J Math Chem (2012) 50:1161–1178
DOI 10.1007/s10910-011-9960-y

ORIGINAL PAPER

Centroid origin shift of quantum object sets and molecular point clouds description and element comparisons

Ramon Carbó-Dorca · Emili Besalú

Received: 23 November 2011 / Accepted: 16 December 2011 / Published online: 25 December 2011

Abstract Shifting the origin of a known quantum object set (QOS) or of some discrete molecular point cloud (MPC) by choosing the centroid of such sets, leads to the way to produce quantum similarity matrices (SM) and tensors, *which can be systematically referred to a canonical origin*, whatever their nature, dimension or cardinality. In this paper the source, significance and properties of such centroid origin shift and the characteristics of the resultant shifted SM are discussed in deep. From such an analysis some interesting applications emerge; as, for instance, a new collection of MPC ordering possibilities. In addition, although all the procedures in this work are described by means of a quantum similarity theoretical background, based on QOS structure within the space of molecules, everything can be also easily implemented in a classical QSPR framework made of molecular numerical images attached to discrete molecular vectors, constructed with well-defined descriptor parameters.

Keywords Density functions · Shape functions · Centroid origin shift · Quantum object sets (QOS) · Molecular point clouds (MPC) · Molecular point huts (MPH) · Quantum similarity · Quantum similarity matrices (SM) · Molecular dissimilarity Euclidian distances · Molecular similarity Carbó indices (CI) · Space of molecules · Space of descriptors

R. Carbó-Dorca (✉) · E. Besalú
Institute of Computational Chemistry, University of Girona, Girona 17071, Catalonia, Spain
e-mail: quantumqsar@hotmail.com

1 Introduction

In a previous paper [1] one of us proposed a new way to solve practical problems in the field of quantitative structure-properties relationships (QSPR)[1] and at the same time circumvent the computational situation, which was there called *dimensionality paradox*,[2] constituting a distinctive characteristic, almost always present in classical QSPR statistical model building, see for example two classical recent references [2,3]. The procedures developed for this mentioned purpose were applied in a classical molecular descriptor vectorial environment, as a way to show the general application of a quantum QSPR (QQSPR) based methodology to both classical and quantum computational backgrounds. Besides, the cited paper has to be considered at the same time as the first contribution of a series of *notes in QSPR* and as the last contribution of a long list of previous papers published in various sources [4–7], studying possible dimensionality paradox free QQSPR mathematical processes dealing with the *space of molecules*[3] instead of being based upon the *space of parameters*, the usual working background of classical QSPR. The present contribution has to be considered as a contextual study for further development in such research direction.

It must be also said that molecular quantum similarity [10–12] theoretical background fundaments the general dimensionality paradox free working scheme in the space of molecules, proposed in the previously mentioned reference [1], even if this previous paper is practically structured in a classical molecular descriptor way. Essentially, the working framework of this initial paper was relying in a symmetrical matrix constructed with the Euclidean scalar products of the discrete molecular descriptor vectors, which in turn have been considered rough approximations of the infinite

[1] In this paper the term QSPR is used preferably, as it encompasses both the most usual QSAR and the less employed QSTR acrostics. In order to distinguish QSPR procedures based upon quantum similarity from the empirical QSPR parametric framework, the terms: quantum QSPR and classical QSPR will be respectively employed.

[2] Contemporary examples of the *dimensionality paradox* can be evidenced in some relatively recent papers chosen among quite a large quantity of publications like: (a) [8], where more than 700 molecules have been studied within a classical QSPR model set up, resulting in an equation bearing less than 10 descriptor parameters and (b) [9] where a set of 101 molecules was studied with QSPR 4 descriptor models.

[3] Often in the QSPR or QSAR common literature the term *chemical space* appears. This concept it is not currently used with the same sense as a possible *molecular space* notion or the equivalent term: *space of molecules* given here, but in a generic fuzzy manner, sometimes even using meaningless and preposterous literary ways. The authors would like to find within the QSPR literature a sharp definition of such *chemical space* term, but until the present time found none. So, in order to employ the terminology of the present work as far away as possible from the usual classical QSAR fuzzy one employed by some authors, the term "space of molecules" has been chosen and will be used here, although "molecular space" might be also synonymously used. This concept has been and will be also employed here as a companion of the term: *space of parameters* or the synonymous concept:*space of descriptors*, which is also a well-defined notion with a precise meaning, commonly used in the task to build up a set of discrete linearly independent vectors, empirically describing a MPC. From such a vectorial descriptor space construct it is normally obtained a QSPR linear equation intently employed for molecular unknown property estimation purposes. The statistical techniques, employed upon the vectors of the space of descriptors, within the empirical QSPR equation search algorithms, which drastically reduce the dimension of molecular descriptors' space, generate the *dimensionality paradox*.

dimensional quantum molecular density functions (DF), the ultimate descriptors containing all the information about the elements of quantum object sets (QOS)[4] [13,14].

Such a matrix, which one can call similarity matrix (SM), made of scalar products acts as a metric or Gram matrix, whose columns in order to be operative need to be considered a priori as linearly independent vector elements, generating the space of molecules in a discrete way and thus possessing a dimension equal to the number of known molecular structures to be studied.

The previous work and the provided initial considerations of the present study have put in evidence that it might be essential to explore the possibility to perform an origin translation of the vectors composing the QOS, and as a consequence the attached *molecular point cloud* (MPC)[5] vectors in doing such a translation can generally will possess a well-defined origin.

Furthermore, such a procedure will allow the study of the shifted MPC vector sets properties and might be taken as the initial step in order to analyze subsequent possible computational applications.

After some, as simpler as possible but canonical, origin shift definition, the new shifted SM might be further immediately used as a completely new background building blocks to computationally obtain the already defined QQSPR in [1] or a simpler set of Quantum Qualitative Structure-Properties Relationships (QQlSPR),[6] by means of MPC elements ordering; peruse, for example, references [11,15] for computational details on ordering algorithms.

Hence, whatever the nature, dimension and cardinality of the problem to be handled could be, the main purpose of such centroid origin shift was intended to describe an unsophisticated procedure, applicable to any QOS. Such a procedure must permit to refer whichever QOS or its attached MPC to a unique canonical origin, obtaining in this way a single reference framework for all the possible choices of QOS.

Manipulations of quantum SM with various purposes have been formerly described, for instance the stochastic transformation [16,17] and more recently other possibilities, using the positive definite nature of the original matrices, have been also theoretically discussed [18].

However, no discussion has been put forward about the fact that QOS or the attached MPC corresponds to an inhomogeneous description of their vectorial origin, nor have been analyzed the advantages one can foresee towards a possible common definition

[4] A *quantum object set* (QOS) is a particular case of a composite tagged set, whose elements are quantum systems, the objects, attached to the appropriate quantum mechanical DF, acting as tags. See below in the next sections and peruse the literature [12–14] for more details.

