Molecular Quantum Similarity and the Fundamentals of QSAR

EMILI BESALÚ, XAVIER GIRONÉS, LLUÍS AMAT, AND RAMON CARBÓ-DORCA*

Institute of Computational Chemistry, University of Girona, Campus Montilivi, 17071 Girona, Catalonia, Spain

Received June 26, 2001 (Revised Manuscript Received December 10, 2001)

ABSTRACT

A general overview on quantum similarity and applications to QSAR is presented. The concepts regarding quantum similarity from its theoretical foundation and consecutive development, involving mathematical formulation and similarity measures, are presented and complemented with application examples. The practical part, based on the well-known Crammer 31 steroids set, covers approximate quantum similarity calculations, molecular superposition, and statistics. In this way, the reader will find both basic general information and applicability of quantum similarity.

Introduction

Since the middle of the nineteenth century,[1] several authors have studied structure−property relationships. Today known as quantitative structure−activity relationships (QSAR), this field has generated a large amount of literature. For example, see the contributions in references 2 and 3 Not long ago, a more general application landscape emerged from initial QSAR ideas, providing the concept of quantitative structure−property relationships (QSPR).[4−6] More recently, the conceptual and practical use of quantitative structure-toxicity relationships (QSTR) has been frequently appearing in current literature.[7]

Since 1970, one of us (R.C.-D.) has been interested in the QSPR field.[8] Such an interest has promoted the first work in the field of molecular quantum similarity (MQS).[9] In papers associated with the initial development of the MQS ideas, some crude, graph-oriented QSPR examples were provided along with the theoretical background.[10] A clear picture was emerging from the steady development of MQS studies. The application of geometry attached to quantum mechanics, along with the quantum mechanical postulates,[11] allowed an extension of quantum theory, within practical chemical problems, into a set of procedures possessing a large applicability in many chemical fields.[12] Around 1985, the use of MQS to QSPR and QSAR was well-founded, both in theory and practice.[13] A large amount of work has been performed to obtain a new theoretical QSPR point of view and to establish coherent mathematical and physical support[14] for both MQS and the application of the described formalism. When the development of MQS reached a steady pace, a volume was published[15] in which a summary of the applications, as well as new developments, of MQS measures (MQSM) could be found.

Among all developments, the most important one has been the description of a fundamental quantum QSPR (QQSPR or Q^2SPR) equation's existence, which also demonstrates that the empirical QSPR models can be generally founded into a well-defined relationship[16] of quantum mechanical origin.

This Account consists of a theoretical introduction to the QS field, followed by an application example.

Theoretical Background

The Role of the Quantum Mechanical Density Function (DF). Perusing old literature on quantum mechanics, such as the book by von Neumann,[17] we found the old Born idea,[18] also developed by Dirac.[19] Such a proposal admits that any microscopic system wave function set, conveniently transformed into a square module, produces a set of probability density functions (DFs), and it is this DF the adequate tool that has to be used for interpreting the experimental observable behavior of particle systems, such as atoms and molecules.

This statistical fundamental for the interpretation of quantum mechanics has to be considered the basic tool of the theory to be used in studies concerning chemical problems. However, such a DF character seems to possess a secondary role in the application of quantum mechanics in chemistry. Maybe this quantum mechanical DF ancillary position has somehow caused, through the chemical

Emili Besalú is the secretary of the Institut de Química Computacional of the Universitat de Girona and since 1994 has been a professor of the chemical physics area of the Department of Chemistry. He performed his Ph.D. tasks at the Universitat de Girona under the direction of Professor R. Carbó-Dorca. His scientific work contains studies of perturbation theory, molecular integrals over exponential-type orbital functions, quantum similarity, and quantum mechanical algorithms in parallel. He has published over 60 papers on these subjects. He was the recipient of the 1997 Antoni Martí i Franquès prize of the Institut d'Estudis Catalans.

Xavier Gironés finished his degree in chemistry in 1997. Since then, he has been working on his Ph.D., devoted mostly to molecular quantum similarity. His other research topics are molecular surface representation from fitted densities, QSAR, and molecular superposition.

Lluís Amat received his degree in chemistry in 1993 and is working on his Ph. D. thesis. His research covers the following topics: molecular electronic density fitting; QSAR; and scientific programming, both serial and parallel processing.

Ramon Carbó-Dorca is director of the Institut de Química Computacional of the Universitat de Girona, which he founded after his arrival in the city of Girona after leaving in the mid 80's the Barcelona's Institut Químic de Sarrià and the Universitat Autònoma, where he had been a professor since 1965. He has published several books on quantum chemistry, linear algebra, and quantum similarity as a result of his teaching and research experience in these fields. He has been working on the Hückel method, electrostatic molecular potential, and self-consistent field and perturbation theories, as well as working on the application of elementary Jacobi rotations in quantum chemistry. In 1980, he introduced the concepts structuring the theoretical body associated with quantum similarity and quantum quantitative structure−property relationships. Several of his contributions are aimed at integrating quantum chemistry with appropriate algorithms that can be automatically programmed for computations in parallel. Starting with a first paper in 1965, which also was the first quantum chemistry paper ever published in Catalonia, he has published more than 200 papers on the aforementioned subjects, including several contributions to explain his group's current scientific work. He has been honored with the Narcís Monturiol medal of the Generalitat de Catalunya.

