Journal of Computer-Aided Molecular Design, 9 (1995) 331–340
ESCOM

J-CAMD 302

Applications of momentum-space similarity

Peter T. Measures[a], Katherine A. Mort[b], Neil L. Allan[a,*] and David L. Cooper[b,*]

[a]*School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, U.K.*
[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 [1], topological analysis of the three-dimensional shapes of charge densities [2], and comparisons of electrostatic potentials [3] or position-space electron densities [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_3XCH_3 (X=O, S, CH_2) and the hydrofluoromethanes [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 [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 [11,12].

Momentum-space molecular similarity

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

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

then the corresponding $\psi(\mathbf{p})$ is given by:

$$\psi(\mathbf{p}) = \sum_\alpha c_\alpha \Phi_\alpha(\mathbf{p}) \exp(-i\mathbf{p} \cdot \mathbf{R}_\alpha) \tag{2}$$

in which $\Phi_\alpha(\mathbf{p})$ are the Fourier transforms of $\phi_\alpha(\mathbf{r})$. The relationship in p-space between the wave function and the

*To whom correspondence should be addressed.

0920-654X/$ 6.00 + 1.00 © 1995 ESCOM Science Publishers B.V.

total electron density $\rho(\mathbf{p})$ is exactly the same as in r-space. For an SCF wave function, the contribution to $\rho(\mathbf{p})$ from an electron in molecular orbital $\psi(\mathbf{p})$ is just $\psi^*(\mathbf{p})\psi(\mathbf{p})$. The momentum density $\rho(\mathbf{p})$ 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, $\rho(\mathbf{p})$ 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:

$$I_{AB}(n) = \int p^n \rho_A(\mathbf{p})\, \rho_B(\mathbf{p})\, d\mathbf{p} \qquad (3)$$

in which $\rho_A(\mathbf{p})$ and $\rho_B(\mathbf{p})$ 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:

$$R_{AB}(n) = 100\, I_{AB}(n) / \sqrt{I_{AA}(n) I_{BB}(n)} \qquad (4)$$

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

$$T_{AB}(n) = 100\, I_{AB}(n) / \left(I_{AA}(n) + I_{BB}(n) + I_{AB}(n)\right) \qquad (5)$$

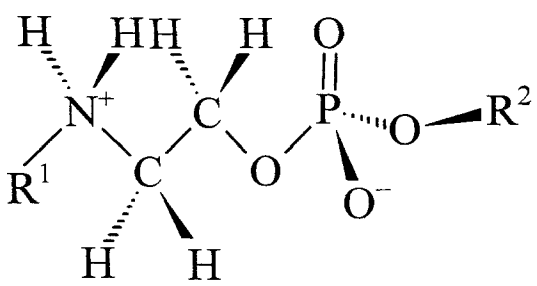
As discussed previously [9], the Tanimoto index is usually the most appropriate for p-space similarity work, since it is the most discriminating. However, the Carbó index can be very useful when the shape of the density function is the major consideration, because it is independent of the (non-zero)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

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, $D_{AB}(n)$:

$$D_{AB}(n) = 100 \times \left(I_{AA}(n) + I_{BB}(n) - 2I_{AB}(n)\right) \qquad (6)$$

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.



Scheme 1. General formula of phospholipids.

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

TABLE 1
EXPERIMENTAL ED_{50} VALUES (FROM REF. 10) FOR THE INHIBITION OF HIV-1 IN C8166 T-LYMPHOBLASTOID CELLS FOR A SERIES OF PHOSPHOLIPIDS