[5] By a *molecular point cloud* (MPC) can be understood a set of homogeneous mathematical objects, whose elements are constituted by discrete tensorial descriptions [15] of a known molecular set. Here it is supposed that such a molecular discrete description elements are constructed by means of quantum similarity procedures theoretical background; although they can be also constructed from the usual building up methodology of descriptors usually employed in classical QSPR, as it has been done in reference [1] as commented before. See, for example, a recent résumé [12] of the most common concepts applied in quantum similarity and quantum QSPR.

[6] In order to distinguish quantitative from qualitative structure-properties relationships it will be used the acrostic QSPR for its usual meaning and QlSPR for the second acception, Ql obviously meaning qualitative. Both acrostics can have an extra initial letter Q meaning quantum. Thus, it might be written QQlSPR, quantum qualitative structure-properties relations.

feature, encompassing *all* possible QOS and their attached MPC, whenever they can be associated to a well-defined universal, canonical origin.

Therefore, according to these previous considerations, this paper will start first defining a straightforward centroid origin shift procedure associated to the tags of a given QOS, providing the background description of a dimensionality paradox free algorithm in order to evaluate unknown molecular properties[7] if needed. Next will be discussed the centroid origin shift over the Hilbert semispace of the density functions tags of a QOS, considered as infinite dimensional descriptor vectors of a given MPC. Later on it will be debated the possible transformations of the shifted elements of SM, constructing new similarity elements to order MPC constituents, which in turn provide the background for applying QQlSPR procedures, by invoking the Mendeleev conjecture [19]. Provided next, another section will describe how to use the centroid origin shifted quantum SM to build higher order Carbó indices (CI) [20–22], considered as a way to obtain generalized tensorial descriptions of molecules. In this manner the MPC structure might be redefined using higher dimensional tensor spaces. Next, as a final section, it will be given a discussion about Euclidian distances and scaled distance dissimilarity indices [23].

2 Quantum object sets (QOS)

A tagged set: $Z = \{z_I \,|I = 1, N\}$, being N the set cardinality, is the Cartesian product of two sets: $Z = O \times T$, where: $O = \{o_I \,|I = 1, N\}$ is the set of the objects and $T = \{t_I \,|I = 1, N\}$ the set of tags. Therefore one can write: $\forall I = 1, N : z_I = (o_I; t_I) \in Z$. See for more details on tagged sets references [13,14].

A QOS is a tagged set: $Q = O \times P$ made by a set of submicroscopic objects and a set of quantum mechanical DF: $P = \{\rho_I \,|I = 1, N\}$, as elements of the tag set.

3 Hilbert semispace source of centroid origin shifts over QOS

After this previous QOS definition, one can use the Hilbert semispace tag set P and define a central averaged DF by means of the trivial expression:

$$\rho_C = N^{-1} \sum_I \rho_I \rightarrow \langle \rho_C \rangle = N^{-1} \sum_I \langle \rho_I \rangle = N^{-1} \sum_I \nu_I = \nu_C \tag{1}$$

with the DF Minkowski norms being defined as:

[7] The advantage of dimensionality paradox free QSPR procedures, when working within the space of molecules, consists into the fact that there cannot be present the modeling problems recently described in a clear and exhaustive manner [24], which appear when working within the descriptors' space. Although one can be sure that other, not yet disclosed, difficulties surely might appear within the space of molecules when used in QSPR studies, with respect to the designed algorithms for evaluating molecular unknown properties.

$$\forall I = 1, N : \langle \rho_I \rangle = \int_D \rho_I (\mathbf{r})\, d\mathbf{r} = \nu_I . \tag{2}$$

Then, the centroid DF can be seen as a function describing the arithmetic average of the number of particles ν_C of all QO's present in the QOS.

The same can be done with a QOS, where the tag set can be made of the corresponding shape functions (ShF) tag set S associated to the DF set P, which can be easily defined as:

$$P = \{\rho_I \,|I = 1, N\} \rightarrow \forall \langle \rho_I \rangle = \nu_I \wedge \sigma_I = \nu_I^{-1} \rho_I \wedge \langle \sigma_I \rangle = 1 \rightarrow S = \{\sigma_I \,|I = 1, N\}$$

Then one can write the shape centroid function as:

$$\sigma_C = N^{-1} \sum_I \sigma_I \rightarrow \langle \sigma_C \rangle = N^{-1} \sum_I \langle \sigma_I \rangle = N^{-1} N = 1, \tag{3}$$

proving that, while the centroid DF has an averaged number of particles $\nu_C \in \mathbf{R}^+$ as Minkowski norm, the ShF centroid can be considered as another ShF bearing a unit norm and therefore being an element of the unit shell of the vector semispace containing ShF.

Thus, from this point of view it seems that centroid ShF's are Minkowski norm-coherent functions, while the centroid DF's presents such a norm producing some real non integer number of particles. Despite of this, in this paper it will be currently used the centroid DF, as the corresponding ShF driven mathematical development will become practically identical, just substituting the DF by ShF symbols. More than that, the whole reasoning collection which follows here in this contribution also becomes perfectly valid within a classical description of the QOS; that is: with the DF or ShF substituted by discrete molecular vectors, constructed by chosen descriptor parameters.[8] In this latter case it is only need to consider the DF symbols employed here as terms associated to the classical molecular discrete vectorial description.

Once defined the QOS centroid DF as done in Eq. (1), one can shift every DF QOS tag in the following way:

$$\forall I = 1, N : \theta_I = \rho_I - \rho_C \rightarrow \mathrm{P}^{(C)} = \{\theta_I \,|I = 1, N\};$$

and ShF can be handled straightforwardly in the same manner.

Similar results will be obtained when, instead of the average centroid DF shown in Eq. (1), a convex linear combination is used as an origin shift, that is:

$$\rho_C = \sum_I \alpha_I \rho_I \wedge \{\alpha_I\} \subset \mathbf{R}^+ \wedge \sum_I \alpha_I = 1, \tag{4}$$

[8] In order to have a better matching between classical QSPR descriptor vectors and quantum DF or ShF functions used in QQSPR or QQlSPR, perhaps it will be convenient to transform *a priori* such vectors into positive non-negative, conveniently normalized vectors.

taking into account Eq. (4), then the centroid DF as defined in Eq. (1) can be associated to a particular convex combination, where: $\forall I : \alpha_I = N^{-1}$.

