Applications of momentum-space similarity

Peter T. Measures\textsuperscript{a}, Katherine A. Mort\textsuperscript{b}, Neil L. Allan\textsuperscript{a,*} and David L. Cooper\textsuperscript{b,*}

\textsuperscript{a}School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, U.K.
\textsuperscript{b}Department of Chemistry, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, U.K.

Received 4 April 1995
Accepted 28 May 1995

Keywords: Molecular similarity; Momentum space; Structure–activity relationships; Drug design; HIV inhibition; Hyperpolarisabilities

Summary

Momentum-space similarity indices were used in studies linking chemical structure to observed activity. These included (a) the biological activity of various molecules that are of interest due to their capacity for HIV inhibition; and (b) the hyperpolarisabilities of series of conjugated molecules. Study (a) included comparisons of the total valence densities of different molecules or the densities associated with particular molecular fragments. Study (b) involved, for each molecule, a comparison of the momentum-space densities of the highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals. The momentum-space approach, which is most sensitive to features of the long-range valence electron density, turned out to be particularly useful for cases such as these, in which the physical property or biological activity has no obvious dependence on the bonding topology of the molecules.

Introduction

Structure–activity relationships are a useful tool in the study of unknown or complex processes. In one way or another, such approaches often involve classifying molecules according to how similar they are to one another. There are many ways of quantifying similarity, including database searching \cite{1}, topological analysis of the three-dimensional shapes of charge densities \cite{2}, and comparisons of electrostatic potentials \cite{3} or position-space electron densities \cite{4,5}. In previous work we have proposed a novel similarity scheme, simple and easy to apply, based on the comparison of momentum-space electron distributions, and we have reported successful applications to model compounds, such as CH\textsubscript{3}XCH\textsubscript{3} (X=O, S, CH\textsubscript{2}) and the hydrofluoromethanes \cite{6-9}. In this paper, we present the results of further studies which involve much larger systems, for which structure activity relationships and molecular similarity concepts could play important roles in rationalising the variation of physical or biological behaviour in series of molecules. The first of these relates to anti-HIV molecules – series of phospholipids and non-nucleoside reverse transcriptase inhibitors. As with many biological processes, the action of these molecules is very complicated or even unknown. Preliminary results for the phospholipids have been presented previously \cite{10}. The second study involves the prediction of molecular hyperpolarisabilities of conjugated systems, such as disubstituted benzenes, styrenes, stilbenes and other diphenyl compounds. Although the methodology for the direct calculation of hyperpolarisabilities is well established, in practice this is difficult and computationally expensive \cite{11,12}.

Momentum-space molecular similarity

The momentum-space wave function can be obtained straightforwardly via Fourier transformation of the position-space wave function \cite{13}. If a molecular orbital \(\psi(r)\) is formed from the overlap of atomic basis functions \(\phi_\alpha(r)\) centred on nuclei \(\alpha\) at positions \(\mathbf{R}_\alpha\), Eq. 1:

\[
\psi(r) = \sum_\alpha c_\alpha \phi_\alpha(r - \mathbf{R}_\alpha)
\]

then the corresponding \(\psi(p)\) is given by:

\[
\psi(p) = \sum_\alpha c_\alpha \Phi_\alpha(p) \exp(-ip \cdot \mathbf{R}_\alpha)
\]

in which \(\Phi_\alpha(p)\) are the Fourier transforms of \(\phi_\alpha(r)\). The relationship in p-space between the wave function and the
total electron density $\rho(p)$ is exactly the same as in $r$-space. For an SCF wave function, the contribution to $\rho(p)$ from an electron in molecular orbital $\psi(p)$ is just $\psi^*(p)\psi(p)$. The momentum density $\rho(p)$ falls off sharply with $p$ [14], so that it is dominated by low values of $p$ which correspond to the slowly varying outer-valence $r$-space electron density. Unlike the position-space density, which is determined largely by the core electrons, $\rho(p)$ emphasises some of the chemically most interesting parts of the electron density.

