A Momentum-Space Approach to Molecular Similarity

Neil L. Allan*

School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, U.K.

David L. Cooper

Department of Chemistry, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, U.K.

Received June 22, 1992

We review briefly a novel approach to the estimation of the similarity of two electron distributions. The **momentum-space**, rather than the much more familiar **position-space**, electron densities are compared. Many of the problems associated with the usual position-space procedures are avoided. Recent applications involving (i) the model series CH_3OCH_3, CH_3SCH_3, and $CH_3CH_2CH_3$ and (ii) the C–F and C–H bonds in the series of fluoromethanes CH_xF_y ($0 \leq x \leq 4$; $0 \leq y \leq 4$) are discussed. The latter also makes use of an efficient localization scheme.

INTRODUCTION

It is clear from the wide range of papers in this volume that there are a large number of diverse approaches to the question of molecular similarity. These include database searching,[1,2] topological analysis,[3] and quantum mechanical methods.[4-12] The last of these, associated, for example, with Carbó,[4,5] Richards,[6-8] and Ponec[9,10] and their respective co-workers, typically involves the use of quantum chemical calculations to compare, for example, the position-space (r-space) electron densities or the electrostatic potentials of two (or more) molecules.

We have proposed[11,12] a novel variation of this quantum chemical approach, involving the comparison of the **momentum-space** (p-space) electron densities of the systems of interest. Momentum-space concepts, in which the electron density is expressed as a function of the momenta of the electrons, are common in solid-state physics but are encountered much less frequently in chemistry.

The momentum-space wavefunction is simply the Fourier transform of the position-space function. Molecular orbitals $\psi(\mathbf{r})$ (MOs) are formed by the overlap of atomic basis functions $\phi_\alpha(\mathbf{r})$ centered on nuclei α at positions $\mathbf{R}_\alpha$:

$$\psi(\mathbf{r}) = \sum_\alpha c_\alpha \phi_\alpha(\mathbf{r} - \mathbf{R}_\alpha) \qquad (1)$$

The corresponding momentum-space MO, $\Psi(\mathbf{p})$, is

$$\Psi(\mathbf{p}) = \sum_\alpha c_\alpha \Phi(\mathbf{p}) \exp(-i\mathbf{p}\cdot\mathbf{R}_\alpha) \qquad (2)$$

where the $\Phi(\mathbf{p})$ are the Fourier transforms of the $\phi_\alpha(\mathbf{r})$:

$$\Phi(\mathbf{p}) = (2\pi)^{-3/2} \int \phi_\alpha(\mathbf{r}) \exp(-i\mathbf{p}\cdot\mathbf{r}) \, d\mathbf{r} \qquad (3)$$

The Fourier transform preserves direction so that it is possible to refer to components of the total momentum in any particular direction such as parallel or perpendicular to a bond or plane of symmetry. The contribution to the total self-consistent field (SCF) p-space (momentum) electron density $\rho(\mathbf{p})$ from MO $\Psi(\mathbf{p})$ is simply $\Psi(\mathbf{p})\Psi^*(\mathbf{p})$.

The momentum density falls off rapidly with $p = |\mathbf{p}|$. Consequently, it is dominated by low values of p which correspond to the slowly varying outer valence electron density in r-space. The form of the electron density in r-space is determined largely by the core electrons and consequently by the nuclear positions. $\rho(\mathbf{p})$, in contrast, emphasizes the

chemically interesting features of the valence electron distribution.

In previous work we have discussed in detail the form of the p-space electron density of a range of small molecules.[13] The process of bond formation is shown in a striking way in the momentum-space representation. We have also studied the momentum-space properties of large polyenes.[14] Figure 1 shows plots of the r-space, and Figure 2 shows p-space electron densities for the occupied π-orbitals of lowest and highest binding energy for a *trans*-polyene containing 20 carbon atoms ($C_{20}H_{22}$), calculated using SCF wavefunctions generated using the GAMESS program[15] with the 3-21G basis set.[16] The position-space densities, which are plotted in the plane $z = 1$ bohr, have characteristic nodes precisely as expected. The momentum-space densities are plotted in the plane $p_z = 1$ atomic unit with the origin ($p_x = p_y = 0$) at the CENTER of each view and are very different in appearance from those in r-space. The p-space plots show sharp wedges peaking very sharply at one particular value of $|p_x|$, i.e., at one value of the momentum component ALONG the carbon chain. As expected from Hückel theory,[14] only the π-orbital with the highest binding energy has non-zero momentum at $p_x = 0$. As the binding energies of the π-orbitals decrease, the wedges move further away from the origin. The wedges furthest from the origin occur for the highest occupied molecular orbital (HOMO). The momentum density formalism is also particularly suitable for comparing the electron distributions of systems with different nuclear frameworks; "difference density" plots in position space are complicated by the different positions of the nuclei.