* To whom correspondence should be addressed. Phone: +34 972 418367. Fax: +34 972 418356. E-mail: quantum@iqc.udg.es.

10.1021/ar010048x CCC: $22.00
Published on Web 04/04/2002

literature, a delay to enlarge its applicability perspectives. Another possibility of the low application profile of DF may be reflected in the fact that literature trends have usually dealt with chemical systems per se, but very seldom in relationship with other parent structures. Chemical language, however, is full of expressions that compare two or more molecules, and experimental chemistry, since the initial analysis of the atomic properties, which has led to the construction of the periodic table of the elements, tends to produce information about chemical properties by reasoning and by comparative thinking.

Admitting the von Neumann,[17] Born,[18] and Dirac[19] interpretations of quantum mechanics, one can accept that the DF of a chemical system, constructed in a precise internal energy state, is the recipient of all the observable information, which can be extracted from such a system. Then it becomes logical to consider the possibility of using quantum mechanical DFs to develop the tools that will allow the comparison of two or more molecules.

DF can be considered at the same time as functions and operators.[20] Thus, nothing prevents consideration of the way two DFs, attached to different systems or states, can be employed to extract the numerical figures, meaning the similarity degree between the compared systems. This can be performed like a statistical expectation value technique, associated with the manner for obtaining it within a unique system state.

Similarity Measures. Once two molecular systems are known, the definition of a quantum similarity measure (QSM) becomes effortless to construct. A comparison of two molecules can be easily constructed using their corresponding DFs. Both DFs can be multiplied and integrated over the respective electronic coordinates in a convenient domain, weighted by a positive definite operator $\Omega(\mathbf{r}_1,\mathbf{r}_2)$. That is,

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

This rule, which when the operator is chosen as the Dirac's function, $\delta(\mathbf{r}_1 - \mathbf{r}_2)$, is called an overlap QSM, permits a large number of applications[21−24] and generalizations.[12,25] Another widely explored possibility[26−28] is the use of Coulomb operator, $|\mathbf{r}_1 - \mathbf{r}_2|^{-1}$, defining a Coulomb QSM. Integral 1, because of the presence of the positive definite operator and DF, always results in a positive definite real number. When relating a system to itself, by means of eq 1, that is, when computing z_{AA}, a quantum self-similarity measure (QS-SM) is obtained. Practical computation of the integral (1) may become unaffordable when the involved DFs correspond to large molecules or have been calculated at high computational levels. To overcome this problem, the promolecular atomic shell approximation (PASA)[29−31] has been defined, modeling the ab initio DF as a linear combination of 1*S* functions, decreasing in this manner the computational requirements and enlarging the potential applications of QSM.

Provided a set of N molecules, there is always the possibility of computing the whole array of QSMs between molecular pairs, producing a symmetric ($N \times N$) matrix: $\mathbf{Z} = \{z_{IJ}\}$, the similarity matrix (SM) of the set. Each column (or row) of the SM: $\mathbf{z}_I$, can be considered as the collection of all the QSMs between the I-th molecule and each element of the set, including itself. Consequently, every vector $\mathbf{z}_I$ is interpreted as a discrete N-dimensional representation of the I-th structure. Such collections of vectors can be considered as a set of molecular descriptors.

However, the SM column collection does not constitute just another set of object descriptors, such as those generally used to theoretically describe a given molecule. From the previous discussion, it can be stated that every descriptor $\mathbf{z}_I$ is:

1) Universal in the sense that it can be obtained from any molecular set and for any molecule in the set.

2) Unbiased, because in the building process, there are no other choices than those provided by the knowledge of the involved DFs and the QSMs, as described in eq 1.

QSMs can be transformed in order to enlarge their application, as will be discussed next.

Similarity Indices. QSM, like any of the off-diagonal elements of the SM, z_{AB}, involving the QSM between molecules A and B, can be easily transformed into a number lying within the interval (0;1], just by using

$$r_{AB} = \frac{z_{AB}}{\sqrt{z_{AA} z_{BB}}} \quad (2)$$

producing the so-called Carbó similarity index (CSI).[32,33] The CSI, as defined in eq 2, corresponds to a cosine of the angle subtended by the involved DFs, considered in turn as vectors. When the CSI approaches unity, the involved molecules can be considered to be more similar, and as the CSI approaches 0, the more dissimilar the compared structures become. The exact unity value is only obtained when A = B.

Stochastic Transformation. Besides CSI, another possible scaling can be performed by means of a stochastic transformation.[34] Such SM transform can be defined by means of

$$s_{AB} = z_{AB}\left(\sum_{C=1}^{N} z_{AC}\right)^{-1} \quad (3)$$

providing a stochastic SM, $\mathbf{S} = \{s_{AB}\}$, where the sum of the elements of each row has been used as a scale factor. This procedure creates an alternative nonsymmetric SM whose columns can also be used as new descriptors for a given molecular set and can be interpreted as discrete probability distributions.

