J Math Chem (2011) 49:1558–1572
DOI 10.1007/s10910-011-9841-4

ORIGINAL PAPER

Quantum similarity and discrete representation of molecular sets

Luz Dary Mercado · Ramon Carbó-Dorca

Received: 8 February 2011 / Accepted: 9 May 2011 / Published online: 11 June 2011

Abstract This study describes how quantum similarity theoretical framework can provide arbitrary molecular sets with discrete, general, unbiased, systematic and homogeneous numerical descriptions of their elements. Necessarily associated to quantum objects, such discrete descriptions can be related to a sequence of tensor collections of increasing rank. When choosing a specific tensor rank description of any molecular set, such a molecular descriptor set can be also considered a set of linearly independent arrays, which can act in turn as a basis set of some vector space of the appropriate dimension, equal to the cardinality of the supporting molecular set. Straightforward manipulation of the metric matrix of such tensor basis sets permits the ordering of the molecules in the set, among other interesting possibilities. Assorted Kruskal tree examples on Cramer steroid set are given.

Keywords Quantum objects · Quantum similarity (QS) · Quantum QSPR · Molecular descriptors · Molecular QS tensors · Ordering of molecular sets · Kruskal trees · Cramer steroid set

1 Introduction

In principle, the usual practice in the field of classical discrete numerical molecular description consists of what one can call molecular descriptors choice, a crucial subject, which along the evolution of the question appears nowadays to be still somewhat and somehow arbitrary. Molecular descriptors are usually made by scalar values of quite varied origin and constitute an increasingly growing family of parameters, as can be deduced from the exhaustive molecular descriptor compendium of two volumes,

L. D. Mercado · R. Carbó-Dorca (✉)
Institut de Química Computacional, Universitat de Girona, 17071 Girona, Catalonia, Spain
e-mail: quantumqsar@hotmail.com

reference [1]. When applied to describe a molecular structure, the chosen molecular descriptors are customarily ordered in a vector array, before to proceed for further manipulation, habitually for QSPR purposes.[1]

On the other hand, from a quantum mechanical point of view, molecular structures can be described in a continuous manner, like by means of the well-known density function (DF), which since the report of Hohenberg-Kohn theorem [2] has acquired a relevant position in modern quantum chemistry. Moreover, the theoretical structure within the field of quantum similarity (QS) employs the DF of any given submicroscopic system as a basic element, in order to compare it with other parent quantum systems, see for example references [3–10].

It is from such QS point of view that molecular set description can be dealt in a general, systematic, unbiased and homogeneous non-arbitrary way [10,11]. Moreover, as it will be shown later on, such a discrete QS molecular description is well established in a precise causal background by means of the so-called quantum QSPR (QQSPR) fundamental equation, see for example references [12,13]. As an obvious theoretical result, QS theoretical structure can attach each element contained within a given molecular set into a tensor of a chosen rank, that is: an array of any many-dimensional structure. Although such a performance could have been explicitly described and analyzed from the early stages of the QS theoretical development, this has not been the case, and for multiple reasons such a characteristic and useful possibility was not previously studied with the sufficient profundity which merits. The analysis of this kind of QS molecular tensorial representations will be the subject of the present study.

Thus, the aim of this paper is essentially focused into the description of the collection of the QS discrete mathematical forms, which can be associated to every element of a molecular set. According to such a main purpose, the present study will start with the description of quantum object sets (QOS), in order to provide the appropriate background for the discrete molecular description via this kind of mathematical constructs. The definition of QS measures as the result of the fundamental QQSPR equation will be studied next. From there the tensor description of every molecule in a QOS will be put forward creating a set of possible discrete QOS (DQOS). The theoretical frame will finish with a discussion on the way these tensor descriptions can be handled to construct metric matrices, which can be further manipulated to extract information from the QOS via a DQOS. A set of several examples to illustrate the previous theoretical presentation and to provide a hint for possible simple ordering applications will finish the present study.

2 Quantum object sets

The mathematical general background of QS can be found in a recent résumé [14] review. Therefore this paper will be provided with the basic elements only. A needed

[1] The authors prefer to use the term *quantitative structure-properties relationships*; QSPR in a short spelling way, which defines a general working platform, trying to obtain some functional (usually linear) between molecular structure descriptors and molecular properties, thus encompassing the usual QSAR acrostic.

elementary and preliminary definition consists in the quite general notion of tagged set (TS) [16–18]. Any TS: S can be considered as the Cartesian product of two sets: 1) the object set: O and 2) the tag set: T, that is: $\mathrm{S} = \mathrm{O} \times \mathrm{T}$. This just means that every element of any TS is an ordered pair made of an object and a tag:

$$\forall s \in \mathrm{S} : s = (w; t) \rightarrow w \in \mathrm{O} \wedge t \in \mathrm{T}. \tag{1}$$

Then, a quantum object set (QOS) Q is a particular kind of TS, where the elements of the object set M are well-defined molecular structures and the tag set P consists of the attached quantum mechanical DFs. This can be so as quantum mechanics permits to proceed in such a way that to every molecule: m_I in a well-defined state and configuration, one can attach a unique DF: ρ_I, which according to quantum mechanical lore it contains all the associated molecular information. That is, one can write:

$$\mathrm{Q} = \mathrm{M} \times \mathrm{P} \Rightarrow \forall q_I \in \mathrm{Q} : q_I = (m_I; \rho_I) \rightarrow m_I \in \mathrm{M} \wedge \rho_I \in \mathrm{P}. \tag{2}$$

Therefore the cardinality M of the QOS Q is coincident with the number of molecular structures in M and DFs in P.