In order to apply momentum-space concepts to the question of molecular similarity, we define a family of molecular dissimilarity indices $D_{AB}(n)$ in terms of the momentum densities of the two molecules:

$$D_{AB}(n) = 100[1/2(I_{AA}(n) + I_{BB}(n)) - I_{AB}(n)] \qquad (4)$$

where

$$I_{PQ} = 2 \int p^n \rho_P(\mathbf{p}) \rho_Q(\mathbf{p}) \, d\mathbf{p} \qquad (PQ = AA, BB, \text{ or } AB) \qquad (5)$$

and the two molecules are labeled A and B. The integrand in eq 5 is always positive.

The relationship between position and momentum spaces is such that it would be possible, of course, to transform eq
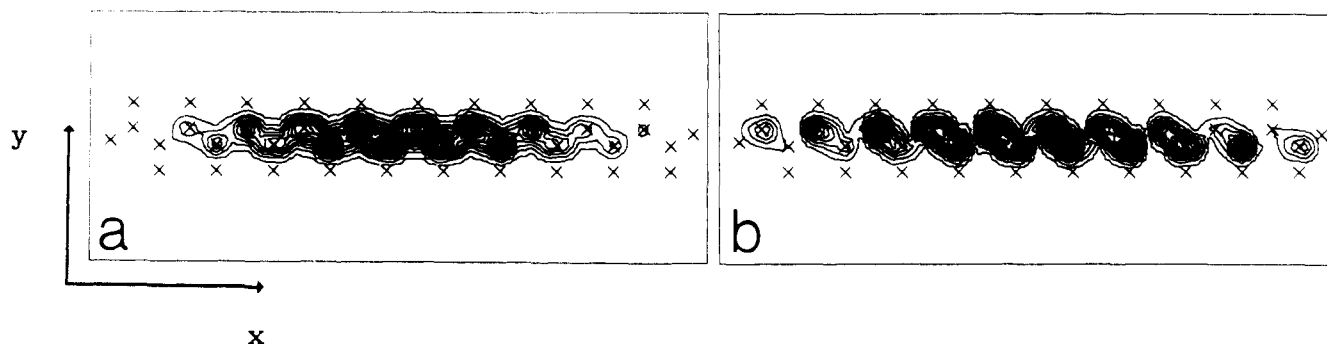
Figure 1. Contour plots (in the plane $z = 1$ bohr) of the r-space density for the occupied π-orbitals of highest (panel a) and lowest (panel b) binding energy for *trans*-$C_{20}H_{22}$. The crosses mark the positions of the nuclei (in the plane $z = 0$).
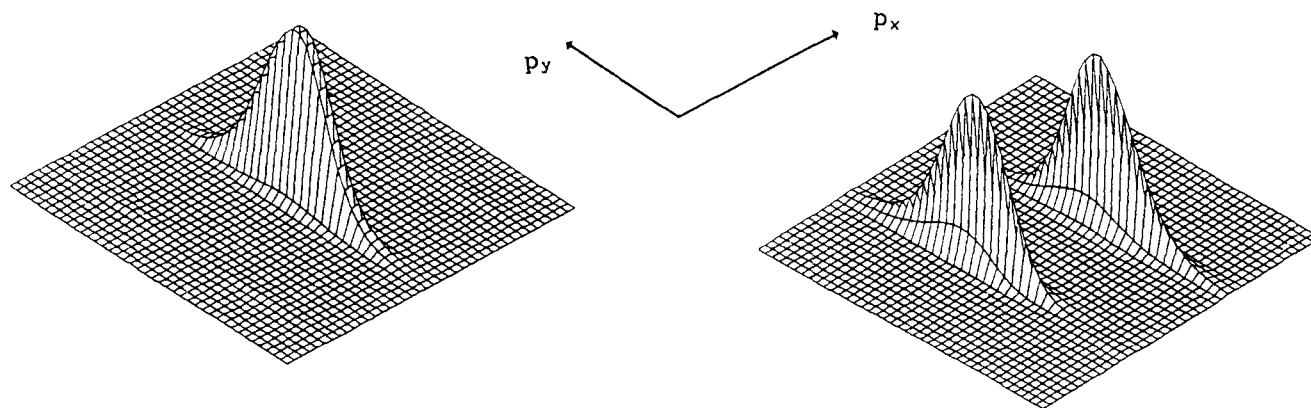

Figure 2. Momentum-space density in the plane $p_z = 1$ atomic unit for the occupied π-orbitals shown in Figure 1, panels a and b, respectively. The origin is at the CENTER of each plot. Each extends to ± 2 atomic units in the p_x and p_y directions. The x, y, and z labels refer to the same orientation of the axes as in Figure 1a.