Fundamental Quantum QSAR Equation. The possibility opened by the SM manipulation over a molecular set, although appealing, would constitute a very limited application of the QSM framework. The praxis of the theoretical findings has conducted the application of the SM, considered as a set of structural descriptors, to QSPR model construction.[35] From the initial results, the possible existence of a sound reason for the general applicability of both QSMs or the CSI set has been deduced to obtain

Table 1. Structures and Biological Activity for a Set of 31 Steroids

N	steroid	CBG	*N*	steroid	CBG
1	aldosterone	−6.279	17	pregnenolone	−5.225
2	androstanediol	−5.000	18	hydroxypregnenolone	−5.000
3	androstenediol	−5.000	19	progesterone	−7.380
4	androstenedion	−5.763	20	hydroxyprogesterone	−7.740
5	androsterone	−5.613	21	testosterone	−6.724
6	corticosterone	−7.881	22	prednisolone	−7.512
7	cortisol	−7.881	23	cortisolacetate	−7.553
8	cortisone	−6.892	24	4-pregnene-3,11,20-trione	−6.779
9	dehydroepiandrosterone	−5.000	25	epicorticosterone	−7.200
10	deoxycorticosterone	−7.653	26	19-nortestosterone	−6.144
11	deoxycortisol	−7.881	27	16a,17a-dihydroxyprogesterone	−6.247
12	dihydrotestosterone	−5.919	28	17a-methylprogesterone	−7.120
13	estradiol	−5.000	29	19-norprogesterone	−6.817
14	estriol	−5.000	30	2a-methylcortisol	−7.688
15	estrone	−5.000	31	2a-methyl-9a-fluorocortisol	−5.797
16	etiocholanone	−5.225			

QSPR models that are quite accurate. It became apparent that such a relationship was the consequence of a simple quantum mechanical application involving the concept of the expectation value attached to a general property. Following quite a lengthy procedure and taking into account the previous definitions, it can be shown that the expectation value of any given property can be written in terms of a linear combination of QSMs.[36]

That is, imagine a structure A and the expectation value of a given property for this molecule: $\langle \pi_A \rangle$, which can be associated with an experimental value of the given property. The following approximate relationship can be found.[37]

$$\langle \pi_A \rangle \approx w_1 z_{1A} + w_2 z_{2A} + ... + w_A z_{AA} ... + w_N z_{NA} \quad (4)$$

where the set $\{z_{IA}\} = \mathbf{z}_A$ is just constructed by the collection of the QSM N-dimensional descriptor of system A. In eq 4, there is also present another N-dimensional vector, $\{w_I\} = \mathbf{w}$, which is a set of coefficients to be computed to optimally fit all the known property values of the set. This is done in the same way as the problem is solved in empirical QSPR models. There are no mathematical differences present between eq 4 and the usual QSPR models except the descriptor origin; however, the fundamental Q^{2}SPR eq 4 can be deduced from the QSM definitions, as in eq 1, plus the definition of the quantum mechanical expectation value concept. In this sense, not only a linear relationship between molecular properties and generally constructed, unbiased molecular descriptors is proved in this context, but also eq 4 provides the possible existence of a causal relationship between properties and QSM descriptors. Thus, if in order to obtain the vector **w** coefficients, some statistical procedure has to be sought, usually related to a least-squares technique or some connected procedures,[38] the final Q^{2}SPR model contains the seed of a causal connection, obtained by means of quantum theory, between the structure, represented by QSM, and the properties of any molecular set.

Application Examples: Modeling a Steroid Set

To provide a visual picture of the use of MQSM in the QSAR field, the Cramer's 31 steroid set[39,40] will be studied using two different methodologies dealing with MQS as the starting point. This molecular set has been widely employed as a benchmark to test novel QSAR procedures[39−42] and constitutes a suitable example for demonstration purposes. The biological activity related to this molecular set, listed in Table 1, is the affinity for corticosteroid binding globuline (CBG).

This molecular set will be studied using two different protocols:

- Full MQSM, using regularly fitted DF[29−31] weighted by the Coulomb operator, and

- Topological quantum similarity indices.

Statistical Considerations. Statistical procedures, like multiple linear regression (MLR)[43] and partial least squares (PLS),[44] are used here to construct QSAR models. All correlations are evaluated by means of goodness-of-fit (r^2),[45] standard deviation of errors in prediction (s)[45] and goodness-of-fit in cross-validation (r^2_{cv}) from a leave-n-out procedure.[46,47] The authors prefer the usage of r^2_{cv} instead of the r^2 for prediction (q^2).[45] This choice was made because of the existence of negative values in the q^2 definition[48,49] and, in addition, because of ambiguous implementations of this index[2].

