Quantum theory of QSAR

Ramon Carbó-Dorca* and Emili Besalú

Institut de Química Computacional. Universitat de Girona

Abstract

Ways of developing the formalism where Quantum Similarity Measures (QSM) become a natural product issuing from a specific mathematical framework related to quantum theory are discussed. This fact is used to establish a fundamental connection between Quantum Theory and QSAR, which is analysed in turn within the realm of discrete quantum chemistry. In order to achieve such an objective several theoretical tools are revised in a previous step. The first section is devoted to constructing the concept of the Tagged Set. Next, the definition of Quantum Object (QO) is clarified by means of Quantum Theory background ideas and the previous Tagged Set formalism. In the definition of QO, Density Functions (DF) are shown to play a fundamental role and a possible simplified mathematical picture is presented for possible computational purposes. In the process of preparing the problem-solving tools, convex sets become prominent, and the notion of Vector Semispace appears as a consequence. The Transformation Rule, a device to connect Wavefunctions with DF, is defined in a new step. Various products of this preliminary discussion are described, among them the concept of Kinetic Energy distributions, issuing from the background concept of extended Hilbert and Sobolev spaces. QSM as a source of discrete representation of molecular structures is made evident in this context. Further theoretical development undertakes precise study of discretization, that is, the transformation of infinite-dimensional functional spaces into n-dimensional ones. This result adds new perspectives to the discrete representation of QO, because a) It provides a source of new QO descriptors, b) It describes the QSAR theoretical background enabling the construction of adequate models like tuned-QSAR, and c) It allows the construction of sound and general alternatives of Hammet's σ or log P parameters. In this context, QSM appear to produce QSAR models constructed with unbiased descriptors, deducible from quantum theory and thus providing these models with a solid causal background. A stochastic transformation of quantum similarity matrices is de-

* Author for correspondence: Ramon Carbó-Dorca, Institut de Química Computacional, Universitat de Girona. 17071 Girona, Catalonia (Spain). Tel. 34 972 418357, Fax: 34 972 418356. Email: quantum@iqc.udg.es

Resum

Es discuteix aquí la forma de desenvolupar un formalisme on les mesures de semblança quàntiques (QSM) es transformen en un producte natural, que sorgeix d'un marc de treball específic relacionat amb la teoria quàntica. Aquesta fita s'empra per establir una connexió fonamental entre la teoria quàntica i les QSAR, que s'estudien més endavant des del punt de vista de la química quàntica discreta. A fi d'assolir aquest objectiu es revisen en un primer pas diverses eines teòriques. D'aquesta manera la primera secció s'associa a la construcció del concepte de *conjunt etiquetat*. Més tard, la definició d'*objecte quàntic* (QO) s'aclareix emprant tant el rerefons de la teoria quàntica com els conceptes previs, que formen part del formalisme de conjunt etiquetat. Per definir un QO, es demostra que les funcions de densitat (DF) tenen un paper principal i es presenta una possible forma matemàtica simplificada amb propòsits computacionals. En el camí de preparar les eines per dilucidar el problema, els conjunts convexos resulten ser prominents, mentre que la noció de *semiespai vectorial*, apareix com a conseqüència. Les regles de transformació, un aparell dissenyat per connectar les funcions d'ona amb les DF, es defineixen en un proper pas. També es descriuen diversos aspectes d'aquest tipus de discussió preliminar, entre altres el concepte de *distribucions d'energia cinètica*, que apareixen dins la definició dels espais de Hilbert generals i els espais de Sobolev. Les QSM, com una font de la representació discreta de les estructures moleculars, es fan evidents dins d'aquest concepte. Un desenvolupament posterior de la teoria intenta estudiar els processos de discretització; això és: la transformació dels espais funcionals d'infinites dimensions en espais n-dimensionals. Aquest resultat afegeix noves perspectives a la representació discreta de QO, ja que: a) esdevé una font de nous descriptors, b) descriu el fonament de les QSAR, cosa que permet la construcció de models adequats com els anomenats *QSAR sintonitzats*, c) permet la construcció d'alternatives generals i molt refinades en paràmetres empírics com els associats a la σ de Hammet o al log P. Dins d'aquest context, les QSM poden donar models QSAR construïts amb descriptors sense biaix, deduïbles de la teoria quàntica: per aquesta raó, les QSM forneixen aquests models d'una sòlida base causal. Es descriu, igualment, una transformació estocàstica de les matrius de

scribed and its role in the field of QSAR analyzed. The possible interest in manipulating a quantum similarity matrix computed on a quantum object set in such a way is diverse. First, any quantum similarity matrix column or row can easily become a discrete probability distribution, associable to a corresponding quantum object density function. Second, in order to ease its subsequent use, the resulting quantum stochastic matrix can easily be symmetrised, by means of any usual procedure or, as described here, by an inward matrix product algorithm. Third, the final matrix transform can be considered as a new quantum similarity index and can be used as a new quantum object descriptor in QSAR models. Fourth, such symmetric stochastic transforms can acquire an interesting role in the approximate solution of the fundamental quantum QSAR (QQSAR) equation under various assumptions. Finally, a new algorithm, based on inward matrix product algebra, to obtain strictly positive constrained solutions of the fundamental QQSAR equation, is described.

semblança quàntica i s'analitza el seu paper en el camp de les QSAR. El possible interès per manipular d'aquesta manera una matriu de semblança quàntica, calculada sobre un conjunt d'objectes quàntics, es divers. Primer, qualsevol fila o columna d'una matriu de semblança quàntica esdevé una distribució de probabilitats discreta, connectable a la DF de l'objecte quàntic associat. Segon, per fer-ne més fàcil l'ús posterior, una matriu estocàstica es pot fàcilment fer simètrica tot emprant un procediment usual o, tal com es descriu aquí, per mitjà de l'algorisme del producte matricial cap endins. Tercer, la forma final de la matriu transformada es pot interpretar com un nou índex de semblança quàntica i pot ser emprat com un nou descriptor d'objectes quàntics en els models QSAR. Quart, la transformada estocàstica simètrica pot adquirir un paper important en la resolució aproximada, tenint en compte diversos supòsits, de l'equació fonamental de les QSAR quàntiques (QQSAR). Finalment, es descriu un nou algorisme, basat en l'àlgebra dels productes matricials cap endins, per obtenir solucions, restringides a ser estrictament positives, de l'equació fonamental QQSAR.

1. Introduction

a) Preliminary Considerations

Similarity is a widespread concept used from the early times of human thought development. From everyday life up to sophisticated philosophical discussions similarity has somehow played and continues to play a leading role. The reason for this fact can be found in the human psychological need to compare objects, situations, sentences, structures,as a mechanism of knowledge acquisition through growing experience accretion. The human mind, in part, uses similarity to recognize current information as forming the body of past stored memories or constituting fresh brand new structures which in turn merit being gathered. Similarity might also be used to connect related sets of previously stored information. In this way, the many facets of the enveloping world, present around the human observer and structured inside the measuring mind, can be comprehended. Thus, similarity concepts and related algorithms cannot be alien to the scientific method.

The ability to measure the amount of similarity between a given pair of objects of any kind may provide the observer with the power to construct schematically new objects. These will also have the potential to be as adequate for any further purpose as the former well-known initial ones. This mechanism leads toward two fundamental needs, which can

be practically stated as: 1) comparable objects must be described in some quantitative way, as simply as possible. 2) from the feasible initial quantitative object description a *Similarity Measure*, that is, a positive real number, counting and resuming the amount of similarities among the compared objects, should also be unambiguously defined. These two points are indispensable tools to allow the observer to grasp common characteristic features between compared objects. If this prospect is feasible to being implemented as an abstract procedure, object ordering may be somehow possible and either object interpolation or object extrapolation can be available as routine procedures. New objects can be constructed in this way. Then, the essential problem of how to define objects in a suitable fashion arises. According to the previous statements, only quantitatively describable objects will possess a potential scientific interest from the similarity point of view. The quantitative, numerical, object description must be of such a nature as to allow some *a posteriori* manipulation leading to the possible comparison between object pairs or higher order object groups.

The most crucial consequence of this scheme appears to be condensed into the possibility of describing in a quantitative manner objects which have not yet been observed in nature. That is, it should be possible to know some entities which do not need to physically exist as natural elements, but which can possess a conceptual form as real as that of

the objects initially chosen at the outset of the similarity study. The objects of interest for similarity measure manipulations are those which can be described as a set of quantitative measurements, obtained in some experimental, empirical or computational way. These measurements can be collected into an ordered *n*-tuple, forming an *n*-dimensional vector. As the number of essentially different, linearly independent, descriptors over every element of the studied object family grows, the dimension of the vector space containing the results of such object descriptor variables may increase in the same amount. Construction of an abstract model of the gathered information describing the objects is a necessary step between the *observational* initial phase and the *comparisonal* step. The objects described in such an abstract vector-like manner can be considered as points usually contained into some *n*-dimensional Euclidean space. As a consequence of this procedure, a set of objects hereafter described as points may from now on be referred to as a *point-cloud*, and the objects contained as elements within the point-cloud may be named *point-objects*. Construction of such point-object sets might be considered, from a computational point of view, as the art of building up the structure of some kind of *hypervirtual reality* entity, in the form of a complex polyhedron, contained within some *n-dimensional cyberspace*. Thus, mathematical manipulations may allow the study of a point-cloud as some real solid form, which is transformed, projected, into some part of the real *three*-dimensional anthropomorphic world. Moreover, a similarity measure can always be defined over a point-cloud. To schematise the feasible procedure associated with this last fundamental statement, it is only necessary to connect a similarity measure with some of the already described vector operations producing scalar values, which are well known and precisely defined in all kinds of *n*-dimensional Euclidean spaces. The point-objects, the studied objects ordered as descriptive elements, which have been schematically discussed so far, could be considered as coming from any suitable observable and measurable source of research. But, when chemistry is concerned, the objects of interest comprise electronic systems structured in the form of atoms and molecules, and can also easily be extended as well to aggregates of other microscopic particles such as those forming nuclei. The aim of the present work is essentially the construction of a practical theoretical basis with easy computational implementation and wide practical application, allowing the comprehensive study of molecular, or even nuclear, systems as point-objects in the sense described above. Starting from the chemical concept of a molecule, there are many ways to produce a point-object structure. However, the ideal procedure permitting this construct will be the one which appears the least arbitrary of them all. That is, starting from the ideas associated with the first work on Molecular Similarity done in our Laboratory, using a Quantum Mechanical description of the molecular universe.

According to quantum mechanics, to every microscopic system, such as a molecule, may be attached a mathematical descriptor constituting the *System State Wavefunction*.

These functions can be precisely obtained for a given chemical system in a given state by solving the Schrödinger equation, previously built up adequately for the studied molecular structure. Wavefunctions have many appealing characteristics when scrutinized from the point of view of molecular similarity. The most promising feature, connected to the need of quantitative object description, is the fact that a wavefunction may be considered a vector belonging to some ∞-dimensional space. Quantum Mechanics in this manner fulfils the connection between an object, an electronic system in this case, and an element belonging to a given vector space, a point-object. Wavefunctions, however, in this initial bulk form, as solutions of the Schrödinger equation, are not suitable for similarity measures computation. This is a consequence of the characteristic mathematical form they possess. This is so because electronic wavefunctions are not appropriately well-defined functions, since, in general, wavefunctions are usually neither positive-definite or negative-definite. It is well known that integrals involving not well-defined functions in some cases may become zero systematically. In fact, this is the general feature of quantum mechanical wavefunctions: they form a complete orthonormalized set for a given electronic system. Thus, integrals involving diverse wavefunction pairs are zero systematically. This characteristic may erode any piece of information, which can be obtained from the raw wavefunctions, when looking for a general similarity measure algorithm. Fortunately, this annoying feature can be easily avoided. When accepting *Quantum Mechanical Postulates* and when using them to study a molecular structure through a *Quantum Similarity Point of View*, a definition of a new best-suited mathematical function can also be considered. It can be chosen as an alternative to the wavefunction concept, in order to obtain a convenient quantum object ∞-dimensional representation, i.e., some sort of generally well-behaved quantum point-object. In fact, Quantum Mechanical theoretical foundations are based not only on the wavefunction structure itself but on its square module: the system's density function. According to the usual admitted doctrine, the *density function* has the power to contain *all the information*, which can be gathered from measuring all the possible observables of the system. More than this, the density function has the form of a *probability* density distribution in the statistical sense. Consequently, it is a *positive definite* function, normalized in such a way as to subtend a finite unit volume. Thus, *electronic density function* has the ideal structure to be manipulated in order to construct a similarity measure attached to a completely general computational procedure.

The logical pattern that configures the present discussion now seems clear. Molecules are the objects of interest in Chemistry studies. Molecules can be described from a theoretical point of view as electronic structures. Molecular electronic systems in a given energy state can be uniquely described by means of quantum density functions. Density functions contain all the information that could be extracted from the attached system by the usual statistical procedures. Similarity Measures can be computed upon density

functions as a part of this statistical manipulation. Molecular structures can then be described within a discrete *n*-dimensional representation using similarity measures computed over a finite number of molecular quantum systems. This is so, because once the similarity measure formalism is established, then to every pair of molecules, say, there can be attached a positive real number. When a given molecule is compared with a set of companion structures, including the studied structure itself, a set of numerical values is obtained which can be rebuilt in the form of an *n*-tuple. When such a discrete representation of molecular structures has been set, they can be compared and ordered using standard procedures over *n*-dimensional vector spaces. If molecular sets can be ordered, unknown properties of as yet un-synthesized, but computer-designed molecules, can be estimated. Finally, the discrete *n*-dimensional molecular description can be used in a *Quantitative Structure-Properties Relationships* framework (QSPR; other related terms are, for example, the so-called QSAR and Quantitative Structure-Toxicology Relationships, QSTR) and many chemical features may be interpreted in this manner. Proceeding in this way, Quantum QSPR (QQSPR) can be easily defined. Starting from here, various alternative definitions and several applications of *Quantum Similarity Measures* (QSM) can be deduced. Before proceeding, however, an observation must be made which will be stressed throughout this study and carefully taken into account in every circumstance: QSMs are able to produce a dual mathematical structure attached to electronic or many-particle systems. On one hand, because of the molecular quantum mechanical postulates, some ∞-dimensional representation of a molecule may always be present; on the other hand, as a result of the QSM algorithm application, a discrete *n*-dimensional description appears to be also attached to the molecule or quantum system under consideration.

One can adopt these schematic ideas as the foundation of *Quantum Similarity* and QQSPR. The present paper will be devoted to developing the associated theoretical body.

b) Initial definitions

Since the publication of the first paper on the comparison of a molecular pair by means of a quantum mechanical density function [1], a great deal of literature has been devoted to the refinement, extensions and theoretical structure of the so– called *quantum similarity* (QS). Various authors have been active in the field, starting from the early applications and further developments of Richards et al. [2-6], who were the first to understand the implications and usefulness of the QS theory. They have been followed by the use of the momentum space formalism of N. L. Allan and D. L. Cooper [7-10] as well as the density matrix extension of J. Cioslowski and co-workers [11-16]. The efforts to perceive the many facets and application of QS have been numerous and the work of R. Ponec et al. [17-23], P. G. Mezey [24,25], and other researchers [2,26-36] must also be taken into consideration?

Several collective volumes have been published. See, for example, the contents of references [37-42], for a general review of the broad spectrum contained in QS techniques.

The present authors have been active in this branch of applied quantum mechanics since 1980. They contributed several publications dealing with MO taxonomy [43], molecular classification [44-46], general QS *Measures* (QSM) and QS *Indices* (QSI) [47-51], parallel computation and algorithm structure of QS [52], definition of accurate approximations of density functions [53-59], the origin of QSAR and QSPR [60], the theoretical meaning of QSI [60-66], extension of QSM to other quantum functions like electronic energies and electrostatic potentials [67], the use of QSM to obtain unbiased molecular superpositions [68], and the role of density functions to generate higher order QSM [47,49,51], among other work [69].