5 into an equivalent r-space integral. It is however MUCH more straightforward to work directly in p-space with eq 5.

$D_{AB}(n)$ is greater than or equal to zero. Values of 0 indicate that the momentum-space electron densities of the two molecules are identical. There is no upper limit on $D_{AB}(n)$. More positive values of $D_{AB}(n)$ indicate greater dissimilarity. The values of n introduce different powers of p in the integrands and so emphasize different regions of the momentum-space valence electron density. Typical values of n are -1, 0, 1, and 2.

$D_{AB}(n)$ depends only on the relative orientation of the two molecules in position space but not on the distance between them. Most of the usual problems associated with the superposition of the two molecules are therefore avoided. The values of the indices are also insensitive to the core electron densities. Small displacements of heavy atoms in r-space have negligible effects on $D_{AB}(n)$. The calculation of $D_{AB}(n)$ from the electron densities is not computationally expensive compared to the cost of ab initio calculations.

APPLICATIONS

Our first example is the set of model compounds CH_3-OCH_3, CH_3SCH_3, and $CH_3CH_2CH_3$ which have been considered in a range of studies of molecular similarity (for example, refs 6 and 7). The use of -S- as a replacement for -CH_2- is a common procedure in drug design since it often leads to comparable activity. The same is not usually true of -O-. The problem of relative orientation is particularly straightforward for these systems. In each case the z-direction was taken to be that of the C_2-axis, with the y-axis in the σ'_v plane. Semiempirical MNDO wavefunctions were obtained for each molecule using the MOPAC program.[17]

The values of $D_{AB}(n)$ for $n = -1$, 0, 1, and 2 are collected together in Table I. It is clear that the greatest similarity

Table I. Values of $D_{AB}(n)$ for CH_3OCH_3, CH_3SCH_3, and $CH_3CH_2CH_3$ ($n = -1$, 0, 1, or 2)

	n	CH_3SCH_3	$CH_3CH_2CH_3$
CH_3OCH_3	-1	161	137
	0	120	93
	1	109	81
	2	118	85
CH_3SCH_3	-1		5
	0		4
	1		7
	2		12

[smallest $D_{AB}(n)$] is between $CH_3CH_2CH_3$ and CH_3SCH_3 for all these values of n.

We have also compared these molecules of the general formula $(CH_3)_2X$ with H_2O and with H_2S.[11] As expected, these are much less similar, and there is considerable variation of the similarity indices with n. One possible criterion for molecular similarity is that this must be indicated by the values of SEVERAL indices, each sampling different regions of the electron density.

It is particularly pleasing that the same trends in $D_{AB}(n)$ are apparent[12] if the values are calculated from ab initio SCF electron densities (4-31G basis). Naturally, the values are not the same, reflecting both the different qualities of the two wavefunctions and that MNDO is a "valence-electron" (as opposed to "all-electron") technique. The routine study of a large series of large molecules would require the use of the much cheaper semiempirical, rather than ab initio, methods.

It is easy to conceive situations in which it would be more pertinent to compare individual orbitals rather than total electron densities. It would also be useful to transform to momentum space only those parts of the r-space wavefunction which describe the most chemically significant regions of the

Table II. Values of $D_{AB}(n)$ for C–F Bonds in CH_3F, CH_2F_2, CHF_3, and CF_4

	n	CH_2F_2	CHF_3	CF_4
CH_3F	−1	0.2	0.4	1.5
	0	0.1	0.3	0.8
	1	0.2	0.3	0.7
	2	0.2	0.4	0.7
CH_2F_2	−1		0.2	1.2
	0		0.2	0.6
	1		0.2	0.5
	2		0.1	0.5
CHF_3	−1			0.8
	0			0.5
	1			0.4
	2			0.4

Table III. Values of $D_{AB}(n)$ for C–H Bonds in CH_4, CH_3F, CH_2F_2, and CHF_3

	n	CH_3F	CH_2F_2	CHF_3
CH_4	−1	1.1	3.9	13.1
	0	0.7	1.8	4.5
	1	0.5	1.1	2.3
	2	0.4	0.9	1.6
CH_3F	−1		1.9	9.2
	0		0.9	3.3
	1		0.7	1.8
	2		0.5	1.3
CH_2F_2	−1			4.5
	0			1.7
	1			1.1
	2			0.8

molecule, e.g., a particular functional group or side chain. This requires the use of a reliable and computationally cheap scheme to localize the occupied semiempirical or ab initio molecular orbitals.

