SCIENTIA gerundensis, 21: 145-152 (1995)

QUANTUM SIMILARITY TOPOLOGICAL INDICES: DEFINITION, ANALYSIS AND COMPARISON WITH CLASSICAL MOLECULAR TOPOLOGICAL INDICES

Emili Besalú and Ramon Carbó
Institut de Química Computacional. Universitat de Girona. 17071 Girona.

RESUM

Es mostra que, gràcies a una extensió en la definició dels Índexs Moleculars Topològics, s'arriba a la formulació d'índexs relacionats amb la teoria de la Semblança Molecular Quàntica. Es posa de manifest la connexió entre les dues metodologies: es revela que un marc de treball teòric sòlidament fonamentat sobre la teoria de la Mecànica Quàntica es pot connectar amb una de les tècniques més antigues relacionades amb els estudis de QSPR. Es mostren els resultats per a dos casos d'exemple d'aplicació d'ambdues metodologies.

RESUMEN

Se muestra cómo, gracias a una extensión en la definición de los Índices Moleculares Topológicos, se llega a la formulación de índices relacionados con la teoría de la Semejanza Molecular Cuántica. Se pone de manifiesto la conexión entre las dos metodologías: se revela cómo un marco de trabajo teórico sólidamente fundamentado en la teoría de la Mecánica Cuántica se puede conectar con una de las técnicas más antiguas relacionadas con los estudios de tipo QSPR. Se muestran resultados de dos casos ejemplo de aplicación de ambas metodologías.

ABSTRACT

It is shown how an extension of the definition of the Molecular Topological Indices can lead to indices related to the Molecular Quantum Similarity theoretical framework. The connection between both approaches is studied. It is described how a solid quantum-mechanical theoretically based framework as the Molecular Quantum Similarity Theory can be connected to one of the oldest techniques related to QSPR. Two application examples are also reported.

Keywords: Molecular Graphs, Molecular Quantum Similarity, Quantum Similarity, Quantum Similarity Indices, Topological Molecular Indices

---

INTRODUCTION

Our Laboratory has been working lately on Molecular Quantum Similarity theory [1], which is a robust quantum-mechanical theoretical framework related to Molecular design and QSPR techniques. Some of the classical techniques which have the same purpose are related to the Topological approach of the Molecular Similarity [2,3]. Here, two families of compounds will be studied under two points of view: the classical topological approach and the Molecular Quantum Similarity related methods.

TOPOLOGICAL AND QUANTUM SIMILARITY RELATED INDICES

Some molecular properties can be correlated to various indices. Two groups of indices can be defined separately: Classical Topological Indices (CTI) and those which will be called here Quantum Similarity Related Indices (QSRI). With respect to the CTI, five classical indices are used here. They can be directly applied over alkanes and are the Wiener Path and Harary Numbers [2,3] as well as the Wiener, Randić and Schultz indices [2,3]. The definition of each kind of index is given in **Table 1**. The summation symbol indices which appear in the definitions of **Table 1** run over all the vertices of the related molecular graphs.

Some results which can be already found in reference [2] will be reproduced here using these five CTI's. The bibliographic results of those indices will be compared with others which can be considered, in some manner, extensions of the former ones. They constitute the indices we have called QSRI. Their definition can be found in **Table 2**. The basic criteria which has been applied when defining some of them was to replace the discontinuous and graph related (planar) parameters entering in the corresponding classical definition by continuous and tridimensional ones. Following this philosophy, **Table 3** collects some of the replacements which must be applied to the classical indices in order to obtain the QSRI. The Q_1 and Q_2 in **Table 3** indices are pure Quantum Similarity Indices [1]. The symbols $\{q_i\}$ appear in their definition and stand for the effective charge of atom number i. The computation of these indices requires to perform the full geometry optimization of the molecules and the computation of the effective charges. This has been obtained by means of the AMPAC program [4] using the AM1 methodology [5].