Also, as defined in Eq. (3), the centroid ShF can be seen as a particular linear combination of the DF tag set P, where the coefficients have been chosen according to the rule:

$$\forall I = 1, N : \alpha_I = (N\nu_I)^{-1}.$$

Moreover, one must be aware of the fact that the centroid shifted set $\mathrm{P}^{(C)} = \{\theta_I \,|I = 1, N\}$ is no longer composed of DF but of DF *differences*, providing a set of functions which cannot be positive definite anymore, consequently not possessing the usual non-negative definition DF or ShF characteristics. Thus, the Minkowski norm of the origin shifted DF is no longer a norm but a pseudonorm, which can be written as:

$$\forall I : \langle\theta_I\rangle = \langle\rho_I\rangle - \langle\rho_C\rangle = \nu_I - N^{-1}\sum_{J=1}^{N}\nu_J = \left(1 - N^{-1}\right)\nu_I - N^{-1}\sum_{J\neq I}\nu_J.$$

Therefore, it is better to compute the Euclidian norm rather than Minkowski's[9] of the centroid shifted DF set, that is

$$\begin{aligned}\forall I : \left\langle|\theta_I|^2\right\rangle &= \langle\theta_I|\theta_I\rangle = \left\langle|\rho_I - \rho_C|^2\right\rangle = \langle\rho_I|\rho_I\rangle - 2\langle\rho_I|\rho_C\rangle + \langle\rho_C|\rho_C\rangle \\ &= Z_{II} - 2N^{-1}\sum_J Z_{IJ} + N^{-2}\sum_J\sum_K Z_{JK} \\ &= Z_{II} - 2N^{-1}\langle|\mathbf{z}_I\rangle\rangle + N^{-2}\langle\mathbf{Z}\rangle\end{aligned}$$

a result which appears to be coincident with the diagonal elements of the centroid shifted similarity matrix: $\mathbf{Z}^{(C)}$. See the next paragraph below for a definition of quantum similarity matrices and the centroid shifted ones.

Finally, there remains to analyze the linear independence of the centroid shifted DF. In fact, while the original DF tag set P might be canonically linearly independent, the centroid shifted tag set $\mathrm{P}^{(C)}$ might be not. To test this linear independence property, it is usual to find out the outcome for the coefficients of the linear combination of the elements of $\mathrm{P}^{(C)}$ equaled to zero, which provides the sequence:

$$\begin{aligned}&\sum_{I=1}^{N}\alpha_I\theta_I = 0 \rightarrow \sum_{I=1}^{N}\alpha_I\left(\rho_I - \rho_C\right) = 0 \rightarrow \sum_{I=1}^{N}\alpha_I\rho_I - N^{-1}\sum_{I=1}^{N}\alpha_I\sum_{J=1}^{N}\rho_J \\ &= 0 \rightarrow \sum_{I=1}^{N}\alpha_I\rho_I - \lambda\sum_{I=1}^{N}\rho_I = 0\left(\Leftarrow N^{-1}\sum_{J=1}^{N}\alpha_J = \lambda\right) \rightarrow \sum_{I=1}^{N}\left(\alpha_I - \lambda\right)\rho_I\end{aligned}$$

[9] In fact, the Euclidian norm of a given function coincides with the Minkowski norm of the function module squared. It is just a matter of from which point of view one is considering such mathematical operations.

$$= 0 \Rightarrow \forall I : \alpha_I - \lambda = 0 \Rightarrow \forall I : \alpha_I = \lambda \Rightarrow \forall I : \alpha_I = (N-1)^{-1} \sum_{J \neq I} \alpha_J$$

If the origin shifted DF set $\mathrm{P}^{(C)}$ is linearly independent then the starting equation only admits the trivial solution: $\forall I : \alpha_I = 0$. However, the result obtained after substituting the shifted DF by the difference of the original ones and the centroid DF, means that every member of the last shifted coefficient collection has to be zero: $\forall I : \alpha_I - \lambda = 0$, supposing the initial DF set P linearly independent. However, this last result is the same as to consider the original coefficient set, associated to the initial linearly independent DF, might be non-zero, that is: $\forall I : \alpha_I = (N-1)^{-1} \sum_{J \neq I} \alpha_J$, which leads to a contradiction. Therefore, the centroid origin shifted DF set $\mathrm{P}^{(C)}$ has to be linearly dependent.

For instance, within the simplest dimension, say: $N = 2$, this property will be reduced to a pair of coefficient equalities, which further reduces to a unique equation, that is:

$$\alpha_A = 2^{-1} (\alpha_A + \alpha_B) \wedge \alpha_B = 2^{-1} (\alpha_A + \alpha_B) \rightarrow \alpha_A = \alpha_B$$

which admits, besides the trivial solution, an infinite number of fulfillments. Consequently, centroid shifted sets of type $\mathrm{P}^{(C)}$ are linearly dependent, even if the original set P is linearly independent. The meaning of this can be associated to the loss of a degree of freedom when starting at P and ending, via centroid origin shifting, into: $\mathrm{P}^{(C)}$. Therefore if the original DF tag set generates an N-dimensional subspace, the centroid shifted tag set $\mathrm{P}^{(C)}$, generates an $(N-1)$-dimensional one.

4 Quantum similarity matrix (SM)

The QOS DF tag sets can easily generate Gram matrices via their scalar products, which originally become metric matrices. A quantum SM, $\mathbf{Z}$, which can be associated to a $[N \times N]$ metric associated to a QOS tag set made of quantum mechanical DF $\mathrm{P} = \{\rho_I \,|I = 1, N\}$, has equivalent row and column possible partitions, which can be written as follows:

$$\mathbf{Z} = \begin{bmatrix} \langle \mathbf{z}_1 | \\ \langle \mathbf{z}_2 | \\ \vdots \\ \langle \mathbf{z}_N | \end{bmatrix} = \begin{bmatrix} |\mathbf{z}_1\rangle \; |\mathbf{z}_2\rangle \cdots |\mathbf{z}_N\rangle \end{bmatrix}$$

Such row and column partitions can be in principle structured using the trivial definitions, associated to the overlap quantum similarity integrals:[10]

[10] Quantum similarity integrals can be defined more generally than the overlap of two density functions, several positive definite weighting operators can be used, see for more details [10]. Overlap similarity measures have been chosen here for simplicity, though. The integral symbols can be taken as a simplified way to represent a general quantum similarity integral structure.

$$\begin{aligned}\mathbf{Z} = \{Z_{IJ} = \langle \rho_I | \rho_J \rangle \,|I, J = 1, N\} \\ \wedge |\mathbf{z}_I\rangle = \{Z_{JI} \,|J = 1, N\} \\ \wedge \langle \mathbf{z}_I| = \{Z_{IJ} \,|J = 1, N\}.\end{aligned} \tag{5}$$

The matrix $\mathbf{Z}$ is also symmetrical, hence it can be also written:

$$\mathbf{Z} = \mathbf{Z}^T \rightarrow \forall I, J : Z_{IJ} = Z_{JI}.$$

Moreover, both row and column dual vector sets: $R = \{\langle \mathbf{z}_I|\} \wedge C = \{|\mathbf{z}_I\rangle\}$ represent in a discrete form an attached set of N *quantum objects*:$\mathrm{O} = \{o_I\}$, which in turn have as original tags the DF elements of the set P. That is, one can define a *discrete* QOS [12] by means of one of the couple of Cartesian products:

$$\mathrm{Q}_R = \mathrm{O} \times \mathrm{R} \vee \mathrm{Q}_C = \mathrm{O} \times \mathrm{C}.$$

Besides, if the quantum object set O is made of molecules, the discrete tag set elements, considered as a whole, permit to form a $N-$dimensional *molecular point cloud* ($N-$D MPC) [12]; thus, transforming in this manner a set theoretical concept into a geometrical object one.