Our second example is a comparison of the C–F and C–H bonds in the series of fluoromethanes CH_xF_y ($0 \leq x \leq 4$; $0 \leq y \leq 4$). MNDO often produces large errors for highly fluorinated compounds,[18] so we have used the semiempirical method AM1 instead (see ref 17). AM1 wavefunctions were calculated for the series CH_3F, CH_2F_2, CHF_3, and CF_4, with each molecule oriented with the z-axis along a C–F bond. Localized molecular orbitals (LMOs) were generated by the population localization procedure due to Pipek and Mezey[19] in which the quantity

$$Z = \sum_A \sum_i |P_A(i)|^2$$

is maximized, where $P_A(i)$ is the contribution made by each electron in MO ψ_i to the Mulliken population on center A.

We have discussed in detail previously[20] how C–F bonds at primary sites in fluoroalkanes differ remarkably little from one molecule to another. The variation in the C–F bond length in our series is very small. The AM1 method suggests almost identical Mulliken charges on F for CH_3F, CH_2F_2, and CHF_3. The calculated charge on the F atoms in CF_4 is slightly less negative. The Mulliken charges on C increase markedly with increasing fluorination. The values of $D_{AB}(n)$ in Table II are very small for all n and show that all the C–F bonds are very similar. Nevertheless the C–F bonds in "adjacent" molecules in the series CH_3F, CH_2F_2, and CHF_3 are the most similar. Where one of the partners is CF_4, the values of $D_{AB}(n)$ are slightly larger; the largest values are for $n = -1$ so that the electron densities differ most at low momentum. It is clear that $D_{AB}(n)$ can differentiate between the C–F bonds in fluoromethanes.

Turning to the C–H bonds, AM1 wavefunctions for the molecules CH_4, CH_3F, CH_2F_2, and CHF_3 were generated, with each molecule orientated such that one C–H bond lay along the z-axis. LMOs were then produced using the population localization procedure. According to both experiment and AM1, the C–H bond length is insensitive to increasing fluorination. Nevertheless, the C–H bonds show significant variation in chemical reactivity, as is evident, for example, from the rate constants and activation energies for H-atom abstraction by free radicals. The reaction with the OH radical is of particular significance because this is the dominant tropospheric loss process for hydrofluorocarbons. Several models predict much longer residence times in the troposphere for CHF_3 than CH_2F_2 or CH_3F.[21,22] Assuming that there is some correlation between properties of the

molecules themselves and those of the corresponding transition states, we would expect the $D_{AB}(n)$ for the C–H bonds to be larger and also exhibit a larger variation with n than those for the C–F bonds. These trends are clearly apparent from the values of $D_{AB}(n)$ listed in Table III. The p-space electron densities differ most at low momentum. The largest $D_{AB}(n)$ occur for comparisons with CHF_3 (compare also the $D_{AB}(-1)$ values in Table I). The activation energy for the reaction with OH is very much larger, and the rate constant (at 298 K) is very much smaller for CHF_3 than for the other hydrofluoromethanes.[23] The LMO dissimilarities highlight strikingly this much lower reactivity of CHF_3.

CONCLUSIONS

In this short paper, we have proposed that the momentum density formalism is a valuable, but largely unexplored, way of visualizing the valence electron density in large molecules. A formulation of molecular similarity in terms of the momentum-space densities circumvents many difficulties with position-space definitions and is not computationally expensive. It is also possible to compare individual orbitals; electron densities associated with individual bonds can be compared with the use of a suitable localization scheme.

The use of semiempirical wavefunctions and the localization procedure means that we are now in a position to study a large series of large molecules, and such investigations, where the aim is to establish new structure–activity relations, are now in progress.