Data Set. Molecular geometries used here were supplied by Gasteiger's research group, who showed[39] that the original work[40] contained some mistakes. The experimental activities are due to Dunn.[42]

Example I: Full MQSM Using a Coulomb Operator. The first results involve the use of a Coulomb operator in definition 1. The necessary molecular pairwise superpositions were computed according to the TGSA algorithm[50] (see Figure 1), which aligns the molecules according to their maximal common substructure and permits the easy computation of the optimal QSM. The molecular electronic DF was constructed from previously computed parameters according to the PASA[29−31] fitted to the 3-21G basis set. All measurements among the 31 steroids were collected into a SM and scaled using a CSI, as in eq 2. The columns of the new SM were used as molecular descriptors in a PLS[44] routine to construct the QSAR models.

A number of four descriptors was chosen among a set, made of eight, to construct a final QSAR model, for

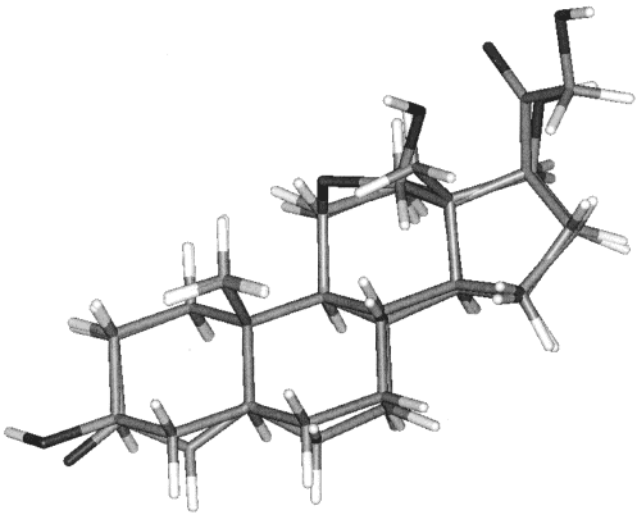

FIGURE 1. Superposition of aldosterone and androstanediol using TGSA.[50]

example,

$$\mathrm{CBG} = 1.196f_1 + 10.79f_2 + 17.62f_3 + 7.993f_4 \quad (5)$$

$$[r^2 = 0.824 \;\; r_{cv}{}^2 = 0.727 \;\; s = 0.447]$$

where $\{f_i\}$ stands for the factors derived of the application of the PLS routine to both the descriptor matrix and the activity data, and $r_{cv}{}^2$ is the goodness-of-fit of the predicted values arising from a leave-one-out procedure. Results of model 5 can be visualized in Figure 2, where cross-validated versus experimental values are plotted. To verify the model validity, a random test[51] has been carried out, randomly permutating the activity vector and constructing models between this random-ordered data and the whole original SM. This procedure, repeated 1000 times, presents the following results:

$$\text{mean } r^2 = 0.321 \quad \text{mean } r_{cv}{}^2 = 0.076 \; (r_{cv} = -0.095)$$
$$\max r^2 = 0.644 \quad \max r_{cv}{}^2 = 0.398$$

Thus, the random models do not achieve significant results. It should be noted that the value of $r_{cv}{}^2$, which may arise from a negative value of r_{cv}, reinforces the nonvalidity of the permutated models, because it implies that those correlations are inverting the data tendency. Figure 3 presents random test results, where a clear separation between original and random models can be distinguished.

Example II: Topological Quantum Similarity Indices. Another successful QSAR/QSPR[2−6] approach relies on the topological paradigm.[52−55] Within this classical approach, the molecular skeleton is represented by means of an undirected graph.[55,56] This graph is codified in terms of a topological matrix (TM),[56−58] **T**, also called connectivity or an adjacency matrix. The TM is symmetric and has dimension *n*X*n*, *n* being the number of atoms in the molecule. The element T_{ij} is 1 if the related atoms *i* and *j* are connected by a graph line and is 0 otherwise. Several numerical indices can be defined from the information

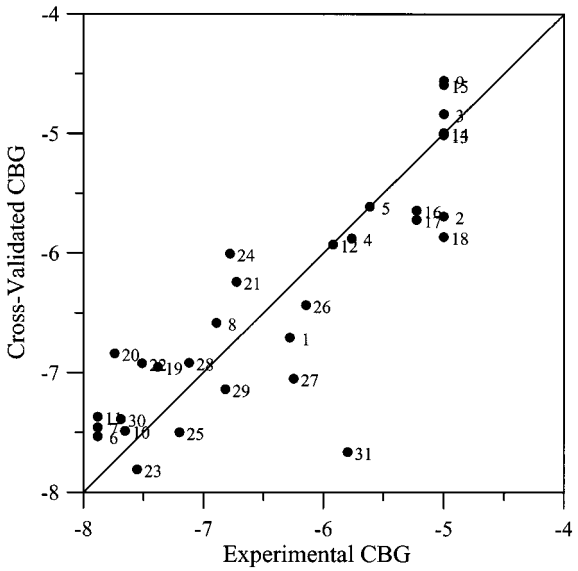

FIGURE 2. Predicted vs experimental values for a set of 31 steroids using Coulomb MQSM.

