A FUNDAMENTAL RELATION OF MOLECULAR INFORMATICS ON THE INFORMATION-CARRYING PROPERTIES OF DENSITY FUNCTIONS

Paul G. MEZEY[a,b]

[a] *Canada Research Chair in Scientific Modeling and Simulation, Department of Chemistry and Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, NL, Canada A1B 3X7; e-mail: pmezey@mun.ca*

[b] *Institute for Advanced Study, Collegium Budapest, Szentháromság u. 2, H 1014 Budapest, Hungary; e-mail: paul.mezey@gmail.com*

Received October 20, 2006
Accepted December 6, 2006

Dedicated to the memory of Professor Jaroslav Koutecký.

The molecular electron density carries the complete information about the molecule. This information is stored in the shape and more general topological features of molecular electron densities. A fundamental relation of molecular informatics, building on the Hohenberg–Kohn theorem, is the holographic electron density theorem: any nonzero volume part of a molecular electron density in a non-degenerate electronic ground state contains the complete information about all properties of the entire molecule. This fundamental feature of all molecules applies to all exhibited and also to all latent molecular properties, where latent properties are those not normally exhibited, only in response to some external stimulus. Recently it has become feasible to compute ab initio quality electron densities and approximate forces acting on individual nuclei in large molecules, even those beyond the thousand atom range, such as proteins. The newly expanded size range where reliable modelling methods can be also applied extends the role of detailed molecular shape analysis to macromolecules. In this context, it has become possible to study how the fundamental information-carrying properties of electron density take a newly recognized role influencing the predominance of specific nuclear conformations within the family of astronomically many potentially stable conformations of some macromolecules. Some special problems and results are discussed.

Keywords: Molecular informatics; Holographic electron density theorem; Latent molecular properties; Stoichiometric molecular families; Excited states; Symmetry.

Some of the fundamental properties of molecular electron densities are rather similar in all molecules[1], yet the fine details of the electron density cloud yield a truly astonishing array of molecular properties, and a rich enough variety to build up complex molecular systems necessary for life.

Collect. Czech. Chem. Commun. 2007, Vol. 72, No. 2, pp. 153–163

doi:10.1135/cccc20070153

The complete information content of molecules is stored in the molecular electron density. The shape and more general topological features of molecular electron densities are important clues to this information[2,3]. The computation of ab initio quality electron densities and approximate forces acting on individual nuclei in large molecules, such as proteins, has become practical in recent years[4,5]. One important relation concerning molecular information is the holographic electron density theorem[6,7], stating that any nonzero volume part of a molecular electron density in a nondegenerate electronic ground state contains the complete information about all properties of the entire, boundaryless molecule. This fundamental property of all molecules, a formal strengthening of the Hohenberg–Kohn theorem[8], and an extension of an earlier result on artificial Coulomb systems with closed boundaries[9], provides a framework for the emerging field of molecular informatics. In this context, the holographic principle for latent molecular properties[10,11], and some additional results concerning information on nuclear rearrangements are of special importance. These results may provide new tools for the analysis of conformational preferences of certain macromolecules such as proteins.

THE HOLOGRAPHIC PROPERTY OF ELECTRON DENSITIES

How is the information concerning the molecule stored within the electron density? The conventional reasoning usually points out that the location and nature (such as the nuclear charge) of nuclei is fully reflected in the electron density. Since molecules contain no other material but nuclei and the electron density cloud, and the electron density fully reflects the nuclear arrangement, the electron density must contain the complete information, by default. In fact, the celebrated Hohenberg–Kohn theorem of molecular physics[8] is a more formal statement of the above: in the nondegenerate electronic ground state of any molecule, the electron density determines the total energy, and through the Hamiltonian, all other molecular properties.

