Similarity of Atoms in Molecules

Jerzy Cioslowski* **and Asiri Nanayakkara**

Contribution from the Department of Chemistry and Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida 32306-3006

Received July 1, 1993[⊗]

Abstract: Similarity of atoms in molecules is quantitatively assessed with a measure that employs electron densities within respective atomic basins. This atomic similarity measure does not rely on arbitrary assumptions concerning basis functions or "atomic orbitals", is relatively inexpensive to compute, and has straightforward interpretation. Inspection of similarities between pairs of carbon, hydrogen, and fluorine atoms in the CH_4, CH_3F, CH_2F_2, CHF_3, CF_4, C_2H_2, C_2H_4, and C_2H_6 molecules, calculated at the MP2/6-311G** level of theory, reveals that the atomic similarity is greatly reduced by a change in the number or the character of ligands (i.e. the atoms with nuclei linked through bond paths to the nucleus of the atom in question). On the other hand, atoms with formally identical (i.e. having the same nuclei and numbers of ligands) ligands resemble each other to a large degree, with the similarity indices greater than 0.95 for hydrogens and 0.99 for non-hydrogens.

Introduction

The concept of atoms is of paramount importance to all branches of chemistry. The approximate transferability of atomic and group properties makes it possible to understand, rationalize, and predict chemical behavior of molecules. This qualitative approach of classical chemistry can be contrasted with the quantitative techniques of quantum mechanics that provide tools for (at least in principle) exact predictions of molecular properties through solution of the Schrödinger (or Dirac) equation. However, the original postulates of quantum mechanics do not allow atoms to be discerned within molecules.

Fortunately, only one additional quantum-mechanical postulate is needed to identify atoms in molecules and calculate their properties. This postulate, which states that atoms in molecules are delineated by zero-flux surfaces (i.e. surfaces that are parallel everywhere to the electron density gradient and do not pass through any of the nuclei), was first formulated by Bader.[1] The partitioning of the Cartesian space spanned by electrons of a given molecule into atomic basins is not at all arbitrary, as only the zero-flux surfaces afford atoms that have all the characteristics of well-defined subsystems conforming to all laws of quantum mechanics (such as the virial and Ehrenfest theorems).

Bader's theory not only quantifies the concept of atoms in molecules, but also provides the means for identifying chemical bonds and defines atomic contributions to one-electron properties.[1] In particular, the number of electrons in atom A that are a part of molecule X is given by

$$N_{A(X)} = \int_{\Omega_A} \rho_X(\vec{r})\, d\vec{r} \qquad (1)$$

where Ω_A is the basin of A and $\rho_X(\vec{r})$ is the electron density of X. Extensions of the original theory rigorously define such important chemical concepts as the covalent bond order,[2] steric crowding,[3] electronegativity, bond hardness, and the charge-transfer component of the bond energy.[4]

The concept of similarity of atoms, functional groups, and entire molecules is as important to classical chemistry as the concept of atoms in molecules. The first quantum-mechanical definition of molecular similarity has been given by Carbo et al.,[5] who proposed the index

$$R_{XY} = \int \rho_X(\vec{r})\, \rho_Y(\vec{r})\, d\vec{r} \Big/ \Big[\int \rho_X^{\,2}(\vec{r})\, d\vec{r} \int \rho_Y^{\,2}(\vec{r})\, d\vec{r}\Big]^{1/2}, \qquad 0 < R_{XY} \leq 1 \quad (2)$$

to quantify the extent to which the molecules X and Y resemble each other. The right-side equality in eq 2 is attained only when X and Y are identical. A similar index, called H_{XY}, has been introduced by Hodgkin and Richards.[6]

Calculation of molecular similarity requires maximization of the similarity index by adjusting the mutual orientation of the molecules. Such adjustment involves varying three components of the translation vector and three Euler angles and requires repeated computation of the similarity index. Because both R_{AB} and H_{AB} are very expensive to evaluate in an *ab initio* manner, rigorous calculations involving those indices have been rare. To rectify this problem, a new molecular similarity measure called NOEL (which stands for the number of overlapping electrons) has been introduced.[7] Employing overlap of one-electron density matrices instead of electron densities in calculations of NOEL results in very low computational cost and simplicity of interpretation. Several interesting applications of NOEL have been reported,[8–10] including a quantitative formulation of the Hammond postulate.[8]

Despite the aforedescribed progress in rigorous definitions of atoms in molecules and molecular similarities, relatively little attention has been paid in the chemical literature to quantifying the similarity of atoms in molecules. Bader and Becker[11] have used the theorem of Riess and Münch[12] to prove that atoms cannot be perfectly transferable between different molecules, i.e. atoms that belong to different molecules cannot be 100% similar. However, no restriction is placed on how closely the perfect similarity can be approached. In fact, as is well-documented by experimental data and confirmed by electronic structure calculations, properties of functional groups and sometimes individual

---

[⊗] Abstract published in *Advance ACS Abstracts,* November 1, 1993.