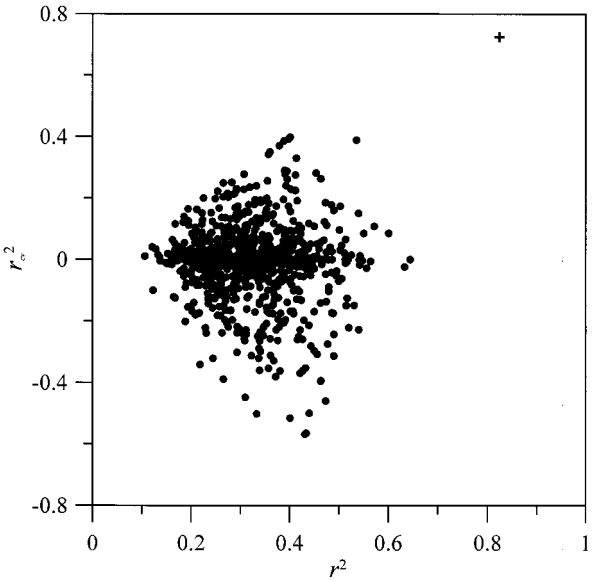

FIGURE 3. Random test for a QSAR model of a 31-steroid set: original point, +; permuted models, ●. (The $r_{cv}{}^2$ from negative values of r_{cv} have kept the sign for clarity.)

contained in the TM. Numerical correlation between classical topological indices and physicochemical or biological properties usually yields acceptable results.[52] Unfortunately, spatial and chemical information may be partially lost in TM.

The topological quantum similarity indices theory extends classical topological theory[25] by redefining the TM in different ways. This is accomplished considering measure 1, assuming that indices A and B refer to atoms in a molecule. In this way, if a density function is attached to each atom,[25,56] integral 1 allows definition of different kinds of TMs, depending on the nature of the operator Ω. For a molecule with *n* atoms, a symmetric *n*X*n* topological quantum similarity matrix (TQSM), $\mathbf{Z} = Z(\Omega) = \{Z_{ij}\}$, can be obtained by collecting the measures of type 1 coming for every atomic pair.

Given a TQSM, an equivalent computation, as in the classical case, can be considered in order to produce the quantum topological indices (TI).[14,59] This point of view allows a redefinition of the classical TIs: Wiener index; Wiener path number; Randić, Schultz, Balaban, and Hosoya indices; Harary number; the generalized connectivity indices of order o and kind k ($^o\chi_k$) of Kier and Hall;[54] and so on.[52,56] In the TQSM framework, the topological distance matrices can be substituted by the 3-dimensional Euclidean distances.[14,56] The full sets of redefined parameters are called topological quantum similarity indices (TQSI), and their origin and the computational details are given in references 14 and 59. Consequently, the general QSM theory leads to the generation of new ab initio molecular descriptors. Several interesting results have been obtained concerning the numerical correlation between indices derived from the TQSM and molecular properties.[14,25,59]

In the present example, two kinds of TMs have been considered: the classical one, **T**, and the TQSM Cioslowski-like matrix **C**, defined as $\mathbf{C} = \{S_{ij}^2\}$, where S_{ij} is the elements of the overlap matrix, which are calculated between pairs of spherical functions centered at atoms i and j.[59] For every TM, a set of 40 indices were computed, and the full data was sent to a multiple linear cross-validation[45,60] and regression program.

In the present application example, all of the combinations of 2, 3, 4, and 5 TQSIs entering a multiple linear model were generated, and the descriptor sets attached to the highest values of the r_{cv}^2 are reported in Table 2.

It is also assumed that a leave-one-out procedure overestimates the predictive capabilities of the tested method,[61,62] so new trends are focused on obtaining results attached to leave-n-out procedures.[46] It has also been demonstrated that when performing linear cross-validation procedures, it is straightforward to obtain the predicted values in a general leave-n-out algorithm.[63]

The final QSAR model using four topological parameters is presented in eq 6. This model involves three kinds of connectivity indices: path (p), cluster (c) and path-cluster (pc)[54] of diverse orders.

$$\mathrm{CBG} = -11.309\ {}^3\chi_c(\mathbf{T}) + 0.431\ {}^7\chi_{pc}(\mathbf{T}) + 7.243\ {}^3\chi_c(\mathbf{C}) - 1.629\ {}^4\chi_p(\mathbf{T}) + 6.857 \quad (6)$$

$$[r^2 = 0.867\ \ r_{cv}{}^2 = 0.803\ \ s = 0.475]$$

Table 2 presents the correlation coefficients for the cross-validation computations that have been carried out for the Cramer set. Leave-n-out tests (n = 1,5) have been performed. Figure 4 shows the 4-descriptors model result, according to eq 6, arising from a leave-2-out procedure. In general, after a leave-n-out test has been performed over a molecular family, each molecule has an attached set of $N_p = \binom{N-1}{n-1}$ predicted values. Except in the case of a leave-one-out test (for which $N_p = 1$), a statistical distribution of predicted values can be analyzed. This distribution may be considered to be Gaussian. In Figure 4, the circles indicate for each molecule the mean value of the corresponding predictions, and the bars in each side are as long as three times the data standard deviation attached to the method dispersion.

