises include the following:

-Conformers: rotation around rotatable bonds introduces changes both in the intermolecular distances and in the interatomic similarities. As an example, Table 3 presents a

**Table 1.** Comparison of the Adjacency ( $T$  and  $Z$ ) and Distance ( $D_{\text{int}}$  and  $D_{\text{float}}$ ) Matrices for Glycine Using Classical and Quantum Similarity-Based Approaches

classical	quantum similarity-based
$T = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	$Z = \begin{pmatrix} 0 & 2.3213 & 2.5302 & 2.1032 & 0.3961 \\ 2.3213 & 0 & 0.2313 & 0.3015 & 0.1237 \\ 2.5302 & 0.2313 & 0 & 0.3198 & 0.0047 \\ 2.1032 & 0.3015 & 0.3198 & 0 & 2.7186 \\ 0.3961 & 0.1237 & 0.047 & 2.7186 & 0 \end{pmatrix}$
$D_{\text{int}} = \begin{pmatrix} 0 & 1 & 1 & 1 & 2 \\ 1 & 0 & 2 & 2 & 3 \\ 1 & 2 & 0 & 2 & 3 \\ 1 & 2 & 2 & 0 & 1 \\ 2 & 3 & 3 & 1 & 0 \end{pmatrix}$	$D_{\text{float}} = \begin{pmatrix} 0 & 1.2372 & 1.3897 & 1.4982 & 2.4173 \\ 1.2372 & 0 & 2.2766 & 2.3715 & 2.6165 \\ 1.3898 & 2.2766 & 0 & 2.5022 & 3.7236 \\ 1.4982 & 2.3715 & 2.5022 & 0 & 1.4675 \\ 2.4173 & 2.6165 & 3.7236 & 1.4675 & 0 \end{pmatrix}$

**Table 2.** Compilation of Classical and QS-Based Indices Formulation

index	classical	QS-based
Randić ( $\chi$ ) <sup>9</sup>	$\sum_{\text{bonds}} \frac{T_{ij}}{\sqrt{\partial_i^T \partial_j^T}}$	$\sum_{\text{bonds}} \frac{Z_{ij}}{\sqrt{\partial_i^Z \partial_j^Z}}$
MTI <sup>24,25</sup>	$\text{Sum}(\partial^T \times (T + D_{\text{int}}))$	$\text{Sum}(\partial^Z \times (Z + D_{\text{float}}))$
Zagreb (M1 and M2) <sup>26,27</sup>	$M_1 = \sum_i \partial_i^T$	$M_1 = \sum_i \partial_i^Z$
	$M_2 = \sum_i \sum_j \partial_i^T \partial_j^T$	$M_2 = \sum_i \sum_j \partial_i^Z \partial_j^Z$
generalized connectivity (order $m$ , type $t$ ) <sup>3</sup>	${}^m\chi_t^T = \sum_i \prod_{j=1}^{m+1} (\partial_i^T)^{-1/2}$	${}^m\chi_t^Z = \sum_i \prod_{j=1}^{m+1} (\partial_i^Z)^{-1/2}$
largest eigenvalue <sup>27</sup>	$x_1 = \max(\text{EigenValues}(T))$	$x_1 = \max(\text{EigenValues}(Z))$
Xu <sup>28</sup>	$Xu = n^{1/2} \log \left( \frac{\sum_i ((\sum_j T_{ij})(\sum_j D_{\text{int},ij})^2)}{\sum_i ((\sum_j T_{ij})(\sum_j D_{\text{int},ij}))} \right)$	$Xu = n^{1/2} \log \left( \frac{\sum_i ((\sum_j Z_{ij})(\sum_j D_{\text{float},ij})^2)}{\sum_i ((\sum_j Z_{ij})(\sum_j D_{\text{float},ij}))} \right)$

**Table 3.** Conformers and Associated QS–Randić Indices for Glycine


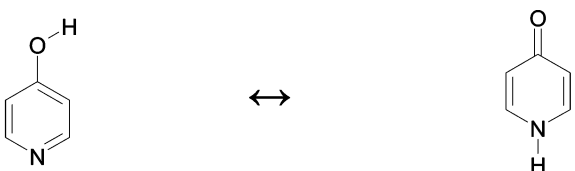
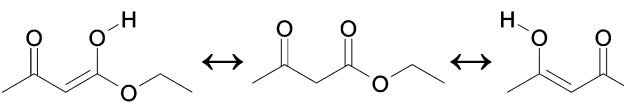
		