[1] Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Clarendon Press: Oxford, 1990.

[2] Cioslowski, J.; Mixon, S. T. *J. Am. Chem. Soc.* **1991**, *113*, 4142.

[3] Cioslowski, J.; Mixon, S. T. *J. Am. Chem. Soc.* **1992**, *114*, 4382.

[4] Cioslowski, J.; Mixon, S. T. *J. Am. Chem. Soc.* **1993**, *115*, 1084.

[5] Carbo, R.; Leyda, L.; Arnau, M. *Int. J. Quantum Chem.* **1980**, *17*, 1185.

[6] Hodgkin, E. E.; Richards, W. G.; *Int. J. Quantum Chem.* **1987**, *S14*, 105.

[7] Cioslowski, J.; Fleischmann, E. D. *J. Am. Chem. Soc.* **1991**, *113*, 64.

[8] Cioslowski, J. *J. Am. Chem. Soc.* **1991**, *113*, 6756.

[9] Cioslowski, J.; Sauer, J.; Hetzenegger, J.; Karcher, T.; Hierstetter, T. *J. Am. Chem. Soc.* **1993**, *115*, 1353.

[10] Cioslowski, J.; Fleischmann, E. D. *Croat. Chem. Acta*, in press.

[11] Bader, R. F. W.; Becker, P. *Chem. Phys. Lett.* **1988**, *148*, 452.

[12] Riess, I.; Muench, W.; *Theor. Chim. Acta* **1981**, *58*, 295.

0002-7863/93/1515-11213$04.00/0

atoms usually do not change significantly from one molecule to another.[13] This approximate transferability extends to the shapes of atoms and functional groups in molecules that are invariant enough to make it possible to construct larger molecules (such as polypeptides) by docking fragments taken from smaller systems.[14]

In principle, the shapes of atoms in molecules can be assessed for similarity with the descriptors developed by Mezey.[15] However, a superior quantitative definition of atomic similarity measure can be obtained only by making it directly dependent on the electron densities within the atoms in question. In this paper, such a similarity measure is described and several examples of its application are presented.

Quantitative Measure of Atomic Similarity

A quantitative measure of atomic similarity should satisfy several conditions. First of all, it should be universally applicable to any atom in any molecule and derivable from the respective electronic wave functions without resorting to arbitrary assumptions concerning basic functions, "atomic orbitals", etc. Second, computations of such a measure should not be excessively expensive. Third, the calculated values of atomic similarity should be easily interpretable, preferably with the value of 1 corresponding to identical atoms, and should depend on the mutual orientation of the atoms in question. The last condition, which rules out similarity measures that depend on simple differences of atomic properties (such as energies, etc.), assures that optimal orientations of atomic basins can be obtained by maximizing the atomic similarities.

A similarity measure that satisfies all of the above conditions is given by

$$S_{A(X),B(Y)} = \left[\int_{\Omega_{AB}} \rho_X(\vec{r})\, d\vec{r} \Big/ \int_{\Omega_A} \rho_X(\vec{r})\, d\vec{r}\right]\left[\int_{\Omega_{BA}} \rho_Y(\vec{r})\, d\vec{r} \Big/ \int_{\Omega_B} \rho_Y(\vec{r})\, d\vec{r}\right] \quad (3)$$

where $\Omega_{AB} \equiv \Omega_{BA} = \Omega_A \cap \Omega_B$ is the intersection (in the set-theoretical sense) of the atomic basins Ω_A and Ω_B. The above definition assumes that the atomic basins have their attractors overlapping and their mutual orientation (parametrized by three Euler angles) is such that $S_{A(X),B(Y)}$ is maximized. The so-defined atomic similarity measure is always positive. It equals 1 only if A and B are identical and attains values <1 otherwise, making the interpretation of the computed $S_{A(X),B(Y)}$ straightforward.