In 1981 Riess and Münch studied bounded molecular models confined to a finite volume, that is, fictitious molecules with closed boundaries, and they demonstrated that for such a system the Hohenberg–Kohn theorem applies for any positive subvolume of the electron density[9]. Whether a similar statement holds or not for real, boundaryless molecules remained an open question until 1999, when Mezey has shown[6,7] that the electron densities of real, boundaryless molecules possess the so-called holographic property: any positive volume of the molecular electron density of a molecule in a non-degenerate electronic ground state contains the complete information about the entire molecule. One important assumption of the proof of this holographic electron density theorem was the continuity of the electron density function, an assumption that follows from the usual formulation of the postulates of quantum mechanics and is rather universally accepted.

In fact, the holographic electron density theorem underlines the peculiar properties of electron density continua, and the particular way the molecular information is stored within molecules. Any given positive volume of the molecular electron density cloud (of a molecule in a non-degenerate ground state), no matter how small, has sufficiently rich information content to fully characterise the complete molecule. On the mathematical level, continuum cardinality plays a role, where the size of the given continuum is irrelevant when its equivalence with another continuum is considered.

One may formulate this theorem and some of its consequences by considering a formal extension operator X that generates a unique extension of the electron density function $\rho_D(\mathbf{r})$ in a finite positive volume D to the complete electron density function $\rho(\mathbf{r})$ of a non-degenerate electronic state over the three-dimensional space

$$X\rho_D(\mathbf{r}) = \rho(\mathbf{r}). \tag{1}$$

One relevant result is the formal proof of the non-existence of local-only symmetries[7]: for any symmetry operator R that belongs to a symmetry element of the local electron density $\rho_D(\mathbf{r})$ in a finite volume D,

$$R\rho_D(\mathbf{r}) = \rho_D(\mathbf{r}) \tag{2}$$

the operators R and X must commute,

$$XR = RX \tag{3}$$

and the same symmetry element must also be a symmetry element of the complete electron density, that is, of the complete molecule,

$$R\rho(\mathbf{r}) = \rho(\mathbf{r}). \tag{4}$$

Hence, if there exists a symmetry for the local electron density $\rho_D(\boldsymbol{r})$, as represented by the symmetry operator *R*, then the same symmetry must also be present for the complete electron density $\rho(\boldsymbol{r})$. In other words, local symmetry *R* implies global symmetry *R* for the complete molecule, and local-only symmetry cannot exist[7].

Here we shall consider this implication of the holographic electron density theorem from a different perspective: how local symmetries enhance each other within the global electron density of the complete molecule.

THE LOCAL SYMMETRY ACCUMULATION PROPERTY OF GLOBAL ELECTRON DENSITIES

Although the information content of a symmetry description of a molecule is rather limited when compared to the complete molecular information, nevertheless, symmetry is a fundamental property and the associated information is of special importance in many applications, including spectroscopy. Here we shall discuss the relations between local and global symmetries of molecular electron densities, and describe a symmetry information accumulation property of the global electron density when compared to local electron densities. Whereas the conclusions we discuss are rather elementary, nevertheless, from the perspective of regarding symmetry as a constraint of geometry, the results indicate that in molecular local and global symmetry relations a peculiar compatibility of constraints plays an important role.

In most physical systems, if several constraints are applied, it is common to find that some of the constraints are contradictory and incompatible, rendering the requested new physical system impossible to realize. It is not automatically guaranteed that two or more constraints, when applied together, hence generating a more restrictive set of conditions, are forming a compatible set of new, stronger constraints. Yet the relations between local and global symmetry properties of electron densities are exhibiting this compatibility property, and the associated molecular information can be formulated in rather simple algebraic terms.

We briefly review the elementary symmetry consequence of the holographic electron density theorem, in the present case formulated specifically for the treatment of symmetry accumulation.

We start with a simplified proof of the original symmetry implication[7].

