ORIGINAL PAPER

Multimolecular polyhedra and QSPR

Ramon Carbó-Dorca

Abstract A multimolecular polyhedron (MP) is defined when studying molecular sets. When any molecular set is described using some chosen N-dimensional vector tags attached to every molecular object in the set, thus forming a tagged set, then the tag set also forms a MP. In the particular case when the molecular tags are density functions, then he quantum object set formed in this manner also defines some MP in ∞-dimensional space. In principle, MP are defined with respect an arbitrary origin, but once constructed one can shift it in two natural ways: (a) by each vertex at a time and (b) once for all by means of the polyhedron centroid. The quantum QSAR frames, which can be constructed following these two origin shift procedures, provide two not so different results when used for property evaluation. The differences between both shifting techniques are here revealed and discussed. A single density function for a given molecule is also studied from this point of view as a sequel of the previous study of MP.

1 Introduction

When defining a set of N molecular structures, therefore by forming any molecular object set (OS) within any descriptive background: quantum (QOS) or classical (COS), the result becomes the same as to construct some N-dimensional polyhedron, with

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

vertices well defined by the vector tags associated to the molecular objects, see for example reference [1] for an up to date discussion. One can call such a construct a *multimolecular polyhedron* (MP). In fact, this polyhedral picture corresponds to the previously described molecular point clouds, see for example reference [1], adopting here a more geometrical allure.

When considering any QOS, then the MP is constructed by means of ∞-dimensional vertices, which can be associated to the density function (DF) tags attached to each quantum object (QO): $P = \{\rho_I \,|\, I = 1, N\}$, say. It can be understood, within such a specific construction, that one is facing a quantum MP (QMP). The same can be considered when a COS is defined, and then the DF are substituted by a set of descriptor vectors, see for example [2] for a discussion on the discrete description of molecular sets, defined in turn within a finite N-dimensional space, one can say that a classical MP (CMP) is constructed.

With the corresponding tags and supposing every molecular structure into the MP different from the rest, one can built up a $(N \times N)$ similarity matrix, $\mathbf{Z}$, which might be simply considered in this case as a metric matrix, associated to the subspace subtended by the MP vertices. In the QOS case, one can write the following algorithm for the similarity matrix build up:

$$\forall I, J = 1, N : \mathbf{Z} = \{z_{IJ}\} \leftarrow z_{IJ} = \langle \rho_I \rho_J \rangle = \int_D \rho_I(\mathbf{r})\rho_J(\mathbf{r})d\mathbf{r}. \tag{1}$$

An equivalent algorithm can be described for the COS case, then the scalar products between the tag vectors can be considered to substitute the integrals in Eq. (1).

2 Vertex shift in a QQSPR framework

In general, the QMP vertices forming the tag set P are defined in principle with respect an arbitrary origin. However, they can be shifted in turn about any of the vertices in the set. Choosing the K-th vertex, for instance, the origin shift produces a new set, where the chosen vertex is now at the origin, then one can write:

$$\exists K : P \Rightarrow \Xi^K = \left\{ \xi_I^K = \rho_I - \rho_K \,|\, I = 1, N \right\} \rightarrow \xi_K^K = 0\,(\mathbf{r}).$$

When dealing within classical QSPR procedures instead of the QQSPR, the DF tag set is substituted by the set of descriptor vectors. The involved procedures and algorithms remain essentially the same, see for example reference [2].

In case that a property set: $\Pi = \{\pi_I \,|\, I = 1, N\}$ is attached to each object, such extra information, when considering a coherent origin shift, connected with the vertex shift, could be translated into a new property set using the K-th object property; like:

$$\exists K : \Pi \Rightarrow X = \left\{ \chi_I^K = \pi_I - \pi_K \,|\, I = 1, N \right\} \rightarrow \chi_K^K = 0.$$

Now, from the quantum point of view one can suppose that there exists a Hermitian quantum QSPR (QQSPR) operator Ω such that,[1] when applied to every QO DF, yields the corresponding property as an expectation value:

$$\forall I : \langle \Omega \rho_I \rangle = \pi_I.$$

In the origin shifted QMP situation one can also write:

$$\exists K : \left\langle \Omega \xi_I^K \right\rangle = \chi_I^K. \tag{2}$$

When adopting QQSPR framework postulates, the operator Ω might be basically expressed as a first approximation, by constructing a linear combination, see for example reference [3] for more details, of the elements of the shifted MP:

$$\Omega \approx \sum_J \omega_J \xi_J^K; \tag{3}$$

where the set $O = \{\omega_J | J = 1, N\}$ is a coefficient set to be determined [3]. Therefore, the approximate Eq. (3) can provide an estimate of the shifted property by means of the expectation value expression (2):

$$
\begin{aligned}
\exists K : \chi_I^K &\approx \sum_J \omega_J \langle \xi_J^K \xi_I^K \rangle = \sum_J \omega_J \langle (\rho_J - \rho_K)(\rho_I - \rho_K) \rangle \\
&= \sum_J \omega_J (\langle \rho_J \rho_I \rangle + \langle \rho_K \rho_K \rangle - (\langle \rho_J \rho_K \rangle + \langle \rho_K \rho_I \rangle)) \\
&= \sum_J \omega_J (z_{IJ} + z_{KK} - (z_{JK} + z_{IK})) = \sum_J \omega_J a_{IJ}^K
\end{aligned}
\tag{4}
$$