REFERENCES AND NOTES

(1) Pepperrell, C. A.; Poirrette, A. R.; Willett, P.; Taylor, R. Development of an atom-mapping procedure for similarity searching in databases of three-dimensional chemical structures. *Pestic. Sci.* **1991**, *33*, 97–111.
(2) Pepperrell, C. A.; Willett, P. Techniques for the calculation of three-dimensional structural similarity using inter-atomic distances. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 455–474.
(3) For example, Duane-Walker, P.; Artera, G. A.; Mezey, P. G. A Complete Shape Characterization for Molecular Charge Densities Represented by Gaussian-Type Functions. *J. Comput. Chem.* **1991**, *12*, 220–230.
(4) Carbó, R.; Leyda, L.; Arnau, M. How Similar is a Molecule to Another? An Electron Density Measure of Similarity between Two Molecular Structures. *Int. J. Quant. Chem.* **1980**, *17*, 1185–1189.
(5) Carbó, R.; Calabuig, B. Molecular Quantum Similarity Measures and N-Dimensional Representation of Quantum Objects. I. Theoretical Foundations. *Int. J. Quantum. Chem.* **1992**, *42*, 1681–1693; II. Practical Applications. *Ibid.* **1992**, *42*, 1695–1709.
(6) Bowen-Jenkins, P. E.; Cooper, D. L.; Richards, W. G. Ab Initio Computation of Molecular Similarity. *J. Phys. Chem.* **1985**, *89*, 2195–2197.
(7) Richards, W. G.; Hodgkin, E. E. Molecular similarity. *Chem. Br.* **1988**, *24*, 1141–1144 and references cited therein.
(8) Burt, C.; Richards, W. G. Molecular similarity: The introduction of flexible fitting. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 231–238.

590 *J. Chem. Inf. Comput. Sci., Vol. 32, No. 6, 1992*

ALLAN AND COOPER

(9) Ponec, R.; Strnad, M. Electron Correlation in Pericyclic Reactivity: A Similarity Approach. *Int. J. Quant. Chem.* **1992**, *42*, 501–508.

(10) Ponec, R.; Strnad, M. Topological aspects of chemical reactivity. Evans–Dewar principle in terms of molecular similarity approach. *J. Phys. Org. Chem.* **1991**, *4*, 701–705.

(11) Cooper, D. L.; Allan, N. L. A novel approach to molecular similarity. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 253–259.

(12) Cooper, D. L.; Allan, N. L. Molecular Dissimilarity: A Momentum-Space Criterion. *J. Am. Chem. Soc.* **1992**, *114*, 4773–4776.

(13) Cooper, D. L.; Allan, N. L. Bond Formation in Momentum Space. *J. Chem. Soc., Faraday Trans.* 2 **1987**, *83*, 449–460.

(14) Cooper, D. L.; Allan, N. L.; Grout, P. J. Momentum Space Studies of Large Polyenes. *J. Chem. Soc., Faraday Trans.* 2 **1989**, *85*, 1519–1529.

(15) Guest, M. F; Kendrick, J. GAMESS user manual CCP1/86/1; Daresbury Laboratory, SERC: Daresbury, 1986.

(16) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods. 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265–3269 and references cited therein.

(17) Stewart, J. J. P. MOPAC: A semi-empirical molecular orbital program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1–103.

(18) Dewar, M. J. S.; Rzepa, H. S. Ground states of molecules. 40. MNDO results for molecules containing fluorine. *J. Am. Chem. Soc.* **1978**, *100*, 58–67.

(19) Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* **1989**, *90*, 4916–4926.

(20) (a) Cooper, D. L.; Allan, N. L.; Powell, R. L. Theoretical Studies of Fluorocarbons. I. Small Perfluoroalkane Molecules. *J. Fluorine Chem.* **1990**, *46*, 317–337. (b) Cooper, D. L.; Wright, S. C.; Allan, N. L.; Winterton, N. Theoretical Studies of Fluorocarbons. II. Fluorine and Chlorine Substituted Alkanes. *J. Fluorine Chem.* **1990**, *47*, 489–507. (c) Cooper, D. L.; Allan, N. L.; Powell, R. L. Theoretical Studies of Fluorocarbons. III. Primary, Secondary, Tertiary and Quaternary Centres. *J. Fluorine Chem.* **1990**, *49*, 421–432.

(21) Derwent, R. G.; Volz-Thomas, A.; Prather, M. J. In UNEP/WMO Scientific Assessment of Stratospheric Ozone: Appendix. AFEAS Report. AFEAS: 1989; Chapter V, p 123.

(22) (a) Cooper, D. L.; Allan, N. L.; McCulloch, A. Reactions of hydro-fluorocarbons and hydrochlorofluorocarbons with the hydroxyl radical. *Atmos. Environ.* **1990**, *24A*, 2417–2419 and 2703. (b) Cooper, D. L.; Cunningham, T. P.; Allan, N. L.; McCulloch, A. Tropospheric lifetimes of potential CFC replacements: rate coefficients for reaction with the hydroxyl radical. *Atmos. Environ.* **1992**, *26A*, 1331–1334.

(23) Hampson, R. F.; Kuryio, M. J.; Sander, S. P. In UNEP/WMO Scientific Assessment of Stratospheric Ozone: Appendix. AFEAS Report. AFEAS: 1989; Chapter III, p 47.