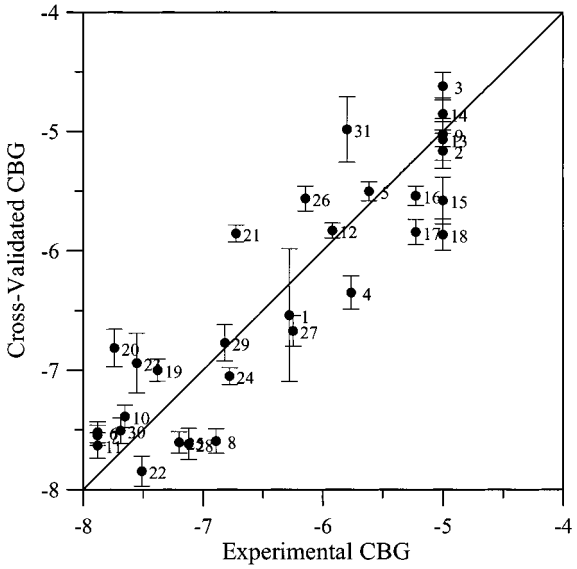

FIGURE 4. Representation of the predicted versus the experimental values obtained for the family of 31 steroids and according to a linear model involving 4 optimized descriptors following a leave-2-out procedure. Error bars are as long as 3σ, the standard deviation attached to a Gaussian distribution of $N_p = 30$ points.

Table 2. Computed r_{cv}^2 Values for Linear Models Involving Several Descriptors and Obtained Following Leave-*n*-Out Procedures

	n in the leave-*n*-out procedure				
no. descriptors	1	2	3	4	5
1	0.394	0.394	0.394	0.394	0.394
2	0.551	0.551	0.551	0.551	0.551
3	0.593	0.592	0.591	0.590	0.590
4	0.803	0.803	0.803		
5	0.848	0.848			

Finally, a random test has also been performed, yielding results similar to the previous example.

Conclusions

It has been shown how the general theory of quantum similarity fundamentals, in some of its extensions, and an application to QSAR theory produces a Q^{2}SAR equation. Despite the general, abstract, and mathematical underlying concepts, numerical and "front end" results can be also reached. In this way, the whole review presented here embraces both the theory and its practical implementations.

As an application example, the Crammer 31 steroids set has been tested within two different, quantum-similarity-related, methodologies.The obtained results are satisfactory, and according to the resulting statistical parameters, the presented QSAR models can be considered sufficiently valid and comparable to other literature proposals.[39−41]

This research has been supported by the Investigation Project No. SAF2000-0223-C03-01 from the Spanish Ministerio de Ciencia

y Tecnología. A University of Girona grant for the investigation project (session date 3/00) is acknowledged by E.B. X.G. benefited from a predoctoral fellowship from the University of Girona. Financial support from the Fundació Maria Francisca de Roviralta is also warmly acknowledged.

References

(1) In Borman, S. New QSAR techniques eyed for environmental assessments. *Chem. Eng. News* **1990**, *68*, 20−23.

(2) Kubinyi, H., Ed. *3D QSAR in Drug Design. Theory, Methods and Applications*; ESCOM Science Publishers: Leiden, 1993.

(3) Kubinyi, H. Quantitative Structure−Activity Relationships in Drug Design. 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 and Sons, Ltd.: Chichester, 1998; Vol. 4, pp 2309−2319.

(4) Charton, M., Ed. *Advances in Quantitative Structure−Property Relationships*; JAI Press: London, 1996; Vol. 1.

(5) Van de Waterbeemd, H., Ed. *Structure−Property Correlations in Drug Research*; Academic Press: San Diego, 1996.

(6) Jurs, P. C. Quantitative Structure−Property Relationships (QSPR). 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 and Sons, Ltd.: Chichester, 1998; Vol. 4, pp 2320−2330.

(7) Boethling, R. S., Mackay, D., Eds. *Handbook of Property Estimation Methods for Chemicals. Environmental and Health Sciences*; Lewis Publishers: London, 2000.

(8) Carbó, R.; Martín, M.; Pons, V. Application of Quantum Mechanical Parameters in Quantitative Structure−Activity Relationships. *Afinidad* **1977**, *34*, 348−353.

(9) Carbó, R.; Arnau, J.; Leyda, L. How similar is a molecule to another? An electron density measure of similarity between two molecular structures. *Int. J. Quantum Chem.* **1980**, *17*, 1185−1189.

(10) Carbó, R.; Calabuig, B. Molecular Similarity and Quantum Chemistry. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; John Wiley: New York, 1990, Chapter 6.

(11) Eyring, H.; Walter, J.; Kimball, G. E. *Quantum Chemistry*; John Wiley: New York, 1944.

(12) Carbó, R.; Calabuig, B.; Vera, L.; Besalú, E. Molecular Quantum Similarity: Theoretical Framework, Ordering Principles, and Visualization Techniques. *Adv. Quantum Chem.* **1994**, *25*, 253−313.

(13) Carbó-Dorca, R.; Besalú, E.; Amat, Ll.; Fradera, X. Quantum Molecular Similarity Measures: Concepts, Definitions, and Applications to Quantitative Structure−Property Relationships. In *Advances in Molecular Similarity*; Carbó-Dorca, R., Mezey, P. G., Eds.; JAI Press: London, 1996; Vol. 1, Chapter 1.