In fact, CTI are not associated directly to a molecular structure but to a graph representation in a hydrogen suppressed manner [2,3]. This is the principal difference between this kind of indices with respect to the new QSRI defined here. Another important characteristic must be noted: as it was signaled above, the correct manner to proceed when working with the CTI ignores the presence of molecular hydrogen atoms. This is not strictly the correct way to proceed when working with the QSRI. In this case *all the atoms* of the molecule must be considered when computing the indices. Although, this characteristics were, in the following examples, changed in order to test other possibilities. QSRI's contain tridimensional information of the molecular structure (euclidean distances between atoms) and, also, chemical information due to the fact that the overlap matrix elements have been obtained using a basis set depending of the nature of the atoms entering into the molecule. In this context, the QSRI may distinguish rotamers, conformers ... contrarily to the CTI ones which cannot.

In fact, every atom within a molecule is represented by a normalized *n*S-STO function. The kind of functions attached to the atoms treated in this work can be found in **Table 4**. This information, together with the knowledge of the interatomic distances, is relevant to obtain the overlap matrix, $\{S_{ij}\}$, and the related parameters which are needed to define some QSRI. The same can be said with respect to the effective charge parameters appearing in the definition of indices Q_1 and Q_2. More details about these features will be explained below.

The methodology followed in this work computes every index for a given molecule and correlates all the values associated to a molecular family to the physicochemical properties. Usually, a multiple regression analysis is done over several indices but, here, only linear regressions will be considered.

A CLASSICAL EXAMPLE: A FAMILY OF ALKANES

It has been chosen, as a first example, a family of 13 alkanes. **Table 5** describes the properties which are considered for this family and the corresponding experimental values [6] are reported. Tables **6** and **7** show the correlation coefficient and the standard deviation attached to every property when compared with the indices. In order to stablish an equal footing comparison between CTI and QSRI, in **Table 6** both sets of indices were computed *without* hydrogens whereas in **Table 7** both indices sets were computed with hydrogens. It can be seen in **Table 6** how the properties correlate with all the CTI and with the W_1^q and x_2^q ones. In **Table 7** more QSRI correlates with the experimental values and, in some cases, CTI give best results than those which can be seen in **Table 6**.

A FAMILY OF BENZENE DERIVATIVES

As in **Table 7**, Tables **8** and **9** were obtained using all the molecular atoms. It has been studied another 13-member family: a set of benzene derivatives. This family was chosen due to the particularity that some molecules bear heteroatoms in their structure. In this case, the Boiling and Melting Points [7] are the studied properties and are listed in **Table 8** together with the molecular specifications. The same treatment as in the case of the alkane family was done over the corresponding indices attached to the benzene derivatives. It can be seen in **Table 9** how the properties only correlate slightly with the QSRIs Q_1 and Q_2. The presence of heteroatoms invalidates the efficiency of the classical indices but the Q_1 and Q_2 QSRIs give a slight statistical significance.

Of course, all the presented results can be improved because there are special classical topological methods and indices which can be applied to molecules having heteroatoms. Also, there are more sophisticated techniques which can be applied when working with QSRI [1]. But, as this work constitutes a preliminary study only, no other method has been applied.

CONCLUSIONS

It has been shown how the Molecular Quantum Similarity theoretical framework can be related to Molecular Topological techniques. It has been proved how the same kind of results can be obtained when dealing with a family of alkanes. Moreover, when heteroatoms are present in the studied molecules, as occurs with the family of benzene derivatives, some advantage can be taken when the problem is studied from the Molecular Quantum Similarity point of view. In both cases a simple correlation is sought, just to see how well both index sets behave.

ACKNOWLEDGEMENTS

This work has been financed by a grant: #QFN91-4206 from CICYT-CIRIT, within the "Programa de Química Fina de la Generalitat de Catalunya".