The scope of the present review focuses on the schematic mathematical interpretation of the ideas associated with QSM in order to understand the context where this kind of measures should be considered. At the same time, it is our intention to obtain as much information as possible on the theoretical implications of the whole discussion. The definition of Quantum Object (QO), given in previous studies with more or less rigour, and successfully used afterwards [44-48,51,70], will constitute the starting point of the present formalism. This definition requires a previous fundamental one, which may be related to the Fuzzy Set conceptual structure [71,72] but which can be redefined as a new collection of more general mathematical devices: the Tagged Sets [73]. QOs are not separable from *Density Functions* (DF), and for QSM practical purposes first order DFs are good candidates for use in molecular comparisons, although higher order DFs can be used as well [47,49,51] and other quantum systems can be studied in the same way as molecular structures [52,74-76]. Part of the present development will be formulated within the possibility of obtaining accurate DF by means of the so-called Atomic Shell Approximation (ASA) functions [51,53-59], and thus the implications of this kind of functional form will be studied in depth. Tagged Sets, QOs, ASAs and DFs open the way to easy definition of the structure of QSM and QSI, from purely theoretical grounds up to practical algorithms. Once the appropriate theoretical framework has been established, the immediate possibility of using QSM as a sound tool to discretize the QO description appears as a natural consequence. Two interesting new features will also be added to the general theory.

Not long ago, Sen and the present authors [77,78] suggested the possible use, in theoretical quantum chemistry applications and subsequent mathematical developments, of an elementary matrix operation: the *inward matrix product* (IMP). Such a product is present, within the usual Fortran 95 compilers, as a set of intrinsic procedures [79], among a related collection of matrix manipulation functions. The literature already referred to on IMP [77,78] must be considered, almost in full, cast in this previous QS and DF theoretical development. Recently, Klein described in detail the potential uses of *partially ordered sets* (posets) in the field of theoretical chemistry [80]. Partial order within a set arises when some object descriptor, associated with the elements of a

given set, is not symmetric with respect to the involved object order. An appropriate example of such an occurrence can be linked to the matrices arising from a nonsymmetric topological similarity measure, proposed by Mezey [24]. In this particular case the topological measure T(A,B), involving two objects A and B, was defined in such a way as to yield a different value when computed in reverse order, that is, T(A,B)≠T(B,A). At the same time, in recent years, it has been demonstrated in our laboratory that *quantum similarity measures* (QSM) constitute a universal unbiased source of quantum object descriptors [60,81-85]. However, the QSM Z(A,B), involving two QOs, A and B, are symmetric measures, that is, Z(A;B)=Z(B;A), and consequently, the associated quantum *similarity matrices* (SM) generated in such manner, become uninteresting for poset construction over QO *sets* (QOS). See [80] and references [47,49-51,86-92], for more details on the definitions used in this paragraph. On the other hand, the structure and construction algorithms leading to the precise definition of quantum *stochastic SM* (SSM) [69] have been recently discussed. Due to the non-symmetry of such SSM, they seem to constitute a potential missing link between QSM and poset definition over QOS. Alternatively, it appears at first sight difficult to promote such nonsymmetric SSM as candidates for a QO descriptor source, *within* a *quantitative structure-activity relationship* (QSAR) framework. This is so because the QSAR dimension reduction problem is based on the descriptor matrix eigenvectors [91], and usually nonsymmetric matrix eigensystems are not so easy to handle as their symmetric counterparts. See, for example references [93,94]. But new insight has proven that this is not so in this particular case [32].

c) Organisation

The present contribution will start with the mathematical interpretation and further development of the ideas associated with Quantum Similarity Measures (QSM) [1-43] which, among other possibilities, can be used to construct discrete *n*-dimensional mathematical representations of molecular structures. The Quantum Object definition given in preceding studies [52,86,87,92] and frequently used afterwards [86,87,91,92,95-107] will constitute the axis of the present formalism. In this way, the theoretical discussion shows the QO to be a concept inseparably coupled to Density Functions . However, several new possibilities become apparent along the path of this theoretical development, among others, Kinetic Energy and Electrostatic Potential density distributions as well as DF transformations. Thus, Tagged Sets, DFs and QOs open the way to an easy definition of QSM structures and their generalisation.

In order to understand the evolution of the concepts, relative to the connection between quantum chemistry and QSAR, leading to the formulation of Quantum QSAR (QQSAR), the originating ideas can now be discussed. In this effort, Extended Hilbert and Sobolev spaces appear as excellent tools to polish the problems that have arisen. So, a schematic review of both themes is provided. Several years ago, Bell [108] presented various proposals related to the attempt to polish some ambiguous theoretical aspects of Quantum Mechanics. One of Bell's suggestions was to produce a set of clear background definitions, where the theory could be easily developed. Continuing in this spirit, the present work is structured along a set of definitions, intended to propose a sound formal basis encompassing the whole area, starting from the basic aspects and ending with the final applications. Thus, this paper is organised in the following way: In the first instance Tagged Sets and Vector Semispaces will be defined. Next, Density Functions, Quantum Objects and Convex Sets will be discussed. Quantum Similarity Measures will follow, and the related subjects of Extended Hilbert spaces, Sobolev spaces and discretization and Similarity Matrices will be analysed. This previous description leads us towards a study of the theoretical background of QSPR and illuminates several varied aspects of the problem. Finally, inward matrix product, Stochastic Similarity Matrices and the solution of the QQSPR and QQSAR problems will be dealt with.

2. Tagged Sets

Consider a collection of objects of arbitrary nature forming a set, and a collection of mathematical elements: Boolean strings, column or row vectors, matrices, functions, forming another set, which can be generally taken to be completely independent in nature from the initial set. Both sets can be related by means of a new composite set construction, according to the following definition.

Definition 1: Tagged Sets.

Let us suppose a given set to be known, the Object Set, **S**, and another set, made of some chosen mathematical elements, which will hereafter be called tags, forming a Tag Set, **T**. A Tagged Set, **Z**, can be constructed by the ordered product: **Z**=**S**×**T**:

$$\mathbf{Z} = \left\{ \forall \theta \in \mathbf{Z} \middle| \exists s \in \mathbf{S} \wedge \exists t \in \mathbf{T} \rightarrow \theta = (s,t) \right\}$$

Tagged Sets constitute a mathematical structure, present in a frequent manner within chemical information handling. Atomic or molecular parametric description may be made and studied inside this general but simple Tagged Set construction. Indeed, a primitive unnoticed Tagged Set structure started when chemistry was born. Later, molecular structure as a chemical object of study has attached a collection of attributes, growing with time. A paradigmatic example is seen in the Mendeleev periodic table of the elements. Such a situation may be generalised by setting a Tagged Set formal building-up rule, where molecules become parts of the Object Set and their ordered attributes can be considered to be the Tag Set elements.

2.1. Boolean Tagged Sets

Tagged Sets may be seen as sets simultaneously taking into account both their own elements and any kind of coherent information to describe them. The simplest of Tagged Sets can be defined whenever the Tag Set part elements can be transformed into or expressed by n-dimensional Boolean or bit strings. Every n-bit string could be easily attached to any of the 2^n vertices of an n-dimensional unit length hypercube, H_n. Thus, any set of objects, possessing some kind of information linked to them, can be structured as a Tagged Set, using the appropriate n-dimensional cube vertices as the Tag Set elements [52,73,87,92].

Moreover, there appears to be present a characteristic feature, which will reappear throughout this paper, associated with the definition of any hypercube vertex Tag Set. Boolean Tags are formed by unit length n-dimensional cube vertices which, due to the bit-like nature of their components, can be considered directed and included into a *Positive Definite* (PD) hyper-quadrant section belonging to some n-dimensional space. It is also obvious that other Tagged Sets can be transformed into such a Boolean form: Consider a Tag Set made up of n-tuples of rational numbers as a quite common and general example. The nature of the molecular information precludes the possibility of easily transform chemical Tagged Sets into Boolean molecular Tagged Sets. In general, the effortless transformation of a Molecular Tagged Set into a Boolean one is propitiated by two converging circumstances. First, the natural intrinsic PD rational character of the experimental or theoretical information gathered into the Molecular Tag Set; second, the peculiar structure of modern electronic computational tools.

From the above point of view, Boolean Tagged Sets can be considered as a sort of canonical form, which can describe, in some ultimate way, any kind of discrete rational information orderly attached to a chosen Object Set.

2.2. Functional Tagged Sets

Until now, Tagged Sets in this discussion have been supposed to be implicitly constructed employing n-dimensional vector-like Tag Sets. However, there is no need to circumscribe Tag Set parts to finite-dimensional space subsets.

Another crucial point has to be considered before going ahead in the description of Tagged Sets. It is related to Boolean Tagged Sets, and appears when an ∞-dimensional hypercube vertex subset is taken as the Tag Set part. A parallelism then naturally appears between the ∞-dimensional vertices and the elements of the [0,1] segment acting as tags. Finally, we must consider the possible use of Boolean Matrices of arbitrary dimension (m×n) or, still more generally, Boolean Hyper-matrices, as sound candidates for inclusion in the Tag Set part. All these multiple possibilities, associated with Boolean Tagged Sets, open the way to considering the possible definition of still more general Tagged Sets. In the same way, Tag Sets can also be made of elements coming from ∞-dimensional spaces. Any practical function space can be considered as belonging to the ∞-dimensional class of spaces. Moreover, among all the possible function families, possessing appropriate homogeneous properties, the most appealing candidate, from the molecular point of view, corresponds to a subset of some probability density functional space. Two reasons point towards this kind of choice. First, Probability DFs (PDFs) are normalizable: they are PD functions too, yielding values within the [0,1] segment, and they may behave like some isomorphic infinite-dimensional limit form of a Boolean Tag Set. Second, according to the interpretation given by von Neumann [109] or Bohm [110], PDFs formed by the squared module of quantum mechanical state wavefunctions, constitute mathematical elements attached to the descriptive behaviour of quantum systems. Recent [111], and not so recent [108], discussions signal towards this descriptive role of the quantum DF too. It seems that, in this case, and from a quantum mechanical perspective, a PDF must of necessity be used, if one is willing to take into account the whole information attached to a given molecular structure. From the preceding ideas, PDF Tagged Sets appear to be the natural ∞-dimensional extension connected to the discrete n-dimensional Boolean Tagged Sets to be used in quantum chemical applications. Actually, there is no need to search for any new mathematical structure: Definition 1 as it is still holds, even when any ∞-dimensional space subset is employed as the Tag Set part.

2.3. Vector Semispaces

This broad catalogue of Tag Set candidates: Boolean strings, PD n-dimensional vectors and functions, permit a great flexibility when a particular Molecular Tagged Set needs to be defined. At the same time, it is very convenient to discuss which kind of Tag Sets can be chosen as candidates to fill the gap between a Boolean hypercube and a PDF space. A natural choice may be constituted by n-dimensional vector spaces with some appropriate restrictions. The following auxiliary definition could be used accordingly.

Definition 2: Vector Semispaces.

A Vector Semispace (VSS) over the PD real field $\boldsymbol{R}^+$, is a Vector Space (VS) with a structure of an Abelian semigroup associated to the vector addition.

By an additive semigroup [112] is understood an additive group without the presence of reciprocal elements. So, the VSS structures do not admit negative vectors and no vector differences are defined. All VSS elements can be considered, in the same way as Boolean hypercube vertices are, directed towards the region of a positive axis hyperquadrant. The present paper also accepts that *null elements* are included in both the scalar field and in the VSS structure. VSS linear combinations are to be considered as created with positive coefficients; thus, vector coordinates are always positive or in some special cases null. Metric VSS may be constructed in the same manner as the usual metric VS, taking into account that scalar products will also become PD, and consequently no negative cosines of vector angles could be obtained. However, no classically defined dis-

tances are allowed, these being the result of vector difference norms. The inverse of the cosine of the angle subtended by two vectors can be used instead.

In fact, any VSS could be taken as attached to some VS, if some procedure could be defined describing the transformation of the VS elements into a VSS. The relevant VSS generating rules will be discussed below. Nothing stops any VSS Tag Set part from being constituted by normalised vectors, whose elements will then be numbers belonging to the unit segment [0,1]. Thus, in this way, VSS Tag Sets could be associated to Boolean Tag Sets as described in section 2.1.

3. Density Functions as Object Tags

Discussion in the following section will deal more precisely with the background ideas subtending the relationship between quantum systems and Tagged Sets. DFs have been, since the early days of Quantum Mechanics, an indispensable tool for defining mechanical systems at the adequate microscopic level. Therefore, Tag Set parts made of quantum DFs ought to be associated with a quantum system's information. Consequently, the quantum theoretical structure fits perfectly into the Tagged Set formalism and permits the definition of valuable new data.

3.1. Quantum Objects
The idea of a QO without a well-designed definition has been used frequently in the literature [47,48,51]. Moreover, the background mathematical structure leading towards the recently published [52,86,89,92] definition of QO is to be found in the previous section. In order to obtain a sound QO definition, some preliminary considerations are necessary.

The QO concept is used here instead of the more restrictive concept of molecule, because the attached quantum mechanical background of the QQSAR models permits us to connect the structure and properties of any microscopic system capable of being described by quantum mechanics.

3.1.1. Expectation Values in Quantum Mechanics
Our starting point is the fact that a quantum study of microscopic systems is essentially associated with the following algorithm:

Algorithm 1: Quantum Mechanics.

1) Construction of the Hamilton operator, $\boldsymbol{H}$.
2) Computation of the state energy-wavefunction pairs, $\{\boldsymbol{E}, \Psi\}$, by solving the Schrödinger-like equation: $\boldsymbol{H}\Psi = \boldsymbol{E}\Psi$.
3) Evaluation of the state DF, $\rho = |\Psi|^2$.

If we know the state DF, all observable property values of the system, ω, can be formally extracted from it, as expectation values, $\langle \omega \rangle$, of the associated Hermitian operator, Ω, acting over the corresponding function, ρ. In the same way as in theoretical statistics, it can be written:

$$\langle \omega \rangle = \int \Omega(\boldsymbol{r})\rho(\boldsymbol{r})d\boldsymbol{r} \qquad (1)$$

where $\boldsymbol{r}$ represents a p-dimensional particle coordinate matrix. It is important to note that equation (1) can be interpreted as some scalar product or linear functional: $\langle \omega \rangle = \langle \Omega | \rho \rangle$, defined within the space where both the involved $\Omega(\boldsymbol{r})$ and $\rho(\boldsymbol{r})$ p-particle operators belong.

A typical example of the scheme described above may be constituted by the electronic part of Electrostatic Molecular Potentials (eEMP), first employed by Bonaccorsi, Scrocco and Tomasi [0]. eEMP evaluated at the position $\boldsymbol{R}$ in 3-dimensional space, $V(\boldsymbol{R})$, computed over first order DF, $\rho(\boldsymbol{r})$, is defined using equation (1) as:

$$\Omega(\boldsymbol{r}) = |\boldsymbol{r} - \boldsymbol{R}|^{-1} \wedge V(\boldsymbol{R}) = \int |\boldsymbol{r} - \boldsymbol{R}|^{-1}\rho(\boldsymbol{r})d\boldsymbol{r} \qquad (2)$$

Not taking into account the electron charge sign, eEMP acts as a PD distribution, with maxima located at the molecular nuclei.

3.1.2. Quantum Object Definition and Generating Rules
Thus, after these preliminary considerations, the next definition can readily be made.

Definition 3: Quantum Object.

A Quantum Object is defined as an element of a Tagged Set, made by Quantum Systems in well-defined states taken as the Object Set part and the corresponding Density Functions constituting the Tag Set part.

The interesting fact lies in the leading role the DF plays in quantum mechanical systems description and, as a consequence, in the QO definition. DF generation in varied wavefunction environments has been studied since the early times of quantum chemistry [114,115]. The most appealing aspect of this situation corresponds to the way DF, ρ, are constructed, starting from the original system's wavefunctions, Ψ. This formation process has been called a *generating rule* [96], which can be shortened by using the symbol: $\boldsymbol{R}(\Psi \rightarrow \rho)$. A generating rule can easily be written, summarising the three steps of quantum mechanical Algorithm 1:

$$\boldsymbol{R}(\Psi \rightarrow \rho) = \left\{ \forall \Psi \in \boldsymbol{H}(\boldsymbol{C}) \rightarrow \exists \rho = \Psi * \Psi = |\Psi|^2 \in \boldsymbol{H}(\boldsymbol{R}^+) \right\} \qquad (3)$$

In equation (3) there are given explicitly the wavefunction Hilbert VS [0], $\boldsymbol{H}(\boldsymbol{C})$, and the DF VSS, $\boldsymbol{H}(\boldsymbol{R}^+)$, defined over the complex and the PD real fields, respectively.