Consider a local domain *D* of a molecular electron density and a symmetry operator *R*, that corresponds to a symmetry element present in the local electron density ρ_D in the domain. We also consider the operator of unique extension X; if X is applied to a non-degenerate ground state local electron density ρ_D within a finite volume D, it uniquely reproduces the complete electron density ρ of the molecule, as implied by the holographic electron density theorem. Since ρ_D has the local symmetry R,

$$R\rho_D = \rho_D. \tag{5}$$

From the holographic property one obtains

$$X\rho_D = \rho. \tag{6}$$

If one applies the operator R to the whole space, that generates a new, R-image version of the entire space, including the R-image versions of everything in the original space, in particular, the R-image version of the original electron density. Note that the proof of the holographic theorem involved no symmetry considerations, hence the same holographic theorem is applicable within this new, artificial space as well, with the same proof. Consequently,

$$XR\rho_D = R\rho. \tag{7}$$

By recognizing the left-hand side of Eq. (*5*) in the left hand side of Eq. (*7*) one obtains

$$X\rho_D = R\rho \tag{8}$$

and comparison of the left-hand sides of Eqs (*6*) and (*8*) gives

$$\rho = R\rho \tag{9}$$

that is, the complete electron density ρ must also have the same symmetry element of operator R as that of the local electron density ρ_D.

Our main goal in this derivation has been to show that Eq. (*9*) follows from Eq. (*4*), however, one may also notice that substitution of ρ from Eq. (*6*) into the right-hand side of Eq. (*7*) gives

$$XR\rho_D = RX\rho_D \tag{10}$$

hence on the set of non-degenerate ground state electron densities and density fragments the operators R and X indeed commute

$$XR = RX. \tag{11}$$

Since the local symmetry element and the corresponding symmetry operator are not restricted to any particular ones from those present for local density ρ, we conclude that the entire local symmetry group G_D of local density fragment ρ_D is also a symmetry group present for the complete electron density ρ.

Take now a different local density fragment $\rho_{D'}$ from a different domain D' of the same global electron density ρ, and assume that it has some local symmetry element corresponding to a symmetry operator R', and repeat the above arguments, leading to the statement that R' is also a symmetry operator corresponding to a global symmetry element for the complete electron density ρ, and the local symmetry group $G_{D'}$ present for the local electron density $\rho_{D'}$ is also a symmetry group present for the complete molecular electron density ρ.

We conclude that each local symmetry group G_D must be a subgroup of the symmetry group G of the complete molecule. Alternatively, one may say that the global symmetry G of the complete electron density is the highest symmetry present anywhere within the molecule.

Considering an abstract space with all local electron density fragments ρ_D and the global electron density ρ of a given molecule as elements, the complete electron density ρ is an accumulation point of all the symmetries present in this abstract space.

LATENT PROPERTIES AND LATENT SYMMETRIES

The holographic electron density theorem has been extended to the so-called "latent" molecular properties[10,11], that is, to those properties which are not exhibited by the actual conformation containing the local domain of the electron density cloud serving as the "predictor" of the rest of the molecule. Such latent properties include the properties of various electronically excited states: clearly, when defining a specific non-degenerate electronic excited state of the molecule, the only extra information in addition to the information evidently present in the non-degenerate ground state electron density is the energy value (a single scalar quantity) of the actual quantum, or the sum of energy values of the set of quanta in a multistep process, that results in the transition from the ground state to the given excited state. This single scalar value serves only as the selector of the actual electronic excited state, and all the essential information about this excited state must already be present in the non-degenerate ground state electron density[11].

Here we are concerned with another set of latent properties: the properties which belong to conformations (nuclear arrangements) different from the conformation that is the source of the local domain of electron density for the given molecule A, used as the "predictor". In fact, one may proceed one step further, and consider all the nuclear arrangements which belong to the given stoichiometry S of the molecule A, that is, one may take the entire potential energy hypersurface $E(K)$ that belongs to the electronic ground state of all systems composed of the given set of nuclei and the given number of electrons[12]. In this notation variable K symbolizes the nuclear configuration of molecule A, where K is an element of the "reduced" nuclear configuration space M describing all possible internal configurations of the nuclei. Note that this space M is a metric space but it can never be a formal vector space[12], a feature that is often overlooked and has led to various misconceptions.