2.007	2.019	2.018
2.021	2.018	2.018
2.023		

figure of 7 different conformers, aligned along with the associated QS–Randić index, found by conformational analysis of glycine,<sup>29</sup> where the values can be compared to the classical value as in eq 4, which only depends on the connectivities. So, even for such a small molecule, there is some variation, which is to increase with larger, more flexible, molecules. To deal with this fact, it is suggested that, in the absence of any kind of empirical coordinates, a conformational analysis should be carried out in order to pick the lowest energy conformer or to average the results among the conformers found.

–Tautomerism: keto–enolic interconversions introduce not only variations in the distances but also in the atomic connectivity. As an example, Table 4 presents two molecules with different tautomeric forms and the associated indices. In this case, it is also suggested to make sure which the relevant structure is or to make some averaging otherwise.

–Stereoisomerism: changes in stereocenters, except for purely chiral molecules, also introduce changes in the interatomic distances. Therefore it is important to deal with the correct stereo configuration or apply some kind of averaging over all possible configurations if the chirality is undefined.

**Table 4.** Examples of Tautomers and Associated Classic and QS–Randić Indices

		
$\chi' = 1.069$		$\chi' = 0.947$
$\chi' = 2.762$		$\chi' = 2.811$
		
$\chi' = 3.492$	$\chi' = 3.475$	$\chi' = 3.492$
$\chi' = 3.483$	$\chi' = 3.472$	$\chi' = 3.444$

## RESULTS: QSPR APPLICATIONS

In this section, some QSPR applications of these newly defined indices are exemplified. Far from trying to define them as universal QSPR descriptors, the intention of these examples is to simply demonstrate how they can be used in real problems. Starting with some initial considerations, this section presents correlation examples with physical properties, biological activity, and toxicity.

**Molecular Modeling Considerations.** All molecules presented in these examples were originally coded as SMILES<sup>30</sup> and fed into Scitegic Pipeline Pilot, where for each molecule a conformational analysis<sup>29</sup> was performed and the lowest energy conformer selected. From the final geometries, an in-house script<sup>31</sup> was used to derive the topological descriptors.

**Statistical Considerations.** Correlation search between the TQSI and the different molecular properties was done using multiple linear regression, generating all possible combinations of descriptors, and the model with best predictivity ( $q^2$ ) was kept for each different dimension.

**Energies of Formation of Alkanes.** This application example deals with the heat of formation ( $H_f$ , in kcal/mol)

**Table 5.** Heat of Formation ( $H_f$ ) for a Set of 60 Hydrocarbons

molecules	Hf	molecules	Hf
ethane	-20.24	1,3,5-triethylbenzene	-3.84
propane	-24.82	1,3-triethylbenzene	-5.22
butane	-30.15	1,4-diethylbenzene	-5.32
2-methylpropane	-32.15	1,2,3,4-tetramethylbenzene	-10.02
pentane	-35.00	1,2,3-triethylbenzene	-16.25
2-methylbutane	-36.92	1,2,4-triethylbenzene	-16.99
2,2-dimethylpropane	-39.67	1,3,5-triethylbenzene	-17.86
hexane	-39.96	1,2,3,4-tetraethylbenzene	-29.46
2-methylpentane	-41.66	1,2,3,5-tetraethylbenzene	-29.36
3-methylpentane	-41.02	1,2,4,5-tetraethylbenzene	-29.46
2,2-dimethylbutane	-44.35	1-methylnaphthalene	27.93
2,3-dimethylbutane	-42.49	2-methylnaphthalene	27.75
heptane	-44.88	1-ethylnaphthalene	23.10
3-methylhexane	-45.96	2-ethylnaphthalene	22.92
2,2-dimethylpentane	-49.27	1,2-dimethylnaphthalene	19.97
2,2,3-trimethylbutane	-48.95	1,3-dimethylnaphthalene	19.55
3-methylheptane	-50.82	1,4-dimethylnaphthalene	19.72
4-methylheptane	-50.69	1,5-dimethylnaphthalene	19.55
2,2-dimethylhexane	-53.71	1,6-dimethylnaphthalene	19.72
2,3-dimethylhexane	-51.13	1,7-dimethylnaphthalene	19.55
2,4-dimethylhexane	-52.44	2,3-dimethylnaphthalene	19.97
2,5-dimethylhexane	-53.21	2,6-dimethylnaphthalene	19.72
3,3-dimethylhexane	-52.61	2,7-dimethylnaphthalene	19.72
1,3-dimethylbenzene	4.12	1-propylnaphthalene	17.85
1,2-dimethylbenzene	4.54	2-propylnaphthalene	17.65
1,4-dimethylbenzene	4.29	2-ethyl-3-methylnaphthalene	15.72
1-methyl-3-ethylbenzene	-0.46	2-ethyl-6-methylnaphthalene	14.65
1-methyl-4-ethylbenzene	-0.78	2-ethyl-7-methylnaphthalene	14.65
1,2,3-trimethylbenzene	-2.29	1-butylnaphthalene	12.68
1,2,4-trimethylbenzene	-3.33	2-butylnaphthalene	12.50