3.2. Density Functions
The way the DF can be variable reduced is well known [114,115]. Integrating the raw DF definition, which appears in the generating rule of equation (3) or as the third step of Algorithm 1, over the entire system particle coordinates, except r of them, produces an r-th order DF. This kind of reduction

has been studied in many ways [52,117,118] and will not be repeated here.

When practical implementation of QSM was considered in this laboratory, a simplified manner of constructing the first-order DF form [53-59] was also proposed and named Atomic Shell Approximation (ASA) DF. A procedure has recently been described [57], bearing the correct necessary conditions to obtain PD ASA DF possessing appropriate probability distribution properties. This will be discussed below. The ASA DF development could be trivially related to the first-order DF form as expressed within MO theory, because the first-order MO DF structure can be defined as:

$$\rho(\boldsymbol{r}) = \sum_i w_i |\varphi_i(\boldsymbol{r})|^2 \qquad (4)$$

although this MO DF can be written in a general way, as a double sum of products of function pairs, coupled with a set of matrix coefficients [119]. However, a simple matrix diagonalization, followed by a unitary MO basis set transformation, can revert DF to the formal expression in equation (4), see for instance [114] or [120] for more details. The coefficient set: $W = \{w_i\} \subset \boldsymbol{R}^+$, usually interpreted as MO occupation indices, corresponds in any case to a collection of positive real numbers. The original MO function set: $f = \{\varphi_i(\boldsymbol{r})\} \subset H(\boldsymbol{C})$ belongs to a Hilbert VS, but appears when used within DF in a squared modular form, that is, $F = \{|\varphi_i(\boldsymbol{r})|^2\} \subset H(\boldsymbol{R}^+)$. This new basis is a set of PD functions belonging to an ∞-dimensional VSS. The result of the PD linear combination of PD functions is a PD first order DF, $\rho(\boldsymbol{r})$. Here, a unit norm convention has been adopted:

$$\text{If } \int |\varphi_i|^2 \, d\boldsymbol{r} = 1; \forall i \ \Rightarrow \int \rho(\boldsymbol{r}) \, d\boldsymbol{r} = \sum_i w_i \int |\varphi_i|^2 \, d\boldsymbol{r} = \sum_i w_i = 1 \quad (5)$$

And this result can be interpreted considering the coefficient set $W=\{w_i\}$, as a discrete probability distribution.

As W is a PD real numerical set, it can be organised as an n-dimensional vector, for instance: $\boldsymbol{w}=(w_1, w_2, ..., w_n)$. Moreover, either W or $\boldsymbol{w}$ can be generated using a complex coefficient set: $X=\{x_i\} \subset \boldsymbol{C}$. A new set of coefficients can be obtained using the modules of the X set elements, as: $w_i = |x_i|^2; \forall i$. Supposing a column vector to be defined with these elements of X: $\boldsymbol{x}=\{x_i\}$, then the norm of such a vector is forced to be: $\langle\boldsymbol{x}|\boldsymbol{x}\rangle=\boldsymbol{x}^+\boldsymbol{x}=1$, and corresponds to the last condition in equation (5). All of this defines a device in close parallelism and similar to the previous quantum mechanical ∞-dimensional generating rule, presented in equation (3). For this purpose a discrete generating rule can be described as follows:

$$R_n(\boldsymbol{x} \to \boldsymbol{w}) = \left\{ \begin{array}{l} \forall \boldsymbol{x} \in \mathbb{v}_n(\boldsymbol{C}) \to \exists \boldsymbol{w} = \left\{ w_i = x_i^* x_i = |x_i|^2 \right\} \in \mathbb{v}_n(\boldsymbol{R}^+) \\ \wedge \ \boldsymbol{x}^+ \boldsymbol{x} = \sum_i x_i^* x_i = \sum_i |x_i|^2 = 1 \to \langle\boldsymbol{w}\rangle = \sum_i w_i = 1 \end{array} \right\} \quad (6)$$

As used in previous equations, the following definition will now continue to hold:

Definition 4: Elements sum of a (m×n) matrix A.

Known a (m×n) matrix $\boldsymbol{A} = \{a_{ij}\}$, by the symbol $\langle\boldsymbol{A}\rangle$ it is meant:

$$\langle\boldsymbol{A}\rangle = \sum_{i=1}^m \sum_{J=1}^n a_{ij}$$

It must be noted here that when $\boldsymbol{A}=Diag(a_i)$, then $\langle\boldsymbol{A}\rangle = Tr|\boldsymbol{A}|$. Also, the matrix operation $\langle\boldsymbol{A}\rangle$ can be considered a linear transformation from the $(m×n)$ matrix vector space to the background definition field.

If an equivalent set of conditions like those shown in equation (5) holds for some r-th order DF basis functions, a discrete generating rule, such as the one described above, can be extended to DF of arbitrary r-th order too.

3.3. Convex Conditions and ASA Fitting
3.3.1. General Discussion

Optimisation of the coefficient vector, $\boldsymbol{w}$, in order to obtain an approximate function completely adapted to *ab initio* DF, must be restricted within the boundaries of some VSS: $V_n(\boldsymbol{R}^+)$ and the element sum, $<\boldsymbol{w}>$, will be unity according to conditions obtained in equation (5). This feature can be cast into a single symbol, which can be referred to as *convex conditions*, $K_n(\boldsymbol{w})$, applying over the n-dimensional vector $\boldsymbol{w}$ and written as:

$$K_n(\boldsymbol{w}) \equiv \left\{ \boldsymbol{w} \in V_n(\boldsymbol{R}^+) \wedge \langle\boldsymbol{w}\rangle = \sum_i w_i = 1 \right\} \quad (7)$$

In a similar notation, the elements of the set $\boldsymbol{W},$ or those of the vector, $\boldsymbol{W}=\{w_i\}$, can be used instead in the symbol defining the convex conditions, that is:

$$K_n(\{w_i\}) \equiv \left\{ \forall i: w_i \in \mathsf{R}^+ \wedge \sum_i w_i = 1 \right\} \quad (8)$$

Together, equations (7) and (8) can be considered the discrete counterparts of the continuous convex conditions, defining a convex DF:

$$K_\infty(\rho) \equiv \left\{ \rho \in H(\boldsymbol{R}^+) \wedge \int \rho(\boldsymbol{r}) \, d\boldsymbol{r} = 1 \right\} \quad (9)$$

Convex sets [121,122] play a leading role in optimisation problems. Recently, they have been introduced as an important mathematical structure to deal with chemical problems [24]. Thus, it is not strange that convexity may be attached to the definition of QO. Moreover, vector coefficients may be easily transformed by means of norm conserving orthogonal transformations, like Elementary Jacobi Rotations (EJR) [123]. EJR can be applied over a generating vector to obtain the desired optimal coefficients, while preserving convex conditions [57,91,96].

As mentioned at the end of section 3.2., the DF form shown in equation (4) can be used to build up new DF elements, preserving $K_\infty(\rho)$. If $\boldsymbol{w}$ is taken as a vector, assuming

the convex conditions $K_n(\boldsymbol{w})$, while the set $\boldsymbol{P} = \{\rho_i(\boldsymbol{r})\} \subseteq \boldsymbol{H}(\boldsymbol{R}^+)$, is used as a given set of homogeneous order DF, then the linear combination:

$$\rho(\boldsymbol{r}) = \sum_i \boldsymbol{w}_i \rho_i(\boldsymbol{r}) \in \boldsymbol{H}(\boldsymbol{R}^+) \tag{10}$$

produces a new DF with the same order and characteristic properties as the elements in the set $\boldsymbol{P}$. It can be said that convex conditions over vector coefficients, affecting DF superpositions, are the way to allow the construction of new DF of the same nature, bearing the same properties. Quite a considerable proportion of chemical computations, performed over numerous molecular systems, is based on such a principle.

3.3.2. ASA Fitting

Although a recent paper [57] gives the complete details of ASA fitting, recent algorithm developments and new atomic fitting tables will be of help to readers interested in applications of the ASA DF [58,59].

Essentially, the ASA fitting algorithm can be divided into three well-defined parts: a) generation of ASA exponents using *even-tempered* geometric sequences [124]; b) optimization of coefficients using an elementary Jacobi rotation (EJR) technique [123]; c) exponent optimization using a Newton method [125]. Only part b) is schematically discussed here. For more details see the references above.

Optimal sets of ASA coefficients and exponents are obtained by minimizing the integral:

$$\varepsilon^{(2)} = \int \left| \rho(\boldsymbol{r}) - \rho^{ASA}(\boldsymbol{r}) \right|^2 d\boldsymbol{r} \tag{11}$$

which corresponds to the common definition of the quadratic error integral function between ab initio, $\rho(\boldsymbol{r})$, and ASA, $\rho^{ASA}(\boldsymbol{r})$, electronic DF, subject to the convex conditions described in equations (7) and (8). Substituting the ASA DF defined in equation (4) and using matrix notation, the integral $\varepsilon^{(2)}$ can be written now as a function of the *n*-dimensional vector $\boldsymbol{w}$:

$$\varepsilon^{(2)}(\boldsymbol{w}) = \boldsymbol{Z} + \boldsymbol{w}^T \boldsymbol{Z} \boldsymbol{w} - 2\boldsymbol{b}^T \boldsymbol{w} \tag{12}$$

where: $Z = \int \left| \rho_i(\boldsymbol{r}) \right|^2 d\boldsymbol{r}$, can be interpreted as an ab *initio quantum* self-similarity measure, see section 4.1. below for more details, and the elements of the matrix $\boldsymbol{Z} = \{Z_{ij}\}$ as well as these of vector $\boldsymbol{b} = \{b_i\}$ are given respectively, by the integrals:

$$Z_{ij} = \int \left| \varphi_i(\boldsymbol{r}) \right|^2 \left| \varphi_j(\boldsymbol{r}) \right|^2 d\boldsymbol{r} \tag{13}$$

$$b_i = \int \left| \varphi_i(\boldsymbol{r}) \right|^2 \rho(\boldsymbol{r}) d\boldsymbol{r} \tag{14}$$

As explained in section 3.2, the set of PD real coefficients $\{w_i\}$ can be substituted employing a complex coefficient set, $\{x_i\}$, using a discrete generating rule as in equation (6). This transforms the function $\varepsilon^{(2)}$ into the expression:

$$\varepsilon^{(2)} = Z + \sum_{i,j \in a} \boldsymbol{x}_i^2 \boldsymbol{x}_j^2 Z_{ij} - 2\sum_{i \in a} \boldsymbol{x}_i^2 \boldsymbol{b}_j \tag{15}$$

Variation of the quadratic error integral function employing EJR has been slightly modified with respect to the methodology described in a previous paper [57]. When an EJR is applied over a vector [126], this is the same as to apply an orthogonal transformation over the vector, which can be identified as $\boldsymbol{J}_{pq}(\alpha)$ and described by the equations:

$$\begin{aligned} \dot{x}_p &\leftarrow c\, x_p - s\, x_q \\ \dot{x}_q &\leftarrow s\, x_p + c\, x_q \end{aligned} \tag{16}$$

where only the elements p and q of the vector $\boldsymbol{x}$ are modified. The symbols c and s, appearing in equation (16) determine the cosine and sine of the EJR angle α. After some straightforward manipulations [57], a quartic equation with respect to s and linear in c is obtained:

$$\delta\varepsilon^{(2)} = E_{04}s^4 + E_{13}cs^3 + E_{02}s^2 + E_{11}cs \tag{17}$$

where the sine and cosine coefficients are functions of both the Z_{ij} and b_i integrals and the vector $\boldsymbol{x}$ elements.

The optimal sine, s^*, related to the EJR procedure can be obtained imposing the null gradient condition $\dfrac{d\,\delta\varepsilon^{(2)}}{d\,s} = 0$.

However, the procedure for obtaining the optimal angle: α^*, can be greatly improved using straightforward Taylor expansions [127,128] in order to replace the s and c expressions, present in equation (16). Finally, a simpler formulation than in equation (16) is obtained for the coefficient set variation:

$$\begin{aligned} \dot{x}_p &\leftarrow c_{opt}\, x_p \text{-} s_{opt}\, x_q = \left(1 - \frac{\alpha_+^2}{2}\right) x_p - \alpha_+ \left(1 - \frac{\alpha_+^2}{6}\right) x_q \\ \dot{x}_q &\leftarrow s_{opt}\, x_p + c_{opt}\, x_q = \alpha_+ \left(1 - \frac{\alpha_+^2}{6}\right) x_p + \left(1 - \frac{\alpha_+^2}{2}\right) x_q \end{aligned} \tag{18}$$

In this manner, Taylor expansions for the definition of the sine and cosine of the EJR transformation, eliminate the need to follow an iterative procedure employed to obtain α^*. As a consequence, a very small amount of computational time is required for ASA fitting procedures, which can be applied to any quantum system [58,59].

3.4. Positive Definite Operators

3.4.1. General considerations on PD operators

The DF themselves may be considered as elements of a VSS or also, alternatively, as members of a PD Operator Set, which can be collected in turn into another isomorphic VSS, whose elements may be considered PD Operators.

The most relevant thing to be noted in the context of PD Operator VSS, as well as in the isomorphic VSS companions structure, is the *closed nature* of such VSS, when appropriate PD coefficient sets are known. That is: PD linear combinations of PD Operators remain PD Operators. Discrete matrix representations of such PD Operators are PD too, and PD linear combinations of PD matrices will remain PD in the

same way. These properties can be expressed in a compact and elegant way, using convex conditions symbols, as previously discussed in section 3.3.1: If $\{K_\infty(\rho_i); \forall i\}$ and $K_n(\boldsymbol{w})$ hold, then equation (10) is a convex function fulfilling $K_\infty(\rho)$.

3.4.2. Differential Operators and Kinetic Energy

Another interesting question, not yet discussed as it deserves to be in the literature could be related to the interpretation of the differential operators role. They can be considered as momentum representatives within the framework of classical quantum mechanics when the position space point of view is chosen, which constitutes the usual, most frequent, computational chemistry option.

A formal puzzle appears to be present here, when one tries to connect a second order differential operator, representing the QO kinetic energy (KE), using the expression of an expectation value. KE expectation values do not fulfil the usual statistical formalism represented by equation (1), but they possess a kind of expression, which adequately transformed and, avoiding scalar factors, looks like a norm, when writing the equalities:

$$2\langle K \rangle = -\int \Psi^* \nabla^2 \Psi dV = \int (\nabla\Psi)^* (\nabla\Psi) dV \qquad (19)$$

where the change of sign can be attributed to Green's first identity [128].

Available textbooks do not explain or even mention this interesting possibility, see for a recent example reference [129]. However, the current literature presents it as a *de facto* characteristic and the usual trend is to classify this oddity within the fuzziness of quantum mechanical postulates. Discussed since the formulation of quantum theory, the presence of quantum mechanical postulates is characterised by quite a large choice of interpretations. This can be evidenced by perusal of any textbook of the type listed in references [130-137]. Interpretations range there from no postulate description at all [130] up to quite large lists of them. The suggestion can even be found that postulates should be substituted by sound definitions [108].