In a QOS every quantum object element can be represented by the attached DF tag. If every one of the molecular elements is different from the rest, as it should be to avoid redundancy when studying internal QOS relationships between its elements, then this implies that every DF tag has to be different from the rest, a condition which involves a property such that every DF tag has to be linearly independent from the rest.[2]

3 Fundamental QQSPR equation

Given a QOS, then one can consider another kind of tag set, which can be attached not necessarily to all the elements of the object set. The object set of a molecular QOS can be further divided in two parts. Some molecules in the QOS can bear a known molecular property value and some not. This fact divides the molecular object set in two disjoint sets accordingly:

$$\mathrm{M} = \mathrm{C} \cup \mathrm{U} \wedge \mathrm{C} \cap \mathrm{U} = \emptyset. \tag{3}$$

Where a) the set: C collects all the molecular structures with an attached known property value and b) the set: U contains the molecules with unknown property values. In any case the cardinalities of both sets are related to the total number of quantum objects:

$$M_C \wedge M_U \rightarrow M = M_C + M_U.$$

[2] This is also equivalent as to consider that no DF in the QOS tag set can be expressed as a linear combination of the rest, or what is the same: $\sum_{I=1}^{M} w_I \rho_I = 0 \rightarrow \forall I : w_I = 0$.

In fact every one of these sets is a QOS by itself. For more details see the references [14–18].

Whatever the case may be, from the quantum mechanical point of view, one can suppose that even a complicated molecular property: π, can be attached to an Hermitian operator: Ω in such a way that knowing the associated DF of the quantum object: ρ, it can be written the integral expectation value expression:

$$\pi = \langle \Omega \rho \rangle = \int_D \Omega(\mathbf{r})\, \rho(\mathbf{r})\, d\mathbf{r}, \tag{4}$$

which permits to compute an estimate of the property, considered as a quantum mechanical statistical expectation value. Via the usual quantum mechanical observable operator building rules, the observable physical molecular properties certainly do possess, a clearly defined Hermitian operator, see for example [19], which can be employed in Eq. (4) in order to obtain a computed property value. However, when dealing with complex (in the sense of complicated) molecular properties, like boiling point, biological activity or toxicity, it is not possible to obtain an immediate Hermitian operator form, like those dealing with dipole or quadrupole moments, say.

In order to overcome this problem when studying QOS and complicated observable molecular properties, it is always possible to write an approximate expression of the attached Hermitian operator, which will be called QQSPR operator from now on; in the simpler form and up to second order, it can be written as:

$$\Omega(\mathbf{r}) \approx \Gamma(\mathbf{r}) + \sum_I w_I \rho_I(\mathbf{r}) + \sum_I \sum_J w_I w_J \rho_I(\mathbf{r})\, \rho_J(\mathbf{r}) + O(3), \tag{5}$$

where $\Gamma(\mathbf{r})$ is a known gauge Hermitian operator, which can be overridden if necessary, $\{w_I\}$ a set of unknown coefficients and finally appear the elements of the known tag set DFs: $\{\rho_I\} \subseteq$ P.

The terms present in Eq. (5) correspond to the zero-th, first and second order of the expansion, which in principle can be written up to any arbitrary order, see for example references [20,21]. The definition of the QQSPR operator (5) after substituting it in Eq. (4), permits to rewrite the quantum mechanical expectation value now for a specific molecular structure m_K associated to a tag DF ρ_K as:

$$\pi_K \approx \langle \Gamma \rho_K \rangle + \sum_I w_I \langle \rho_I \rho_K \rangle + \sum_I \sum_J w_I w_J \langle \rho_I \rho_J \rho_K \rangle + O(3) \tag{6}$$

Thus, Eq. (6) allows writing the property π_K for some molecule m_K belonging to the QOS object set up to a third order approximation. The molecules of the set C can be employed, with its known property values, to evaluate the unknown coefficients: $\{w_I\}$, see for example references [22–26]. Besides, one can say that Eq. (6) expresses the so-called fundamental QQSPR equation up to third order.

The QQSPR objective consists to determine in one way or another the coefficients $\{w_I\}$, using the known properties of the set C. Accepting this prospect in order to be able to obtain an estimate, as accurate as possible, of the unknown property values of

the set U; for more details see references [22–26]. This process, which corresponds to a new way to look at the QSPR problem, from the point of view of systematic molecular QS discrete description, avoids the so-called classical QSPR dimensional paradox [26], which haunts the traditional procedures.

4 Quantum similarity measures (QSM) and QS tensors

Equation (6) presents three sets of integrals involving the DF elements of the QOS tag set, namely:

$$\forall K : \langle \Gamma \rho_K \rangle = \int_D \Gamma(\mathbf{r})\, \rho_K(\mathbf{r})\, d\mathbf{r} \rightarrow \mathbf{Z}^{(0)} = \left\{ z_K^{(0)} = \langle \Gamma \rho_K \rangle \right\}, \tag{7}$$

$$\forall I, K : \langle \rho_I \rho_K \rangle = \int_D \rho_I(\mathbf{r})\, \rho_K(\mathbf{r})\, d\mathbf{r} \rightarrow \mathbf{Z}^{(1)} = \left\{ z_{IK}^{(1)} = \langle \rho_I \rho_K \rangle \right\}, \tag{8}$$

$$\forall I, J, K : \langle \rho_I \rho_J \rho_K \rangle = \int_D \rho_I(\mathbf{r})\, \rho_J(\mathbf{r})\, \rho_K(\mathbf{r})\, d\mathbf{r}$$
$$\rightarrow \mathbf{Z}^{(2)} = \left\{ z_{IJK}^{(2)} = \langle \rho_I \rho_J \rho_K \rangle \right\}. \tag{9}$$

Equations (7), (8) and (9) provide the sequence of QS zero-th, first and second order terms corresponding to a first, second and third rank tensors respectively, whose elements are the QSM involving one, two and three DF.