Calculation of $S_{A(X),B(Y)}$ involves three steps. First, electronic wave functions of the molecules in question, X and Y, are computed and the corresponding electron densities are obtained. Second, the atomic basins of A and B are determined with standard methods[16] and the numbers of electrons, given by the integrals over Ω_A and Ω_B in eq 3, are calculated. Starting with a reasonable guess, the three Euler angles that describe the mutual orientation of A and B are optimized. The optimization requires repeated calculations of the integrals over $\Omega_{AB} \equiv \Omega_{BA}$ that enter eq 3 and their gradients with respect to the Euler angles. Computation of these quantities would be prohibitively expensive with the old integration algorithm,[16] but the recent substantial progress in evaluation of atomic properties[17] usually makes the calculations of atomic similarity (including the initial guess for the Hessian that is obtained by finite differentiation of the gradients and full optimization that yields $S_{A(X),B(Y)}$ accurate to within 10^{-6}) fast enough to be completed in single minutes of CPU time on a single-processor of a CRAY Y-MP supercomputer.[18]

Table I. Numbers of Electrons in Atoms of Molecules under Study

molecule	atom	N_A	molecule	atom	N_A
C_2H_2	C	6.151	CH_2F_2	C	4.806
	H	0.849		H	0.957
C_2H_4	C	6.044		F	9.640
	H	0.978	CHF_3	C	4.168
C_2H_6	C	5.973		H	0.912
	H	1.009		F	0.640
CH_4	C	6.016	CF_4	C	3.476
	H	0.996		F	9.631
CH_3F	C	5.411			
	H	0.984			
	F	9.637			

Table II. Similarity of Carbon Atoms in Fluoro-Substituted Methanes

	CH_4	CH_3F	CH_2F_2	CHF_3	CF_4
CH_4	1.000				
CH_3F	0.909	1.000			
CH_2F_2	0.816	0.890	1.000		
CHF_3	0.711	0.776	0.865	1.000	
CF_4	0.623	0.667	0.737	0.838	1.000

Results and Discussion

In order to demonstrate the potential usefulness of the atomic similarity measure, we calculated its values for several pairs of carbon, hydrogen, and fluorine atoms. The electron densities were computed for the CH_4, CH_3F, CH_2F_2, CHF_3, CF_4, C_2H_2, C_2H_4, and C_2H_6 molecules. The MP2/6-311G** level of theory was used in both calculations of electron density and geometry optimizations. All calculations were carried out with the GAUSSIAN 90 suite of programs[19] running on a CRAY Y-MP supercomputer.

The calculated number of electrons, eq 1, is compiled in Table I. As expected, F atoms in fluoromethanes are negatively charged because of their strong electron-withdrawing character, resulting in charges on both the C and H atoms that become progressively more positive with the increasing fluorine substitution. The decrease in number of electrons in the H atoms is also observed upon replacing the single C–C bond in ethane by multiple bonds in ethene and ethyne, giving rise to the electric charges for hydrogens that are slightly negative in the first molecule but positive in the others.

For the discussion of atomic similarities, it is convenient to define as ligands all the atoms with the nuclei linked through bond paths to the nucleus of the atom in question. This leads to a definition of formally identical atoms as those with the same nuclei and number of ligands (the ligands themselves do not have to be formally identical). Thus, for example the C atom in CHF_3 has four ligands and is formally identical with the C atom in CH_2F_2, but not with that in C_2H_2. However, the hydrogens in all three of these molecules are formally identical and each have one ligand.

As inspection of Table II reveals, the formally identical carbon atoms in fluoromethanes are actually quite dissimilar. The dissimilarity increases monotonically with the difference in the number of fluorine ligands, resulting in only ca. 62% similarity between the C atoms in CH_4 and CF_4. Substitution of one hydrogen by fluorine decreases the atomic similarity of the central carbon by an amount that increases with the degree of fluorine

---

(13) For a recent review, see: Bader, R. F. W.; Popelier, P. L. A.; Keith, T. A. *Angew. Chem., Int. Ed. Engl.*, in press.

(14) Chang, C.; Bader, R. F. W. *J. Phys. Chem.* **1992**, *96*, 1654.

(15) Mezey, P. G. *J. Comput. Chem.* **1987**, *8*, 462. Mezey, P. G. *J. Math. Chem.* **1988**, *2*, 299, 325.

(16) Biegler-Koenig, F. W.; Bader, R. F. W.; Tang, T. H. *J. Comput. Chem.* **1982**, *3*, 317–328.

(17) Cioslowski, J.; Nanayakkara, A.; Challacombe, M. *Chem. Phys. Lett.* **1993**, *203*, 137. Cioslowski, J. *Chem. Phys. Lett.* **1992**, *194*, 73.

(18) The program SIMATO that carries out the atomic similarity calculations is available from the corresponding author upon request.