Bibliography

1. a) CARBÓ, R., ARNAU, M. and LEYDA, L., Int.J.Quantum Chem., *1980*, 17, 1185.
   b) CARBÓ, R., and ARNAU, C., "Molecular Engineering: a general approach to QSAR", in: de las Heras, F.G., and Vega, S., Eds., "Medicinal Chemistry Advances", Pergamon Press, Oxford, 1981.
   c) CARBÓ, R., and DOMINGO, L., Int.J.Quantum.Chem., *1987*, 23, 517.
   d) CARBÓ, R., and CALABUIG, B., Comp. Phys. Commun., *1989*, 55, 117.
   e) CARBÓ, R., and CALABUIG, B., "Molecular Similarity and Quantum Chemistry", Chap. 6, p. 147, in Johnson M.A. and Maggiora G., Eds., *Concepts and Applications of Molecular Similarity*, John Wiley & Sons Inc, New York, 1990.
   f) CARBÓ, R., CALABUIG, B., "Proceedings del XIX Congresso Internazionale dei Chimici Teorici dei Paesi di Espressione Latina, Roma, Italy, September 10-14, 1990", J.Mol.Struct. (Teochem), *1992*, 254, 517.
   g) CARBÓ, R., and CALABUIG, B., J.Chem.Inf.Comput.Sci., *1992*, 32, 600.
   h) CARBÓ, R., and CALABUIG, B., "Quantum Similarity", in Fraga S., Ed., "Structure, Interactions and Reactivity", Elsevier Pub., Amsterdam, 1992.
   i) CARBÓ, R., and CALABUIG, B., Int.J.Quantum Chem., *1992*, 42, 1681 and 1695.
   j) CARBÓ, R., CALABUIG, B., BESALÚ, E., and MARTÍNEZ, A., Molecular Engineering., *1992*, 2, 43.
   k) CARBÓ, R., BESALÚ, E., CALABUIG, B., and VERA, L., Adv.Quan.Chem., *1994*, 25, 253-313.

2. KIER, L.B., and HALL, L.H., Molecular Connectivity, in "Chemistry and Drug Research", Ac. Press., New York, 1976.

3. MIHALIC, Z., and TRINAJSTIC, N., J.Chem.Educ., *1992*, 69:9, 701-712.

4. LIOTARD, D.A., HEALY, E.F., RUIZ, J.M., and DEWAR, M.J.S., AMPAC-version 2.1. Quantum Chemistry Program Exchange, program 506, QCPE Bull. 9, 1989.

5. DEWAR, M.J.S., ZOEBISCH, E.G., HEALY, E.F., and STEWART, J.J.P., J. Am. Chem. Soc. *1985*, 107, 3902.

6. GAKH, A.G., GAKH, E.G., STUMPER, B.G., and NOID, D.W., J. Chem. Inf. Comp. Scci., *1994*, 34, 832-839.

7. LIDE, D.R., Ed., Handbook of Chemistry and Physics, 72ᴺᴰ edition, Boca Raton, 1991-1992.

Table 1. Definition of some CTI [2,3]. Parameters are: D_{ij}: graph topological distance between atoms i and j; p_3: number of carbon atoms of the molecule separated by three bonds; T_{ij}: elements of the topological matrix associated to the molecular graph; V_i: topological valence associated to the vertex (atom) i, i.e., the sum of the topological matrix elements placed in ith row. The Shultz Index is the sum of the elements of a vector defined as the product of the vector of the graph vertices valencies (V) by the matrix resulting from the sum of the adjacency (A) and distance (D) matrices [3]

Index Name	Definition
Wiener Path Number	$W_o = \sum_i \sum_{j>i} D_{ij}$
Wiener Index	$w_o = W_o + p_3$
Randić Index	$\chi_o = \sum_i \sum_{j>i} \frac{T_{ij}}{\sqrt{V_i V_j}}$
Schultz Index	$S_o = \sum_i \{ V\,[A+D] \}_i$
Harary Number	$H_o = \sum_{i=2}^{n} \sum_{j>i} \frac{1}{D_{ij}^2}$

Table 2. Definition of some QSRI. See text for meaning.