4.1 Zero-th order terms

The gauge operator Γ can be arbitrarily chosen as it only provides an origin shift of the evaluation of the property. Thus, it can be selected as the unit operator as simplest choice; then, the zero-th order terms will become the number of electrons N_K of the respective molecules m_K:

$$\forall K : z_K^{(0)} = \langle \Gamma \rho_K \rangle = \langle I \rho_K \rangle = \langle \rho_K \rangle = \int_D \rho_K(\mathbf{r}) d\mathbf{r} = N_K. \tag{10}$$

In Eq. (10), the integral involving any DF, corresponds to its Minkowski norm. This normalization factor can be employed on DFs to bring them into *shape functions*, see for example [27], by using a homothetic transformation like:

$$\sigma(\mathbf{r}) = \langle \rho \rangle^{-1} \rho(\mathbf{r}).$$

4.2 First order terms

The first order tensorial terms constitute a symmetric matrix or what is the same, a second rank symmetric tensor:

$$\forall I, K : z_{IK}^{(1)} = \langle \rho_I \rho_K \rangle = \langle \rho_K \rho_I \rangle = z_{KI}^{(1)} \rightarrow \mathbf{Z}^{(1)} = \left(\mathbf{Z}^{(1)} \right)^T , \tag{11}$$

which has been customarily called the QS matrix, see for example [8,10,11]. The double density integrals involved in these terms of Eq. (11), correspond to a usual scalar product involving two DFs. A set of DFs Euclidian norms can be obtained within the diagonal terms of the QS matrix:

$$\forall I : z_{II}^{(1)} = \langle \rho_I \rho_I \rangle ,$$

which can provide a new kind of homothetic DF, not described until now; for the occasion will be named *second rank volume function*:

$$\upsilon^{(2)} (\mathbf{r}) = \langle \rho\rho \rangle^{-\frac{1}{2}} \rho (\mathbf{r}) . \tag{12}$$

4.3 Second order terms

While the set of second order terms appearing into the fundamental QQSPR equation create a supersymmetric third rank tensor:

$$\begin{aligned} \forall I, J, K : \langle \rho_I \rho_J \rho_K \rangle &= \langle \rho_I \rho_K \rho_J \rangle = \langle \rho_K \rho_I \rho_J \rangle \\ &= \langle \rho_J \rho_I \rho_K \rangle = \langle \rho_J \rho_K \rho_I \rangle = \langle \rho_K \rho_J \rho_I \rangle . \end{aligned} \tag{13}$$

These third rank tensors have been scarcely employed in QS molecular manipulations [28,29]. Their elements have been habitually (and obviously) named as triple density QS measures. A study in deep of triple density measures will be published elsewhere. Here will be also commented the nature of the integrals in Eq. (13), which can be supposed as a first term of a generalized scalar product involving three DF [30–32]. A triple density norm will correspond to integrals like:

$$\forall I : z_{III}^{(2)} = \langle \rho_I \rho_I \rho_I \rangle , \tag{14}$$

which correspond to the hyperdiagonal of the third rank QS tensor: $\mathbf{Z}^{(2)}$. The triple density norm in Eq. (14), can be employed as a normalization factor to define a third rank volume function, like the Euclidian norm has been employed leading to the homothetic volume function found in Eq. (12):

$$\upsilon^{(3)} (\mathbf{r}) = \langle \rho\rho\rho \rangle^{-\frac{1}{3}} \rho (\mathbf{r}) .$$

Such a way of constructing volume functions, can be completed using the obvious redefinition of the shape function into a first rank volume function, that is:

$$\upsilon^{(1)}(\mathbf{r}) = \sigma(\mathbf{r}).$$

Therefore the fundamental QQSPR equation in its simpler form gives rise to a sequence of increasing rank supersymmetric tensors, whose elements constitute the so-called QS measures and at the same time defines a sequence of volume functions, which are simply built as a homothetic transformation of the DF.

4.4 General rank QSM and volume functions

The integral tensorial elements of any rank, which can be easily inferred from enlarging the fundamental QQSPR equation series (6) to higher terms, see for example [20,21]:

$$z^{(N)}_{I_1 I_2 \ldots I_N} = \left\langle \rho_{I_1} \rho_{I_2} \ldots \rho_{I_N} \right\rangle = \int_D \rho_{I_1}(\mathbf{r}) \rho_{I_2}(\mathbf{r}) \cdots \rho_{I_N}(\mathbf{r}) \, d\mathbf{r} \geq 0, \tag{15}$$

are constructed at any order or rank, by integrals involving positive (in fact non-negative) DFs; thus necessarily have to yield positive definite values and as a consequence can be properly called measures, constituting some kind of generalized volume computation.

The QS tensor family: $\left\{\mathbf{Z}^{(P)} \,|\, P = 0, 1, 2 \ldots, N, \ldots\right\}$ can be also considered as hypersymmetric hypermatrices with positive definite elements, thus belonging to vector semispaces[3] of appropriate dimension [14].