(14) Carbó-Dorca, R.; Amat, Ll.; Besalú, E.; Gironés, X.; Robert, D. Quantum Molecular Similarity: Theory and Applications to the Evaluation of Molecular Properties, Biological Activities and Toxicity. In *Fundamentals of Molecular Similarity*; Carbó-Dorca, R., Gironés, X., Mezey, P. G., Eds.; Kluwer Academic/Plenum Press: New York, 2001; Chapter 12.

(15) Carbó-Dorca, R.; Robert, D.; Amat, Ll.; Gironés, X.; Besalú, E. Molecular Quantum Similarity in QSAR and Drug Design. *Lecture Notes in Chemistry* **2000**, *73*, Springer-Verlag (Berlin).

(16) Carbó-Dorca, R. Quantum Quantitative Structure−Activity Relationships (QQSAR): A Comprehensive Discussion Based on Inward Matrix Products, Employed as a Tool to Find Approximate Solutions of Strictly Positive Linear Systems and Providing s QSAR-Quantum Similarity Measures Connection. In *Proceedings of the European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2000)*; CDROM edited by Facultat d'Informàtica de Barcelona (FIB)−Universitat Politècnica de Catalunya (UPC)−International Centre for Numerical Methods in Engineering (CIMNE): Barcelona, 2000; Computational Chemistry Section, Chapter 12.

(17) von Neumann, J. *Mathematical Foundations of Quantum Mechanics*; Princeton University Press: Princeton, NJ, 1955.

(18) Born, M. *Atomic Physics*; Blackie and Son: London, 1945.

(19) Dirac, P. A. M. *The Principles of Quantum Mechanics*; Clarendon Press: Oxford, 1983.

(20) Mc. Weeny, R. *Methods of Molecular Quantum Mechanics*; Academic Press: London, 1978.

(21) Fradera, X.; Amat, Ll.; Besalú, E.; Carbó-Dorca, R. Application of Molecular Quantum Similarity to QSAR. *Quant. Struct.-Act. Relat.* **1997**, *16*, 25−32.

(22) Amat, Ll.; Carbó-Dorca, R.; Ponec, R. Molecular Quantum Similarity Studies as an Alternative to Log P Values in QSAR Studies. *J. Comput. Chem.* **1998**, *14*, 1575−1583.

(23) Ponec, R.; Amat, Ll.; Carbó-Dorca, R. Molecular Basis of Quantitative Structure-Properties Relationships (QSPR): A Quantum Similarity Approach. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 259−270.

(24) Mezey, P. G.; Ponec, R.; Amat, Ll.; Carbó-Dorca, R. Quantum Similarity Approach to the Characterization of Molecular Chirality. *Enantiomer* **1999**, *4*, 371−378.

(25) Carbó-Dorca, R.; Amat, Ll.; Besalú, E.; Lobato, M. Quantum Similarity. In *Advances in Molecular Similarity*; Carbó-Dorca, R., Mezey, P. G., Eds.; JAI Press: London, 1998; Vol. 2, Chapter 1.

(26) Robert, D.; Gironés, X.; Carbó-Dorca, R. Quantification of the Influence of Single Point Mutations on Haloalkane Dehalogenase Activity: A Molecular Quantum Similarity Study. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 839−846.

(27) Robert, D.; Amat, Ll.; Carbó-Dorca, R. Quantum Similarity QSAR: Study of Inhibitors Binding to Thrombin, Trypsin and Factor Xa, Including a Comparison with CoMFA and CoMSIA Methods. *Int. J. Quantum Chem.* **2000**, *80*, 265−282.

(28) Gironés, X.; Amat, Ll.; Robert, R.; Carbó-Dorca, R. Use of Electron−electron Repulsion Energy as a Molecular Descriptor in QSAR and QSPR. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 477−485.

(29) 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.

(30) Amat, Ll.; 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.

(31) 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.

(32) Carbó, R.; Besalú, E.; Amat, Ll.; Fradera, X. On Quantum Molecular Similarity Measures (QMSM) and Indices (QMSI). *J. Math. Chem.* **1996**, *19*, 47−56.

(33) Besalú, E.; Amat, Ll.; Fradera, X.; Carbó, R. An Application of the Molecular Quantum Similarity: Ordering of Some Properties of the Hexanes. In *QSAR and Molecular Modeling: Concepts, Computational Tools and Biological Applications, Proceedings of the 10th European Symposium on SAR, QSAR and Molecular Modeling*; Sanz, F., Giraldo, J., Manaut, F., Eds.; Prous Science: Barcelona, 1995, Section VI, 396−399.

(34) Carbó-Dorca, R. Stochastic Transformation of Quantum Similarity Matrixes and Their Use in Quantum QSAR (QQSAR) Models. *Intl. J. Quantum Chem.* **2000**, *79*, 163−177.