**Table 6.** Correlation and Prediction Results for the Heat of Formation of Alkanes

no.	equation	$r^2$	$q^2$
1	$H_f = 0.22(\pm 0.02) \cdot M_1 - 46.77(\pm 3.40)$	0.719	0.701
2	$H_f = -14.86(\pm 0.95) \cdot {}^0\chi + 0.35(\pm 0.01) \cdot M_1 + 17.06(\pm 4.33)$	0.947	0.942
3	$H_f = 27.10(\pm 3.10) \cdot \chi - 26.63(\pm 1.48) \cdot {}^0\chi + 0.20(\pm 0.02) \cdot M_1 + 13.31(\pm 2.88)$	0.978	0.974
4	$H_f = 30.33(\pm 2.15) \cdot \chi - 24.53(\pm 1.05) \cdot {}^0\chi - 12.74(\pm 1.58) \cdot {}^2\chi + 0.16(\pm 0.01) \cdot M_1 + 16.70(\pm 2.01)$	0.990	0.988

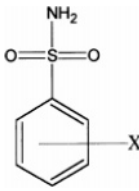
of alkanes.<sup>32</sup> Alkanes are a typical benchmark case for novel topological indices; therefore, an example with these simple molecules is provided. Structures and  $H_f$  are presented in Table 5, and the correlation and prediction results, up to 4 descriptors, are presented in Table 6.

**Inhibitory Activity of Benzenesulfonamides.** This application example is composed of 29 benzenesulfonamide carbonic anhydrase inhibitors,<sup>33</sup> which are validated therapeutic agents to treat glaucoma, gastric ulcers, neurological alterations, convulsions, altitude sickness, and tumors among others. Structures and inhibitory activity, expressed in terms of  $\log K_c$ , are displayed in Table 7, whereas Table 8 summarizes the QSPR results.

**Environmental Toxicity of Aromatic Compounds.** This application example consists of a series of aromatic compounds and their related aquatic toxicity to the species *Tetrahymena pyriformis*,<sup>34</sup> expressed in terms of pC (logarithm of the necessary concentration to reduce the population to 50% in 96 h). All data regarding the compounds are presented in Table 9, and the correlation results are presented in Table 10.

As seen from the statistical results, all three systems are nicely correlated with remarkable predictive capacity. This statement is not really surprising for the first alkane case,

**Table 7.** Carbonic Anhydrase II Inhibitory Activity ( $\log K_c$ ) for a Set of 29 Sulfonamides

		
X		$\log K_c$
H		6.69
4-Me		7.09
4-Et		7.53
4-nPr		7.77
4-nBu		8.3
4-nAm		8.86
4-COOMe		7.98
4-COOEt		8.5
4-COOPr		8.77
4-COOBu		9.11
4-COOAm		9.39
4-COOHe		9.39
4-CONHMe		7.08
4-CONHEt		7.53
4-CONHPr		8.08
4-CONHBu		8.49
4-CONHAm		8.75
4-CONHHe		8.88
4-CONHHp		8.93
3-COOMe		5.87
3-COOEt		6.21
3-COOPr		6.44
3-COOBu		6.95
3-COOAm		6.86
2-COOMe		4.41
2-COOEt		4.8
2-COOPr		5.28
2-COOBu		5.76
2-COOAm		6.18