Any $N-$D MPC can be scaled and transformed by various means [16–18]. Among many possibilities one can choose to divide each SM element by the number of electrons of each involved molecular structure, then the MPC transforms into an equivalent *molecular point hut* (MPH) [12], and can be considered made with the column or row vector of a SM constructed in turn with the aid of ShF tag set. The transformation of a DF based SM into a ShF based SM can be done in a simple way. Suppose a diagonal matrix, whose elements are the particle numbers, associated to the original QOS DF tag set by means of the Minkowski norms:

$$\mathbf{N} = Diag\,(v_I \,|I = 1, N)$$

the inverse of this matrix can be also obtained without effort:

$$\mathbf{N}^{-1} = Diag\left(v_I^{-1} \,|I = 1, N\right),$$

and with such a diagonal matrix one can perform the SM transformation, providing the matrix:

$$\begin{aligned}\mathbf{S} = \mathbf{N}^{-1}\mathbf{Z}\mathbf{N}^{-1} \\ \rightarrow \forall I, J : S_{IJ} = (v_I v_J)^{-1} Z_{IJ} = (v_I v_J)^{-1} \langle \rho_I | \rho_J \rangle = \langle \sigma_I | \sigma_J \rangle,\end{aligned} \tag{6}$$

which is nothing else but the quantum SM obtained employing the ShF. Of course, the transformation is reversible and therefore one can write:

$$\mathbf{Z} = \mathbf{NSN}. \tag{7}$$

Distances computed from both DF and ShF SM separately will not be invariant. Because it can be written, for instance:

$$D^2_{Z;IJ} = Z_{II} + Z_{JJ} - 2Z_{IJ} = \nu_I^2 S_{II} + \nu_J^2 S_{JJ} - 2\nu_I \nu_J S_{IJ}$$
$$\neq D^2_{S;IJ} = S_{II} + S_{JJ} - 2S_{IJ} = \nu_I^{-2} Z_{II} + \nu_J^{-2} Z_{JJ} - 2\nu_I^{-1}\nu_J^{-1} Z_{IJ},$$

which also proves that transformation from DF to ShF or vice versa, cannot be even associated to an homothecy of the DF or ShF originally employed QOS tag sets.

Both MPC and MPH can be considered in general as polyhedrons in some $N-$D vector semispace, see for more details [12]; that is: $N-$D differently shaped geometrical bodies but with their vertex set bearing positive definite coordinates.

Taken as a cloud or hut of points, or a polyhedron, scaled or transformed in some way or another, essentially the SM row or column sets, R or C, need to be understood as referred to a common coordinate origin.

Moreover, in any case, considering the mentioned R or C sets as $N-$dimensional coordinate vectors, one can describe a convenient origin shift affecting the whole $N-$D MPC or MPH, without destroying the metric vicinity structure of their elements.

5 Quantum SM over centroid shifted DF tag sets

When using the DF centroid shifted functions like the ones defined in Eq. (1), one must be aware that the new SM $\mathbf{Z}^{(C)}$, and therefore the associated MPC or MPH, would vary accordingly. One can easily write in this case:

$$\begin{aligned}\forall I, J : Z^{(C)}_{IJ} &= \langle \theta_I | \theta_J \rangle = \langle \rho_I - \rho_C | \rho_J - \rho_C \rangle \\ &= \langle \rho_I | \rho_J \rangle - \langle \rho_I | \rho_C \rangle - \langle \rho_J | \rho_C \rangle + \langle \rho_C | \rho_C \rangle \\ &= Z_{IJ} - N^{-1} \sum_K Z_{IK} - N^{-1} \sum_K Z_{JK} + N^{-2} \sum_K \sum_L Z_{KL}. \end{aligned} \quad (8)$$

Using the following symbols to describe the complete sums[11] of the columns and the whole of the original SM:

$$\langle |\mathbf{z}_I\rangle \rangle = \sum_K Z_{IK} \wedge \langle |\mathbf{z}_J\rangle \rangle = \sum_K Z_{JK} \wedge \langle \mathbf{Z} \rangle = \sum_K \sum_L Z_{KL}$$

it is easy to rewrite Eq. (8) as:

$$\forall I, J : Z^{(C)}_{IJ} = Z_{IJ} - N^{-1} \langle |\mathbf{z}_I\rangle + |\mathbf{z}_J\rangle \rangle + N^{-2} \langle \mathbf{Z} \rangle \quad (9)$$

[11] A *complete sum* of a vector or matrix corresponds to the sum of all its elements. For instance, if the N-dimensional vector $|\mathbf{a}\rangle = \{a_I \,|I = 1, N\}$ is defined, then the complete sum is written as: $\langle |\mathbf{a}\rangle \rangle = \sum_{I=1}^{N} a_I$. The complete sum of a $(M \times N)$ matrix, $\mathbf{A} = \{A_{IJ} \,|I = 1, M; J = 1, N\}$ can be easily defined as follows:
$\langle \mathbf{A} \rangle = \sum_{I=1}^{M} \sum_{J=1}^{N} A_{IJ}$. In infinite dimensional vector semispaces, the complete sum will become a Minkowski norm like the already defined one in Eq. (2).