A remark is in order here about the absence of a scalar imaginary unit factor accompanying the nabla operator in expression (19) and in the following equations. It seems no longer to be necessary now to use this imaginary scalar factor, and indeed in what follows this also seems to be the case. The imaginary factor will appear compulsively only if one needs to stress the unitary nature of its matrix representation. Moreover, equation (19) states that KE will of necessity become a DP quantity, as it is the result of a norm-like expression. An in-depth discussion of this subject has been given elsewhere [138] and the relationship of this problem to quantum mechanical equations has been recently set forth [139]

3.5. Extended Hilbert Spaces

KE integrands, like those that appear in equation (19), being represented in modular form, can formally be considered to be behaving like DF. They can thus also be supposed to be-

long to a given VSS. A possible way to consider this problem will be brieffuly described in terms of our view of things as previously stated. Supposing that the original Hilbert space, $\boldsymbol{H}(\boldsymbol{C})$, where wavefunctions belong, is modified into another extended one, $\boldsymbol{H}^{(\nabla)}(\boldsymbol{C})$, which also contains the wavefunction's first derivatives, the quantum mechanical momentum representation, that is:

$$\forall \Psi \in \boldsymbol{H}(\boldsymbol{C}) \Rightarrow \Psi \in \boldsymbol{H}^{(\nabla)}(\boldsymbol{C}) \wedge \exists \nabla(\Psi) \in \boldsymbol{H}^{(\nabla)}(\boldsymbol{C})$$

Then, considering the attached VSS, $\boldsymbol{H}(\boldsymbol{R}^+)$, where the DF belong, one can also accept that:

$$\forall \rho = |\Psi|^2 \in \boldsymbol{H}(\boldsymbol{R}^+) \Rightarrow \exists \kappa = |\nabla\Psi|^2 \in \boldsymbol{H}(\boldsymbol{R}^+)$$

to every DF, ρ, there exists in this way a momentum DF or what one might perhaps more descriptively term this kind of distribution: KE DF, κ, belonging to a Hilbert VSS. The KE DF when integrated provides the expectation value of the QO KE. The following sequence, developing details appearing in equation (19), will shed light on the proposed question:

$$2\langle K \rangle = \int \kappa \, dV = \int |\nabla\Psi|^2 \kappa \, dV = \int (\nabla\Psi)^* \cdot (\nabla\Psi) \, dV =$$
$$-\int \Psi^* |\nabla|^2 \Psi \, dV = -\langle |\nabla|^2 \rangle \equiv -\int \Psi^* \Delta\Psi \, dV = -\langle \Delta \rangle$$

where the minus sign appears as a consequence of Green's first identity [128], as mentioned in regard to equation (19).

KE can then be considered related to the norm of momentum, the QO wavefunctions gradient. As a consequence it could be interesting to obtain KE DF, κ, maps or images in the same way as they are customarily obtained for the DF, ρ [24,25]. Complementary information on electronic DF will surely be obtained from these representations. Similar behaviour of both functions at large distances from the molecular nuclei is to be expected, but with very different behaviour near the nuclei. See references [52], [138] and [139] for a complementary discussion.

3.5.1. Generating Rules

To obtain a coherent picture, with KE occupying a sound place, among other quantum mechanical structures, then the Hilbert VS, $\boldsymbol{H}^{(\nabla)}(\boldsymbol{C})$, could be defined as not only containing wavefunctions but their first derivatives too. This allows us to construct the associated DF VSS, $\boldsymbol{H}^{(\nabla)}(\boldsymbol{R}^+)$, as containing not only DF but also KE DF. The elements of this peculiar Hilbert VS, where both wavefunctions and their gradients are taken into account simultaneously, can be ordered in the form of column vectors, like:

$$|\Phi\rangle = |\Psi; \nabla\Psi\rangle \in \boldsymbol{H}^{(\nabla)}(\boldsymbol{C}) \qquad ,$$

this form can be attached to a scalar to vector transformation using a vectorial operator, involving the gradient, such as:

$$|1; \nabla\rangle[\Psi] = |\Psi; \nabla\Psi\rangle = |\Phi\rangle \qquad .$$

Or by a diagonal transformation, employing the same elements:

$$Diag(1, \nabla)|\Psi\rangle = Diag(|\Psi\rangle, \nabla|\Psi\rangle) = |\Phi\rangle \qquad (20)$$

In the case of one particle QO, the necessary quadrivector structure, adopted by the extended wavefunctions, acquire a qualitative similarity to relativistic spinors [140,141]. In order to obtain mathematical coherence, even in non-relativistic Quantum Mechanics, it seems that wavefunctions could be easily attached to a vector-like representation, originated due to the presence of momentum and thus of KE differential operators. See references [138,139,142] for more details.

The generating rule within the extended wavefunction domain, can now be written as:

$$\mathbf{R}\left(|\Phi\rangle \to |\rho; \kappa\rangle\right) = \begin{cases} \forall |\Phi\rangle = |\Psi; \nabla\Psi\rangle \in \mathbf{H}^{(\nabla)}(\boldsymbol{C}) \to \\ \exists \rho = \Psi^*\Psi = |\Psi|^2 \wedge \exists \kappa = (\nabla\Psi)^*(\nabla\Psi) \\ = |\nabla\Psi|^2 \Rightarrow |\rho; \kappa\rangle \in \mathbf{H}^{(\nabla)}(\boldsymbol{R}^+) \end{cases} \qquad (21)$$

The DF ρ can be considered normalised, according to equation (5). The KE DF κ, can be normalised too, the gradient density norms being in absolute value twice the kinetic energy expectation value, $\langle K \rangle$. This amounts to the same as considering the extended wavefunction, $|\Phi\rangle$: $\langle \Phi | \Phi \rangle = (1 + 2\langle K \rangle)$, normalised. This is a consequence of the characteristics of the spaces containing both the wavefunction and their gradient, whose elements, then, should be considered as square summable functions as discussed next.

3.5.2. Projectors

The projectors associated with the extended quantum mechanical wavefunctions will possess a matrix structure like:

$$|\Phi\rangle\langle\Phi| = \begin{pmatrix} |\Psi|^2 & \Psi^*(\nabla\Psi) \\ (\nabla\Psi)^*\Psi & |\nabla\Psi|^2 \end{pmatrix} = P$$

then, using symmetrisation: $Q = \frac{1}{2}(P^+ + P)$, the new projector could be written as the matrix:

$$Q = \begin{pmatrix} \rho & |j\rangle \\ \langle j| & \kappa \end{pmatrix}$$

so : $Tr(Q) = Tr(P) = \rho + \kappa$, and the off-diagonal elements are:

$$|j\rangle = \frac{1}{2}\left(\Psi^*(\nabla\Psi) + (\nabla\Psi)^*\Psi\right)$$

3.6. Extended Sobolev Spaces

Actually, the definition of extended Hilbert spaces [138] has led to several fruitful applications [92,142]. Among these is a natural set-up of the Schrödinger equation within the energy expectation value in approximate form, as well as the appearance of the well-known KE DF [138,143,144]. KE DF can be seen, within the extended Hilbert space formalism as a component of a total QO density, involving for a given QO

the sum of the KE density part and the particle position probability density.

Extended Hilbert spaces can be used, along with the concepts associated with the theoretical foundations of QSM, to put in evidence the quantum mechanical origin of the so-called quantitative structure-properties relationships (QSPR) [145-147] giving rise to the framework which can be called quantum QSPR (QQSPR). This is discussed below.

The most recent development of QQSPR theory and of the subsequent applications of extended Hilbert spaces has been the evidence of their connection with Sobolev spaces [112,148]. The exploitation of these findings has been recently described [149], within the algorithm of SCF theory, as a procedure to solve the involved generalized secular equations.

This section pretends to broadly re-structure the previous formalism of extended Hilbert spaces as well as to generalize the formal structure of Sobolev spaces.

3.6.1. Sobolev spaces

The resultant structure of the extended Hilbert space norms as defined in the previous section can be associated with the usual form of a Sobolev space [112]. In Sobolev spaces [148], the norm of any of their elements, Ψ say, is defined within a general formulation as:

$$\|\Psi\|_n^m = \sum_{p=0}^{m} \left\|\nabla^p\Psi\right\|_n \qquad (22)$$

where the first term of the sum in equation (22) has to be considered as a way to write the bulk wavefunction:

$$\nabla^0\Psi \equiv \Psi \qquad (23)$$

It is obvious that the extended Hilbert space functions, from the point of view of the above-defined norms, can be considered a Sobolev space with the norm defined as:

$$\langle \Phi | \Phi \rangle \equiv \|\Psi\|_2^1 \qquad (24)$$

3.6.2. Generalisation of Sobolev Spaces

A trivial generalization of Sobolev spaces can be readily described, taking as a starting point the extended Hilbert spaces as defined above. Suppose that the Sobolev definition (22) can be generalized in such a way that it can be written, among other possibilities as:

$$\|\Psi\|_n^{r;s} = \sum_{p=1}^{r} \left\||\Psi|^p\right\|_n + \sum_{q=1}^{s} \left\|\nabla^q\Psi\right\|_n \qquad (25)$$

In a symmetrical way with the first term in equation (25), one can also define the following norm:

$$\|\Psi\|_n^{r;s;t} = \sum_{p=1}^{r} \left\||\Psi|^p\right\|_n + \sum_{k=1}^{t}\sum_{q=1}^{s} \left\||\nabla^q\Psi|^k\right\|_n \qquad (26)$$

so, in fact equation (26) transforms into (25), when t=1, in the same way as equation (25) transforms into the original Sobolev definition (22) when r=1.

Taking both this and the conventional meaning of the zeroth order gradient symbol into account, as already described in equation (23), one can first try to reach an extended Hilbert space whose functions can be associated to a generalized Sobolev space (25) with the following norm structure:

$$\langle \Phi | \Phi \rangle \equiv \|\Psi\|_2^{2;1} = \|\Psi\|_2 + \left\||\Psi|^2\right\|_2 + \|\nabla\Psi\|_2$$

So, the vector form of the extended wavefunctions can be newly written as a three-dimensional column matrix:

$$|\Phi\rangle = \begin{pmatrix} \Psi \\ |\Psi|^2 \\ \nabla\Psi \end{pmatrix} = \begin{pmatrix} \Psi \\ \rho \\ \nabla\Psi \end{pmatrix} \qquad (27)$$

taking into account that the density function ρ in the expression (27) is the squared module of the initial wavefunction form. The explicit extended function norm will now be written as:

$$\langle \Phi | \Phi \rangle = 1 + \langle \rho | \rho \rangle + 2\langle K \rangle \qquad (28)$$

The Hilbert norm of the total density form appears in equation (28) as another term in the extended Hilbert space norm. The new positive definite term corresponds to an integral with the form:

$$\langle \rho | \rho \rangle = \int \rho^2 d\boldsymbol{r} = \left\langle |\Psi|^2 \left| |\Psi|^2 \right. \right\rangle = \int |\Psi|^4 d\boldsymbol{r} \qquad (29)$$

Such kinds of integrals are well known in the field of quantum similarity and correspond to a quantum self-similarity overlap-like measure [92] involving the N-th order density matrices. They are the generalised form of the integrals studied in the next section, 4.1.1.

The possible use of extended Sobolev spaces to build up a nonlinear Schrödinger equation has been recently discussed [150] and the computation of the $\langle \rho | \rho \rangle$ integrals have been also studied [151].

3.7. Diagonal Hamilton Operators

There is only one final point to underline: The calculation of energy expectation values, within the extended Hilbert space framework. This can be done, for example, using the Born-Oppenheimer approach, defining an electronic Hamilton operator with a diagonal matrix structure and, in addition, supposing $|\Psi\rangle$ normalised:

$$\begin{aligned} \boldsymbol{H} = Diag\left(V; \tfrac{1}{2}\boldsymbol{I}\right) &\wedge \langle\Psi|\Psi\rangle = 1 \Rightarrow \boldsymbol{E} = \langle\Phi|\boldsymbol{H}|\Phi\rangle = \\ \langle\Psi|V|\Psi\rangle &+ \tfrac{1}{2}\langle\nabla\Psi|\nabla\Psi\rangle \equiv \langle V\rangle + \langle K\rangle \end{aligned} \qquad (30)$$

In the diagonal Hamilton operator definition, V is the potential part and $\boldsymbol{I}$, a unit matrix with the appropriate dimension. In any case, the KE inverse mass factors, if needed, could be supposed implicitly inserted into the gradient symbols. This allows the possible use of standard variational procedures, even in the extended space $\boldsymbol{H}^{(\nabla)}(\boldsymbol{C})$.

Not only the Hamilton operator but also the elements of

the extended Hilbert space could be written as a diagonal matrix, as has been shown in equation (20). So, the system energy in equation (30) can also be written from this alternative point of view as a trace of a diagonal matrix:

$$E = \langle\Phi|\boldsymbol{H}|\Phi\rangle = \left\langle Diag\left(\langle\Psi|V|\Psi\rangle; \tfrac{1}{2}\langle\nabla\Psi|\nabla\Psi\rangle\right)\right\rangle$$

using the symbol previously described in definition 4.

The same ideas can be extended to other operators, which can substitute the gradient and they can be used in the same manner in an ASA framework [138,139].

4. Discrete QO Representations

If studied from the computational side, quantum chemistry possesses an essentially discrete numerical structure. Or, at least one can say that quantum chemical descriptions are based on a mixture of continuous functions and discrete coefficients, with a heavy emphasis on the latter. This fact poses no great problem, as discontinuity in space should not be a constraint as pointed out by Dedekind [152]. The quantum description of molecular systems ought to be necessarily associated with this discrete environment. The origin of this situation must be found in the approximate nature of the Schrödinger equation solutions for atomic and molecular systems. In this section, the possible issues of this fact will be discussed. The mathematical definitions and properties given so far and the ones provided next are directly applicable to molecules, but they can also be effortlessly generalised to any quantum system, nuclei for instance [74,76]. Hence, the repeated use of the QO general concept throughout this work, as mentioned above.

4.1. Quantum Similarity Measures

Suppose a Tagged Set $\boldsymbol{T}$, formed by QO, belonging to an Object Set, M, made up of microscopic systems and taking the Tag Set, $\boldsymbol{P}$, as the collection of the systems DF in a given state and computed within a uniform order for every system, that is, $\boldsymbol{T} = M \times \boldsymbol{P}$. Then, choose a PD operator, Ω, provided with the appropriate homogeneous dependence on the Tag Set DF coordinates. The following definition can be used to describe QSM [49,52,70].

Definition 5: Quantum Similarity Measures.

Suppose a known Quantum System set M, and a chosen DF Tag Set $\boldsymbol{P}$. A QSM, $Z(\Omega)$, weighted by a PD operator Ω, is an application of a Quantum Object Tagged Set, $\boldsymbol{T} = M \times \boldsymbol{P}.$, direct product: $\boldsymbol{T} \otimes \boldsymbol{T}$, into the PD real field, $\boldsymbol{R}^+$, such as $Z(\Omega) : \boldsymbol{T} \otimes \boldsymbol{T} \rightarrow \boldsymbol{R}^+$.

4.1.1. Some QSM Forms

In practice, this can be translated into an integral measure computation involving two QO $\{A,B\} \in \boldsymbol{T}$:

$$\boldsymbol{z}_{AB}(\Omega) = \iint \rho_A(\boldsymbol{r}_1)\Omega(\boldsymbol{r}_1,\boldsymbol{r}_2)\rho_B(\boldsymbol{r}_2) d\boldsymbol{r}_1 d\boldsymbol{r}_2 \in \boldsymbol{R}^+ \qquad (31)$$

where $\{\rho_A, \rho_B\} \in \boldsymbol{P}$, are the respective Tag Set DF of the involved QO pair. The Tag Set DF can be considered here in a very broad sense, following the previous discussion on the possible extension of the DF concept. This form, as presented in equation (31), is the one currently quoted in the literature [44, 46, 97-107]. The values of the integral $\boldsymbol{z}_{AB}$ are always PD and real: they are the result of using all the integrand elements as PD functions or operators. In equation (31), the QSM can be interpreted as a weighted scalar product between the DF associated with the involved QO. When both QO, {A,B}, in equation (31) are the same, the QSM is called a Quantum Self-Similarity Measure (QS-SM). QS-SM can be considered nothing else than a norm of the involved DF, as well as equation (31) can be considered a scalar product. Finally, the DP nature of all the involved integrands, providing the structure of a measure to this kind of integrals, can also be marginally interpreted as some kind of generalised molecular volume.

The usual choice for the PD weight operator in equation (31) has been the Dirac delta function $\delta(\boldsymbol{r}_1 - \boldsymbol{r}_2)$. This transforms the general QSM definition into the so-called Overlap-like QSM:

$$\boldsymbol{z}_{AB}(\Omega) = \iint \rho_A(\boldsymbol{r}_1)\delta(\boldsymbol{r}_1,\boldsymbol{r}_2)\rho_B(\boldsymbol{r}_2)d\boldsymbol{r}_1 d\boldsymbol{r}_2 = \int \rho_A(\boldsymbol{r})\rho_B(\boldsymbol{r})d\boldsymbol{r} \quad (32)$$

A choice of a third DF tag as the PD weight operator: $\Omega(\boldsymbol{r}_1, \boldsymbol{r}_2) \equiv \rho_c(\boldsymbol{r})$, transforms the general definition (31) into a Triple QSM [0]:

$$\boldsymbol{z}_{AB;C} = \int \rho_A(\boldsymbol{r}) \, \rho_C(\boldsymbol{r})\rho_B(\boldsymbol{r})d\boldsymbol{r} \quad (33)$$

and in the same way multiple QSM can be defined [47,49,51,69].