This result permits to set up and solve a matrix system in order to evaluate the QQSPR operator coefficients, which can be easily obtained as:

$$^K \mathbf{A} |\mathbf{w}\rangle = |^K \mathbf{x}\rangle \rightarrow |\mathbf{w}\rangle = {}^K \mathbf{A}^{-1} |^K \mathbf{x}\rangle,$$

which in turn precludes that the vector $|\mathbf{w}\rangle \equiv |^K \mathbf{w}\rangle$ will depend on the specified vertex shift employed.

The involved matrices will have the K-th row and column null, so the information about the vertex, which serves to perform the shift, amounts the same as being overridden. The summations in the definition of the QQSPR operator might be restricted to: $J \neq K$, and thus the problem becomes attached to a dimension $N - 1$. This circumstance has been already described in a previous set of published papers [4–6].

---

[1] Within classical QSPR framework and at a first order approximation, a vector has to be supposed to exist, instead of the QQSPR operator, see for instance [3].

3 Average vertex shifting

In any case one might average the previous Eq. (4) providing the estimated expectation values. Using the parameter: $v = \sum_J \omega_J$, one can write, for instance, the complete sum of values provided in every vertex shift like:

$$
\forall I : \langle \chi_I \rangle = \sum_K \chi_I^K \approx \sum_K \sum_J \omega_J (z_{IJ} + z_{KK} - (z_{JK} + z_{IK}))
$$

$$
= \sum_K \sum_J \omega_J z_{IJ} + \sum_K \sum_J \omega_J z_{KK} - \sum_K \sum_J \omega_J (z_{JK} + z_{IK})
$$

$$
= N \sum_J \omega_J z_{IJ} + v \sum_K z_{KK} - \sum_K \sum_J \omega_J z_{JK} - v \sum_K z_{IK}
$$

So, one can finally obtain for the average expectation value of a property, attached to some specific molecule, the following expression:

$$
\forall I : N^{-1} \langle \chi_I \rangle = \sum_J \omega_J z_{IJ} + N^{-1} \sum_K \left(v(z_{KK} - z_{IK}) - \sum_J \omega_J z_{JK} \right)
$$

$$
= \sum_J \omega_J z_{IJ} + v N^{-1} Tr|\mathbf{Z}| - N^{-1} \left(v \langle\langle \mathbf{z}_I | \rangle + \sum_J \omega_J \langle\langle \mathbf{z}_J | \rangle \right) \quad (5)
$$

where one can define the rows of the similarity matrix $\mathbf{Z}$ as forming the set: $\{\langle \mathbf{z}_I \| I = 1, N\}$. In this way one can see how the shifted vertices property average will influence the final result. The first term is the unmodified non shifted expectation value, the second term can be considered a constant, depending on the QMP and the attached property set, the final term corresponds to the origin shift correction.

4 Average vertex shift QQSPR operator evaluation

One can easily be aware, when using the alternative vertex shifts of the QMP, that the problem could be the dependence of the QQSPR operator coefficients on the specific shift chosen, as previously commented. However, this problem might be also obviated, in the same way as in the expectation value evaluation, by means of the origin shifted similarity matrix construction:

$$
\forall K, I, J : a_{IJ}^K = z_{IJ} + z_{KK} - (z_{JK} + z_{IK}) \rightarrow {}^K\mathbf{A} = \left\{ a_{IJ}^K \right\},
$$

then an average matrix might be also constructed: $\mathbf{A} = N^{-1} \sum_K {}^K\mathbf{A}$, needing an extra property vector average: $|\mathbf{x}\rangle = N^{-1} \sum_K |{}^K\mathbf{x}\rangle$. Finally, this permits to obtain

the QQSPR operator coefficients by means of a unique linear system solved by the matrix equation:

$$|\mathbf{w}\rangle = \mathbf{A}^{-1}|\mathbf{x}\rangle.$$

5 Centroid shift

The averaged QMP vertex shifting, as discussed up to now, could be seen as similar to define a DF centroid:

$$\rho_C = N^{-1}\sum_K \rho_K$$

and then computing a general unique QMP origin shift:

$$P \Rightarrow \Xi^C = \left\{ \xi_I^C = \rho_I - \rho_C | I = 1, N \right\}.$$

This is easy to see, as when averaging every QMP vertex origin shifted by means of the QMP vertices, one obtains:

$$\forall I : \overline{\xi_I} = N^{-1}\sum_K \xi_I^K = N^{-1}\sum_K (\rho_I - \rho_K)$$

$$= \rho_I - N^{-1}\sum_K \rho_K = \rho_I - \rho_C = \xi_I^C$$

Despite of this property related with the DF vertices origin shift, one can try to compute the average expectation value over the centroid shifted DF QMP vertices. It can be written now directly as:

$$\chi_I^C \approx \sum_J \omega_J \langle \xi_J^C \xi_I^C \rangle = \sum_J \omega_J \langle (\rho_J - \rho_C)(\rho_I - \rho_C) \rangle$$

$$= \sum_J \omega_J \left(\langle \rho_J \rho_I \rangle + \langle \rho_C \rho_C \rangle - (\langle \rho_J \rho_C \rangle + \langle \rho_C \rho_I \rangle) \right)$$

$$= \sum_J \omega_J \left(z_{IJ} + N^{-2}\sum_K\sum_L z_{KL} - N^{-1}\sum_K (z_{JK} + z_{IK}) \right)$$

$$= \sum_J \omega_J \left(z_{IJ} - N^{-1}\sum_K z_{JK} \right) + \upsilon N^{-1}\left(N^{-1}\sum_K \left(\left(\sum_L z_{KL}\right) - z_{IK} \right) \right)$$

$$= \sum_J \omega_J z_{IJ} + \upsilon N^{-2}\langle \mathbf{Z} \rangle - N^{-1}\left(\upsilon \langle \langle \mathbf{z}_I | \rangle + \sum_J \omega_J \langle \langle \mathbf{z}_J | \rangle \right)$$

where it has been used the complete sum of the similarity matrix, symbolized by:
$$\langle \mathbf{Z} \rangle = \sum_K\sum_L z_{KL}.$$

Then one can see that the expectation values of the centroid shifted QMP DF vertices, yield a different result than in the previous expectation value average (5), obtained within the QMP vertex shift framework. However both procedures just provide a outcome with a unique subtle difference.

Such a difference proves that while the centroid DF can be obtained via a simple average of the vertex shifted DF set, it is not the same to obtain an average of the associated properties of vertex shifted DF; or the property directly issued from the centroid shifted DF. The difference indicates that both procedures correspond to yield the same structure, but provided with two different estimated property origin shifts, which will depend on the QMP vertices similarity matrix form.

6 Centroid shifts in intramolecular polyhedra

The centroid shift technique applied in the previous sections on MP can be employed in kind of intramolecular set of cases, when shifting the molecular DF with respect of intramolecular shape functions (SF) adequately defined.

a. DF and SF in a MO framework

Suppose a molecular structure and known an attached set of MO: $M = \{\varphi_I | I = 1, N\}$. With them one can easily construct a set of MO shape functions (SF): $P = \{\sigma_I = |\varphi_I|^2 | I = 1, N\}$. Because of the Euclidian normalization associated to the MO set:

$$\forall I : \langle \varphi_I | \varphi_I \rangle = \langle |\varphi_I|^2 \rangle = \int_D |\varphi_I|^2 dV = 1,$$

then the MO SF set is normalized in the sense of Minkowski:

$$\forall I : \langle \sigma_I \rangle = 1.$$

With the MO SF set one can also construct a density function (DF) as a linear combination:

$$\rho = \sum_I \omega_I \sigma_I,$$

where use is made of the set of occupation numbers: $\{\omega_I | I = 1, N\} \subset \mathbb{R}^+$, which has the property of summing up to the number of electrons ν, that is: $\nu = \sum_I \omega_I$.

Therefore, the Minkowski norm of the DF corresponds to the number of electrons:

$$\langle \rho \rangle = \sum_I \omega_I \langle \sigma_I \rangle = \sum_I \omega_I = \nu.$$

As it is well-known one can also describe the associated SF, just Minkowski normalizing the DF:

$$\sigma = v^{-1}\rho \rightarrow \langle \sigma \rangle = 1.$$

The set M can be considered as an intramolecular polyhedron.

b. Centroid SF and origin shifted SF

The MO SF set in any case forms a SF polyhedron. Thus a centroid function can be easily defined:

$$\sigma_C = N^{-1}\sum_I \sigma_I$$

which is naturally Minkowski normalized:

$$\langle \sigma_C \rangle = N^{-1}\sum_I \langle \sigma_I \rangle = N^{-1}N = 1$$

and thus can be considered bearing a SF structure.