The practical transformation from the original SM $\mathbf{Z}$ to the centroid shifted one $\mathbf{Z}^{(C)}$, say, can be performed just by defining a vector whose elements are the complete sums of the columns:

$$\left|\mathbf{z}^{(S)}\right\rangle = \left\{z_I^{(S)} = \langle|\mathbf{z}_I|\rangle\right\} \to \langle\mathbf{Z}\rangle = \left\langle\left|\mathbf{z}^{(S)}\right\rangle\right\rangle$$

in this way one can rewrite Eq. (9) as:

$$\forall I, J : Z_{IJ}^{(C)} = Z_{IJ} - N^{-1}\left(z_I^{(S)} + z_J^{(S)}\right) + N^{-2}\langle\mathbf{Z}\rangle \tag{10}$$

6 Centroid origin shift translations of MPC

The translations, called here MPC *centroid origin shifts*, which one can achieve over any $N-$D MPC they are obviously not unique and only some possibilities will be described here, as other kinds of concrete ways of performing this task can be defined and will not be given here, but elsewhere if needed. For instance, the $N-$D MPC column centroid can be defined as the uniform linear convex combination of the columns of the SM $\mathbf{Z}$:

$$|\mathbf{c}\rangle = N^{-1}\sum_{K=1}^{N}|\mathbf{z}_K\rangle$$

and then one can redefine the $N-$D MPC associated new origin shift in the obvious manner:

$$\forall I = 1, N : |\mathbf{a}_I\rangle = |\mathbf{z}_I\rangle - |\mathbf{c}\rangle\,. \tag{11}$$

While the initial origin might be situated far away from a bulk MPC, such centroid translation, as defined above, corresponds to assign a new origin situated nearest and within the MPC vertex component set. It might have interesting potential possibilities as well, in reference to further MPC information gathering.

The elements of the centroid shifted MPC, collected into the matrix $\mathbf{A} = \{|\mathbf{a}_I\rangle\,|I = 1, N\}$ can be written as:

$$\forall I, J : A_{JI} = Z_{JI} - N^{-1}\sum_K Z_{JK} = Z_{JI} - N^{-1}\langle\langle\mathbf{z}_J|\rangle = Z_{JI} - N^{-1}\langle|\mathbf{z}_J\rangle\rangle;$$

however the resulting matrix $\mathbf{A}$ is not symmetrical anymore, as it is the one obtained from the centroid shifted QOS tag set DF. To obtain a symmetrical matrix it is sufficient to use the well-known average, providing the matrix: $\mathbf{C} = \frac{1}{2}\left(\mathbf{A} + \mathbf{A}^T\right)$, whose elements are described as:

$$\begin{aligned}\forall I, J : 2C_{IJ} &= A_{IJ} + A_{JI} = 2Z_{IJ} - N^{-1}\sum_K (Z_{KI} + Z_{KJ})\\ &= 2Z_{IJ} - N^{-1}\left(\langle|\mathbf{z}_I\rangle\rangle + \langle|\mathbf{z}_J\rangle\rangle\right)\end{aligned}$$

which is not completely coincident with the previous centroid shifted QOS tag set result (10), although both have some common part.

Alternatively, and in a better point of view, one can start from the column shifted matrix $\mathbf{A}$ and perform over the rows of it, defined as: $\mathbf{A} = \{\langle \mathbf{a}_I | \, | I = 1, N\}$ another origin shift, providing a new row centroid origin shifted matrix:

$$\mathbf{B} = \{\langle \mathbf{b}_I | \, | I = 1, N\} \leftarrow \forall I : \langle \mathbf{b}_I | = \langle \mathbf{a}_I | - \langle \mathbf{d}| \leftarrow \langle \mathbf{d}| = N^{-1} \sum_I \langle \mathbf{a}_I |;$$

then, the elements of the new origin shifted matrix, using the symmetry of the original SM $\mathbf{Z}$, can be written as:

$$\begin{aligned} \forall I, J : B_{IJ} &= A_{IJ} - N^{-1} \sum_I A_{IK} \\ &= Z_{IJ} - N^{-1} \sum_K Z_{KJ} - N^{-1} \sum_K \left(Z_{IK} - N^{-1} \sum_L Z_{KL} \right) \\ &= Z_{IJ} - N^{-1} \sum_K (Z_{KI} + Z_{KJ}) + N^{-2} \sum_K \sum_L Z_{KL} \end{aligned}$$

providing an interesting result, which demonstrates that: $\mathbf{B} = \mathbf{Z}^{(C)}$. That is: when centroid shifting the original SM twice, first by its columns and by the resultant rows afterwards or vice versa, one ends into the SM obtained from the centroid shifted QOS DF tag set.

Perhaps, another one of the most interesting origin shift manipulations can be done, when one can also use the above result to reconstruct a Gram matrix with the row and column elements of the centroid shifted matrix $\mathbf{Z}^{(C)}$, that is:

$$\forall I, J : G_{IJ} = \left\langle \mathbf{z}_I^{(C)} | \mathbf{z}_I^{(C)} \right\rangle \rightarrow \mathbf{G} = \left(\mathbf{Z}^{(C)} \right)^T \mathbf{Z}^{(C)} = \left(\mathbf{Z}^{(C)} \right)^2$$

which will have the structure of the Gram matrix associated to the dual sets of the columns $\mathrm{C}^{(C)} = \left\{ \left| \mathbf{z}_I^{(C)} \right\rangle \right\}$ or rows $\mathrm{R}^{(C)} = \left\{ \left\langle \mathbf{z}_I^{(C)} \right| \right\}$ of $\mathbf{Z}^{(C)}$.

The definition of the centroid shifted SM can be subject to the already discussed transformations which have been described for the original SM [21–23].

Thus the centroid origin shift performed upon the original DF tag set or the symmetrical origin shift performed upon rows and columns of the original SM lead to the same translated MPC. However, as the centroid origin shifted DF set is no longer linearly independent, the corresponding SM $\mathrm{C}^{(C)}$ or $\mathrm{R}^{(C)}$ sets which are linearly dependent too, spanning some $(N-1)$ −D vector space. Therefore, the symmetrical origin shifted SM is no longer a metric matrix, but has to be considered solely a Gram matrix of the origin shifted tag set: $\mathrm{P}^{(C)} = \{\theta_I \, | I = 1, N\}$.

7 Invariance of similarity indices upon origin shift

In any case though, one can realize that differences between pairs of the elements, chosen from the original SM components, will become invariant when performed on the translated ones. This is so, because:

$$\forall I, J : \theta_I - \theta_J = \rho_I - \rho_C - (\rho_J - \rho_C) = \rho_I - \rho_J .$$

Thus, Euclidian or higher order distances, see for example references [25–27] between elements of the $N-$D MPC will become invariant too. Such a property has to be expected, constituting a well-known general property of vector space translations, in turn related to Euclidian norm invariance. Such a property it is also related to the fact that, when any $N-$D MPC, considered as a polyhedron, is translated as a whole, the distances between its vertices will be invariant, as the polyhedron shape will not be modified.