Index Name	Definition
Wiener Path Number #1	$W_1^q = \sum_i \sum_{j>i} d_{ij}$
Wiener Path Number #2	$W_2^q = \sum_i \sum_{j>i} d_{ij}\, S_{ij}$
Randić Index #1	$\chi_1^q = \sum_i \sum_{j>i} \frac{S_{ij}}{\sqrt{V_i V_j}}$
Randić Index #2	$\chi_2^q = \sum_i \sum_{j>i} \frac{S_{ij}}{\sqrt{v_i v_j}}$
Harary Number #1	$H_1^q = \sum_{i=2}^{n} \sum_{j>i} \frac{S_{ij}}{d_{ij}^2}$
Quantum Similarity Index #1	$Q_1 = \sum_i \sum_j q_i\, q_j\, S_{ij}$
Quantum Similarity Index #2	$Q_2 = \sum_i \sum_j \lvert q_i\, q_j \rvert\, S_{ij}$

Table 3. Replacement rules leading from CTI to QSRI.

from CTI	to QSRI
D_{ij}	d_{ij}: tridimensional euclidean distance between atoms i and j.
T_{ij}	S_{ij}: overlap matrix element of the orbitals attached to the atoms i and j.
V_i	v_i: sum of the overlap matrix elements of ith row.

Table 4. Kind of *n*S-STO functions attached to every atom when computing Molecular Quantum Similarity Related Indices.

Atom	Orbital type (STO)	STO Exponent
H	1s	1.2000
C	2s	1.6083
O	2s	2.2458
Cl	3s	2.3561

Table 5. Experimental property values for the studied alkanes [6]. [a]Heat capacity (J/mol grad) at 300 °K; [b]Density (Kg/m^3) at 25 °C; [c]Refractive index at 25 °C; [d]Gibbs energy (KJ/mol) at 300°K; [e]Enthalpy (KJ/mol) at 300°K; [f]Boiling point (°C).

	Property					
Alkane	C[a]	d[b]	n[c]	G[d]	H[e]	BP[f]
7	166.00	679.50	1.3851	9.50	33.56	98.40
3M6	164.50	682.88	1.3861	6.60	30.71	91.85
3E5	166.80	693.92	1.3911	12.7	31.71	93.48
33MM5	166.70	689.16	1.3884	4.80	29.33	86.04
2M6	165.40	674.34	1.3823	4.90	31.21	90.03
24MM5	171.70	668.23	1.3788	4.90	29.58	80.47
23MM5	161.80	690.81	1.3895	7.60	28.62	89.75
22MM5	167.70	669.48	1.3800	2.10	29.50	79.17
223MMM4	164.20	685.64	1.3869	6.30	28.28	80.86
3M5	140.88	659.76	1.3739	-2.12	26.32	63.28
2M5	143.01	648.52	1.3687	-4.05	26.61	60.27
23MM4	140.21	657.02	1.3723	-1.77	24.77	57.99
22MM4	142.26	644.46	1.3660	-7.42	25.40	49.74

Table 6. Correlation coefficients and standard deviations for the alkanes properties with respect to the molecular indices. Molecules where treated in a hydrogen suppressed way. See text for meanig.

	Property											
	C		d		n		G		H		BP	
Index	r	s	r	s	r	s	r	s	r	s	r	s
W_0	0.9136	4.88	0.7475	10.84	0.7878	0.0051	0.8645	2.89	0.9598	0.73	0.9583	4.47
w_0	0.9273	4.49	0.8117	9.53	0.8466	0.0044	0.9025	2.48	0.9491	0.82	0.9788	3.20
x_0	0.8488	6.34	0.8064	9.65	0.8390	0.0045	0.9108	2.38	0.9619	0.71	0.9793	3.17
S_0	0.9199	4.70	0.7435	10.91	0.7840	0.0051	0.8576	2.96	0.9559	0.76	0.9535	4.71
H_0	0.9282	4.46	0.8308	9.08	0.8419	0.0044	0.7623	3.73	0.6405	1.99	0.7788	9.81
W_1^q	0.8850	5.59	0.7303	11.15	0.7717	0.0052	0.8476	3.06	0.9633	0.7	0.9483	4.96
x_2^q	0.9731	2.76	0.8341	9.	0.8571	0.0042	0.8348	3.17	0.7826	1.62	0.8923	7.06

Table 7. Correlation coefficients and standard deviations for the alkane properties with respect to the molecular indices. All the atoms were considered.