Mnemonic	R^1	R^2	ED_{50} (μM)
HX1	methyl	*n*-hexyl	>200
DD1	methyl	*n*-dodecyl	>200
OD1	methyl	*n*-octadecyl	25
EG1	methyl	ethyl glycolate	110
OL1	methyl	oleyl	10
HX2	*t*-butyl	*n*-hexyl	40
DD2	*t*-butyl	*n*-dodecyl	10
OD2	*t*-butyl	*n*-octadecyl	3
EG2	*t*-butyl	ethyl glycolate	200
OL2	*t*-butyl	oleyl	3
HX3	hydrogen	*n*-hexyl	>200
DD3	hydrogen	*n*-dodecyl	4
OD3	hydrogen	*n*-octadecyl	3.5
OL3	hydrogen	oleyl	0.5

semiempirical MNDO geometry optimisations using the MOPAC code [15]. No search for the global minimum conformation was carried out. The conformation used for the framework common to all systems studied is as shown in the general formula in Scheme 1.

Since it is likely that the membrane lipids must be mimicked by the phospholipids, it seems appropriate to compare total densities for the complete molecules. In our previous work, all molecules were compared with the inactive compound HX1 and with the highly active compound OL2. We found that active compounds have a high similarity to OL2 and a low similarity to HX1. Since the $T_{AB}(-1)$ values for the series with a free amine group were calculated before virology data became available, we were able to predict correctly that OL3 and OD3 would have small ED_{50} values, but that HX3 would be inactive. We have now calculated values of $T_{AB}(-1)$ for all pairs of molecules, with the phospholipids aligned so as to match as closely as possible the positions of atoms common to all the systems. The resulting matrix is represented pictorially in Fig. 1, in which we have replaced the plethora of numerical results by different degrees of shading. Black denotes very high similarity (≥91%) and white denotes low similarity (≤60%). The bars next to the labels indicate that the activity is high ($ED_{50} \leq 10$ μM), medium, or low ($ED_{50} \geq 110$ μM). The same figure is obtained whether the molecules are compared in zwitterionic or neutral form. All numerical values are available from the authors on request [16]. In general, with the exception of the DD series which is shown separately, active molecules are similar to each other and dissimilar to inactive molecules. Inactive molecules are similar to each other and dissimilar to active molecules. It is clear from Fig. 1 that HX2, displaying medium activity, occupies an intermediate position between active and inactive sets. The results for the DD series are not as clear-cut. Although active DD3 and DD2 are similar to other active molecules, so is inactive DD1. These results suggest that an experimental re-examination of the activity of DD1 could be very worthwhile.