4.1.2. Coulomb energy as a QS-SM
However, other possibilities are open to the QSM definition, reverting in the end to a formal structure such as the one appearing in Definition 5. This may be illustrated by the expectation value of the Coulomb energy for a p-particle system, which may be written, employing equation (1), as:

$$\boldsymbol{R} = (\boldsymbol{r}_1, \boldsymbol{r}_2, ... \boldsymbol{r}_p) \wedge \Omega(\boldsymbol{R}) = \sum_{i<j} \boldsymbol{r}_{ij}^{-1} : \langle \boldsymbol{C} \rangle = \int \Omega(\boldsymbol{R})\rho(\boldsymbol{R})d\boldsymbol{R} > 0 \quad (34)$$

where the density matrix is evaluated using the generating rule (3) over the total particle coordinates. It is well known that this produces, for example, in the framework of the MO theory closed-shell mono-configurational case an expression, where Coulomb $\{\langle ii|jj \rangle\}$ and exchange $\{\langle ij|ij \rangle\}$ two-electron integrals play a leading role [154]:

$$\langle \boldsymbol{C} \rangle = \sum_i \sum_j (2\langle ii|jj \rangle - \langle ij|ij \rangle) > 0 \quad (35)$$

Although both parts of the expression can be used by themselves as self-similarity measures [43] over MO's, the positive definite nature of the molecular quantum Coulomb energy, as a whole, can be considered such a similarity measure too. The same can be said when observing the

multi-configurational equivalent of equation (35), see for example [154]. This opens the way to the potential use as a molecular descriptor of this self-similarity measure, which customarily appears computed in the available quantum chemical programs. In equation (35) a negative sign is present, which can be associated to the determinantal structure of electronic wavefunctions, a consequence of Pauli's exclusion principle.

4.1.3. Extension of QSM
As was pointed out above, Coulomb energy may also be seen as a QSM if we adopt another point of view. Equation (34) can be used as starting point. After this, using equation (31) and considering A=B, a QS-SM involving a square DF, may be rewritten as:

$$z_{AA}^{(2)}(\Omega) = \int \Omega(\boldsymbol{R})\rho_A^2(\boldsymbol{R})d\boldsymbol{R} \quad (36)$$

Thus, nothing stops this last second-order DF integral from being generalised to an n-th order DF form:

$$z_{AA}^{(n)}(\Omega) = \int \Omega(\boldsymbol{R})\rho_A^n(\boldsymbol{R})d\boldsymbol{R} \quad (37)$$

in this manner, a first order QS-SM form could easily be written as a particular case:

$$z_{AA}^{(1)}(\Omega) = \int \Omega(\boldsymbol{R})\rho_A(\boldsymbol{R})d\boldsymbol{R} \quad (38)$$

Then, if the Ω operator structure as given in equation (34), is used in equation (38), it is easy to see that: $Z_{AA}^{(1)}(\Omega) = \langle \boldsymbol{C} \rangle$

Continuing the above discussion and definitions, in Section 4.1.4., a generalised QSM structure might be constructed but this will not be discussed here. The interested reader is referred to references [47,49,86,91,92].

4.2. Discrete Representation of Quantum Objects: Similarity matrices
4.2.1. Theoretical considerations
The possibility of obtaining multiple relationships between the appropriate number of QOS elements, via their DF tags, in terms of QSM, as discussed in the previous section, gives rise to other interesting consequences, besides the calculation of the possible relationship between QOs. The most relevant one constitutes the potential representation of a QO as a discrete vector or matrix.

Definition 6 : Similarity Matrices.

Suppose a QOS: $\boldsymbol{T}=M \times \boldsymbol{P}$ of cardinality n. The symmetric $(n \times n)$ matrix: $\boldsymbol{Z} = \{z_{IJ}\}$ whose elements are computed using QSM between pairs of QO in $\mathbf{T}$, will be called a Similarity Matrix (SM).

By construction, provided that all the involved QO are different, any SM could be considered a PD metric matrix, belonging to some matrix VSS: $\boldsymbol{Z} \in M_{(n \times n)}(\boldsymbol{R}^+)$ [51]. Such a matrix

can also be interpreted as the representation of the PD operator, Ω, in the basis set defined by the QO DF. Considering the SM column vectors: $\boldsymbol{Z} = \{z_I\}$, this set also belongs to some n-dimensional VSS: $\forall I : \boldsymbol{z}_I \in V_n(\boldsymbol{R}^+)$. Moreover, every column, Z_I, of the SM can be associated to an n-dimensional discrete representation of the I-th QO in the Tagged QOS. The set of columns of the SM was also referred to in earlier papers [44-48], in an obvious descriptive manner as a Molecular Point Cloud.

Discrete mathematics can be of much help in the description of Tagged Sets and the relationships of their elements, several aspects of the basic information on this subject can be found in reference [87].

4.2.2. Similarity Matrices and Carbó Index

The overall relationships between the elements of a molecular data set can be expressed in matrix form, yielding the SM. Any SM is symmetric, indicating that the QSM between two molecules is identical independently of the order of the comparison of the QO. The order of magnitude of the different types of MQSM is highly connected to the structural form of the molecule, and to the presence of heavy atoms. Due to the particular construction of the SM, the diagonal elements of SM bring out information on the size of the QO.

Several transformations of SM can be performed, yielding the so-called molecular QS indices (MQSI), which scale or normalise the SM. In particular, a normalisation of the MQSM, known in the literature as the Carbó index, [1] can be defined as:

$$C_{IJ} = Z_{IJ}\left(Z_{II}Z_{JJ}\right)^{\frac{-1}{2}} \qquad .$$

The overall set of Carbó indices can also be expressed in a matrix form. The Carbó index can be interpreted as the cosine of the angle subtended by both involved DF in ∞-dimensional space, and so it ranges from zero to one. When the Carbó index is closer to one, more similar can be considered both QO. That is, the similarity between the two QOs increases as the index number approaches one. Therefore, for two identical QO, that is, the main diagonal Carbó index elements, the value of one is found, irrespective of the analysed QO. More details on MQSI can be found in references [41-42, 60].

5. Discrete representations and QSAR

Discrete QO Tagged Sets may be defined at the same time as the original ∞-dimensional ones. Suppose a QO Tagged Set, $\boldsymbol{T}=M\times\boldsymbol{P}$, is known, and a Similarity Matrix, $\boldsymbol{Z}=\{z_I\}$, considered as a hypermatrix, is formed by column vectors as elements, evaluated using the Definition 6 procedure.

Definition 7 : Discrete Quantum Object Sets

A Discrete Quantum Object Set can be constructed as a new Tagged Set: $\boldsymbol{Z}=M\times\boldsymbol{Z} \wedge \boldsymbol{Z}=\{z_I\}$, with the same Object Part as the original QOS Tagged Set, $\boldsymbol{T}$, but with the Tag Part formed by the columns of the Similarity Matrix, $\boldsymbol{Z}$.

5.1. Discrete Expectation values

The point of view appearing in Definition 7 is the same as finding out the way to project a set of points, defined in some ∞-dimensional VSS, into an n-dimensional vector structure. Any Similarity Matrix collecting a similarity relationship between any studied QO and a set of parent QO structures is a source of an *unbiased* QO representation in the form of n-dimensional discrete information. SM columns are to be considered molecular descriptors chosen in such a way that the remaining arbitrariness of choice corresponds to the nature of the weight operator appearing in the MQSM calculation.

Obviously, from this point of view discrete QO representations will depend on the PD weight operator, Ω. Suppose that a collection of Discrete QO Tagged Sets is formed using various PD operators, $\{\Omega_i\}$, producing a collection of PD SM: $\{\boldsymbol{Z}(\Omega_i)\}$. A new PD SM, can be obtained, choosing a set of PD scalars, $\{\alpha_i\}\in\boldsymbol{R}^+$, by forming the linear combination:

$$\boldsymbol{Z} = \sum_i \alpha_i\boldsymbol{Z}(\Omega_i) \qquad .$$

The columns of the combined SM, $\boldsymbol{Z}$, can be considered linear combinations with the same coefficients, as those employed to construct the new PD SM. That is:

$$\forall \boldsymbol{z}_I \in \boldsymbol{Z}: \boldsymbol{z}_I = \sum_i \alpha_i z_I(\Omega_i) \qquad .$$

This discrete kind of QO elements can be considered as a source of descriptors, which can be of use in the field of QSAR or QSPR. It has recently been shown how QSM can be used as the origin of multilinear QSAR [60,86]. This can be made by associating a given QO property, π, to the expectation value of some unknown operator in the way of equation (1). Taking into account that the DF and the unknown PD operator belong to the same VSS, then both can possess an associated discrete representation in the appropriate n-dimensional VSS. In this framework, equation (1) adopts the discrete counterpart form:

$$\pi = \langle\omega\rangle \approx \boldsymbol{w}^T\boldsymbol{z} = \sum_k w_k z_k \qquad (39)$$

Where $\boldsymbol{w}$ is an n-dimensional vector attached to the unknown operator, to be determined, in a least-squares sense, and $\boldsymbol{z}$ is an n-dimensional Discrete QO Tag. Equation (39) can be called the fundamental QQSAR equation. In section 5.4, the structure of the fundamental QQSAR equation will be studied in a general manner.

This is not surprising if the origin of equation (39) is taken into account, as given by the ∞-dimensional counterpart in equation (1). Indeed, both equations represent scalar products in conveniently chosen VSS.

It should be noted that QSM, collected as a vector, $\boldsymbol{z}$, associated with a molecular structure, can be considered from a quantum mechanical point of view, as an ultimate way of representing quantum systems in a discrete manner. Empir-

ical QSAR parameters, whatever their origin and number, are considered as more or less successful attempts to simulate such QSM vectorial description. In light of equation (39) and the following in-depth discussion of the QSPR problem, the QSM vectors cannot be considered as another set of molecular descriptors, chosen in the usual arbitrary way. Rather, they are obtained as a result of analysing the final consequences of Quantum Mechanics, applied to the description of QO, as atoms and molecules. It has been shown that even the origin of topological matrices can be traced to similarity matrices. Thus, topological indices can be constructed from the similarity measures contained in them [91].

It can be concluded that any attempt to describe a known QO by using an arbitrary number of parameters of any kind, other than those computed using QSM, should be considered as a rough way to simulate the theoretically correct QO descriptors, represented, in turn, by the discrete QSM vectors $\mathbf{z}$. From this point of view, the usual QSPR techniques, employed in chemistry for more than a hundred years, ought to be accepted as an empirical procedure to obtain approximate expectation values within a discrete framework.

5.1.2. Simple linear QSPR model involving QS-SM
For a given QO in a studied discrete Tagged QOS, equation (39) can be written as:

$$\pi_I \approx \mathbf{w}^T \mathbf{z}_I = \sum_K w_K z_{KI} \qquad (40)$$

Equation (40) can be rewritten isolating the self-similarity part, Diag($\mathbf{Z}$)={Z_{II}}, which is coincident with the diagonal of the SM:

$$\pi_I \approx \mathbf{w}_I z_{II} + \sum_{K \neq I} w_K z_{KI} \qquad (41)$$

The terms: $a = w_I$, $b = \sum_{K \neq I} w_K z_{KI}$, can be considered as constants, or at least varying slowly within a QO homogeneous series, made up of molecular structures, for example. It is then not difficult to see that a linear relationship may be present between some properties and self-similarity:

$$\pi_I \approx a z_{II} + b \qquad (42)$$

This kind of relationship has been used successfully to assess some QSAR and QSPR, see for example [97-105].

This simplified equation and the discussion in the previous section 4.1.2 about Coulomb two-electron energy as some sort of QS-SM can be used together [97-101,106,107]. They permit us to consider that these expectation values, as well QS-SM themselves, constitute very good candidates for competing, under some favourable circumstances within homogeneous molecular series, with empirical parameters, like Hammet σ constant or log$P$, the octanol-water partition coefficient.

5.1.3. Convex superposition of SM
Suppose, now, in another possible QSAR situation, that a discrete QOS is known. The associated discrete Tag Set: $\{z_I\}$ can be constructed by a convex combination of several appropriate elements belonging to the same Tag Set, in such a form that:

$$\mathbf{z}_I = \sum_k \alpha_k Z_I(\Omega_k) = \sum_k \alpha_k \mathbf{Z}_I(k) \wedge K_n(\{\alpha_k\}) \qquad (43)$$

Thus, using equation (39), for every QO property the following relationship holds:

$$\pi_I \approx \mathbf{w}^T \left(\sum_k \alpha_k z_I(k) \right) = \sum_k \alpha_k \left(\mathbf{w}^T \mathbf{z}_I(k) \right)$$
$$= \sum_k \alpha_k \pi_I(k) \qquad (44)$$

Where, $\pi_I(k) \equiv \pi_I(\Omega_K)$ represents the estimated value of the considered property obtained for the I-th QO, using the weight operator Ω_K.

5.2. Quantum QSAR
Within the QSM framework, QSAR can be renamed as *Quantum* QSAR (QQSAR), in order to distinguish between the models obtained by using SM or SI elements as descriptors, as previously discussed from the empirical ones obtained employing, usually for molecules, other kinds of parameters coming from several heterogeneous origins. The reason for this is simple: QQSAR are obtained from the considerations that a QSAR model, relating Discrete QO Sets (DQOS) elements with some experimental property, constructed within a discrete quantum framework like the one described in QSM theoretical background, becomes completely equivalent to the evaluation of an expectation value [60,86] as in a typical quantum mechanical measuring procedure, as was pointed out in section 5.1. Here, a deeper insight into the problem will be provided.

5.2.1. Expectation values
Indeed, when studying a quantum mechanical system, the expectation value of some observable ω, in a well-defined QO system state, can be written by means of the integral [138] equivalent to equation (1):

$$\langle \omega \rangle = \int W(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} \qquad (45)$$

where $W(\mathbf{r})$ is an associated Hermitian operator to be determined for a given QOS, and $\rho(\mathbf{r})$ the system state density function.

On the other hand, the expression (45), from a QQSAR point of view can be interpreted as a scalar product, that is:

$$\langle \omega \rangle = \langle W | \rho \rangle \qquad (46)$$

Taking into account the unknown nature of the QQSAR operator $W(\mathbf{r})$, one can consider that it can be decomposed as a product of the operator leading to the expectation value: $W_\omega(\mathbf{r})$, still unknown and to be determined, by a known positive definite weight operator, $\Omega(\mathbf{r}, \mathbf{r}_0)$. Thus, the expecta-

tion value as presented in expression (45) can now be written as:

$$\langle\omega\rangle = \langle W_\omega|\Omega|\rho\rangle \tag{47}$$

This is the same as to transform equation (45) into the equivalent of the more general integral:

$$\langle\omega\rangle = \int W_\omega(\boldsymbol{r})\Omega(,\boldsymbol{r},\boldsymbol{r}_0)\rho(\boldsymbol{r}_0)d\boldsymbol{r}\, d\boldsymbol{r}_0 \tag{48}$$

It is easy to see that the positive definite weight operator $\Omega(\boldsymbol{r},\boldsymbol{r}_0)$ can particularly be chosen as Dirac's delta function: $\delta(\boldsymbol{r}-\boldsymbol{r}_0)$, and, in so doing equation (45) is recovered. In order to distinguish the expectation value general definition, as presented in equation (48) from the usual choice in equation (45), where the weight can be considered a unit operator, the general integral (48) can be named as a *weighted expectation* value integral expression.