One may regard a formal process where one reference nuclear configuration K of molecule A is changed to another nuclear configuration K'. We assume that configuration K is the one that belongs to the electron density that contains the small domain used as predictor in the sense of the holographic electron density theorem. One should note that either or both of these nuclear configurations may, in fact, belong to single molecules, or to a set of two or several molecules with a specified mutual nuclear arrangement, as long as the overall stoichiometry is the same that specified by the reference nuclear configuration K.

In the possession of the above information, the net change of the conformational process

$$K \Rightarrow K' \tag{12}$$

assumed to be confined to the electronic ground state, is uniquely specified by the two sets of nuclear coordinates, that is, by a finite set of scalar numbers. For the given reference configuration K, these numbers serve merely as the selector of the new configuration K', and all the remaining essential information about the electronic density in the new nuclear configuration K' must be already present in the electronic density of the reference configuration K. Since the information about the complete electron density of the reference configuration must be present in the local electron density domain taken as the local predictor in the sense of the holographic electron density theorem, it follows that the same local domain of electron density cloud corresponding to the nuclear configuration K also contains the complete information about the electron density and all other properties of the molecular system of the other nuclear configuration K' as well, once the selection K' is specified.

Since the choice of both K and K' is arbitrary, one may conclude that, at least in principle, any small nonzero volume domain of the electron density cloud of any nuclear conformation K contains the complete information about all molecular species along the entire electronic ground-state potential energy hypersurface $E(K)$, possibly excluding conformations with degenerate electronic states (surface crossings).

In fact, properties of molecular species with conformations different from the reference conformation may be regarded as formal latent properties of the molecular species of the reference conformation itself.

In some sense, any positive volume local electron density domain of any nuclear arrangement can act as an "over-the-horizon radar" along the entire potential energy hypersurface $E(K)$: it contains the complete information about all chemical species along the potential energy hypersurface, including all conformers of the given reference species, and also all products of dissociation and recombination reactions.

All the chemistry information along the complete potential energy hypersurface (with the possible exclusion of conformations with degenerate electronic states corresponding to surface crossings) is contained within any positive volume of electron density of any of the species present along the potential energy hypersurface.

What is the additional information that defines the properties of electronically excited state potential energy hypersurfaces? Again, we may start with a thought-experiment. Consider the most stable nuclear configuration K of the electronic ground-state potential energy hypersurface $E(K)$ for the given overall stoichiometry S. The corresponding molecular species possesses a set of well-defined quantum states, and the only additional information required to obtain another molecular species on a selected electronically excited state potential energy hypersurface $E'(K)$ is the energy of the quantum that is able to trigger the corresponding electronic transition. Hence, using this energy value, a single scalar number, as the selector, all additional information about this excited state species, and by the same argument that has been used in the case of the ground electronic state, all information about all molecular species along the entire electronically excited state potential energy hypersurface $E'(K)$ is then contained within any small nonzero volume of the electronic density cloud of the ground-state molecular species with nuclear configuration K.

Since the choice of the excited state potential energy hypersurface $E'(K)$ was arbitrary, we may conclude that for the given overall stoichiometry S, the information concerning all properties of all molecular species along all electronic state potential energy hypersurfaces is contained within any small nonzero volume of the electronic density cloud of the ground-state molecular species with nuclear configuration K.

Although rather evident from the above arguments, it is worthwhile to point out some of the special relations that apply to chemical reactions along a given potential energy hypersurface, as well as in between different potential energy hypersurfaces. We shall address these problems within the context of sub-stoichiometries.