On the other hand, one can say with respect to CI [28] that they are invariant upon computing them, using any of the matrices (6) or (7); which is the same: irrespective of using DF or ShF [29] the same result will be obtained. However, CI will not remain invariant upon centroid origin shift as Euclidian distances do. This is due that this kind of similarity indices are attached to the cosines of the angles between vectors pairs (or several ones), and the measured angles vary from an origin to another. For instance, to prove this variance of CI upon origin shift one can write over the original SM column or row sets the second (or if needed higher) order CI matrix [22]:

$$\forall I, J : r_{IJ} = \langle \rho_I | \rho_J \rangle \left(\langle \rho_I | \rho_I \rangle \langle \rho_J | \rho_J \rangle \right)^{-\frac{1}{2}} = Z_{IJ} \left(Z_{II} Z_{JJ} \right)^{-\frac{1}{2}} ; \tag{12}$$

then, choosing to compute the CI over the set: $\mathrm{P}^{(C)} = \{\theta_I \,|I = 1, N\}$ will result into the form corresponding to the same algorithm:

$$\begin{aligned} \forall I, J : r_{IJ}^{(C)} &= \langle \theta_I | \theta_J \rangle \left(\langle \theta_I | \theta_I \rangle \langle \theta_J | \theta_J \rangle \right)^{-\frac{1}{2}} \\ &= Z_{IJ}^{(C)} \left(Z_{II}^{(C)} Z_{JJ}^{(C)} \right)^{-\frac{1}{2}} \neq r_{IJ}, \end{aligned} \tag{13}$$

but observing the resultant CI elements of Eq. (13) above, one can see that the CI matrix elements structure, in principle will essentially differ from the original CI form associated to Eq. (12).

This result becomes obvious, already in a QOS constructed by a pair of QO DF's only. Whichever their nature or magnitude considered as vectors, they will provide the same result on centroid shifting: two collinear origin shifted densities. This is so because, whenever one takes two QO into account only: $\{\rho_A; \rho_B\}$ say, then the centroid is defined by: $\rho_C = 2^{-1} (\rho_A + \rho_B)$ and thus the shifted densities become:

$$\begin{aligned} &\theta_A = \rho_A - \rho_C = 2^{-1} (\rho_A - \rho_B) \wedge \theta_B = \rho_B - \rho_C = 2^{-1} (\rho_B - \rho_A) \\ &\text{therefore} : \theta_A = -\theta_B \rightarrow r_{AB}^{(C)} = -1. \end{aligned}$$

As a general consequence, centroid shifted SM associated to cardinality 2 QOS cannot produce any significative discrimination or significative information between whatever defined isolated couple of QO. As a corollary, one can say that: *in order to obtain a starting informative description upon centroid shifting it is necessary to manipulate at least cardinality 3 QOS.* Alternatively, one can also say that comparison of an isolated pair of vertices of a MPC or MPH cannot provide significative information about its similarity.

Resuming what it has been discussed: when transforming DF into ShF, the invariance of CI has been reported some time ago [29], while Euclidian distances are not invariant by such kind of transformations. Such a property is exactly the reverse of the encountered property connected with the origin shift, that is, in the origin shift case: distances are invariant but not CI's. Both invariance-variance properties must be taken into consideration, when either similarity or dissimilarity indices are studied.

8 Transformations of the centroid origin shifted SM leading to similarity indices and ordering procedures

It has been already commented the way CI can be obtained from the shifted SM elements. It is a matter to simply apply the recently described ways to construct distances or generalized CI [22].

It might be interesting to note that Euclidian and higher order generalized distances [25–27] are confined to the possibility of comparing two quantum objects only, while CI can be easily generalized to compare an indefinite number of objects [22]. Also, among the new points of view associated to Euclidian distances one can describe the new dissimilarity index which can be called a scaled distance [23], which has been recently proved to be related to the so-called Hodgkin-Richards similarity index [30]. This kind of issues will be discussed in the next sections.

9 Higher order generalized CI based on centroid shifted similarity matrices

The recently proposed computational possibilities of generalized CI [22] open another way to enlarge the usefulness of the shifted SM columns and rows collections. For example, from the set $\mathrm{C}^{(C)} = \left\{\left|\mathbf{z}_I^{(C)}\right\rangle\right\}$ and using both inward vector product and complete vector sum operations [12], one can construct generalized scalar products precisely writing:

$$\forall I_K\ (K = 1, v) : \left(\left|\mathbf{z}_{I_1}^{(C)}\right\rangle, \left|\mathbf{z}_{I_2}^{(C)}\right\rangle, \ldots \left|\mathbf{z}_{I_v}^{(C)}\right\rangle\right)$$
$$= \left\langle\left|\mathbf{z}_{I_1}^{(C)}\right\rangle * \left|\mathbf{z}_{I_2}^{(C)}\right\rangle * \ldots \left|\mathbf{z}_{I_v}^{(C)}\right\rangle\right\rangle = \sum_J \left(Z_{JI_1}^{(C)} Z_{JI_2}^{(C)} \ldots Z_{JI_v}^{(C)}\right)$$

which can be considered symmetrical v rank tensors and that can be used to define v-th order CI. The associated v-th order norms to be properly computed with the norm characteristics, need to be defined with absolute values when the order is odd [30]; then in general one can define v-th order norms like:

$$\forall I : \left(\left|\mathbf{z}_I^{(C)}\right\rangle, \left|\mathbf{z}_I^{(C)}\right\rangle, \ldots \left|\mathbf{z}_I^{(C)}\right\rangle\right) = \sum_J \left|Z_{JI}^{(C)}\right|^v$$

and a v-th order CI can be computed as:

$$r_{I_1 I_2 \ldots I_v}^{(C)} = \left[\prod_{K=1}^{v} \left(\left|\mathbf{z}_{I_K}^{(C)}\right\rangle, \left|\mathbf{z}_{I_K}^{(C)}\right\rangle, \ldots \left|\mathbf{z}_{I_K}^{(C)}\right\rangle\right)\right]^{-\frac{1}{v}} \left(\left|\mathbf{z}_{I_1}^{(C)}\right\rangle, \left|\mathbf{z}_{I_2}^{(C)}\right\rangle, \ldots \left|\mathbf{z}_{I_v}^{(C)}\right\rangle\right).$$