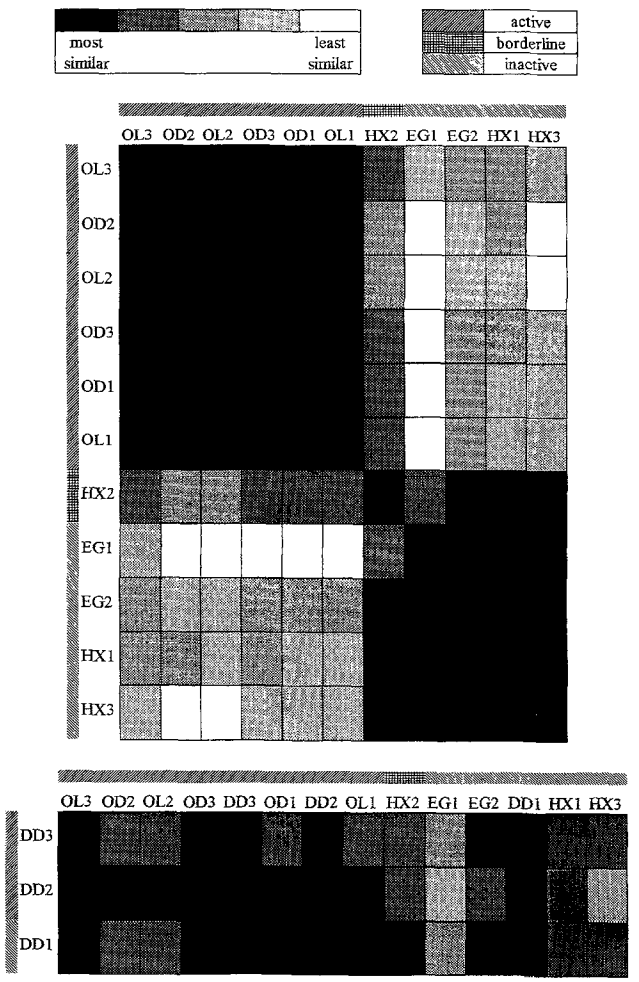


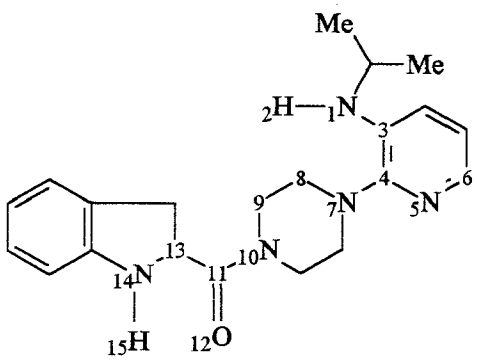
Fig. 1. Values of the molecular similarity index $\overline{T}_{AB}(-1)$ for various phospholipids.

Non-nucleoside reverse transcriptase inhibitors (NNRTIs) of HIV

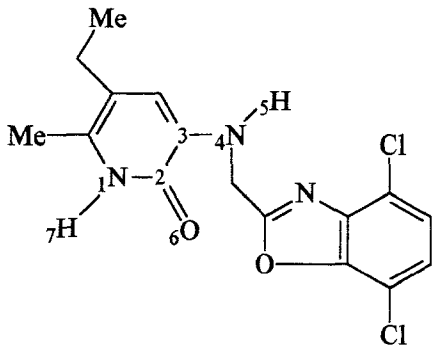
Our second example concerns the molecules shown in Fig. 2, which all inhibit the action of the reverse transcriptase enzyme, in spite of their very different chemical structures [17,18]. It is worthwhile to inquire whether our measures of momentum-space similarity can identify a feature which is common to these various NNRTIs. Initial energy minimisations using the DTMM program [19] were followed by semiempirical MNDO geometry optimisations using the MOPAC code [15], without searching for global minimum conformations. Details of all geometries employed in the present work are available from the authors on request.

All of the molecules feature at least two nitrogen atoms in close proximity and, accordingly, we have examined various groups of atoms which include two or three nitrogens. As in previous work on model series [9], this was done simply by considering the contribution to the total density only from those basis functions associated with the atoms of interest. The relative orientation of the molecules was established as follows: with the first three atoms in the chosen group labelled i, j and k, the vector $\mathbf{r}_{ij}$ defines the z-axis and $\mathbf{r}_{ij} \times \mathbf{r}_{jk}$ the y-axis. Since the groups are very similar to one another, the values of normalised indices such as $T_{AB}(-1)$ are very high indeed, and so it is somewhat more convenient to work with the dissimilarity index $D_{AB}(-1)$. The various results are summarised in Table 2. In those cases where heavy atoms are listed, any directly attached hydrogen atoms not explicitly shown in the figure were also included in the comparisons.

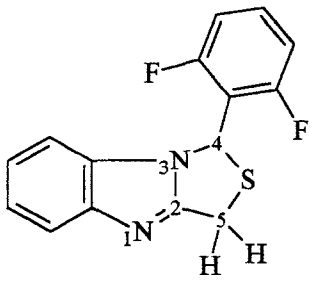
The chosen groups in L-697,661, nevirapene and 'NSC' are all very similar to one another, with values of $D_{AB}(-1)$ less than 16. In the cases of BHAP and TIBO, we exam-



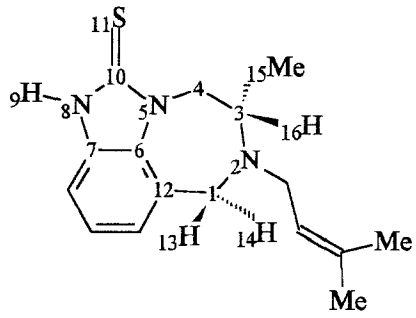
BHAP



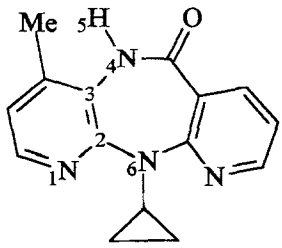
L-697,661



'NSC'



TIBO



Nevirapene

Fig. 2. Non-nucleoside reverse transcriptase inhibitors considered in this work, shown with appropriate atom labels.