For a given overall stoichiometry S, of a specified set of nuclei and a given number of electrons, a sub-stoichiometry S' is one that contains a subset of the nuclei involved in the overall stoichiometry S. Along potential energy hypersurfaces of the overall stoichiometry S (for any electronic state), any substoichiometry S' can be treated easily by considering special arrangements of the nuclei: when focusing on sub-stoichiometry S', one can consider only those regions of the potential energy hypersurface which involve the remaining nuclei "far removed", that is, with arrangements of mutually infinite distances from each other and also from the selected set of nuclei in S'. Evidently, such arrangements form a subset of all nuclear arrangements present within the metric nuclear configuration space M, hence a study of a sub-stoichiometry S' within the metric nuclear configuration space M corresponds to a simple restriction M' of this very metric space M. Note that the same metric of the space M is also a valid metric within M'.

Consequently, all properties of all molecular species of any sub-stoichiomedtry S' that is a part of any (possibly infinitely dissociated) overall set of species along all electronic state potential energy hypersurfaces are also fully determined by the information present within any small nonzero volume of the electronic density cloud of the ground state molecular species with nuclear configuration K.

That is, all properties of products of dissociation reaction or recombination reactions involving the overall stoichiometry S or any sub-stoichiometry S' of S are fully determined by the information present within any small nonzero volume of the electronic density cloud of the ground-state molecular species with nuclear configuration K.

Note that in the above arguments the choice of the most stable nuclear configuration K of the ground electronic state potential energy hypersurface was motivated simply by easy reference; in fact, any other nuclear configuration K'' with non-degenerate electronic state along this hypersurface can be used as reference.

Concerning reactions involving transitions between ground state and excited state potential energy hypersurfaces, the same arguments apply for molecules of the sub-stoichiometry S' as those for the complete potential energy hypersurface of stoichiometry S.

CONCLUSIONS

Molecular informatics is ultimately based on the information stored in molecular electron density clouds. A fundamental property of such electron density clouds is expressed by the holographic electron density theorem, describing the relations between local and global properties. The holographic electron density theorem for latent properties provides an extension that serves as the bridge between different electronic states, different molecular conformations, different chemical species, and reactants and products of chemical reactions. Many of these connections can be described within a comprehensive framework involving potential energy hypersurfaces and the underlying nuclear configuration space, the metric space M, that also serves as a tool for the study of interrelations among a given overall stoichiometry S (involving all molecular species with a common overall set of nuclei and a specified number of electrons) and all possible sub-stoichiometries S' of S, involving only a subset of the nuclei. The holographic electron density theorem for latent molecular properties extends to these cases as well, providing a unified framework for molecular informatics as well as new tools for the study of conformational changes, electronic excitations, and chemical reactions.

REFERENCES

1. Levy M. in: *Recent Developments and Applications of Modern Density Functional Theory, Theoretical and Computational Chemistry* (J. M. Seminario, Ed.), Vol. 4, p. 3. Elsevier Science Publ., Amsterdam 1996.
2. Mezey P. G.: *Shape in Chemistry: An Introduction to Molecular Shape and Topology*. VCH Publishers, New York 1993.
3. Mezey P. G. in: *Encyclopedia of Computational Chemistry* (P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III and P. R. Schreiner, Eds), Vol. 4, p. 2582. John Wiley & Sons, Chichester 1999.
4. Mezey P. G.: *J. Math. Chem.* **1995**, *18*, 141.

5. Mezey P. G.: *Int. J. Quantum Chem.* **1997**, *63*, 39.
6. Mezey P. G.: *Mol. Phys.* **1999**, *96*, 169.
7. Mezey P. G.: *J. Math. Chem.* **1998**, *23*, 65.
8. Hohenberg P., Kohn W.: *Phys. Rev. B* **1964**, *136*, 864.
9. Riess J., Münch W.: *Theor. Chim. Acta* **1981**, *58*, 295.
10. Mezey P. G.: *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 224.
11. Mezey P. G.: *J. Math. Chem.* **2001**, *30*, 299.
12. Mezey P. G.: *Potential Energy Hypersurfaces*. Elsevier, Amsterdam 1987.