10 Hilbert space centroid shifted similarity tensors

Definition of Hilbert space higher order products of centroid shifted functions can be useful to understand the previous description in this section. Similarity matrices involving two quantum objects have been discussed and permitted to define the centroid shifted similarity matrix $\mathbf{Z}^{(C)}$. But triple or higher order similarity matrices have also been described [32], even employed in some applications [33]. It will be illustrative to know how triple scalar products of centroid shifted functions behave as a first example. Then, it must be defined the triple product of the elements of the set $\mathrm{P}^{(C)} = \{\theta_I \,|I = 1, N\}$, taking into account the index permutation equivalences:

$$\begin{aligned} Z_{IJK}^{(C)} &= \left\langle \theta_I^{(C)} \theta_J^{(C)} \theta_K^{(C)} \right\rangle = \langle (\rho_I - \rho_C)(\rho_J - \rho_C)(\rho_K - \rho_C) \rangle \\ &= \langle \rho_I \rho_J \rho_K \rangle - \langle \rho_I \rho_C \rho_K \rangle - \langle \rho_C \rho_J \rho_K \rangle + \langle \rho_C \rho_C \rho_K \rangle \\ &\quad - \langle \rho_I \rho_J \rho_C \rangle + \langle \rho_I \rho_C \rho_C \rangle + \langle \rho_C \rho_J \rho_C \rangle - \langle \rho_C \rho_C \rho_C \rangle \\ &= Z_{IJK} + Z_{KCC} + Z_{ICC} + Z_{JCC} - (Z_{IKC} + Z_{JKC} + Z_{IJC} + Z_{CCC}) \end{aligned}$$

Three kinds of integral structures are relevant in this triple origin shifted DF context, including the centroid DF one, two or three times accordingly. One can write for these generic integrals:

$$\begin{aligned} Z_{IJC} &= N^{-1} \sum_P Z_{IJP} \wedge Z_{ICC} = N^{-2} \sum_P \sum_Q Z_{IPQ} \wedge Z_{CCC} \\ &= N^{-3} \sum_P \sum_Q \sum_R Z_{PQR} \end{aligned}$$

The triple product integral Z_{CCC} involving the centroid DF corresponds to the averaged complete sum of the triple DF tensor: $\mathbf{Z}^{(3)} = \{Z_{IJK}\}$; that is, one can write:

$$Z_{CCC} = N^{-3} \left\langle \mathbf{Z}^{(3)} \right\rangle.$$

Also, the integrals of the generic type Z_{ICC}, can be interpreted as corresponding to the averaged complete sums of the rank two tensor components of $\mathbf{Z}^{(3)}$, which can also be interpreted as the rank 2 tensorial representation of every quantum object, and can be written as the $(N \times N)$ matrices, which can be easily defined as:

$$\forall I = 1, N : \mathbf{Z}_I^{(2)} = \left\{ Z_{IPQ} \,|\, P, Q = 1, N \right\}$$
$$\text{then} : \forall I : Z_{ICC} = N^{-2} \left\langle \mathbf{Z}_I^{(2)} \right\rangle.$$

Finally the same definition can be used to describe the integrals Z_{IJC} as:

$$\mathbf{Z}_C^{(2)} = \{ Z_{IJC} \,|\, I, J = 1, N \} = N^{-1} \sum_P \mathbf{Z}_P^{(2)}.$$

Then, although appearing as a cumbersome task, one can systematically obtain expressions for higher order similarity tensors based on an arbitrary number of centroid shifted DF.

Of course, similarity tensors can be taken as a way to describe quantum objects discretely in higher dimensions and as a potential source of CI involving more than two quantum objects.

11 Euclidian Distances and scaled Euclidian distances

Euclidian distance matrices over the centroid shifted QO can include all the original QO and will be coincident, because of the previous discussion, with the distance matrix computed from the original SM.

Another discussion possibility may be based on the newly defined scaled squared Euclidian distance indices (SDI) [31]. The proposed form of SDI is like:

$$\forall I, J : T_{IJ}^2 = \frac{D_{IJ}^2}{(Z_{II} + Z_{JJ})} = 1 - \frac{2Z_{IJ}}{(Z_{II} + Z_{JJ})} = 1 - H_{IJ}, \tag{14}$$

where H_{IJ} is coincident with the so-called Hodgkin-Richards similarity index (HRI) [30] and D_{IJ}^2 is just the squared Euclidian distance.

The HRI was originally defined as a possible variant of the original version of the CI [31], but lacking of any deep mathematical analysis about its possible attached significance, excepting a previous study which appeared several years ago [34], where the connection of both CI and HRI was attempted.

Equation (14) shows beyond any doubt that HRI is nothing than the result of manipulating in a straightforward way the squared Euclidian distance. The form: $T_{IJ}^2 = 1 - H_{IJ}$ or even better: its square root, is therefore preferable to the sole use of HRI. Thus the SDI's are proved to be dissimilarity indices.

It is interesting to see how SDI and thus HRI will be affected by a double origin shift. One can write in this case:

$$\left(T_{IJ}^{(C)}\right)^2 = 1 - \frac{2Z_{IJ}^{(C)}}{\left(Z_{II}^{(C)} + Z_{JJ}^{(C)}\right)} = \left(1 - H_{JI}^{(C)}\right),$$

where one can observe that the resulting centroid shifted SDI has obviously the same structure as the original one. As in the case of CI, and contrarily to Euclidian distances,

the SDI it is not invariant upon origin shifting. The HRI can be obtained from the original similarity matrix elements as:

$$H_{JI}^{(C)} = \frac{Z_{IJ} + \zeta_{IJ}}{\frac{1}{2}\left(Z_{II} + Z_{JJ}\right) + \zeta_{IJ}}$$

where it has been used the definition:

$$\zeta_{IJ} = -N^{-1}\left(\langle|\mathbf{z}_I\rangle\rangle + \langle|\mathbf{z}_J\rangle\rangle\right) + N^{-2}\langle\mathbf{Z}\rangle\,.$$

One can see from this result that the centroid shifted HRI, corresponds to the quotient elements of the original index shifted by a quantity, which is equal for both numerator and denominator related to the centroid shifting of the corresponding elements.

12 Generalized Euclidian distances

As it has been commented, Euclidian distances are well-defined mathematical objects devised for comparison of two vectors or QO. However, there is an easy way to compare more than a couple of points within QOS, MPC or MPH. One of the possible arguments can be the same as in the generalization of the CI computational environment, which has been developed some sections above and commented in several previous papers, see for example the last one [22]. One can envisage distances not associated to a pair of quantum objects or point vectors only, but to tensorial products of any rank of them.