**Table 8.** Correlation and Prediction Results for the Inhibitory Activity of Benzenesulfonamides

no.	equation	$r^2$	$q^2$
1	$\log K_c = 2607.75(\pm 495.55) \cdot {}^5\chi_c - 10.36(\pm 3.39)$	0.506	0.445
2	$\log K_c = -120.22(\pm 11.84) \cdot {}^5\chi_p + 331.64(\pm 26.51) \cdot {}^6\chi_p + 6.09(\pm 0.49)$	0.858	0.801
3	$\log K_c = -32.73(\pm 10.20) \cdot {}^3\chi_c - 124.53(\pm 10.25) \cdot {}^5\chi_p + 377.01(\pm 26.79) \cdot {}^6\chi_p + 9.50(\pm 1.14)$	0.899	0.860
4	$\log K_c = 99.7(\pm 0.90) \cdot {}^0\chi - 29.11(\pm 6.07) \cdot {}^4\chi_c - 211.72(\pm 13.86) \cdot {}^5\chi_p + 2.2e-3(\pm 0.29e-3) \cdot M_2 - 15.26(\pm 2.25)$	0.922	0.891

where the driving features are molecular size and ramification, and these are properly described by ramification-like indices. The second case is more challenging, as it already involves an activity related to a target, so that descriptors should describe this interaction as well. The last case, which also provides the worst results, reflects the difficulty inherent to describe a property that might be reflected by different modes of actions, such as toxicity is. Nevertheless, even in such cases where a truly predictive correlation cannot be found, the model could be used for simple discriminant or binary classification purposes.

## CONCLUSIONS

A novel procedure to generate Topological Quantum Similarity Indices has been presented. The exposed meth-



**Table 9.** Fish Toxicity (pC) for a Set of 92 Aromatic Compounds

compound name	pC	compound name	pC
phenol	3.45	nitrobenzene	2.97
2-methylphenol	3.77	2-nitrotoluene	3.59
3-methylphenol	3.48	3-nitrotoluene	3.65
4-methylphenol	3.74	4-nitrotoluene	3.67
2,4-dimethylphenol	3.86	2,3-dimethylnitrobenzene	4.39
2,6-dimethylphenol	3.75	3,4-dimethylnitrobenzene	4.21
3,4-dimethylphenol	3.92	2-chloronitrobenzene	3.72
2,3,6-trimethylphenol	4.21	3-chloronitrobenzene	4.01
4-ethylphenol	4.07	4-chloronitrobenzene	4.42
4-propylphenol	4.09	2,3-dichloronitrobenzene	4.66
4-butylphenol	4.47	2,4-dichloronitrobenzene	4.46
4-tert-butylphenol	4.46	2,5-dichloronitrobenzene	4.59
2-tert-butyl-4-methylphenol	4.90	3,5-dichloronitrobenzene	4.58
4-pentylphenol	5.12	2-chloro-6-nitrotoluene	4.52
4-tert-pentylphenol	4.81	4-chloro-2-nitrotoluene	4.44
2-allylphenol	3.96	aniline	2.91
2-phenylphenol	4.76	2-methylaniline	3.12
1-naphthol	4.50	3-methylaniline	3.47
4-chlorophenol	4.18	4-methylaniline	3.72
4-chloro-3-methylphenol	4.33	N,N-dimethylaniline	3.33
4-chloro-3,5-dimethylphenol	4.66	2-ethylaniline	3.21
3-methoxyphenol	3.22	3-ethylaniline	3.65
4-methoxyphenol	3.05	4-ethylaniline	3.52
4-phenoxyphenol	4.58	4-butylaniline	4.16
quinoline	3.63	2,6-diisopropylaniline	4.06
chlorobenzene	3.77	2-chloroaniline	4.31
1,2-dichlorobenzene	4.40	3-chloroaniline	3.98
1,3-dichlorobenzene	4.28	4-chloroaniline	3.67
1,4-dichlorobenzene	4.56	2,4-dichloroaniline	4.41
1,2,3-trichlorobenzene	4.89	2,5-dichloroaniline	4.99
1,2,4-trichlorobenzene	4.83	3,4-dichloroaniline	4.39
1,3,5-trichlorobenzene	4.74	3,5-dichloroaniline	4.62
1,2,3,4-tetrachlorobenzene	5.35	2,3,4-trichloroaniline	5.15
1,2,3,5-tetrachlorobenzene	5.43	2,3,6-trichloroaniline	4.73
1,2,4,5-tetrachlorobenzene	5.85	2,4,5-trichloroaniline	4.92
3-chlorotoluene	3.84	ααα-4-tetrafluoro-3-methylaniline	3.77
4-chlorotoluene	4.33	ααα-4-tetrafluoro-2-methylaniline	3.78
2,4-dichlorotoluene	4.54	pentafluoroaniline	3.69
2,4,5-trichlorotoluene	5.06	2-nitroaniline	4.15
3,4,5-trichlorotoluene	4.60	3-nitroaniline	3.24
pentachlorotoluene	6.15	4-nitroaniline	3.23
benzene	3.09	2-chloro-4-nitroaniline	3.93
toluene	3.13	4-bromoaniline	3.56
2-xylene	3.48	3-benzoyloxyaniline	4.34
3-xylene	3.45	4-hexyloxyaniline	4.78
4-xylene	3.48	4-ethoxy-2-nitroaniline	3.85