Finally, as Eq. (15) can be associated to a general definition of some kind of scalar product involving a definite number of DFs, from there one can obtain the form of a generic *Nth rank volume function* [15] via a homothetic transformation like:

$$\upsilon^{(N)}(\mathbf{r}) = \left\langle \rho^N \right\rangle^{-\frac{1}{N}} \rho(\mathbf{r}) \rightarrow \left\langle \left(\upsilon^{(N)}(\mathbf{r}) \right)^N \right\rangle = 1, \tag{16}$$

yielding a function which will possess a unit norm of the same rank category.

5 Interpretation of the QS tensors as discrete molecular representations: discrete QOS (DQOS)

The QS tensor family constructed by the series development of the fundamental QQSPR Eq. (6), can be also seen as a sequence of discrete descriptions of the molecules involved in the background QOS definition. Suppose any of the QS tensors,

[3] Vector semispaces are vector spaces defined over the positive definite real numbers, whose additive part is a semigroup, a group without reciprocal elements. No negative numbers or subtractions are involved in vector semispace algebra.

with the elements:

$$\mathbf{Z}^{(N)} = \left\{ z^{(N)}_{I_1 I_2 \ldots I_N} \right\}, \tag{17}$$

it is not difficult to imagine that the generic QS tensor (17) is made of QS tensors of a rank one unit inferior, like:

$$\mathbf{Z}^{(N[v])} = \left\{ z^{(N)}_{I_1 I_2 \ldots I_N} \right\} \rightarrow \forall I : \mathbf{Z}^{(N[v-1])}_I = \left\{ z^{(N)}_{I; I_1 I_2 \ldots I_{N-1}} \right\}, \tag{18}$$

where the supraindices within square brackets, informs about the involved tensor rank. With this definition, the former arbitrary and general QS tensor form can be rewritten as a row (or column) vector:

$$\mathbf{Z}^{(N[v])} = \left(\mathbf{Z}^{(N[v-1])}_1 \; \mathbf{Z}^{(N[v-1])}_2 \; \ldots \; \mathbf{Z}^{(N[v-1])}_M \right). \tag{19}$$

Every element of the subtensor set: $\Theta^{(N)} = \left\{ \mathbf{Z}^{(N[v-1])}_I \, | I = 1, M \right\}$ corresponds to an index, which it is attached to some element of the molecular object set, employed to construct the QS tensor elements. Every tensor element of $\Theta^{(N)}$ somehow represents a discrete numerical image of every molecule in the QOS.

5.1 A simple example

In order to provide an illustrative schematic example of the previous QS discrete molecular description, take for instance the QS matrix $\mathbf{Z}^{(1)}$, which following the previous discussion given above, can be supposedly constructed by first rank tensors or vectors: $\mathbf{Z}^{(1)} = \{ |\mathbf{z}_I\rangle \, | I = 1, M \}$, where M is the cardinality of the QOS and the tensor components correspond to the columns (or rows) of the QS matrix: $|\mathbf{z}_I\rangle = \left\{ z^{(1)}_{JI} \, | J = 1, M \right\}$. Such a partition permits to obtain the first order column vector set: $\Theta^{(1)} = \{ |\mathbf{z}_I\rangle \, | I = 1, M \}$.

5.2 DQOS metric matrices

The set of discrete column vectors $\Theta^{(1)}$ represents a linearly independent set of vectors, which can be employed as a basis set in order to generate a vector space of the appropriate dimension. One can also compute with the $\Theta^{(1)}$ elements a new matrix via their scalar products, which can be easily constructed and computed, using the symmetry of the QS matrix $\mathbf{Z}^{(1)}$, as:

$$\mathbf{S} = \{ s_{IJ} \} = \left(\mathbf{Z}^{(1)} \right)^T \mathbf{Z}^{(1)} = \left(\mathbf{Z}^{(1)} \right)^2 \rightarrow \forall I, J : s_{IJ} = \langle \mathbf{z}_I | \mathbf{z}_J \rangle . \tag{20}$$

The matrix (20) is positive definite: **S**>0 and corresponds to the metric matrix of the discrete molecular representation associated to the set: $\Theta^{(1)}$.

In fact, one can also consider that the correspondence: $\forall I : \rho_I \leftrightarrow |\mathbf{z}_I\rangle$ exists.Therefore, one can define as a new TS, a discrete QOS (DQOS): $\mathrm{D}^{(1)} = \mathrm{M} \times \Theta^{(1)}$, which has to be in correspondence with the original QOS, that is: $\mathrm{Q} \leftrightarrow \mathrm{D}^{(1)}$.

5.3 General discrete QOS

This particular construction permits to easily imagine a series of DQOS, associated to the corresponding arbitrary rank QS tensors collection:

$$\mathrm{D}^{(N)} = \mathrm{M} \times \Theta^{(N)} \Rightarrow \mathrm{Q} \leftrightarrow \mathrm{D}^{(N)}.$$

To end this discussion, one must expect that with the DQOS tag parts one can construct metric matrices, which will behave in the same way as the particular QS matrix associated to the second rank QS tensor $\mathbf{Z}^{(1)}$. It is just a matter of defining the appropriate generalized scalar products leading to the corresponding metric.

6 General DQOS and metric matrices

When considering the partition of the general QS tensor described in Eq. (19), taking in consideration the definitions (17) and (18), with the subtensors of the QS tensor it is easy to imagine the possibility to perform scalar products, yielding a metric matrix of the same dimensions as in the QS matrix $\mathbf{Z}^{(1)}$ case.