A quantitative definition of similarity may be obtained by comparing momentum-space density functions, using the analogue of the generalised overlap first proposed by Carbó [4] for position-space densities:

$$I_{AB}(n) = \int p^n \rho_A(p) \rho_B(p) \, dp$$

in which $\rho_A(p)$ and $\rho_B(p)$ are the $p$-space electron densities of molecules A and B. We have included an additional $p^n$ term in the integrand: the inclusion of powers of $p$ allows different regions of momentum space to be emphasised. The value of $n$ is typically chosen as $-1$, as this highlights the slowest moving electrons and gave the most discriminating index in previous work [6].

It is often convenient to scale $I_{AB}(n)$ into the range 0-100%, so that higher values indicate higher similarity. The methods for scaling $I_{AB}(n)$ used here are the Carbó-like index [4,6], Eq. 4:

$$R_{AB}(n) = 100 \frac{I_{AB}(n)}{I_{AA}(n)I_{BB}(n)}$$

and the Tanimoto-like index [9,10], Eq. 5:

$$T_{AB}(n) = 100 \frac{I_{AB}(n)}{(I_{AA}(n) + I_{BB}(n) + I_{AB}(n))}$$

As discussed previously [9], the Tanimoto index is usually the most appropriate for $p$-space similarity work, since it is the most discriminating. However, the Carbó index can be very useful when the shape of the density function is the major consideration, because it is independent of the (non-zero) value of $m$ in $\rho_A(p) = m \times \rho_A(p)$.

For quantities which show very high similarity, such as the $p$-space electron densities for analogous fragments in closely related molecules, it is convenient to dispense with the normalisation of $I_{AB}(n)$ and to use instead a $p$-space dissimilarity index, $D_{AB}(n)$:

$$D_{AB}(n) = 100 \times (I_{AA}(n) + I_{BB}(n) - 2I_{AB}(n))$$

This distance-like index cannot be negative and larger values of $D_{AB}(n)$ imply greater dissimilarity, without an upper bound. The relationships between these different definitions, and that of Hodgkin and Richards [3], are explored in detail in Ref. 9a.

It is possible to determine $I_{AB}(n)$ for total electron densities, total valence electron densities, or the densities associated with one or more orbitals of interest, using wave functions taken from ab initio or semiempirical calculations. Localised orbitals or molecular fragments can be considered [7-9]. When two different molecules are compared, the value of $I_{AB}(n)$ depends on the relative orientation of the two molecules, but not on their relative separation: the momentum density is independent of the choice of $r$-space origin. This is an additional advantage over more conventional position-space measures of molecular similarity.

**Anti-HIV-1 phospholipids and non-nucleoside reverse transcriptase inhibitors**

In this section we consider two series of molecules in relation to their HIV inhibition. These illustrate different uses of the momentum-space (dis)similarity indices, in which we examine the valence electron density associated with the whole molecule or with molecular fragments.

**Lipids**

In our first example, we consider virology data for some phospholipids. The molecules of interest have the general formula given in Scheme 1. Some preliminary results concerning this series have been reported previously [10]. There appears to be no obvious correlation between the $R^1$ and $R^2$ groups and HIV-1 inhibition (see Table 1). At the $R^1$ position, replacement of methyl by tert-butyl results in much greater activity (smaller $ED_{50}$ values) for the most inactive compounds (DD1 and HX1). This substitution also leads to some increase in activity in the cases of OD1 and OL1. On the other hand, EG1 is more active than EG2. When $R^1$ is simply a hydrogen atom, so that there is a free amine group, the compounds tend to have very low $ED_{50}$ values; the exception is HX3 which, unlike HX2, is inactive. Experimental data are not currently available for EG3.

The inhibition mechanism of these phospholipids is not completely understood, although it is thought that they first insert into the membranes of the virus. Molecular similarity concepts are particularly useful in such situations. In view of the size of the molecules, we generated computationally inexpensive $r$-space wave functions from...