TABLE 2
$D_{AB}(-1)$ VALUES FOR NON-NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS OF HIV

System	Fragment	L-697,661	Nevirapene	'NSC'
L-697,661	1–7	0	14.2	15.9
Nevirapene	1–6	14.2	0	11.2
'NSC'	1–5	15.9	11.2	0
BHAP:a	3–7	23.0	12.6	1.0
BHAP:b	10,11,13,14,12,15	25.2	38.9	88.0
BHAP:c	1,3,4,5,2,7	85.5	38.9	31.0
BHAP:d	7–10	409.3		
TIBO:a	12,6–9,5	23.0	8.4	2.6
TIBO:b	8,10,5,6,4,11,9	79.2	86.1	58.7
TIBO:c	8,7,6,5,4,9	80.4	86.1	
TIBO:d	6–11,5	81.3	87.2	
TIBO:e	4–9	81.3	93.0	
TIBO:f	1–5,10,13–16	178.7	190.7	
TIBO:g	1,12,6,5,10	154.0		

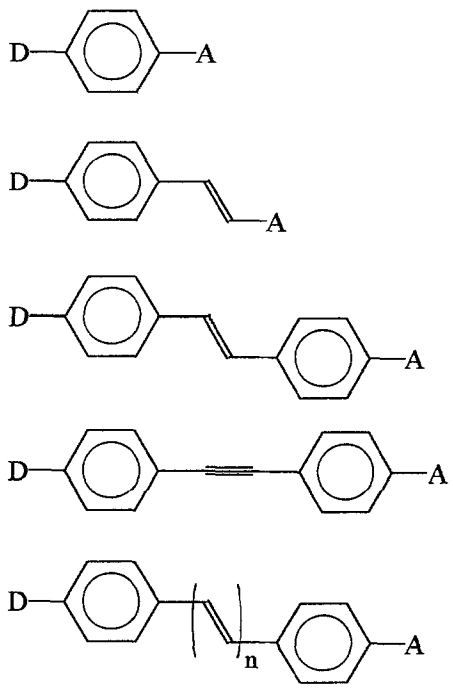
See Fig. 2 for the atom labelling. Blank entries indicate that $D_{AB}(-1)$ was not calculated.

ined various possible choices, each containing two or three nitrogen atoms. Comparisons with L-697,661 (see Table 2) suggest that we should reject BHAP:c and BHAP:d. Further comparisons with nevirapene suggest BHAP:a to be the most appropriate choice; this is confirmed by the $D_{AB}(-1)$ values relative to 'NSC'. Comparisons of the various TIBO possibilities with L-697,661 or nevirapene (see Table 2) suggest that TIBO:a is the most appropriate choice. As expected, we observe a very low value of $D_{AB}(-1)$ between TIBO:a and 'NSC' (see Table 2). Similarly, the dissimilarity between BHAP:a and TIBO:a is $D_{AB}(-1) = 16.0$, comparable to the corresponding values between L-697,661, nevirapene and 'NSC'. We notice that the comparisons with L-697,661 appear to be slightly less discriminating than those with nevirapene or 'NSC' for selecting the preferred groups in BHAP and TIBO.