5.2.2. Fundamental Quantum QSAR equation
When studying a known QOS DF tag set: $\{\rho_I(\boldsymbol{r})\}$, and if a particular QOS element, A say, is chosen, then equation (45) can be written as:

$$\langle\omega_A\rangle = \int W(\boldsymbol{r})\rho_A(\boldsymbol{r})d\boldsymbol{r} \tag{49}$$

Now, whenever both operator and density function in the integral (45) can be considered to belong to the same VSS, and expressing approximately the unknown operator, $W_\omega(\boldsymbol{r})$, as a linear combination of the QOS tag set, acting as a basis:

$$w_\omega(\boldsymbol{r}) \approx \sum_I w_I \rho(\boldsymbol{r}) \tag{50}$$

a new relationship is obtained after substituting expression (50) in equation (49) while taking into account the decomposition (48):

$$\langle\omega_A\rangle = \sum_I w_I \iint \rho_I(\boldsymbol{r})\Omega(\boldsymbol{r},\boldsymbol{r}_0)\rho_A(\boldsymbol{r}_0)d\boldsymbol{r}\, d\boldsymbol{r}_0 = \sum_I w_I z_{IA}(\Omega) \tag{51}$$

which is equivalent to the proposed equation (39). Admitting equation (51) holds for every QO, and that the SM, $\boldsymbol{Z}$, obtained employing the weight operator $\Omega(\boldsymbol{r},\boldsymbol{r}_0)$, is symmetric, then an equivalent equation can be written in matrix form as:

$$\boldsymbol{Z}\boldsymbol{w} = \pi \tag{52}$$

Where: π is a column vector containing the expectation values or the QO property of interest and w is another column vector, containing the unknown coefficient set $\{w_I\}$ of the operator expression (50). In the light of all the discussed aspects of the problem, it seems that equation (52) can from now on be termed the *fundamental* QQSAR equation.

The QQSAR problems of type (52) are associated with the unknown operator, $W_\omega(\boldsymbol{r})$, supposedly formed by many entangled terms, whose coefficient coordinates, computed

with respect to the QOS DF tags acting as a basis set, should be determined from the knowledge of the associated problem expectation values. The usual classical procedure to solve the problem could be based on some algorithm related to the well-known least-squares procedure.

This close relationship between expectation values in continuous and discrete quantum frameworks cannot be so easily deduced as a theoretical consequence, outside of the QSM models of QSAR. So, the term QQSAR is completely justified and could be applied from now on to QSAR models, obtained by means of quantum similarity measures or indices. Several application papers have recently appeared in the literature [97-107] showing how QSM can be successfully used within this QQSAR formalism. The SM appearing in the fundamental QQSAR equation (52) can be constructed in several ways, even using diverse weighting operators when computing the QSM, transforming the SM into SI or even gathering several different SM in convex combination [82], as has been already discussed in section 5.

5.3. QQSAR modeling
The procedure used here for dealing with the QQSAR model, is based on inward matrix products and the definite positive nature of the quantum SM or SI, which can be employed to construct it. The possibility of transforming the experimental activity data into a positive definite vector is also taken into account because many properties possess intrinsically positive definite values or can be scaled and origin shifted to fulfill this property. The whole problem has been studied with detail from various points of view [78,89], so it is merely outlined here.

As discussed in the previous section, the QQSAR models can be written as a matrix equation like expression (52), involving the chosen SM and the QOS experimental data to be described. All known matrices belong to some VSS of the appropriate dimension. In order to keep the model describing positive definite property values, the solution could be forced to be an element of some VSS too. So, using the symbol: $\boldsymbol{A}* > 0$, to indicate that a chosen matrix $\boldsymbol{A}$ has all its values defined in $\boldsymbol{R}^+$, that is, to represent in a shorter form the property: $\boldsymbol{A} = \{a_{ij}\} \rightarrow \forall i,j : a_{ij} \in \boldsymbol{R}^+$. It is not compulsory for this matrix to be symmetric, it will only be necessary for it to be diagonalizable. This mathematical structure also covers the possibility of dealing with stochastic similarity matrices, as will be discussed in the next section. In general, the constrained QQSAR model can be written:

$$\boldsymbol{Z}* > 0 \wedge \pi* > 0 : \boldsymbol{Z}\boldsymbol{w} = \pi \rightarrow \boldsymbol{w}* > 0 \tag{53}$$

5.4. Fundamental QQSAR equation approximate solution
In order to take into account the implication, which can be considered as a constraint, imposed on the solutions to be found when solving the equation (53) linear system, an algorithm can be easily designed as follows.

Not long ago, Sen and one of us [77] suggested the possible use, in theoretical quantum chemistry applications and

subsequent mathematical developments, of an elementary matrix operation: the *inward matrix* product. Such product is present, within the usual Fortran 95 compilers, as a set of intrinsic procedures [78], among a related collection of matrix manipulation functions.

The inward matrix product (IMP) between two matrices, **A** and **B**, bearing the same arbitrary dimension, is defined without problems as another matrix, **C**, with the same dimension form. Using n×m matrices as a typical quite general example, the following straightforward algorithm can be designed for the inward product definition:

$$\mathbf{C} = \mathbf{A} * \mathbf{B} \rightarrow \forall i, j : c_{ij} = a_{ij} b_{ij} \tag{54}$$

Inward products possess a standard set of properties, which have already been described in detail [77,89] and will not be repeated here.

It is obvious that if the restriction $\mathbf{x} * > 0$ on the vector **w** holds, there then exist three real matrices, **T**, **x** and **p**, which can be computed as the inward square root of the system matrices, that is, $\mathbf{T} = \mathbf{Z}^{\left[\frac{1}{2}\right]}$; $\mathbf{x} = \mathbf{w}^{\left[\frac{1}{2}\right]}$; $\mathbf{p} = \pi^{\left[\frac{1}{2}\right]}$, in the same fashion as this inward matrix power was implicitly defined in equation (54). So, equation (53), bearing the corresponding constraint on the unknowns' vector can be written as:

$$(\mathbf{T} * \mathbf{T})(\mathbf{x} * \mathbf{x}) = (\mathbf{p} * \mathbf{p}) \tag{55}$$

This suggests the possibility of constructing an alternative approximate system. Indeed, merely by exchanging the sites of classical and inward matrix products on the left of equation (55) we obtain:

$$(\mathbf{Tx}) * (\mathbf{Tx}) = (\mathbf{p} * \mathbf{p}) \tag{56}$$

leading to the approximate reduced system:

$$\mathbf{Tx} = \mathbf{p} \tag{57}$$

requiring no restriction on the unknown vector elements and thus finally proving solvable, considering the approximate nature of the solution yielding the QQSAR model.

The way chosen here to achieve an approximate solution of the fundamental QQSAR equation takes the following path: First, the eigensystem of the symmetric matrix **T** is obtained, and an approximate spectral decomposition of this matrix is used employing a cutoff value, ε, on the eigenvalues, in order to get rid of numerical noise. This can be written explicitly using a logical Kronecker's delta [52] as:

$$\mathbf{T}^{-1} \approx \sum_i \delta(\tau_i > \varepsilon) \tau_i^{-1} |i\rangle\langle i| \tag{58}$$

where $\{\tau_i\}$ is the spectrum and $\{|i\rangle\}$ the eigenvector system of the matrix **T**. The same can then be done for the evaluation of an approximate inverse:

$$\mathbf{Z}\mathbf{D}^{-1}\mathbf{D}\mathbf{w} = \pi \rightarrow \mathbf{S}^T\mathbf{v} = \pi \tag{59}$$

Finally, the approximate solution can be written using:

$$\mathbf{x} \approx \mathbf{T}^{-1}\mathbf{p} = \sum_i \delta(\tau_i > \varepsilon) \tau_i^{-1} |i\rangle\langle i|\mathbf{p}\rangle \tag{60}$$

so, the approximate solution of the original system (53), $_a\mathbf{w}$, with the adequate constraint, is simply computed using the inward product: $_a\mathbf{w} = \mathbf{x} * \mathbf{x}$ is sufficient to obtain a set of estimated property values:

$$_a\pi = \mathbf{Z}(_a\mathbf{w}) \tag{61}$$

As discussed earlier, this algorithm can be employed over any constrained linear system problem, so the matrix **Z**, may be used directly, and so can any symmetric manipulation of the stochastic transformations, like the one present in section 6.5.

6. Stochastic Transformations of Quantum Similarity Matrices

This section will deal with some application of IMP ideas, among other problems, simultaneously employing the underlying statistical nature of QSM, and outlining the definition of a stochastic transformation of quantum SM. As a consequence of these two possibilities, a new set of QSI will emerge. They can be directly used as discrete quantum object descriptors in classical QSAR studies or can be employed for other purposes: quantum object classification, for instance. Moreover, partial order structure on quantum object sets appears to be easily associated with the asymmetric stochastic matrix structure. Finally, because of the connection of similarity matrices with the fundamental QQSAR equation (39), such a stochastic transformation of any similarity matrix can be shown to play a quite interesting role in obtaining QQSAR models by using, in part, inward matrix product algebra.

Some previous questions should be carefully analyzed before this prospect is developed. In the following sections several points related to the field of QQSAR will be discussed and their connection with classical QSAR procedures studied. At first sight both quantum and empirical QSAR procedures seem to be the variants of the same theoretical structure. However, while keeping clear the interest of the authors for *both* alternatives, they feel that it must be stressed the conceptual differences of both classical and quantum QSAR approaches, present over the respective scientific contents of each procedure.

The computation of SM over a QOS providing a new DQOS structure, as discussed above, produces a set of N-dimensional tags, which can be associated to the original infinite dimensional DF tags. Despite the strict positive definiteness of the SM column set elements, $\{z_i\}$, which appear as a consequence of the QSM definition presented in equation (31), the connection between the *n*-dimensional tags and the DF is not immediately evident. However, it can be deduced after taking into account the nature of the involved

DF tags, which can be considered in turn either as positive definite functions or projection operators. The present section will present an SM transformation, producing a new collection of n-dimensional tags. These are provided with a structure that indicates that they form a discrete probability distribution. This possibility should be expected as a plausible outcome of QSM theory, due to the quantum mechanical origin of all the QO tags employed so far.

N-dimensional quantum SM columns or rows, $\{z_I\}$, or simply, their elements, are made, by construction, of positive definite real, in computational practice rational, numbers. This characteristic property can be summed up by saying that the set of the SM columns or rows, the molecular point cloud, belongs to a *Vector Semispace* (VSS): $\{z_I\} \subset V_N(\mathbf{R}^+)$ as defined in section 2.3 above, see also references [87,88]. Note, for example, that the QO point cloud, defined by the DQOS tag set elements and mentioned in the previous section, defines a characteristic set of points in some N-dimensional VSS since, due to the nature of the QSM definition, all their components are made up strictly of positive numbers.

The VSS structure in general and, in particular, the construction of SM precludes that the sum of every SM row (or column) elements is a positive real number, for example: $\langle z_I \rangle = \sum_J z_{IJ} \in \mathbf{R}^+$. These sums can be used as row (or column) scale factors in order to trivially obtain a new row (or column) set belonging to the same VSS, that is, $s_I = \langle z_I \rangle^{-1} z_I$, but possessing the imposed form of a discrete probability distribution. That is, the following equalities can easily be written:

$$\langle s_I \rangle = \left\langle \langle z_I \rangle^{-1} z_I \right\rangle = \langle z_I \rangle^{-1} \langle z_I \rangle = 1 \qquad .$$

The set of N rows: $\mathbf{S} = \{s_I\}$, ordered forming a square (N×N) matrix $\mathbf{S}$, produce a nonsymmetric *stochastic matrix* [69] as a result.

A trivial compact way to produce the stochastic row matrix may be described by first constructing the diagonal matrix, $\mathbf{D}$, whose elements are made up of the sums of row (or column) SM elements:

$$\mathbf{D} = Diag(\langle z_1 \rangle, \langle z_1 \rangle, ..., \langle z_N \rangle) \qquad (62)$$

producing the matrix product: $\mathbf{S} = \mathbf{D}^{-1}\mathbf{Z}$. In the same manner, a column stochastic matrix will be defined straightforwardly by the transpose of the previous definition: $\mathbf{S}^T = \mathbf{Z}\mathbf{D}^{-1}$, where one must take into account the symmetrical structure of the original SM.

The row $\{\langle s_I |\}$ or column $\{|s_I\rangle\}$ stochastic sets, being associable to a collection of discrete probability distributions, may be even better connected to the DF tag set: $\{\rho_I\}$ of the original QOS, than the rows or columns of the attached SM. In fact, they can be used as a tag set to produce an attached tagged set, which can be called a *Discrete Stochastic QOS* (DSQOS), when combined with the microscopic systems belonging to the object set: $\{\omega_I\}$ of the original QOS.

For instance, taking into account the same considerations as those used before when previously discussing the nature of the SM rows, the connection between the original QOS elements with the stochastic matrix rows: $(\omega_I; \rho_I) \leftrightarrow (\omega_I; \langle s_I |); \forall I$, defines the elements of a DSQOS.

6.1. Inward Symmetrisation of Stochastic Quantum Similarity Matrices and Stochastic Quantum Similarity Indices

The elements of the DSQOS tag set can be directly used as QO discrete descriptors, admitting the actual implications: $\forall I; \omega_I \leftrightarrow \rho_I \leftrightarrow \langle s_I |$, in a manner like that previously discussed, when dealing with the significance of the SM elements. The sole problem lies in the fact that, unlike the SM, $\mathbf{Z}$, the row stochastic matrix, $\mathbf{S}$, is not symmetric. However, this does not constitute a restrictive problem when alternative manipulations of the DSQOS are envisaged, and even has other possible uses, not contained in symmetric QO descriptor structures.

A *partial order* [80] over this kind of DSQOS can easily be associated to this characteristic mathematical matrix structure. Thus, the possibility of constructing the stochastic matrix $\mathbf{S}$ from a trivial manipulation of the SM, $\mathbf{Z}$, generally completes the QSM theory in an elegant way, providing QOS with a partial ordering formalism. Because of the usual symmetric structure of the associated DQOS SM, the possibility of transforming DQOS into partially ordered sets, posets, was still absent from quantum similarity theoretical lore. Thus, QOS can be transformed into DQOS, and even more easily into DSQOS: there a poset structure can be studied and employed, due to the unsymmetrical structure of the tag set stochastic matrix.

6.2. Inward Matrix Product Symmetrisation

In addition to the classical symmetrisation techniques, which customarily use the sum or classical product of the matrix and its transpose, there can also be described a simpler algorithm, involving an IMP.

It is easy to see that the commutative IMP:

$$\mathbf{R} = \mathbf{S}^T * \mathbf{S} = \mathbf{S} * \mathbf{S}^T \qquad (63)$$

of the stochastic matrix $\mathbf{S}$ with its transpose, $\mathbf{S}^T$, produces a symmetric matrix as a result. This can be easily demonstrated because the following relationship, using the first equality of equation (63), holds for all the elements of the matrix $\mathbf{R}$:

$$\forall i, j: r_{ij} = s_{ij}^{(T)} s_{ij} = s_{ji} s_{ij} = s_{ij} s_{ji} = s_{ji}^{(T)} s_{ji} = r_{ji} \qquad (64)$$

that is, $\mathbf{R} = \mathbf{R}^T$. The same result is obtained using the second part of the equation (63).

6.3. Stochastic Similarity Indices

The symmetric matrix $\mathbf{R}$ could be used in the same way as the original SM, but for unit homogenisation purposes with respect to the initial SM, $\mathbf{Z}$, it is preferable to employ the inward square root of the symmetric inward product (63), defined by the algorithm:

$$\mathbf{Q} = \mathbf{R}^{\left[\frac{1}{2}\right]} \rightarrow \forall i, j: q_{ij} = \sqrt{r_{ij}} \qquad (65)$$

One can consider the new symmetric matrix $\boldsymbol{Q}$ as holding *Stochastic Quantum Similarity Indices* (SQSI). In Fortran 95, all the operations starting with the known SM, $\boldsymbol{Z}$, and leading to computation of the matrix $\boldsymbol{Q}$, can be written in an extremely simple and short code.

6.3.1. Relationship between stochastic similarity indices and the Carbó index.

It is worthwhile, at this point, to consider how the new kind of stochastic similarity indices may be related to previous well-defined similarity indices. These, as what we have called the Carbó index [1] (defined in section 4.2.2) have been present in the literature for some time. An expression of the elements of matrix $\boldsymbol{Q}$ in terms of the original symmetric SM, leads to the equality sequence:

$$q_{ij} = \sqrt{r_{ij}} = \sqrt{s_{ij}s_{ji}} = \sqrt{\frac{z_{ij}z_{ji}}{\langle \mathbf{z}_i \rangle \langle \mathbf{z}_j \rangle}} = \frac{z_{ij}}{\sqrt{\langle \mathbf{z}_i \rangle \langle \mathbf{z}_j \rangle}} \tag{66}$$

which resembles the Carbó similarity index continuous form, that is, the index defined over the involved QO DF tag couples, using the associated SM elements, which is written as:

$$K_{ij} = \frac{z_{ij}}{\sqrt{z_{ii}z_{jj}}} \tag{67}$$

If the Carbó index (67) is expressed in a hybrid continuous-discrete form, by means of the norms of the SM columns (or rows):

$$\theta_{ij} = \frac{z_{ij}}{\sqrt{\langle \mathbf{z}_i | \mathbf{z}_i \rangle \langle \mathbf{z}_j | \mathbf{z}_j \rangle}} \quad ,$$

then, as a final alternative, the completely discrete form of the Carbó index could even be written using a scalar product of SM columns (or rows) as:

$$\eta_{ij} = \frac{\langle \mathbf{z}_i | \mathbf{z}_j \rangle}{\sqrt{\langle \mathbf{z}_i | \mathbf{z}_i \rangle \langle \mathbf{z}_j | \mathbf{z}_j \rangle}} \quad .$$

This discussion leaves incomplete the possible ways any SM can be transformed into a similarity index matrix. When discussing below the possible transformations of the fundamental QQSAR equation, a related index to the stochastic similarity one described in equation (66) will appear as a natural consequence of the stochastic scaling of the SM.

