Quantum Similarity of Atoms: a Numerical
Hartree Fock and Information Theory
Approach

A. Borgoo\textsuperscript{a}, M. Godefroid\textsuperscript{b}, K. D. Sen\textsuperscript{c}, F. De Proft\textsuperscript{a}

P. Geerlings\textsuperscript{a,\!*}

\textsuperscript{a}Vrije Universiteit Brussel (Free University of Brussels VUB), Departement of
General Chemistry (ALGC), Pleinlaan, 2, 1050 Brussels Belgium

\textsuperscript{b}Service de Chimie quantique et Photophysique - CP160/09, Université Libre de
Bruxelles, Av. F.D. Roosevelt 50, 1050 Brussels Belgium

\textsuperscript{c}School of Chemistry, University of Hyderabad, Hyderabad 500046 India

Abstract

In this letter Quantum Similarity for Atoms (H-Xe) is investigated using electron
densities and shape functions, looking for patterns of periodicity as in Mendeleev’s
Table. An \textit{LS}-dependent restricted Hartree Fock method is used to obtain the wave
functions from which the electron densities are calculated. Utilizing the quantum
similarity proposed by Carbó a nearest neighbour dominated similarity is retrieved,
masking periodicity. Introduction of the information discrimination concept with
reference to the noble gas atom of the previous row is found to reveal periodicity,
with improved results when densities are replaced by shape functions throughout.
This confirms recent literature on the fundamental role of the shape function as
carrier of information.

Preprint submitted to Elsevier Science 5 October 2004
1 Introduction

For many decades similarity between molecules has, certainly implicitly, been a cornerstone in chemical research: similar molecules, usually (vide infra) considered to be molecules of comparable shape, are indeed expected to display similar properties, in particular similar reactivity patterns, and so are expected to chemically behave in a similar way [1]. Both the ever lasting use of the concept of functional groups in organic chemistry [2], transferable (to a certain extent) from one molecule to another, and the intense QSAR investigations in pharmaceutical chemistry [3] are illustrations of the attempts to master and exploit similarity in structure, physicochemical properties and reactivity of molecular systems.

The question as to define similarity within a quantum mechanical context was addressed relatively late, in the early 1980’s, in the pioneering work of Carbó and coworkers [4,5] and led to a series of Quantum Similarity Measures (QSM) and Indices (QSI). These were essentially based on the electron density distribution of the two Quantum objects (in casu molecules) to be compared. The link between Similarity Analysis and Density Functional Theory [6,7] built on the electron density as the basic carrier of information, and pervading quantum chemical literature at that time, is striking.

Whereas the last 25 years witnessed a multitude of studies on various aspects of quantum similarity of molecules (the use of different separation operators [5], the replacement of the density by more appropriate reactivity oriented functions [6,7], within the context of conceptual DFT [8], the treatment of enan-
tomers [9,10]) the field of similarity between isolated atoms remained, most remarkably, nearly unexplored with the exception of two papers by Carbó and coworkers. In the first paper [11] atomic self-similarity was studied, whereas the second one [12] contains a relatively short study on atomic and nuclear similarity, leading to the conclusion that atoms bear the highest resemblance to their neighbours in the Periodic Table.

On the other hand periodicity in Mendeleev’s Table is probably the most basic concept and most important workhorse in chemistry. The relationship between Atomic Quantum Similarity and periodicity has, to the best of our knowledge not been addressed in the literature yet. For us it was tempting to see whether the electron density based similarity measures can indeed generate periodicity.

Due to the Hohenberg Kohn theorems [13] the electron density can indeed be considered as the basic carrier of information although for many properties (the most evident example being the energy) it is not known how to extract the relevant information from the density. All this prompted us to investigate the quantum similarity of isolated atoms in a systematic way, paying particular attention to the question whether density based QSM and QSI indeed generate chemical periodicity in “a natural way”. It turns out that the shape function $\sigma(r)$, i.e. the density per particle $\frac{\rho(r)}{N}$, [14], which recently received attention as an even more fundamental carrier of information [15–17], appears in a natural way.