It is tempting to believe that the low dissimilarity indices observed for the chosen groups in L-697,661, nevirapene, 'NSC', BHAP:a and TIBO:a have some bearing on the observed biological activity of these structurally very different compounds. It could be very interesting to use values of $D_{AB}(-1)$ to screen new compounds for the relevant electronic feature. However, all of these inhibitors are noncompetitive, in that they do not bind to the substrate binding site of the enzyme but elsewhere, probably altering the secondary structure of the enzyme so that it can no longer function. Once the drug molecule is bound to the enzyme, this deactivation could be due to other electronic structure features in the molecule, or it might be purely steric. It may well be that the common feature we have identified is a *recognition site* for binding to the enzyme, but that it does not determine the steric interference subsequently required for deactivation. This could be one explanation for the fact that our $D_{AB}(-1)$ values for TIBO:a groupings in various TIBO derivatives show no obvious correlation with the observed marked variation in bioactivity [20].

Hyperpolarisabilities

Materials with high nonlinear optical (NLO) properties have many practical applications in laser physics and photonics [21,22]. The development of NLO devices has been limited by a lack of compounds with the desired chemical and physical properties. Materials must possess a crystal structure with the appropriate symmetry, consist of molecules with large NLO coefficients and have suitable mechanical properties. The ability to control the alignment of the chromophores in the medium is relative-



Scheme 2. General formulae of the benzenes, styrenes, stilbenes, diphenylacetylenes and diphenylpolyene oligomers considered in this work. D and A denote donor and acceptor groups, respectively.

TABLE 3
EXPERIMENTAL VALUES[a] OF THE HYPERPOLARISABILITY β AND CALCULATED VALUES OF $R_{AB}(-1)$ AND Ω

Class of compounds	Acceptor	Donor	β (expt.; 10^{-30} esu)	$R_{AB}(-1)$	Ω[b]
Benzenes	CN	Cl	0.8	41.8	0.51
	CN	Me	0.7	43.9	0.53
	CN	NH_2	3.1	46.9	0.65
	CN	NMe_2	5.0	48.8	0.69
	CN	OMe	1.9	44.9	0.59
	CN	OPh	1.2	44.7	0.58
	COH	Me	1.7	46.8	0.57
	COH	NMe_2	6.3	52.9	0.76
	COH	OMe	2.2	48.4	0.63
	COH	OPh	1.9	48.0	0.63
	NO_2	Me	2.1	43.1	0.53
	NO_2	NH_2	9.2	54.6	0.78
	NO_2	NMe_2	12.0	56.7	0.84
	NO_2	OH	3.0	50.3	0.67
	NO_2	OMe	5.1	50.9	0.69
	NO_2	OPh	4.0	50.8	0.68
Stilbenes	CN	NMe_2	36	55.4	1.05
	CN	OH	13	53.7	0.95
	CN	OMe	19	53.9	0.97
	NO_2	Br	14	54.4	0.96
	NO_2	Me	15	55.0	0.99
	NO_2	NH_2	40	57.5	1.17
	NO_2	NMe_2	73	57.9	1.19
	NO_2	OH	17	56.0	1.01
	NO_2	OMe	28	57.3	1.06
	NO_2	OPh	18	55.5	1.03
Diphenylacetylenes	CN	NH_2	20	63.5	1.11
	CN	NHMe	27	63.7	1.16
	CN	NMe_2	29	64.0	1.17
	NO_2	OMe	14	63.8	1.17
	NO_2	Br	10	63.5	1.06
	NO_2	NH_2	40, 24	64.1	1.26
	NO_2	NHMe	46	64.4	1.29
	NO_2	NMe_2	46	64.4	1.29
Styrenes	CN	NMe_2	23	55.1	0.90
	CN	OMe	7.0	52.5	0.86
	COH	Br	6.5	56.2	0.81
	COH	OMe	11	53.6	0.82
	COH	NMe_2	30	56.2	0.92
	NO_2	NMe_2	50	59.1	1.03
	NO_2	OH	18	55.8	0.89
	NO_2	OMe	17	56.1	0.89
Diphenylpolyenes, n = 2	CN	OMe	27	58.5	1.11
	NO_2	OMe	47	60.4	1.23
	NO_2	NMe_2	107	61.8	1.36
Diphenylpolyenes, n = 3	CN	OMe	40	62.5	1.24
	NO_2	OMe	76	63.6	1.36
	NO_2	NMe_2	131	64.8	1.48
Diphenylpolyenes, n = 4	NO_2	OMe	101	65.6	1.45
	NO_2	NMe_2	190	66.7	1.56
Diphenylpolyyne, n = 2	NO_2	NH_2	28	69.9	1.42