	Property											
	C		d		n		G		H		BP	
Index	r	s	r	s	r	s	r	s	r	s	r	s
W_0	0.95928	3.39	0.80774	9.62	0.84224	0.0044	0.88367	2.7	0.92481	0.99	0.96046	4.36
w_0	0.96067	3.33	0.80984	9.57	0.84407	0.0044	0.88397	2.69	0.92278	1.00	0.96000	4.38
x_0	0.98106	2.32	0.85706	8.41	0.88176	0.0039	0.87094	2.83	0.83308	1.44	0.92012	6.13
S_0	0.95423	3.59	0.80035	9.78	0.83574	0.0045	0.88233	2.71	0.93136	0.95	0.96170	4.29
H_0	0.97056	2.89	0.86022	8.32	0.87926	0.0039	0.83936	3.13	0.76281	1.68	0.87409	7.6
W_1^q	0.92076	4.68	0.77434	10.33	0.81240	0.0048	0.86748	2.86	0.94839	0.82	0.95708	4.53
W_2^q	0.97752	2.53	0.86050	8.31	0.88196	0.0039	0.85522	2.98	0.79495	1.58	0.89592	6.95
x_1^q	0.96840	2.99	0.85117	8.56	0.86984	0.0041	0.82660	3.24	0.74960	1.72	0.86104	7.96
x_2^q	0.97812	2.5	0.86240	8.26	0.88832	0.0038	0.89063	2.62	0.85594	1.34	0.93988	5.34
H_1^q	0.97718	2.55	0.84688	8.68	0.86967	0.0041	0.84358	3.09	0.79390	1.58	0.88904	7.16
Q_1	0.96564	3.12	0.86894	8.08	0.89570	0.0037	0.90685	2.43	0.86498	1.30	0.95221	4.78
Q_2	0.97090	2.87	0.86707	8.13	0.89059	0.0037	0.87735	2.76	0.81596	1.50	0.91699	6.24

Table 8. Identification for the members of the benzene derivatives family and the corresponding boiling (BP) and melting (MP) points [7].

Molecule	BP (°K)	MP (°K)
1,4-dimethilbenzene	411.5	286.3
1,3-dimethylbenzene	412.3	225.3
1,2-dimethylbenzene	417.6	247.9
Methylbenzene	383.8	178.2
1,4-dihidroxybenzene	558.1	446.1
1,3-dihidroxybenzene	451.1	389.1
1,2-dihidroxybenzene	518.1	378.1
Hidroxybenzene	455.0	314.0
1,4-dichlorobenzene	447.6	326.2
1,3-dichlorobenzene	446.6	248.4
1,2-dichlorobenzene	453.6	256.1
Chlorobenzene	404.9	227.9
Benzene	353.2	278.7

Table 9. Correlation coefficients and standard deviations for the melting and boiling points of the benzene derivatives. All the atoms were considered.

	Property			
	BP		MP	
Index	r	s	r	s
W_0	-0.14885	53.	-0.17938	74.6
w_0	-0.14810	53.2	-0.17884	74.7
x_0	-0.03378	53.7	-0.05799	75.8
S_0	-0.14918	53.2	-0.17966	74.6
H_0	-0.15938	53.1	-0.20038	74.3
W_1^q	-0.15284	53.1	-0.19239	74.5
W_2^q	-0.32295	50.9	-0.30622	72.2
x_1^q	-0.25231	52.0	-0.20519	74.3
x_2^q	-0.07262	53.6	-0.05457	75.8
H_1^q	-0.36953	50.	-0.32682	71.7
Q_1	0.54149	45.2	0.74239	50.8
Q_2	0.51404	46.1	0.71672	52.9