This Letter is organized as follows: in section 2 we briefly describe how the QSI values were calculated for different atom pairs in an $LS$-dependent numerical Hartree Fock scheme. In section 3 Information Theory based quantities are entering in the discussion yielding the periodicity sought-after. Section 4 contains conclusions and prospects.
2 The Quantum Similarity Index: the nearest neighbour effect

The Quantum Similarity Measure $Z_{AB}$ between two systems $A$ and $B$ has been defined by Carbó [4] as

$$Z_{AB}(\Omega) = \int \rho_A(r_1) \Omega(r_1, r_2) \rho_B(r_2) \, dr_1 dr_2$$  \hspace{1cm} (1)$$

where $\rho_A$ and $\rho_B$ are the (spinless) electron density functions for both systems and where $\Omega(r_1, r_2)$ is a separation operator. Most often $\Omega$ is chosen as the Dirac $\delta$-function $\delta(r_1 - r_2)$ reducing 1 to an overlap type integral, the other most often used operator being $r_1^{-1}$ converting 1 in a Coulomb type integral [5].

Normalisation is taken care of by defining a QSI, $SI_{\Omega}$ as

$$SI_{\Omega} = \frac{Z_{AB}}{\sqrt{Z_{AA}Z_{BB}}}$$  \hspace{1cm} (2)$$

ensuring that $0 \leq SI_{\Omega} \leq 1$.

Evaluation of a similarity index for atoms is straightforward via 1 and 2 using electron densities of sufficient quality for the isolated atoms.

These densities were generated within a numerical $LS$-dependent Restricted Hartree Fock scheme [18] using Froese Fischer’s HF86 program [19]. The single-configuration Hartree-Fock approximation gives reasonably accurate results in many cases thanks to Brillouin’s theorem assuring that part of “radial correlation” is already included at least to first order [18]. On the other hand we deliberately limited ourselves to the first four rows of the Periodic Table, neglecting relativistic effects in our analysis. In the case of incompletely filled subshells spherical averaging was applied, yielding a density of the form
\[ \rho(r) = \frac{1}{4\pi} \sum_n \sum_{l=0}^{n-1} \frac{P_{nl}^2(r)}{r^2} q_{nl} \]  

(3)

where \( q_{nl} \) is the number of electrons in the subshell considered and where the \( P_{nl}(r) \) functions are the radial wave functions \( R_{nl}(r) \) multiplied by \( r \). Note that in a \( LS \)-dependent Hartree Fock scheme the radial wave functions vary, for a given electronic configuration, from one term to another, hence eq. (3) should be read as

\[ \rho(r) = \rho_{\alpha LS}(r) \]  

(4)

where \( \alpha \) represents the configuration, the coupling scheme, and any other quantum numbers required to denote a particular \( LS \) term. For all further purposes the numerical \( P_{nl} \) functions generated by HF86 were manipulated with a home made Mathematica [20] program. For testing the latter, we checked that the integration of the density yields the total number of electrons

\[ 4\pi \int_0^{\infty} \rho(r)r^2 dr = \sum_{nl} q_{nl} = N, \]  

(5)

as expected from the normalization property of the HF numerical one-electron radial functions. As pointed out above, the \( LS \)-dependent restricted Hartree-Fock approximation yields to \( LS \)-dependent Hartree-Fock equations for the open subshells due to the term-dependency of the coulomb interaction. In the single incomplete shell case, corresponding to the ground configurations we are interested in for the present study, the term dependency is usually fairly small. Froese Fischer compared mean radii of the radial functions [18]. The differences in \( \langle r \rangle \) for the outer orbitals between the values obtained from a HF calculation on the lowest term and those for the average energy of the configuration are of the order of 1-5 %. Although the \( LS \)-dependency does not show up explicitly in the direct and exchange potentials of the closed-
subshell radial Hartree-Fock equations, the closed-subshell radial functions are ultimately $LS$-dependent through the coupling between the orbitals in the HF equations to be solved in the iterative procedure, but these relaxation effects turn out to be even smaller.

As explicitly indicated through 4 the density built from the one-electron radial functions could therefore be $LS$-dependent but this issue has not been yet investigated quantitatively to the best of the authors’ knowledge. Combining the term-dependent densities $\rho_A = \rho_{\alpha_A L_A S_A}$ and $\rho_B = \rho_{\alpha_B L_B S_B}$ for the same atom in the same electronic ground configuration, but possibly different states, (and adopting the Dirac $\delta$-function for $\Omega$) for evaluating the Quantum Similarity Measure $Z_{AB}$ of 1, the similarity matrix can be constructed according to (2). Its matrix elements are reported in Table 1 for the $np^2$ configuration of Carbon ($n = 2$) and Silicon ($n = 3$). As expected, the deviation of the off-diagonal elements from 1 is rather small, the HF orbitals for the different terms $^3P$, $^1D$ and $^1S$ being highly similar, although not identical.