[a] Experimental values were taken from Ref. 23.
[b] $\Omega = R_{AB}(-1)/(\varepsilon_1 - \varepsilon_0)^2$.

ly unrefined, so that the bulk of chemical research, both experimental [23,24] and theoretical [25], has been targeted at improving the molecular hyperpolarisabilities of the constituent molecules.

The different orders of linear and nonlinear response of the molecular dipole component, μ_i, to an electric field **E** can be expressed as a power series [26]:

$$\mu_i = \mu_{0i} + \alpha_{ij} E_j + \beta_{ijk} E_j E_k + \gamma_{ijkl} E_j E_k E_l + \ldots \quad (7)$$

where μ_0 is the permanent dipole moment, α the polarisability, and β and γ the first- and second-order hyperpolarisabilities, respectively. The principal concern of this and other studies [25] is the lowest nonlinear response β, which is zero if a molecule has inversion symmetry. Molecules with high values of β often have asymmetric electron distributions, arising from conjugated organic frameworks separating electron donor and acceptor groups. Examples include 1,4-disubstituted benzenes, 4,β-substituted styrenes, 4,4'-disubstituted stilbenes, 4,4'-disubstituted diphenylacetylenes and 4,4'-disubstituted α,ω-diphenylpolyene oligomers. General structures of these compound groups are depicted in Scheme 2. In these structures, A and D denote acceptor and donor groups, respectively. Here we concentrate on the systems, listed in Table 3, for which the hyperpolarisabilities have been measured experimentally by Cheng and co-workers [23].

The first-order hyperpolarisability β can, of course, be calculated from first principles [25], either from the field-modified Hartree–Fock orbitals [12] or by the 'sum-over-states' method [11], which treats the interaction of the dipole with the electric field as a time-independent perturbation. The direct calculation of β by either means is difficult and time consuming and the results are very sensitive to the quality of the wave function. An alternative approach is not to attempt to calculate values of β a priori, but to look instead for an empirical correlation between β and some quantity which can be evaluated easily. Of course, it is desirable that this quantity be relatively insensitive to the quality of the wave function. In looking for a structure–activity relationship, we have been guided by the two-state model, which is applicable when the molecule shows a strong charge-transfer interaction. In such a case, the 'sum-over-states' is dominated by the first excited state and it follows that:

$$\beta \propto \frac{|\langle 0|\mu|1\rangle|^2}{(\varepsilon_1 - \varepsilon_0)^2}\{\langle 1|\mu|1\rangle - \langle 0|\mu|0\rangle\} \quad (8)$$

Cheng et al. [23] have suggested that the two-state model can be applied to disubstituted benzenes and that even when it appears to be inapplicable, the model can still serve as a useful qualitative guide to the variation of β from one system to another. In the simplest treatment, the excited state arises by excitation of an electron from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO), and so it is logical to compare the HOMOs and the LUMOs of the molecules of interest.

All wave functions were generated using semiempirical MNDO and AM1 geometry optimisations, as implemented in the well-established MOPAC program [15]. Some of the geometries calculated by MNDO exhibited serious errors. The coordination of the nitrogen atom in Ph-NMe_2 should be planar [27], whereas MNDO incorrectly predicted a pyramidal geometry. For Ph-OMe and Ph-OH, MNDO erroneously predicted a value of 90° for the C(H)-O-C-C torsion angle. In all these cases, the appropriate experimentally determined values [28] were substituted for the relevant parameters.