In fact, the set: $\Theta^{(N)} = \left\{ \mathbf{Z}_I^{(N[v-1])} | I = 1, M \right\}$ constitute a linearly independent set of some vector space, with vectors possessing the structure of the general rank subtensor set. Thus, a similar correspondence relationship between the DF tag set and the set $\Theta^{(N)}$ elements can be also supposed: $\forall I : \rho_I \leftrightarrow \mathbf{Z}_I^{(N[v-1])}$. The metric matrix, which can be constructed with the set $\Theta^{(N)}$ elements, can be obtained by means of a generalized scalar product definition, which in turn can be formally written in the following manner:

$$\mathbf{S}^{(N)} = \left\{ s_{IJ}^{(N)} = \left\langle \mathbf{Z}_I^{(N[v-1])} * \mathbf{Z}_J^{(N[v-1])} \right\rangle | \forall I, J = 1, M \right\} \tag{21}$$

The generalized scalar products in Eq. (21) are built from the complete sum of a tensor with the aid of the inward product of two tensors. They are easily defined as follows. Given a tensor of rank r : $\mathbf{A}^{[r]} = \left\{ a^{[r]}_{I_1 I_2 \ldots I_r} \right\}$ the complete sum corresponds to the expression:

$$\left\langle \mathbf{A}^{[r]} \right\rangle = \sum_{I_1} \sum_{I_2} \cdots \sum_{I_r} a^{[r]}_{I_1 I_2 \ldots I_r} \tag{22}$$

and the inward product of two tensors of the same rank is a tensor bearing the rank of the product elements:

$$\mathbf{A}^{[r]} * \mathbf{B}^{[r]} = \mathbf{T}^{[r]} \rightarrow \forall I_1, I_2, \ldots I_r : t^{[r]}_{I_1 I_2 \ldots I_r} = a^{[r]}_{I_1 I_2 \ldots I_r} b^{[r]}_{I_1 I_2 \ldots I_r} \tag{23}$$

In this way, definitions (22) and (23) permit to perfectly describe the generalized scalar product (21).

The collection of metric matrices: $\{\mathbf{S}^{(N)}\}$, obtained via the generalized scalar products of the QS tensor collection, upon partition of every rank tensor into a set of subtensors one rank unit less, correspond to a collection of discrete evaluations of the QS matrix $\mathbf{Z}^{(1)}$. They must bear similar information as such double density QSM matrix possesses.

7 Some computational examples

As a working example it has been chosen the Cramer steroid set [33,34], which has also been studied previously in several computational case examples from a QS perspective [35–39]. The QS programming suite defined in a previous study [40] has been employed to construct the QS matrices and the Kruskal trees [4,40] which will be shown in this last section. QS calculations have been performed under the ASA approach [41–45], with atomic overlap populations obtained for the Cramer molecules and geometries optimized at several theory levels under several basis sets [47] QS matrices have been calculated by means of a Monte Carlo technique, which is well documented in reference [40].

Besides of the two biological activities reported in the origin literature of the chosen Cramer steroid set, here we have also included two Kruskal trees with physical properties: the total electronic energy and the dipole moment module. These two scalars appear as a product of the calculations on the studied Cramer steroids and have been chosen as an example of how QS performs with complicated but usual quantum mechanical physical properties ordering. Also this kind of properties have been included to see in this specific case if molecular properties with well-defined quantum mechanical operators could be employed in the context of a QQSPR operator. As there is a large deal of possible current quantum chemical computational variants to solve Schrödinger equation, which can be taken into account, like theory level and basis set chosen for the background calculation of the molecular structures and attached DFs, only a small amount of the obtained pictures in this study will be shown here.

In any case the full amount of the 31 steroids has been chosen in all the drawings; although in the first paper of the molecular family only 21 molecules were shown with attached biological activities. The Cramer steroid structures can be found in any of the previous quoted works, so they will not be repeated here. The results will be presented according to the nature of the molecular property depicted in the tree.

Before the diagrams are shown, some words must be said about the associated discrete metric matrix, which every QS matrix or first order QS tensor has attached, as it has been commented before and described in Eq. (20). Every different QOS QS matrix has been manipulated here within the DQOS structure. However, in all tested cases the generated Kruskal trees have been shown no so invariant in both QOS and DQOS generated QS matrices. Of course, this Kruskal tree variance cannot be taken as a general proof against the plausible isomorphism conjecture one can suppose between QOS and DQOS metric matrices. The present results being so particular, do not clearly contradict the intuition about the existence of such an isomorphic property. In the same