For instance, taking the centroid shifted QOS: $\mathrm{P}^{(C)} = \{\theta_I\,|I = 1, N\}$, one can construct the rank 2 tensor product:

$$\mathrm{T}_2^{(C)} = \mathrm{P}^{(C)} \otimes \mathrm{P}^{(C)} = \{\theta_I\theta_J\,|I;\, J = 1, N\}$$

and then use the Euclidian distance definition over the elements of $\mathrm{T}_2^{(C)}$, that is:

$$D_{IJKL}^2 = \left\langle|\theta_I\theta_J - \theta_K\theta_L|^2\right\rangle = \langle\theta_I\theta_J\theta_I\theta_J\rangle + \langle\theta_K\theta_L\theta_K\theta_L\rangle - 2\langle\theta_I\theta_J\theta_K\theta_L\rangle$$

where the involved quantum similarity integrals can be expressed in general as quadruple integrals:

$$\langle\theta_I\theta_J\theta_K\theta_L\rangle = \int_D \theta_I(\mathbf{r})\,\theta_J(\mathbf{r})\,\theta_K(\mathbf{r})\,\theta_L(\mathbf{r})\,d\mathbf{r}$$

which can be developed in turn with the whole integral structure or by means of particular easier expressions. For instance, using the generalization of an idea put forward by Cioslowski and Fleischmann [35] in earlier stages of the development of quantum similarity, it can be written:

$$\langle \theta_I \theta_J \theta_K \theta_L \rangle_{CF} = \int_D \int_D \theta_I (\mathbf{r}_1) \theta_J (\mathbf{r}_2) \theta_K (\mathbf{r}_1) \theta_L (\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2$$
$$\equiv \langle \theta_I \theta_K \rangle \langle \theta_J \theta_L \rangle = Z_{IK}^{(C)} Z_{JL}^{(C)}$$

producing quadruple centroid shifted QO scalar products in terms of the shifted SM elements, without need to compute any extra integral. However, developing here and at this moment such kind of theory about generalized Euclidian distances in QOS or MPC, will demand a large amount of additional information and discussion. The authors feel it is preferable to present such possible generalization of QOS Euclidian distances elsewhere in another more devoted context.

13 Conclusions

Object origin shift can be coherently described at various theoretical levels, from quantum mechanical Hilbert semispaces up to classical descriptor spaces. The resultant centroid shifted SM in the space of molecules can be used to make available a canonical origin shift, intended for further research and aimed to the computation of similarity integrals and QQSPR or QQlSPR applications, involving any number of QO DF tags. Euclidian distances between QO's will be invariant upon this centroid origin shifts. These results indicate as a first conclusion that Euclidian distances, which have been scarcely employed in quantum similarity literature, can be a good tool to compare elements in arbitrary QOS leading directly to simple QQlSPR procedures. The second conclusion corresponds to the fact that, when using similarity indices like CI or HRI, centroid origin shift is a recommended first transformation to be performed in initial similarity matrices or tensors, in order to have a well-established canonical reference in this alternative framework. In this manner quantum similarity procedures are provided with a universal reference structure.

References

1. R. Carbó-Dorca, A. Gallegos, Á.J. Sánchez, J. Comput. Chem. **30**, 1146 (2008)
2. Y.C. Martin, *Quantitative Drug Design: A Critical Introduction* (CRC Press, Boca Raton, 2010)
3. J.P. Doucet, A. Panaye, *Three Dimensional QSAR: Applications in Pharmacology and Toxicology* (CRC Press, Boca Raton, 2010)
4. R. Carbó-Dorca, SAR QSAR Environ. Res. **18**, 265 (2007)
5. R. Carbó-Dorca, S. Van Damme, Theor. Chem. Acc. **118**, 673 (2007)
6. R. Carbó-Dorca, S. Van Damme, Afinidad. **64**, 147 (2007)
7. R. Carbó-Dorca, S. Van Damme, Intl. J. Quant. Chem. **108**, 1721 (2007)
8. A.R. Katritzky, I.B. Stoyanova-Slavova, D.A. Dobchev, M. Karelson, J. Mol. Graph. Mod. **26**, 529 (2007)
9. J. Xu, L. Wang, L. Wang, X. Shen, W. Xu, J. Comp. Chem. doi:101002/jcc.21907, 3241 (2011)
10. R. Carbó-Dorca, A. Gallegos, *Quantum Similarity and Quantum QSPR (QQSPR)* (Invited contribution) Entry: 176, In Meyers, Robert (Ed.) Encyclopedia of Complexity and Systems Science, vol. **8** (Springer, New York, 2009), pp. 7422–7480.
11. R. Carbó-Dorca, E. Besalú, L.D. Mercado, J. Comp. Chem. **32**, 582 (2011)
12. R. Carbó-Dorca, E. Besalú; J. Math. Chem. doi:10.1007/s10910-011-9906-4 (2011)
13. R. Carbó-Dorca, J. Math. Chem. **23**, 353 (1998)

14. R. Carbó-Dorca, Adv. Molec. Simil. **2**, 43–72 (1998)
15. L.D. Mercado, R. Carbó-Dorca, J. Math. Chem. **49**, 1558 (2011)
16. R. Carbó-Dorca, Int. J.Quantum Chem. **79**, 163 (2000)
17. R. Carbó-Dorca, J. Math. Chem. **27**, 357 (2000)
18. R. Carbó-Dorca, J. Math. Chem. doi:10.1007/s10910-011-9921-5 (2011)
19. R. Carbó, E. Besalú, Afinidad **53**, 77 (1996)
20. R. Carbó, E. Besalú, Ll. Amat, X. Fradera, J. Math. Chem. **18**, 237 (1995)
21. R. Carbó-Dorca, J. Math. Chem. **32**, 201 (2002)
22. R. Carbó-Dorca, J. Math. Chem. **49**, 2109 (2011)
23. R. Carbó-Dorca, J. Math. Chem. doi:10.1007/s10910-011-9920-6 (2011)
24. J.C. Dearden, M.T.D. Cronin, K.L.E. Kaiser, SAR & QSAR Environ. Res. **20**, 241 (2009)
25. P.H.A. Sneath, R.R. Sokal, *Numerical Taxonomy* (W. H. Freeman & Co., San Francisco, 1973)
26. N. Jardine, R. Sibson, *Mathematical Taxonomy* (Wiley, London, 1971)
27. M.M. Deza, E. Deza, *Encyclopedia of Distances* (Springer, Berlin, 2009)
28. R. Carbó, L. Leyda, M. Arnau, Int. J. Quant. Chem. **17**, 1185 (1980)
29. P. Bultinck, R. Carbó-Dorca, J. Math. Chem. **36**, 191 (2004)
30. E.E. Hodgkin, W.G. Richards, Int. J. Quant. Chem. **S14**, 105 (1987)
31. R. Carbó-Dorca, J. Math. Chem. **47**, 331 (2010)
32. R. Carbó, B. Calabuig, E. Besalú, A. Martínez, Mol. Eng. **2**, 43 (1992)
33. D. Robert, R. Carbó-Dorca, J. Chem. Inf. Comp. Sci. **38**, 620 (1998)
34. R. Carbó, E. Besalú, L. Amat, X. Fradera, J. Math. Chem. **19**, 47 (1996)
35. J. Cioslowski, E.D. Fleischmann, J. Am. Chem. Soc. **113**, 64 (1991)