In the remaining part of this discussion all atoms are considered in their lowest energy term. In figure 1 we extract, as a case study from the complete atom QSI-matrix (given as supplementary material) the relevant information for the noble gases. Here the similarities were calculated using the Dirac delta function as separation operator. From these data it is clear that the similarity indices are higher, the closer the atoms are in the periodic table (smallest $\Delta Z$, $Z$ being the atomic number). The tendency noticed by Robert and Carbó in [12] is regained in the present study at a more sophisticated level. It can hence be concluded that the QSI involving $\rho(\mathbf{r})$ and evaluated with $\delta(\mathbf{r}_1 - \mathbf{r}_2)$ as separation operator $\Omega$, does not generate periodicity.
3 Discrimination of Information

In search for a similarity index reflecting periodicity a first possibility would be to opt for a different separation operator, for example putting more weight on the valence electrons. Nevertheless we thought that such type of constructions, besides being to some extent "trial and error" may also lead to a certain bias. So, a more fundamental approach was followed based on information theory which recently gained widespread interest in the quantum chemical community.

Gadre and Sears [21] for example studied the near Hartree-Fock information entropy $S_\rho = - \int \rho(\mathbf{r}) \log \rho(\mathbf{r}) d\mathbf{r}$ and its analogue in momentum space. One of the present authors presented an overview of computations on the Shannon entropy of the 1-normalized density $S_\eta = (S_\rho / N) + \log N$ [22]. Nagy and Parr [23] argued that information entropy gives a measure for the quality of an approximate electron wave function. Finally the important work of Nalewajski and Parr [24] should be mentioned pointing out that the Kulback Leibler entropy deficiency promotes Hirschfeld's [25] stockholder procedure as a natural way to generate atomic charges.

A point of major interest in recent literature on the information theory approach is the information discrimination $\Delta S$, defined below for a continuous probability distribution $P_k(x)$.

$$\Delta S(P_k/P_0) = \int P_k(x) \log \frac{P_k(x)}{P_0(x)} dx$$

Here $P_k(x)$ and $P_0(x)$ are normalized probability density distributions ($P_0(x)$ is known as prior distribution), $\Delta S$ expressing, as described by Kulback [26,27],
the mean information per observation for discrimination in favor of $P_1(x)$ over $P_0(x)$ (also referred to as cross-entropy [28] or entropy deficiency [29]). Use of this necessitates the choice of a reference density $P_0(x)$. Still concentrating on the electron density as an information carrier we chose $P_0$ in a way which bears analogy to the way Sanderson formulated his electronegativity scale [30]. Sanderson based his reasoning on the average electron density of a given atom $A$

$$ED_A = \frac{Z_A}{\frac{4}{3}\pi r_A^3} \tag{7}$$

where $Z_A$ is the atomic number and $r_A$ the covalent radius. The corresponding value for a (hypothetical) noble gas atom with the same number of electrons, obtained by linear interpolation between the noble gases, is denoted as $ED_0$. The ratio

$$S_A = \frac{ED_A}{ED_0} \tag{8}$$

was called the stability ratio for atom $A$ and put in relation with the electronegativity of atom $A$ on the basis of the compactness of the electron cloud, reflecting, in a present-day terminology, its shape. In the same vein we considered as reference density in (6) the renormalized noble gas density of the preceding row as the atom considered. The quantity to be evaluated then becomes

$$\Delta S_A^\rho = \int \rho_A(r) \log \frac{\rho_A(r)}{\frac{N_A}{N_0} \rho_0(r)} \, dr, \tag{9}$$

with $\rho_0(r)$ the density of the reference noble gas, scaled by the factor $\frac{N_A}{N_0}$ to yield upon integration the same number of electrons as in atom $A$, characterised by $\rho_A(r)$. It is easily seen that the integrand can be rewritten in (9) so
the shape functions $\sigma_A(r)$ and $\sigma_0(r)$ appear in $\Delta S_A^\rho$.

$$\Delta S_A^\rho = \int \rho_A(r) \log \frac{\sigma_A(r)}{\sigma_0(r)} dr$$  \tag{10}$$

Figure 2 indicates that a $\Delta S_A^\rho$ versus $Z$ plot reveals periodicity to some extent. Intuitively we can consider the data as reflecting the difference between the information content of an atom when the information of the inner (completely filled) shells has been removed by putting it in the reference density. So essentially information on the valence electrons is displayed, the noble gas considered having the same core as the atom under consideration.