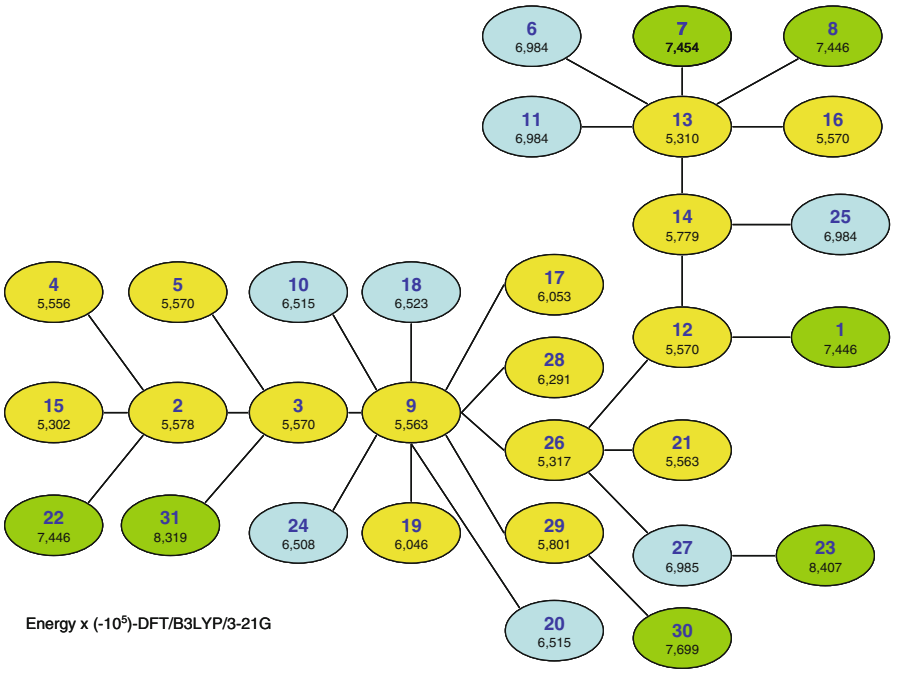

Fig. 1

way, one can expect that the Mendeleev conjecture [46] applies, when use is made of the ordering of a molecular QOS by means of a Kruskal tree or a similar technique based on the attached QS matrix.

The integer numbers in each oval corresponding to the elementary nodes of the Kruskal trees shown in the following pictures, correspond to each steroid in the original Cramer set ordering. The numbers below each ordering number are the studied property values. The colors have been used to show three arbitrary property ranges used to divide each property numerical set.

1. Kruskal tree from a QS matrix computed with a Monte Carlo 70 iteration technique with a background calculation performed at a DFT level using a 3-21G basis set. The numbers are electronic energies in terms of $-10^{-5} Hartrees$ (see Fig. 1).

A large backbone with similar (around -5.5×10^{-5} $Hartrees$) energy values, the yellow ovals are bound together while lesser energy values become marginal endpoints in the tree.

2. Kruskal tree from a QS matrix computed with a Monte Carlo 70 iteration technique using a background calculation performed at a HF level with a 3-21G basis set. The property values correspond to electronic dipole moment modules in Debye units (see Fig. 2).

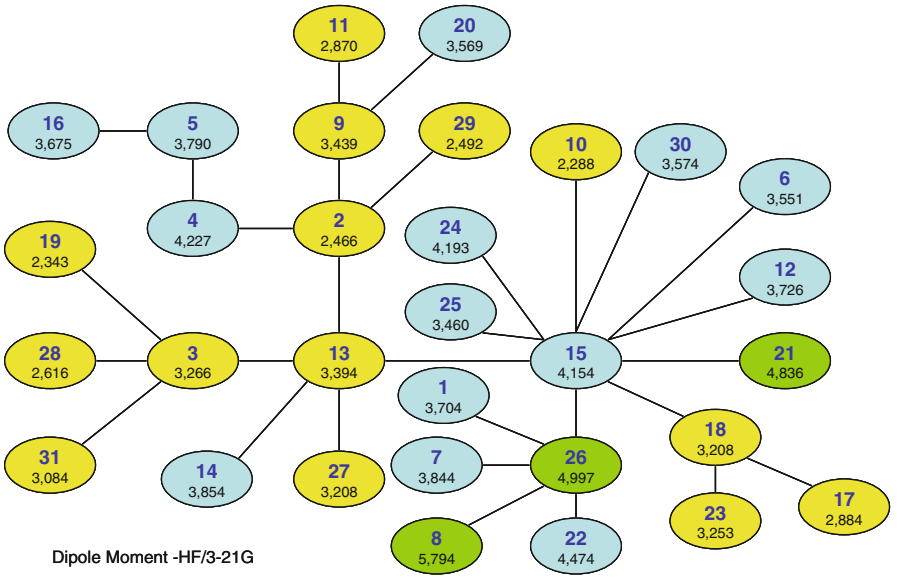

Fig. 2

Lower dipole moment values are with few exceptions forming the larger backbone of the tree. Only steroid 10 appears in a marginal position. Larger dipole molecules can be considered grouped together with two marginal exceptions.

3. Kruskal tree from a QS matrix computed using a Monte Carlo 70 iteration technique with a background calculation performed at a HF level with a 3-21G basis set. The numbers are log(K_a) CBG biological activities (see Fig. 3).

It is interesting here to note that the Kruskal tree with the biological activity depicted here possess a backbone formed by the higher activity values (lower in absolute value). Lesser activity values are terminal nodes of the tree.

4. Kruskal tree from a QS matrix computed with a Monte Carlo 70 iteration technique with a background calculation performed at a DFT level with a 3-21G basis set. The numbers are log(K_a) TBG biological activities (see Fig. 4).

The missing activities of this Kruskal tree (steroids 21 up to 31) are not given in the original Cramer literature source and are depicted in grey here. The positions of the attached steroids in the Kruskal tree can be employed to estimate their corresponding missing values. Usually, the position of the missing activity U set indicates that the TBG values of these steroids may bear values, which can be interpolated or extrapolated from the known C set activity oval values, to which they are connected. For instance, steroids 22, 31 and 24 in the lower left part of the Kruskal tree can be estimated to bear values around −7. Molecule 25 can be estimated to have about −6. Steroids 28, 26 and 29 maybe associated to a −7 value too, while 27, 23 and 30 can be estimated with a value around −9. Obviously this numerical guess of the missing values cannot be equivalent to a QQSPR computation. Neverteless, the estimated U set