**Table 10.** Correlation and Prediction Results for the Toxicity of Aromatic Compounds

no.	equation	r <sup>2</sup>	q <sup>2</sup>
1	pC = 4.80(±0.64)· <sup>3</sup> χ <sub>p</sub> + 2.25(±0.26)	0.382	0.332
2	pC = 1.50(±0.19)· <sup>0</sup> χ - 0.62e-3(±0.1e-3)·M <sub>2</sub> - 0.45(±0.53)	0.477	0.419
3	pC = 6.57(±0.59)· <sup>2</sup> χ - 104.31(±17.04)· <sup>5</sup> χ <sub>c</sub> - 0.4e-3(±6.7e-5)·M <sub>2</sub> - 0.14(±0.35)	0.625	0.549
4	pC = 7.27(±0.82)· <sup>2</sup> χ - 3.38(±2.69)· <sup>4</sup> χ <sub>p</sub> - 114.22(±18.72)· <sup>5</sup> χ <sub>c</sub> - 0.4e-3(±0.6e-5)·M <sub>2</sub> - 0.25(±0.37)	0.631	0.553

odology supersedes the previous lack of a proper definition for the atomic density functions by using fitted densities derived from the Atomic Shell Approximation. The usage of fitted densities allows a proper background for the calculation of similarity-based adjacency matrices as well as allows a considerable speed up compared to the usage of purely ab initio densities.

In addition, successful QSPR examples allow considering TQSI's as useful molecular descriptors in correlation studies as well as suitable complement to existing ones.

## ACKNOWLEDGMENT

One of us (A.G.) benefits from a grant from the Spanish Ministerio de Educación y Ciencia.