Note however that in equation (10) a direct dependency of $\Delta S_A^\rho$ on $N_A$ is observed as $\rho_A(r) = \sigma_A(r) N_A$. This behavior might be at the origin of the similar steepness of the curves belonging to first, second, third and fourth row atoms.

It is tempting to eliminate the $N_A$ dependency, easily justified by rebuilding the theory directly in terms of shape functions, yielding

$$\Delta S_A^\sigma = \int \sigma_A(r) \log \frac{\sigma_A(r)}{\sigma_0(r)} dr$$  \tag{11}$$

Figure 3 now shows that periodicity is more pronounced coupled to the fact that the distance between points in a given period is decreasing gradually from first to fourth row. One hereby regains one of the basic characteristics of the Periodic Table namely that the evolution in (many) properties through a given period slows down when going down in the Table. The decrease in slope of the four curves is a further illustration. In Figure 3, the regular decrease in slope is reminiscent of the standard nuclear packing fraction curves in the negative region [31,32] which also correlate with the nuclear abundance and
stability data as the atomic number increases.

4 Conclusion

$L_S$-dependent numerical Hartree Fock densities for atoms H-Xe combined in a similarity index with a Dirac delta function separation operator yield a nearest neighbour dominated similarity, masking periodicity. Introduction of the information discrimination concept with reference to the noble gas atom of the previous row leads to periodicity, with more pronounced results when densities are replaced by shape functions throughout. The present study also highlights the importance of the choice of the reference used to discriminate information in equation (11), which in our work has been fixed as the noble gas density of the previous row in the periodic table. As shown above the chemically intuitive choice of the reference density plays a key role in unmasking the periodicity in the Table of Mendeleev.

The final result can be translated in the gain of information that valence electrons bear in their shape as compared to that of the noble gas with the same core. The importance of the shape function as carrier of information as recently discussed by Parr, Ayers and the present authors [14–17] is confirmed and strengthened.

Acknowledgments

P. G. and K. D. S. thank the F. W. O. Flanders for financial support for a short stay of K. D. S. in the ALGC group. P. G. thanks G. Boon for carrying out a
preliminary study for this paper. M.G. thanks the Communauté française of Belgium (Actions de Recherche Concertées) and the Belgian National Fund for Scientific Research (FRFC convention) for their financial support.
<table>
<thead>
<tr>
<th>term</th>
<th>$C^3P$</th>
<th>$C^1D$</th>
<th>$C^1S$</th>
<th>$Si^3P$</th>
<th>$Si^1D$</th>
<th>$Si^1S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C^3P$</td>
<td>1.</td>
<td>0.9999</td>
<td>0.9999</td>
<td>1.</td>
<td>1.</td>
<td>1.</td>
</tr>
<tr>
<td>$C^1D$</td>
<td>0.9999</td>
<td>1.</td>
<td>0.9999</td>
<td>1.</td>
<td>1.</td>
<td>1.</td>
</tr>
<tr>
<td>$C^1S$</td>
<td>0.9999</td>
<td>0.9999</td>
<td>1.</td>
<td>1.</td>
<td>1.</td>
<td>1.</td>
</tr>
</tbody>
</table>

Table 1

Similarity between different terms of C and Si in $ns^2np^2$ configuration.
Fig. 1. Similarity Indices (SI) for noble gases, using the Dirac delta function as separation operator. Each noble gas corresponds with a curve comprising the similarity indices with all other atoms.
Fig. 2. Information discrimination (ID), (equation 10) versus $Z$ for atomic densities with the noble gas of the previous row as reference.
Fig. 3. Information discrimination (ID), (equation 11) versus $Z$ for atomic shape functions with the noble gas of the previous row as reference.
References


[31] F.W. Aston, Nature 140 (1937), 149