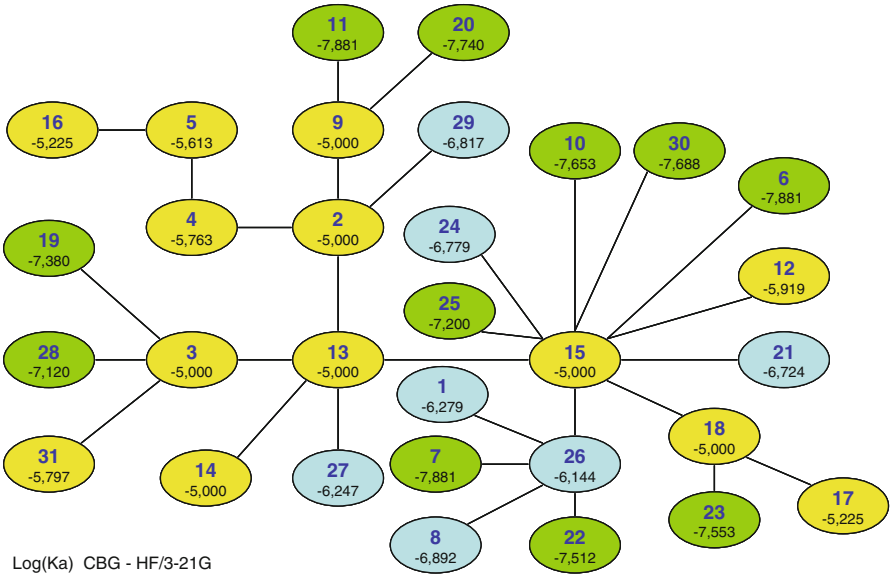
11
-7,881
20
-7,740
16
-5,225
5
-5,613
9
-5,000
29
-6,817
10
-7,653
30
-7,688
6
-7,881
4
-5,763
2
-5,000
24
-6,779
19
-7,380
12
-5,919
25
-7,200
28
-7,120
3
-5,000
13
-5,000
15
-5,000
21
-6,724
1
-6,279
18
-5,000
31
-5,797
14
-5,000
27
-6,247
7
-7,881
26
-6,144
17
-5,225
23
-7,553
8
-6,892
22
-7,512
Log(Ka) CBG - HF/3-21G

Fig. 3

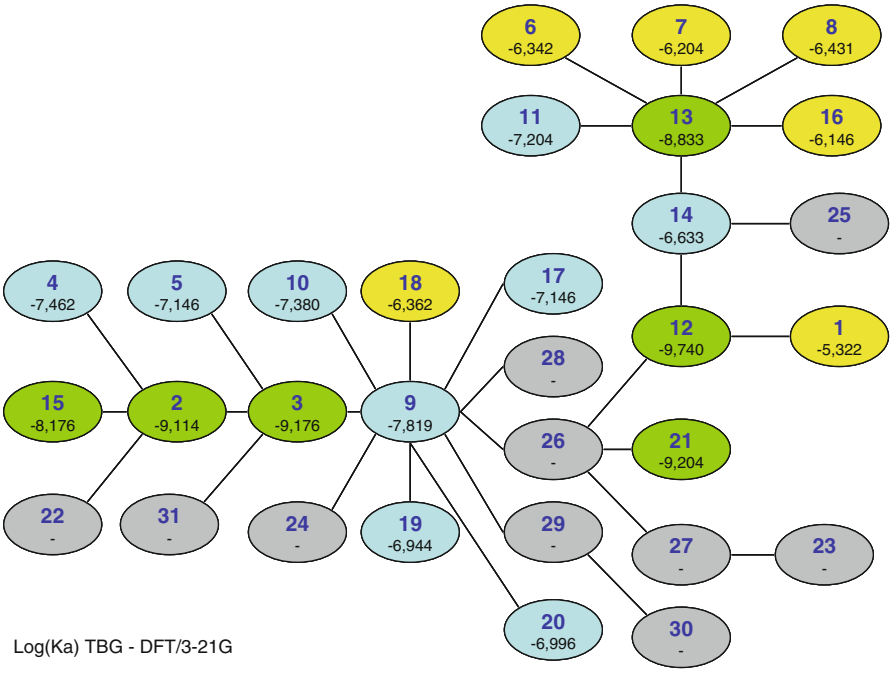
6
-6,342
7
-6,204
8
-6,431
11
-7,204
13
-8,833
16
-6,146
14
-6,633
25
-
4
-7,462
5
-7,146
10
-7,380
18
-6,362
17
-7,146
12
-9,740
1
-5,322
28
-
15
-8,176
2
-9,114
3
-9,176
9
-7,819
26
-
21
-9,204
22
-
31
-
24
-
19
-6,944
29
-
27
-
23
-
20
-6,996
30
-
Log(Ka) TBG - DFT/3-21G

Fig. 4

values can be considered a qualitative estimation which can help to seek for coherence in a further quantitative calculation.

8 Conclusions

Quantum Similarity Techniques permit to obtain a general description of molecular sets, which is unbiased, homogeneous, universal and flexible. As an example of the possibilities of such molecular description mode, ordering of a molecular set by means of Kruskal trees has been performed for several physical properties and biological activities.