6.4. Quantum QSAR and Stochastic Quantum Similarity Indices

In this section the QQSAR fundamental equation (39) is introduced as a linear system, deducible solely from employing quantum principles, based in turn on the concept of expectation value. It must be noted that, contrary to the usual quantum mechanical procedure for obtaining observable expectation values of a given system, which deals with a unique QO, within the QQSAR theoretical background, an expectation value evaluation has to be associated with a given QOS of arbitrary cardinality.

6.5. Stochastic Transformations over fundamental QQSAR equation

Starting from the fundamental QQSAR equation (52), in addition that a straight solution is sought, various equation transformations can be performed. Among other possibilities, if a set of stochastic transformations, as discussed in the previous section, has to be taken into account, then the possible manipulations can be readily described as follows.

Using the diagonal matrix (62), which contains the SM row sums as elements, then the following linear system is obtained:

$$\mathbf{D}^{-1}\mathbf{Z}\mathbf{w} = \mathbf{D}^{-1}\pi \rightarrow \mathbf{S}\mathbf{w} = \mathbf{p} \tag{68}$$

Equation (68), related to a column stochastic transformation definition can be written as:

$$\mathbf{Z}\mathbf{D}^{-1}\mathbf{D}\mathbf{w} = \pi \rightarrow \mathbf{S}^T\mathbf{v} = \pi \tag{69}$$

and both can produce a single symmetric transform:

$$\mathbf{D}^{-1}\mathbf{Z}\mathbf{D}^{-1}\mathbf{D}\mathbf{w} = \mathbf{D}^{-1}\pi \rightarrow \mathbf{A}\mathbf{v} = \mathbf{p} \tag{70}$$

Both equations (68) and (69) can be further symmetrised by multiplying both sides by $\boldsymbol{S}^T$ and $\boldsymbol{S}$, respectively. However, in both cases further simplification leads to equation (70).

On the other hand, equation (70) provides a transformed SM with a Carbó index structure, that is:

$$\mathbf{A} = \mathbf{D}^{-1}\mathbf{Z}\mathbf{D}^{-1} \rightarrow a_{ij} = \frac{z_{ij}}{\langle z_i \rangle \langle z_j \rangle} \tag{71}$$

The resulting transformed matrix, $\boldsymbol{A}$, is a symmetric matrix, but fails to correspond exactly to a stochastic matrix as in previous transformations, instead resembling the matrices defined in equations (63) and (65). The transformed unknown and property vectors coincide here with those of equations (68) and (69) respectively, if the original diagonal matrix transformation is used. The matrix $\boldsymbol{A}$, though not bearing such an evident meaning as the row or column stochastic transformed matrices, may constitute an interesting new similarity index matrix. This connection could be more understandable and evident by taking into account the discussion and remarks started when the stochastic similarity index definition in equation (66) was set.

7. Conclusions

A general discussion on the formal structure of QSM and the surrounding problems has been carried out. A limited number of definitions accurately design the notion of QO and the Tagged Set structure of QOS naturally appears. The attached formalism becomes a rich source of new concepts,

such as the KE DF and other DF forms, which are based on extended Hilbert spaces. This formalism leads towards a matrix formulation of energy expectation values. The new density function forms can be described and visualised too, producing useful suggestive pictures of the molecular structure environment.

Discretization of QO descriptions, via general definitions of QSM, also leads to the definition of SM and discrete QOS. The ascription of all the involved Tag Sets to some VSS, allows the analysis of several aspects of QSM as well as of the QSPR. In this way, it can be said that QSM n-dimensional vectors, associated with a given molecular structure, can be considered, from a quantum mechanical point of view, as an unbiased way to discretely represent microscopic systems. In this QSM context, the user has no way to choose arbitrarily the parameters and, in so doing, possibly interact with the resulting model. Empirical QSAR parameters, whatever their origin, should be considered as examples of more or less successful attempts to simulate such QSM vectorial description.

IMP have a leading role both in the definition of the indices and in the construction of algebraic QQSAR models. QQSAR models present a solid quantum mechanical theoretical background and an unbiased general manner of choosing the quantum objects descriptors, while providing equivalent relationships to the ones obtained using classical empirical QSAR.

No linear hypothesis between QO properties and structure need be admitted in QQSAR: a fundamental QQSAR equation appears because of the structure of quantum mechanics. By means of any statistically bound model, even if the final result is able to pass significance tests, no causal relationship between structure and activity is establishable. However, as the fundamental QQSAR equation is derived from pure quantum mechanical grounds, employing the expectation value concept, then a resulting model derived by solving such a fundamental QQSAR equation can be interpreted as bearing a causal relationship between QO structure, represented by the QSM or QSI matrices, and activity, associated with QO experimental properties.

An approximate way to solve linear systems with elements belonging to VSS, while constraining the solutions to also belong to some VSS, has been described. It is essentially based on IMP techniques and mainly corresponds to the nature of approximate possible solutions of the fundamental QQSAR equation.

The abovementioned characteristics are not the sole differences between QSAR and QQSAR. Another important question of diversity between both procedures appears associated to the QO descriptors employed by each technique. QSAR somehow possess complete freedom in choosing only several from among thousands of descriptors. Such QSAR molecular parameters are coming from a wide variety of sources: from purely binary descriptors, passing through myriad topological indices, up to quantum mechanical molecular properties. Also, at present, point networks are used as molecular descriptors, and these are made up of several breeds of fields, computed around the molecular structures.

On the opposite direction, QQSAR descriptors are fixed by the density functions of the molecular structures involved and the weight operators chosen to compute them. When all is said and done, this QQSAR constraint appears to be caused by molecular wavefunction uniqueness, when a given system state is studied. There is no free molecular descriptor choice in QQSAR: Their values become fixed by the quantum mechanical nature of the QO set under study. This may seem a troublesome restriction when considered from a classical QSAR point of view, but it corresponds to the nature of quantum mechanics, which accordingly proposes that all the studied system information should be contained within the attached state density function. Thus, unlike empirical QSAR, QQSAR models are attached to an unbiased descriptor choice, in the sense that the user cannot choose among a large variety of parameters, but such parameters appear as a theoretical consequence of the QO quantum mechanical description.

Another question, related to the last paragraph, concerns the general nature of QSM QO descriptors. The quantum mechanical general information contents of density functions can be used to generate QQSAR parameters. These can afterwards be employed to describe any quantum system: nucleus, atom or molecule. As another consequence of their quantum mechanical origin, these QSM parameters can be used for the description of any observable molecular property. Thus, to end with this somewhat sketchy preliminary description of QQSAR, for any set of QOs, some descriptor set within QSM techniques can always be found to obtain structure-property relationships via a fundamental QQSAR equation model. QQSAR praxis confirms this obvious behaviour of the theory.

None of these considerations destroys the possibility of using QSM as descriptors for classical empirical QSAR procedures. Moreover, considered from this point of view, QSM can be admitted as new parameters to be taken into account at the moment of using the usual classical QSAR procedures.

Stochastic transformation of quantum similarity matrices, and the subsequent optional symmetrisation of the resulting matrix, yields a new set of quantum similarity indices. These can be employed as the similarity matrix itself in order to obtain QQSAR models. Stochastic quantum similarity matrices provide the missing link in quantum similarity measures theory connecting quantum object sets with the structure of partially ordered sets.

Acknowledgements

Some common abbreviations used throughout this work

ASA   Atomic Shell Approximation
DF   Density Function
DQOS   Discrete Quantum Object Set
DSQOS   Discrete Stochastic Quantum Object Set
EJR   Elementary Jacobi Rotation
EMP   Electrostatic Molecular Potential
IMP   Inward Matrix Product
KE   Kinetic Energy
MO   Molecular Orbital
MQSI   Molecular Quantum Similarity Index
MQSM   Molecular Quantum Similarity Measure
PD   Positive Definite
PDF   Probability Density Function
QO   Quantum Object
QOS   Quantum Object Set
QSI   Quantum Similarity Index
QSM   Quantum Similarity Measure
QQSAR   Quantum Quantitative Structure-Activity Relationships
QQSPR   Quantum Quantitative Structure-Property Relationships
QSAR   Quantitative Structure-Activity Relationships
QSPR   Quantitative Structure-Property Relationships
QSTR   Quantitative Structure-Toxicology Relationships
SM   Similarity Matrix
SQSI   Stochastic Quantum Similarity Indices
SSM   Stochastic Similarity Matrix
VS   Vector Space
VSS   Vector Semispace

References

[1] R. Carbó, M. Arnau, L. Leyda, Int. J. Quant. Chem. 17 (1980) 1185-1189.

[2] E. E. Hodgkin, W. G. Richards, Int. J. Quant. Chem. 14 (1987) 105-110.

[3] C. Burt, W. G. Richards, Ph. Huxley, J. Comp. Chem. 11 (1990) 1139-1146.

[4] A. C. Good, E. E. Hodgkin, W. G. Richards, J.Chem.Inf.Comp.Sci. 32 (1992) 188-191.

[5] A. C. Good, Sung-Sau So, W.G. Richards, J. Math. Chem. 36 (1993) 433-438.

[6] A. C. Good, W. G. Richards, J. Chem. Inf. Comp. Sci. 33 (1993) 112-116.

[7] D. L. Cooper, N. L. Allan, J. Comp. Aid. Mol. Des. 3 (1989) 253-259.

[8] D. L. Cooper, N. L. Allan, J. Am. Chem. Soc. 114 (1992) 4773-4776.

[9] D. L. Cooper, K. A. Mort, N. L. Allan, D. Kinchington, Ch. McGuidan, J. Am. Chem. Soc. 115 (1993) 12615-12616.

[10] N. L. Allan, D. L. Cooper, Topics in Current Chemistry 173 (1995) 85-111.

[11] J. Cioslowski, E. D. Fleishmann, J. Am. Chem. Soc. 113 (1991) 64-67.

[12] J. B. Ortiz, J. Cioslowski, Chem. Phys. Lett. 185 (1991) 270-275.

[13] J. Cioslowski, M. Challacombe, Int. J. Quant. Chem. S25 (1991) 81-93.

[14] J. Cioslowski, S. T. Mixon, Can. J. Chem. 70 (1992) 443-449.

[15] J. Cioslowski, A. Nanayakkara, J. Am. Chem. Soc. 115 (1993) 11213-11215.

[16] J. Cioslowski, B. B. Stefanov, P. Constans, J. Comp. Chem. 17 (1996) 1352-1358.

[17] R. Ponec, M. Strnad, Collect. Czech. Chem. Commun. 55 (1990) 896-902.

[18] R. Ponec, M. Strnad, Collect. Czech. Chem. Commun. 55 (1990) 2583-2589.

[19] R. Ponec, M. Strnad, J. Phys. Org. Chem. 4 (1991) 701-705.

[20] R. Ponec, M. Strnad, J. Math. Chem. 8 (1991) 103-112.

[21] R. Ponec, M. Strnad, Int. J. Quant. Chem. 42 (1992) 501-508.

[22] R. Ponec, J. Chem. Inf. Comp. Sci. 33 (1993) 805-811.

[23] R. Ponec, Overlap Determinant Method in the Theory of Pericyclic Reactions, Lecture notes in Chemistry 65, Springer Verlag, Berlin, (1995).

[24] P. G. Mezey, Shape in Chemistry: an Introduction to Molecular Shape and Topology, VCH, New York, (1993).

[25] P. G. Mezey in Molecular Similarity I, K. Sen (Editor) Topics in Current Chemistry, Vol. 173, Springer Verlag, Berlin, (1995), pp. 63-83.

[26] Ch. Lee, Sh. Smithline, J. Phys. Chem. 98 (1994) 1135-1138.

[27] C. Amovilli, R. McWeeny, J. Mol. Struct. (Theochem) 227 (1991) 1-9.

[28] A. C. Good, J. Mol. Graph. 10 (1992) 144-151.

[29] R. Benigni, M. Cotta-Ramusino, F. Giorgi, G. Gallo, J. Med. Chem. 38 (1995) 629-635.

[30] J. D. Petke, J. Comp. Chem. 14 (1993) 928-933.

[31] A. Riera, J. Mol. Struct. (Theochem) 259 (1992) 83-98.

[32] R.Carbó-Dorca. Quantum QSAR and the Eigensystems of Stochastic Quantum Similarity Matrices. J. Math. Chem. (submitted) IT-IQC-00-19.

[33] F. Fratev, O. E. Polansky, A. Mehlhorn, V. Monev, J. Mol. Struct. 56 (1979) 245-253.

[34] F. Fratev, V. Monev, A. Mehlhorn, O. E. Polansky, J. Mol. Struct. 56 (1979) 255-266.

[35] M. Ho, V.H. Smith, Jr., D.F. Weaver, C. Gatti, J. Chem. Phys. 108 (1998) 5469-5475.

[36] R. Benigni, M. Cotta-Rasmusino, G. Gallo, F. Giorgi, A. Giuliani, M. R. Vari, J. Med. Chem. 43 (2000) 3699-3707.

[37] Concepts and Applications of Molecular Similarity. M. A. Johnson, G. Maggiora (Editors), John Wiley & Sons Inc., New York, (1990).

[38] Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches. R. Carbó

(Editor), Understanding Chemical Reactivity, Kluwer Academic Publishers, Vol. 14, Dordrecht, (1995).

[39] Molecular Similarity. K. Sen (Editor) Topics in Current Chemistry, Vols. 173 & 174, Springer Verlag, Berlin, (1995).

[40] Molecular Similarity in Drug Design, P. M. Dean (Editor), Blackie Academic & Professional, London, (1995).

[41] Advances in Molecular Similarity, Vol. 1, R. Carbó-Dorca, P. G. Mezey (Editors), JAI Press Inc., Greenwich (Conn.), (1996).

[42] Advances in Molecular Similarity, Vol. 2, R. Carbó-Dorca, P. G. Mezey (Editors), JAI Press Inc., Greenwich (Conn.), (1998).

[43] R. Carbó, Ll. Domingo, Int. J. Quant. Chem., 23 (1987) 517-545.

[44] R. Carbó, B. Calabuig, Comp. Phys. Commun. 55 (1989) 117-126.

[45] R. Carbó, B. Calabuig, J. Mol. Struct. (Theochem) 254 (1992) 517-531.

[46] R. Carbó, B. Calabuig in: Computational Chemistry: Structure, Interactions and Reactivity, Vol. A. S. Fraga (Editor), Elsevier, Amsterdam, (1992), pp. 300-324.

[47] R. Carbó, B. Calabuig, L. Vera and E. Besalú, Adv. Quantum Chem. 25 (1994) 253-313.

[48] R. Carbó, B. Calabuig, Int. J. Quant. Chem. 42 (1992) 1681-1709.

[49] R. Carbó, E. Besalú in Molecular Similarity and Reactivity: From Quantum Chemistry to Phenomenological Approaches. R. Carbó (Editor) Kluwer Academic Publishers, Dordrecht, (1995), pp 3-30.

[50] R. Carbó, B. Calabuig, J. Chem. Inf. Comp. Sci. 32 (1992) 600-606.

[51] R. Carbó-Dorca, E. Besalú, L. Amat, X. Fradera in Advances in Molecular Similarity, R. Carbó-Dorca, P. G. Mezey (Editors), JAI Press Inc., Greenwich (Conn.), Vol. 1, (1996). pp. 1-42.