(19) Frisch, M. J.; Head-Gordon, M.; Trucks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Robb, M.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R. L.; Kahn, L. R.; Stewart, J. J. P.; Topiol, S.; Pople, J. A. GAUSSIAN 90, Revision F; GAUSSIAN, Inc.: Pittsburgh, PA, 1990.

Table III. Similarity of Hydrogen Atoms in Fluoro-Substituted Methanes

	CH_4	CH_3F	CH_2F_2	CHF_3
CH_4	1.000			
CH_3F	0.988	1.000		
CH_2F_2	0.978	0.981	1.000	
CHF_3	0.951	0.955	0.972	1.000

Table IV. Similarity of Fluorine Atoms in Fluoro-Substituted Methanes

	CH_3F	CH_2F_2	CHF_3	CF_4
CH_3F	1.000			
CH_2F_2	0.997	1.000		
CHF_3	0.996	0.997	1.000	
CF_4	0.995	0.995	0.997	1.000

Table V. Similarity of Carbon Atoms in Simple Hydrocarbons

	C_2H_2	C_2H_4	C_2H_6	CH_4
C_2H_2	1.000			
C_2H_4	0.873	1.000		
C_2H_6	0.803	0.861	1.000	
CH_4	0.821	0.857	0.963	1.000

substitution. Thus, the similarity between the C atoms in CH_4 and CH_3F is ca. 91%, but that for the CHF_3/CF_4 pair equals only ca. 84%.

The H atoms remain largely unchanged with the fluorine substitution on the carbons, as reflected by the values of atomic similarities that are close to 1 (Table III). Even with three hydrogens of CH_4 replaced by F atoms in the CHF_3 molecule, the remaining H atom retains ca. 95% similarity. This small decrease in similarity can be attributed to the F atoms that, although not directly bonded to the hydrogen in question, "eat up" parts of its basin. The presence of this phenomenon is confirmed by the fact that the similarities between the F atoms remain very high (≥99.5%, Table IV).

In agreement with expectations based on "chemical intuition", the similarities of C atoms in simple hydrocarbons (Table V) are quite sensitive to the change in the number of ligands (i.e. the change in hybridization). Interestingly enough, the character of ligands (hydrogens vs carbons) plays only a secondary role as attested by the similarity of the C atoms in the CH_4/C_2H_6 pair that amounts to more than 96%. This observation is also confirmed by the following comparisons between the values of atomic similarities of the C atoms: the CH_4/C_2H_4 vs the C_2H_6/C_2H_4 pair, and the CH_4/C_2H_2 vs the C_2H_6/C_2H_2 pair.

Table VI. Similarity of Hydrogen Atoms in Simple Hydrocarbons

	C_2H_2	C_2H_4	C_2H_6	CH_4
C_2H_2	1.000			
C_2H_4	0.934	1.000		
C_2H_6	0.923	0.979	1.000	
CH_4	0.927	0.985	0.993	1.000

Changes in the hybridization of carbon atoms also affect the similarities of the hydrogen ligands in simple hydrocarbons (Table VI). This fact is particularly well reflected in the similarity of H atoms in C_2H_2 and C_2H_6 that is only ca. 93%. On the other hand, the hydrogens with the formally identical carbon ligands in CH_4 and C_2H_6 are more than 99% similar.

Conclusions

Similarity of atoms in molecules can be quantitatively assessed with a measure that employs electron densities within respective atomic basins. This atomic similarity measure does not rely on arbitrary assumptions concerning basis functions or "atomic orbitals", is relatively inexpensive to compute, and has straightforward interpretation. Inspection of similarities between pairs of carbon, hydrogen, and fluorine atoms in 8 molecules reveals that the atomic similarity is greatly reduced by a change in the number or the character of ligands. On the other hand, atoms with formally identical ligands resemble each other to a large degree, with the similarity indices greater than 0.95 for hydrogens and 0.99 for non-hydrogens.

Because of its quantitative character, the atomic similarity measure is expected to find several applications in chemistry: quantification of atomic chiralities, assembly of molecules from atomic and functional "blocks", and characterization of atomic surfaces being selected examples. The measure can also be very useful for comparing densities within atomic basins calculated at different levels of approximation, allowing for rigorous assessment of the validity of various electronic structure methods in predicting atomic, bond, and group properties.

Acknowledgment. This work was partially supported by the National Science Foundation under Contract No. CHE-9224806, the U.S. DOE through its Supercomputer Computations Research Institute, and the donors of the Petroleum Research Fund, administered by the American Chemical Society (Grant No. PRF 25076-G6).