References

1. R. Todeschini, V. Consonni, in *Methods and Principles in Medicinal Chemistry*, ed. by R. Mannhold, H. Kubinyi, H. Timmerman Molecular descriptors for chemoinformatics, vol. 41 (Wiley, London, 2009)
2. P. Hohenberg, W. Kohn, Phys. Rev. **136**, B864 (1964)
3. R. Carbó, L. Leyda, M. Arnau, Intl. J. Quant. Chem. **17**, 1185–1189 (1980)
4. R. Carbó, L. Domingo, Intl. J. Quant. Chem. **32**, 517–545 (1987)
5. R. Carbó, B. Calabuig, Intl. J. Quant. Chem. **42**, 1681–1693 (1992)
6. R. Carbó, B. Calabuig, Intl. J. Quant. Chem. **42**, 1695–1709 (1992)
7. R. Carbó, E. Besalú, B. Calabuig, L. Vera, Adv. Quant. Chem. **25**, 253 (1994)
8. R. Carbó, E. Besalú, in *Theoretical Foundation of Quantum Similarity*, ed. by R. Carbó Molecular similarity and reactivity: from quantum chemical to phenomenological approaches, understanding chemical reactivity, vol. 14 (Kluwer Academic Publishers, Amsterdam, 1995), pp. 3–30
9. R. Carbó, E. Besalú, L. Amat, X. Fradera, J. Math. Chem. **19**, 47 (1996)
10. R. Carbó-Dorca, E. Besalú, J. Molec. Struct. (Theochem) **451**, 11 (1998)
11. R. Carbó-Dorca, L. Amat, E. Besalú, M. Lobato, in *Quantum Similarity.* Advances in molecular similarity, vol. 2, (JAI Press Inc, Greenwich (Conn.), 1998), pp. 1–42
12. R. Carbó-Dorca, E. Besalú, Intl. J. Quan. Chem. **88**, 167 (2002)
13. R. Carbó-Dorca, J. Math. Chem. **36**, 241 (2004)
14. R. Carbó-Dorca, E. Besalú, Shells, point cloud huts, generalized scalar products, cosines and similarity tensor representations in vector semispaces IQC Technical Report TC 2001-1
15. R. Carbó-Dorca, Quantum Similarity Volume Functions and Generalized Carbó Indices, IQC Technical Report TC 2001-2
16. R. Carbó-Dorca, J. Math. Chem. **22**, 143 (1997)
17. R. Carbó-Dorca, J. Math. Chem. **23**, 353 (1998)
18. R. Carbó-Dorca, Advances in molecular similarity vol. 2, (JAI Press Inc, Greenwich (Conn.), 1998) p. 43
19. A. Messiah, *Quantum Mechanics* (Dover Publications, New York, 1999)
20. R. Carbó-Dorca, E. Besalú, Intl. J. Quan. Chem. **88**, 167 (2002)
21. R. Carbó-Dorca, A. Gallegos, in *Quantum Similarity and Quantum QSPR (QQSPR)*, ed. by R. Meyers Encyclopedia of complexity and systems science (Springer, New York, 2009), pp. 7422–7480
22. R. Carbó-Dorca, SAR and QSAR Environ. Res. **18**, 265 (2007)
23. R. Carbó-Dorca, S. Van Damme, Theor. Chem. Acc. **118**, 673 (2007)
24. R. Carbó-Dorca, S. Van Damme, Afinidad **64**, 147 (2007)
25. R. Carbó-Dorca, S. Van Damme, Intl. J. Quan. Chem **108**, 1721 (2007)
26. R. Carbó-Dorca, A. Gallegos, A.J. Sánchez, J. Comput. Chem. **30**, 1146 (2008)
27. P. Bultinck, R. Carbó-Dorca, J. Math. Chem. **36**, 191 (2004)
28. R. Carbó, B. Calabuig, E. Besalú, A. Martínez, Mol. Eng. **2**, 43 (1992)
29. D. Robert, R. Carbó-Dorca, J. Chem. Inf. Comp. Sci. **38**, 620 (1998)
30. R. Carbó-Dorca, J. Math. Chem. **32**, 201–223 (2002)
31. R. Carbó-Dorca, J. Math. Chem. **44**, 628 (2008)

32. R. Carbó-Dorca, J. Math. Chem. **47**, 331 (2010)
33. R.D. Cramer III, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. **110**, 5959 (1988)
34. M. Wagener, J. Sadowski, J. Gasteiger, J. Am. Chem. Soc. **117**, 7769 (1995)
35. M. Lobato, L. Amat, E. Besalú, R. Carbó-Dorca, Quant. Struct. Act. Relat. **16**, 1 (1997)
36. D. Robert, L. Amat, R. Carbó-Dorca, J. Chem. Inf. Comp. Sci. **39**, 333 (1999)
37. R. Carbó-Dorca, L. Amat, E. Besalú, X. Gironés, D. Robert, J. Mol. Struct. (Theochem) **504**, 181 (2000)
38. E. Besalú, X. Gironés, L. Amat, R. Carbó-Dorca, Acc. Chem. Res. **35**, 289 (2002)
39. R. Carbó-Dorca, S. Van Damme, Intl. J. Quant. Chem. **108**, 1721 (2008)
40. R. Carbó-Dorca, E. Besalú, L.D. Mercado, J. Comp. Chem. **32**, 582 (2011)
41. R. Carbó-Dorca, J. Math. Chem. **47**, 331 (2010)
42. P. Constans, R. Carbó, J. Chem. Inf. Comp. Sci. **35**, 1046 (1995)
43. P. Constans, X. Fradera, L. Amat, R. Carbó, in *Quantum Molecular Similarity Measures (QMSM) and the Atomic Shell Approximation (ASA). Proceedings of the 2nd Girona Seminar on Molecular Similarity. July 1995.* Advances in molecular similarity, vol. 1. (JAI Press Inc, Greenwich (Conn.), 1996), p. 187
44. L. Amat, R. Carbó-Dorca, J. Comput. Chem. **18**, 2023 (1997)
45. L. Amat, R. Carbó-Dorca, J. Comput. Chem. **20**, 911 (1999)
46. R. Carbó, E. Besalú, Afinidad **53**, 77 (1996)
47. Gaussian 03, Revision B.05, M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople (Gaussian Inc., Wallingford CT, 2004)