(35) Robert, D.; Gironés, X., Carbó-Dorca, R. Molecular Quantum Similarity Measures as Descriptors for Quantum QSAR. *Polycyclic Aromatic Compounds* **2000**, *19*, 51−71.

(36) Carbó, R.; Besalú, E.; Amat, Ll.; Fradera, X. Quantum Molecular Similarity Measures (QMSM) as a natural way leading towards a Theoretical Foundation of Quantitative Structure−Activity Relationships (QSPR). *J. Math. Chem.* **1995**, *18*, 237−246.

(37) Carbó-Dorca, R. Inward Matrix Products: Extensions and Applications to Quantum Mechanical Foundations of QSAR. *Theochem* **2001**, *537*, 41−54.

(38) Neter, J.; Wasserman, W.; Kutner, M. H. *Applied Linear Statistical Models*; R. D. Irwin Inc.: Boston, 1990.

(39) Wagner, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of molecular surface properties for modeling corticosteroid binding globulin and cytosolic Ah receptor activity by neural networks. *J. Am. Chem. Soc.* **1995**, *117*, 7769−7775.

(40) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(41) Robert, D.; Amat, L.; Carbó-Dorca, R. Three-Dimensional Quantitative structure−activity relationships from tuned molecular quantum similarity measures. Prediction of the corticosteroid-binding globulin binding affinity for a steroid family. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 333−344.

(42) Dunn, J. F.; Nisula, B. C.; Rodbard, D. Transport of steroid hormones: binding of 21 endogeneus steroids to both testosterone-binding globuline and corticosteroid-binding globuline in human plasma. *J. Clin. Endocrin. Metab.* **1981**, *53*, 58−68.

(43) Myers, R. H. *Classical and modern regression with applications*; PWS−KENT Publishing Company: Boston, 1990.

(44) Wold, S.; Sjösröm, M.; Eriksson, L. Partial least squares projections to latent structures (PLS) in chemistry. In *Encyclopedia of Computational Chemistry*; von Ragué Schleyer, P., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F. III, Screiner, P. R., Eds.; John Wiley and Sons, Ltd.: Chichester, 1994; Vol. 4, 2006−2021.

(45) Montgomery, D. C.; Peck, E. A. *Introduction to linear regression analysis*; Wiley: New York, 1992.
(46) Whitley, D. C.; Ford, M. G.; Livingstone, D. J. Unsupervised Forward Selection: A Method for Eliminating Redundant Variables. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1160−1168.
(47) Wold, S. Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics* **1978**, *20*, 125−127.
(48) Amat, L.; Besalú, E.; Carbó-Dorca, R.; Ponec, R. Identification of Active Molecular Sites Using Quantum Self-Similarity Measures. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (4), 978−991.
(49) Richon, A. B.; Young, S. S. An Introduction to QSAR Methodology. A WEB paper of the *Network Science Corporation*: http://www.netsci.org/Science/Compchem/feature19.html (page acceded on date 25 May 2001).
(50) Gironés, X.; Robert, R.; Carbó-Dorca, R. TGSA: a molecular superposition program based on topo-geometrical considerations. *J. Comput. Chem.* **2001**, *22*, 255−263.
(51) Wold, S.; Erikson, L. Statistical validation of QSAR results. In *Chemometric methods in molecular design*; Van de Waterbeemd, H., Ed.; VCH: New York, 1995, Section IV, 309−318.
(52) Balaban, A. T. Chemical Graphs: Looking Back and Glimpsing Ahead. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 339−350.
(53) Randić, M.; Jerman-Blazic, B.; Trinajstić, N. Development of 3-dimensional molecular descriptors. *Comput. Chem.* **1990**, *14*, 237−246.
(54) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
(55) Harary, F. *Graph Theory*; Addison-Wesley: Reading, 1971.
(56) Mihalic, Z.; Trinajstić, N. A Graph-Theoretical Approach to Structure−Property Relationship. *J. Chem. Educ.* **1992**, *69*, 701−712.
(57) Gutman, I.; Trinajstić, N. Graph Theory and Molecular Orbitals. *Top. Curr. Chem.* **1973**, *42*, 49−93.
(58) Cvetković, D.; Doob, M.; Sachs, H. *Spectra of Graphs*. Academic: New York, 1980.
(59) Besalú, E.; Gallegos, A.; Carbó-Dorca, R. Topological Quantum Similarity Indices and Their Use in QSAR: Application to Several Families of Antimalarial Compounds. *MATCH—Communications in Mathematical and Computational Chemistry* **2001**, *44*, 41−64.
(60) Allen, D. M. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* **1974**, *16*, 125−127.
(61) So, S.-S.; van Helden, S. P.; van Geerestein, V. J.; Karplus, M. Quantitative Structure−Activity Relationship Studies of Progesterone Receptor Binding Steroids. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 762−772.
(62) Shao, J. Linear-Model selection by cross-validation. *J. Am. Stat. Assoc.* **1993**, *88*, 486-494.
(63) Besalú, E. Fast Computation of Cross-Validated Properties in Full Linear Leave-Many-Out Procedures. *J. Math. Chem.* **2001**, *29*, 191−204.

AR010048X