[52] R. Carbó, E. Besalú, J. Math. Chem. 18 (1995) 37-72.

[53] J. Mestres, M. Solà, M. Duran, R. Carbó, J. Comp.Chem. 15 (1994) 1113-1120.

[54] P. Constans, R. Carbó, J. Chem. Inf. Comput. Sci. 35 (1995) 1046-1053.

[55] P. Constans, L. Amat, X. Fradera, R. Carbó-Dorca in Advances in Molecular Similarity, R. Carbó-Dorca, P. G. Mezey (Editors), JAI Press Inc. Greenwich (Conn.), (1996), Vol. 1, pp. 187-211.

[56] L. Amat, R. Carbó, P. Constans, Scientia Gerundensis 22 (1996) 109-121.

[57] L. Amat, R. Carbó-Dorca, J.Comp.Chem. 18 (1997) 2023-2039.

[58] L. Amat, R. Carbó-Dorca, J. Comput. Chem. 20 (1999) 911-920.

[59] L. Amat, R. Carbó-Dorca, J. Chem. Inf. Comput. Sci., 40 (2000) 1188-1198

[60] R. Carbó, E. Besalú, L. Amat, X. Fradera, J. Math. Chem. 18 (1995) 237-246

[61] R. Carbó, E. Besalú, L. Amat, X. Fradera, J. Math. Chem. 19 (1996) 47-56.

[62] D.Robert, R.Carbó-Dorca, J. Chem. Inf. Comput. Sci., 38 (1998) 469-475.

[63] L.Amat, D.Robert, E.Besalú, R. Carbó-Dorca, J. Chem. Inf Comput. Sci., 38 (1998) 624-631.

[64] D. Robert, R. Carbó-Dorca, J. Chem. Inf. Comput. Sci., 38 (1998) 620-623.

[65] D. Robert, L. Amat, R. Carbó-Dorca, J. Chem. Inf. Comp. Sci. 39 (1999) 333-344.

[66] X. Gironés, L.Amat, R. Carbó-Dorca. SAR & QSAR Envir. Research 10 (1999) 545-556.

[67] R. Carbó, E. Besalú, J. Math. Chem. 20 (1996) 247-261.

[68] P. Constans, L. Amat, R. Carbó-Dorca, J. Comp. Chem. 18 (1997) 826-846.

[69] R. Carbó-Dorca, Int. J. Quantum. Chem. 79 (2000) 163-177.

[70] E. Besalú, R. Carbó, J. Mestres, M. Solà in Molecular Similarity I. K. Sen (Editor), Topics in Current Chemistry, Vol. 173, Springer Verlag, Berlin, (1995), pp. 31-62.

[71] L. A. Zadeh, Inform. Control 8 (1965) 338.

[72] E. Trillas, C. Alsina, J. M. Terricabras, Introducción a la Lógica Difusa, Ariel Matemática, Barcelona, (1995).

[73] R. Carbó-Dorca, J.Math. Chem 22 (1997) 143-147.

[74] D. Robert, R. Carbó-Dorca, J. Math. Chem. 23 (1998) 327-351.

[75] D. Robert, R. Carbó-Dorca, Il Nuovo Cimento A111 (1998) 1311-1321.

[76] D. Robert, R. Carbó-Dorca, Int. J. Quantum Chem. 77 (2000) 685-692.

[77] K. D. Sen, R. Carbó-Dorca, J. Mol. Struc. (Theochem) 501 (2000) 173.

[78] R. Carbó-Dorca; Inward Matrix Products: Extensions and Applications to Quantum Mechanical Foundations of QSAR. Institute of Computational Chemistry Technical Report IT-IQC-99-15. See also: Theochem (in press).

[79] Lahey/Fujitsu Fortran 95. Lahey Computer Systems. Incline Village (Nevada), v5.5 (1999).

[80] J. Klein, D. Babic; J. Chem. Inf. Comput. Scie., 37 (1997) 656.

[81] D.Robert, R.Carbó-Dorca, J. Chem. Inf. Comput. Sci. 38 (1998) 469.

[82] L.Amat, D.Robert, E.Besalú, R. Carbó-Dorca, J. Chem. Inf Comput. Sci., 38 (1998) 624.

[83] D. Robert, R. Carbó-Dorca, J. Chem. Inf. Comput. Sci., 38 (1998) 620.

[84] D. Robert, L.Amat, R. Carbó-Dorca; J. Chem. Inf. Comput. Sci. 39 (1999) 333.

[85] X. Gironés, L.Amat, R. Carbó-Dorca; SAR and QSAR in Environmental Research 10 (1999) 545.

[86] R. Carbó-Dorca, E. Besalú; J. Mol. Struct. (Theochem) 451 (1998) 11-23.

[87] R. Carbó-Dorca, Vol 2, 43-72 in: «Advances in Molecular Similarity », R. Carbó-Dorca, P. G. Mezey (Editors) (1998) JAI Press (London).

[88] R. Carbó-Dorca; J. Math. Chem., 23 (1998) 353.

[89] R. Carbó-Dorca, Ll. Amat, E. Besalú, X. Gironés, D. Robert. in: The Fundamentals of Molecular Similarity, R. Carbó-Dorca, X. Gironés and P. G. Mezey (eds.). Kluwer Acad. Publishers. Dordrecht, 2000.

[90] R. Carbó-Dorca, J.Math. Chem 22 (1997) 143.

[91] R. Carbó-Dorca, L. Amat, E. Besalú, M. Lobato, in Advances in Molecular Similarity, R. Carbó-Dorca, P. G. Mezey (Editors), JAI Press Inc. Greenwich (Conn.), (1998), Vol. 2, pp. 1-42.

[92] R. Carbó-Dorca, Ll. Amat, E. Besalú, X. Gironés, D. Robert, J. Mol. Struct. (Theochem), 504 (2000) 181-228

[93] J.H.Wilkinson, The Algebraic Eigenvalue Problem. Clarendon Press, Oxford (1965).

[94] J.H.Wilkinson, C. Reinsch, Linear Álgebra. Springer Verlag, Berlin (1971)

[95] R. Carbó-Dorca, J. Math. Chem., 23 (1998) 353-364.

[96] R. Carbó-Dorca, J.Math.Chem. 23 (1998) 365-375.

[97] X. Fradera, L. Amat, E. Besalú, R. Carbó-Dorca, Quant. Struct.-Act. Relat., 16 (1997) 25-32

[98] M. Lobato, L. Amat, E. Besalú, R. Carbó-Dorca, Scientia Gerundensis, 23 (1998) 17-28.

[99] M. Lobato, L. Amat, E. Besalú, R. Carbó-Dorca, Quant. Struct.-Act.Relat. 16 (1997) 465-472.

[100] D. Robert, R. Carbó-Dorca, J. Chem. Inf. Comp. Sci., 38 (1998)620-623.

[101] L.Amat,D.Robert,E.Besalú,R.Carbó-Dorca, J.Chem.Inf.Comp.Sci., 38 (1998) 624-631.

[102] R.Ponec,L.Amat,R.Carbó-Dorca, Molecular Basis of Quantitative Structure Property Relationships (QSPR): A quantum similarity approach, IT-IQC-98-07, see also: J. Comput-Aided Mol Des, in press.

[103] L.Amat,R.Carbó-Dorca,R.Ponec, J. Comp. Chem. 19 (1998) 1575-1583.

[104] R,Ponec, L.Amat, R.Carbó-Dorca, J. Phys. Org. Chem. 12 (1999) 447-454.

[105] P.G.Mezey, R.Ponec, L.Amat, R.Carbó-Dorca, Enantiomers 4 (1999) 371-378.

[106] X. Gironés, L. Amat, R. Carbó-Dorca. Using molecular quantum similarity measures as descriptors in quantitative structure-toxicity relationships SAR & QSAR Envir. Research 10 (1999) 545-556

[107] Gironés, X.; Amat,L.; Carbó-Dorca,R.;Robert, D. Use of electron-electron repulsion energy as a molecular descriptor in QSAR and QSPR studies J. Comp.-Aided Molec. Design 14 (2000) 477-485

[108] J.S.Bell, Speakable and Unspeakable in Quantum Mechanics, Cambridge University Press, Cambridge (1993).

[109] J. Von Neumann, Mathematical Foundations of Quantum Mechanics, Princeton University Press., Princeton (N.J.), (1955).

[110] D. Bohm, Quantum Theory, Dover Publications, Inc., New York, (1989).

[111] S. Goldstein, Physics Today, March 1988, 42 and April (1988), 38.

[112] I. M. Vinogradov (Editor) Encyclopaedia of Mathematics, Reidel-Kluwer Academic Publishers, Dordrecht, (1987).

[113] R. Bonaccorsi, E. Scrocco, J. Tomasi, J. Chem. Phys. 52 (1970) 5270-5284.

[114] P. O. Löwdin, Phys. Rew. 97 (1955) 1474-1489.

[115] R. McWeeny, Revs. Mod. Phys. 32 (1960) 335-369.

[116] S. K. Berberian, Introducción al Espacio de Hilbert, Editorial Teide, Barcelona, (1970).

[117] R.McWeeny, Proc.Roy.Soc. A,253(1959)242-259.

[118] E.R.Davidson. Reduced Density Matrices in Quantum Chemistry, Academic Press, New York, (1976).

[119] E.V. Ludeña, in: S.Fraga (Coordinador), Química Teórica, Nuevas Tendencias vol. 4, pp.117-160, Consejo Superior de Investigaciones Científicas, Madrid, (1987).

[120] I. Shavitt in: H. F. Schaefer III (Editor), Methods of Electronic Structure Theory, Modern Theoretical Chemistry, vol. 3, pp. 189-275, Plenum Press, New York (1977).

[121] J. Stoer, Ch. Witzgall, Convexity and Optimization in Finite Dimensions, Die Grundlehren der matematischen Wissenchaften in Einzeldarstellungen, Band 163, Springer Verlag, Berlin, (1970).

[122] P.M.Gruber,J.M.Wills (Editors), Handbook of Convex Geometry, North-Holland, Amsterdam, (1993).

[123] C. G. J. Jacobi, J. Reine Angew. Math. 30 (1846) 51-94.

[124] S. Huzinaga, M. Klobukowski, J. Mol. Struct. (Theochem), 167 (1998) 1-209.

[125] D.A. Pierre, Optimization Theory with Applications, Wiley, New York (1969).

[126] R. Carbó, L. Molino, B. Calabuig, J. Comp. Chem., 13 (1992) 155-159.

[127] M. R. Spiegel. Mathematical Handbook. McGraw-Hill, New York, (1968).

[128] A. Jeffrey, Handbook of mathematical Formulas and Integrals, Academic Press, New York (1995).

[129] P. W. Atkins and R. S. Friedman, Molecular Quantum Mechanics, Oxford University Press, Oxford (1997).

[130] P.A.M.Dirac, The Principles of Quantum Mechanics, Clarendon Press, Oxford (1958).

[131] H.Eyring,J.Walter,G.E.Kimball, Quantum Chemistry, John Wiley, New York (1944).

[132] F.L.Pilar, Elementary Quantum Chemistry,McGraw-Hill Pub. Co.,Princeton (1990).

[133] L.Pauling, E.B.Wilson Jr.,Introduction to Quantum mechanics, Dover, New York (1985).

[134] A.Messiah, Mécanique Quantique, Dunod, Paris (1959).

[135] A.S.Davydov,Quantum Mechanics, Pergamon Press, New York (1965).

[136] H.A.Bethe, R.Jackiw, Intermediate Quantum Mechanics, Benjamin, Menlo Park (1986).

[137] J.C.Paniagua and P.Alemany. Química Quàntica. Llibres de l'Índex. Universitat. Vols 1 and 2, Barcelona, 1999-2000.

[138] R. Carbó-Dorca, E. Besalú, X. Gironés, Advances in Quantum Chemistry, vol 38 (2000) 3-63.

[139] R.Carbó-Dorca, J.Karwowski, «Theoretical and computational aspects of extended wavefunctions», IT-IQC-00-28, submitted to Int. J. Quant. Chem.

[140] R.E.Moss, Advanced Molecular Quantum Mechanics, Chapman and Hall, London (1973).

[141] W.Greiner, Relativistic Quantum Mechanics, Springer Verlag, Berlin (1997).

[142] R. Carbó-Dorca, Mathematical Elements of Quantum Electronic Density Function. Institute of Computational Chemistry Technical Report. IT-IQC-00-03.

[143] N. H. March; Electron Density Theory of Atoms and Molecules. Academic Press (1992) London.

[144] R. Carbó-Dorca; A Discussion on an Apparent MO Theory Paradox. Institute of Computational Chemistry Technical Report. IT-IQC-99-17. J. Math. Chem. (in press).

[145] X. Gironés, A. Gallegos, R. Carbó-Dorca; Modelling Antimalarial Activity: Application of Kinetic Energy Density Quantum Similarity Measures as Descriptors in QSAR. Institute of Computational Chemistry Technical Report. IT-IQC-00-12.

[146] R. Carbó-Dorca, L. Amat, E. Besalú, X. Gironés, D. Robert; Quantum Molecular Similarity: Theory and Applications to the Evaluation of Molecular Properties, Biological Activities and Toxicity. Institute of Computational Chemistry Thechnical Report. IT-IQC-99-16.

[147] R. Carbó-Dorca; Intl. J. Quantum Chem. 79 (2000) 163.

[148] S. L. Sobolev; Math. Sb. 4 (1938) 471.

[149] A. Bach, Ll. Amat, E. Besalú, R. Carbó-Dorca, R. Ponec; Quantum Chemistry, Sobolev Spaces and SCF. Institute of Computational Chemistry Technical Report. IT-IQC-00-24. J. Math. Chem. (in press).

[150] R. Carbó-Dorca, «Extended Sobolev Spaces and Approximate Stationary Solutions for Electronic Systems within Nonlinear Schrödinger Equation». IT-IQC-00-25. Submitted to J. Math. Chem.

[151] E.Besalú, R.Carbó-Dorca and J.Karwowski, «Generalized one-electron spin functions and self-similarity measures». IT-IQC-00-30. Submitted to J.Math.Chem.

[152] R.Dedekind, Essays on the Theory of Numbers, Dover, New York (1963).

[153] R. Carbó, B. Calabuig, E. Besalú, A. Martínez, Molecular Engineering 2 (1992) 43-64.

[154] R.G.Parr, The Quantum theory of Molecular Electronic Structure, W.A.Benjamin Inc. New York (1963).

[155] R.Carbó, J.M.Riera, A general SCF Theory, Lecture Notes in Chemistry, vol. 5, Springer-Verlag, Berlin (1978).

Proceedings of the 4th Girona Seminar on Molecular Similarity. Kluwer (in press)

About the authors

Ramon Carbó-Dorca currently directs the Institut de Química Computacional *of the* Universitat de Girona*, having founded the Institute after leaving Barcelona in the mid 1980's. In Barcelona he has been working at the* Institut Químic de Sarrià *and the* Universitat Autònoma de Barcelona*, where he has been a professor since 1965. He has published several books on Quantum Chemistry, Linear Algebra and Quantum Similarity, as a result of his teaching and research experience in these fields. His work essentially involves research related to the Hückel method, Electrostatic Molecular Potential, Self-Consistent Field and Perturbation Theories, and the application of Elementary Jacobi Rotations in quantum chemistry. In 1980 he originated the concepts structuring the theoretical body associated with Quantum Similarity and Quantum Quantitative Structure-Properties Relationships. Several of his contributions attempt to integrate quantum chemistry with appropriate algorithms capable of being automatically programmed for computations in parallel. Starting with a first paper in 1965, which was also the first quantum chemistry paper ever published in Catalonia, he has since published more than 200 papers on the aforementioned subjects, including several contributions to divulgate his group's current scientific work. He has been awarded the Narcís Monturiol medal of the Generalitat de Catalunya.*

Emili Besalú is the secretary of the Institut de Química Computacional *of the* Universitat de Girona*. He has been a professor there since 1994 in the Chemical Physics Area of the Department of Chemistry. His PhD work was done at the Universitat de Girona under the direction of Professor R. Carbó-Dorca. His scientific work includes studies of Perturbation Theory, molecular integrals over exponential type orbital functions, Quantum Similarity and quantum mechanical algorithms in parallel. He has published over 60 papers on these subjects. He was the recipient of the 1997 Antoni Martí i Franquès prize of the Institut d'Estudis Catalans.*