## REFERENCES AND NOTES

- (1) Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRC: Boca Raton, FL, 1992.
- (2) Balaban, A. T. *Chemical Applications of Graph Theory*; Academic: London, 1976.
- (3) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: New York, 1976.
- (4) Randić, M.; Krais, G. A.; Dzonova-Jerman-Blazic, B. *Chemical Applications of Topology and Graph Theory*; Elsevier: Amsterdam, 1983; pp 192–205.
- (5) Randić, M. Topological Indices. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; pp 3018–3032.
- (6) Estrada, E. Recent Advances on the Role of Topological Indices in Drug Discovery Research. *Curr. Med. Chem.* **2001**, *8*, 1573–1588.
- (7) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- (8) Kier, L. B. A Shape Index from Molecular Graphs. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109–116.
- (9) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (10) Balaban, A. T. A discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
- (11) Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotological State*; Academic Press: New York, 1999.
- (12) Gálvez, J.; García, R.; Salabert, M. T.; Soler, R. Charge indices: New Topological Descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520–525.
- (13) *From Chemical Topology to Three-Dimensional Geometry*; Balaban, A. T., Ed.; Plenum Press: New York, 1997.
- (14) Estrada, E.; Molina, E. 3D Connectivity Indices in QSPR/QSAR Studies. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 791–797.
- (15) Estrada, E.; Ramírez, A. Edge Adjacency Relationships and Molecular Topographic Descriptors. Definition and QSAR Applications. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 837–843.
- (16) Estrada, E. Three-Dimensional Descriptors Based on Electron Charge Density Weighted Graphs. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 708–713.
- (17) Estrada, E.; Delgado, E. J.; Alderete, J. B.; Jaña, G. A. Quantum-connectivity descriptors in modeling solubility of environmentally important organic compounds. *J. Comput. Chem.* **2004**, *25*, 1787–1796.
- (18) Lobato, M.; Amat, L.; Besalú, E.; Carbó-Dorca, R. Structure–Activity Relationships of a steroid family using QSM and topological QS Indices. *Quant. Struct.-Act. Relat.* **1997**, *16*, 465–472.
- (19) Constants, P.; Carbó, R. Atomic shell approximation: electron density fitting algorithm restricting coefficients to positive values. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1046–1053.
- (20) Amat, L.; Carbó-Dorca, R. Quantum similarity measures under atomic shell approximation: first-order density fitting using elementary Jacobi rotations. *J. Comput. Chem.* **1997**, *18*, 2023–2039.
- (21) Amat, L.; Carbó-Dorca, R. Fitted electronic density functions from H to Rn for use in quantum similarity measures: cis-diamminedichloroplatinum(II) complex as an application example. *J. Comput. Chem.* **1999**, *20*, 911–920.
- (22) Amat, L.; Carbó-Dorca, R. Molecular Electronic Density Fitting Using Elementary Jacobi Rotations under Atomic Shell Approximation. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1188–1198.
- (23) ASA repository for different basis sets at the Institute of Computational Chemistry: <http://stark.udg.es/cat/similarity/ASA/funcset.html> (last accessed on July 2004).
- (24) Schultz, H. P. Topological organic chemistry. 1. Graph theory and topological indices of alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227–228.
- (25) Mueller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. Molecular topological index. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 160–163.
- (26) Gutman, I.; Trinajstić, N. Graph theory and molecular orbitals. Total p-electron energy of alternant hydrocarbons. *Chem. Phys. Lett.* **1972**, *17*, 535–538.

- (27) Gutman, I.; Ruščić, B.; Trinajstić, N.; Wilcox, C. F., Jr. Graph theory and molecular orbitals. XII. Acyclic polyenes. *J. Chem. Phys.* **1975**, *62*, 3399–3405.
- (28) Ren, B. A New Topological Index for QSPR of alkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 139–143.
- (29) The conformational analysis has been performed using Scitegic Pipeline Pilot, creating a protocol containing a molecular reader, a 3D generator, a 3D conformer generator, and a simple energy evaluator. Further details on Pipeline Pilot software can be found on the web at <http://www.scitegic.com> (last accessed on July 2004).
- (30) SMILES stands for Simplified Molecular Input Line Entry Specification. Further details can be found on the web at <http://www.daylight.com> (last accessed on July 2004).
- (31) Python scripting language (<http://www.python.org>) was used for this work for its simplicity and easy extensibility. The Numerical module was also used (<http://sourceforge.net/projects/numpy>), which speeds up some operations that might be tedious if direct scripting were used (last accessed on July 2004).
- (32) Mercader, A.; Castro, E. A.; Toropov, A. Maximum Topological Distances Based Indices as Molecular Descriptors for QSPR. 4. Modeling the Enthalpy of Formation of Hydrocarbons from Elements. *Int. J. Mol. Sci.* **2001**, *2*, 121–132.
- (33) Thakur, A.; Thakur, M.; Khadikar, P. V.; Supuran, C. T.; Sudele, P. QSAR study on benzenesulphonamide carbonic anhydrase inhibitors: topological approach using Balaban index. *Bioorg. Med. Chem.* **2004**, *12*, 789–793.
- (34) Roy, K.; Ghosh, G. QSTR with Extended Topochemical Atom Indices. 2. Fish Toxicity of Substituted Benzenes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 559–567.

CI049721Y