A shifted SF can be also defined, yielding a function, which can be written as:

$$\zeta_C = \sigma - \sigma_C = v^{-1}\sum_I \omega_I \sigma_I - N^{-1}\sum_I \sigma_I = \sum_I \theta_I \sigma_I$$

Where the coefficients of the shifted SF can be described in terms of the original DF occupation numbers as:

$$\forall I : \theta_I = v^{-1}\omega_I - N^{-1} = (vN)^{-1}(N\omega_I - v).$$

Also one can deduce without effort that there is a trivial connection between occupation numbers and shifted SF coefficients:

$$\forall I : \omega_I = v(\theta_I + N^{-1}).$$

The corresponding coefficients to the associated centroid shifted DF can be obtained just scaling them with the number of electrons:

$$\forall I : \eta_I = v\theta_I = \omega_I - vN^{-1} \leftrightarrow \forall I : \omega_I = \eta_I + vN^{-1}.$$

It can be also easily deduced that the Minkowski norm of the shifted SF is null:

$$\langle \zeta_C \rangle = \langle \sigma \rangle - \langle \sigma_C \rangle = 0,$$

in fact, the sum of the coefficients associated to the shifted SF add to zero:

$$\sum_I \theta_I = (vN)^{-1} \left(N \sum_I \omega_I - vN \right) = 0.$$

Of course, when dealing with DF defined in a monoconfigurational closed shell situation, then the shifted SF is just the null function.

c. Atomic shell approximation (ASA) DF and SF

Within the ASA framework, a molecular DF is constructed with a set of atomic parts $\{\sigma_A | A = 1, N\}$ which act as an atomic SF set: $\forall A : \langle \sigma_A \rangle = 1$, and are centered at each nucleus of the molecule. The molecular ASA DF is constructed with:

$$\rho = \sum_A Q_A \sigma_A$$

where the coefficient set $\{Q_A | A = 1, N\}$ correspond to atomic populations or just atomic numbers, which fulfill: $\sum_A Q_A = v$. Therefore a molecular ASA SF can be straightforwardly constructed as:

$$\sigma = v^{-1} \rho.$$

The set of atomic SF can be seen as a intramolecular polyhedron, thus as in the MO case a centroid SF can be constructed, having the form:

$$\sigma_C = N^{-1} \sum_A \sigma_A$$

At the same time a shifted ASA SF can be build up:

$$\zeta_C = \sigma - \sigma_C = (vN)^{-1} \sum_A (NQ_A - v)\sigma_A = \sum_A \vartheta_A \sigma_A$$

Which yields a shifted ASA SF which is formally exact as in the MO case, with the appropriate substitutions. The relationships between the involved coefficients can be described in the same way:

$$\forall A : \vartheta_A = (vN)^{-1}(NQ_A - v) \leftrightarrow Q_A = v(\vartheta_A + N^{-1}).$$

The sum of the shifted ASA SF coefficients yields:

$$\sum_A \vartheta_A = (vN)^{-1} \left(N \sum_A Q_A - Nv \right) = 0$$

as in the previous MO case.

It is obvious that the centroid shift corresponds to a general operation which is certainly a useful characteristic of molecular polyhedra.

7 Discussion

The results prove that the correction of the initial expectation value difference appears as just consisting into a similarity matrix trace average term, appearing in the vertex shift procedure. Such term transforms into an alternative complete similarity matrix sum average term, when taking into account the centroid shift.

Centroid shift has been already employed systematically in Quantum QSPR studies providing quite coherent results [3]. Centroid shift provides almost the same algorithm for expectation value evaluation, but with a more involved mathematical process. Apparently it corresponds thus to the fact that only a numerical origin shift separates a procedure from the other.

The origin shift technique can be revealed as an interesting feature which can be applied into molecular DF from the point of view of intramolecular polyhedra present into their definition.

References

1. R. Carbó-Dorca, E. Besalú, J. Math. Chem. **50**, 210 (2012)
2. R. Carbó-Dorca, A. Gallegos, A.J. Sánchez, J. Comp. Chem. **30**, 1146 (2008)
3. R. Carbó-Dorca, J. Comp. Chem. **34**, 766 (2013)
4. R. Carbó-Dorca, E. Besalú, J. Math. Chem. **50**, 1161 (2012)
5. R. Carbó-Dorca, J. Math. Chem. **51**, 289 (2013)
6. R. Carbó-Dorca, J. Math. Chem. **51**, 1397 (2013)
7. R. Carbó-Dorca, Quantum Similarity Chapter 17, in *Concepts and Methods in Modern Theoretical Chemistry*, vol. 1, ed. by S.K. Ghosh, P.K. Chattaraj (CRC Press, Taylor & Francis, 2013), pp. 349–365