In preliminary work [9], we considered only 1,4-benzene derivatives and, for each molecule, calculated the similarity index $R_{AB}(-1)$ between its HOMO and LUMO by numerical integration, considering only contributions from basis functions centred on the structural unit common to all these molecules, i.e., the atoms in the benzene ring. Clearly, there is no problem in deciding the relative orientation in the present case. R_{AB} was chosen as the most appropriate similarity index since, as discussed above, this index is predominantly shape-, rather than size-sensitive. A reasonable correlation was obtained between $R_{AB}(-1)$ and β for 1,4-benzene derivatives. The correlation between the inverse of the square of the HOMO/LUMO energy separation and β was much poorer.

We have now established useful empirical correlations for a much larger set of molecules. For the 1,4-disubstituted benzenes, 4,β-substituted styrenes, 4,4'-disubstituted stilbenes and 4,4'-disubstituted diphenylacetylenes, we have considered a range of donor and acceptor groups (see Table 3). Inspired by the form of Eq. 8, we have divided the values of $R_{AB}(-1)$, obtained as before, by the square of the HOMO/LUMO energy separation:

$$\Omega = \frac{R_{AB}(-1)}{(\varepsilon_1 - \varepsilon_0)^2} \quad (9)$$

In Fig. 3 we present results only for MNDO wave functions, as the values of Ω obtained from AM1 wave functions are similar. All the experimental values of β were taken from Ref. 23. It is clear from Figs. 3a–d that a reasonable correlation exists for each series taken separately. For all the molecules, these correlations are more successful than those between either β and $R_{AB}(-1)$ alone or β and $1/(\varepsilon_1 - \varepsilon_0)^2$. The least-squares fitted curves in Fig. 3 take the following form:

$$\beta = a\,\Omega^2 + b\,\Omega + c \quad (10)$$

with coefficients and root-mean-square (rms) deviations in β as listed in Table 4. It is worth noting, for each of

TABLE 4
COEFFICIENTS FOR EQ. 10 AND RMS DEVIATIONS IN β FOR THE FITTED CURVES IN FIG. 3

Series	a	b	c	rms	Figure
1,4-Disubstituted benzenes	103.784	−106.784	28.5793	0.56708	3a, full curve
4,4'-Disubstituted stilbenes	635.333	−1169.27	551.945	6.6112	3b, full curve
4,β-Substituted styrenes	406.929	−562.463	198.550	2.2161	3c, full curve
4,4'-Disubstituted diphenylacetylenes	301.912	−580.066	288.424	6.3504	3d, full curve
4,4'-Disubstituted diphenylacetylenes	−112.405	415.962	−304.092	0.77580	3d, dashed curve

these series of molecules, and particularly for the benzene derivatives, that there is a wider scatter when calculated values of β, obtained directly using MOPAC and the finite field approach, are plotted versus the corresponding experimental values. A minor point is that the experimental value of β for the outlier diphenylacetylene in Fig. 3d ($A=NO_2$, D=OMe) appears somewhat small, considering the relative position of the corresponding molecules with the same donor and acceptor groups in the other three series. At first sight, the full curve shown in Fig. 3d is the least convincing of all fits. The points marked * relate to the same molecule ($A=NO_2$, $D=NH_2$), but in different solvents. If the measurement with the lower value of β is omitted, as well as the outlier ($A=NO_2$, D=OMe), it is possible to fit the dashed curve in Fig 3d, which has an rms deviation in β of 0.78.

Longer conjugated systems, such as the 4,4'-disubstituted α,ω-diphenylpolyene oligomers, are of considerable interest since some of them have large values of β. We have considered the effect of increasing the conjugation length for three families, each of which consists of a disubstituted benzene, a styrene, and diphenylpolyenes (up to n=4) with a *fixed* choice of donor and acceptor groups. The systems studied, together with the corresponding

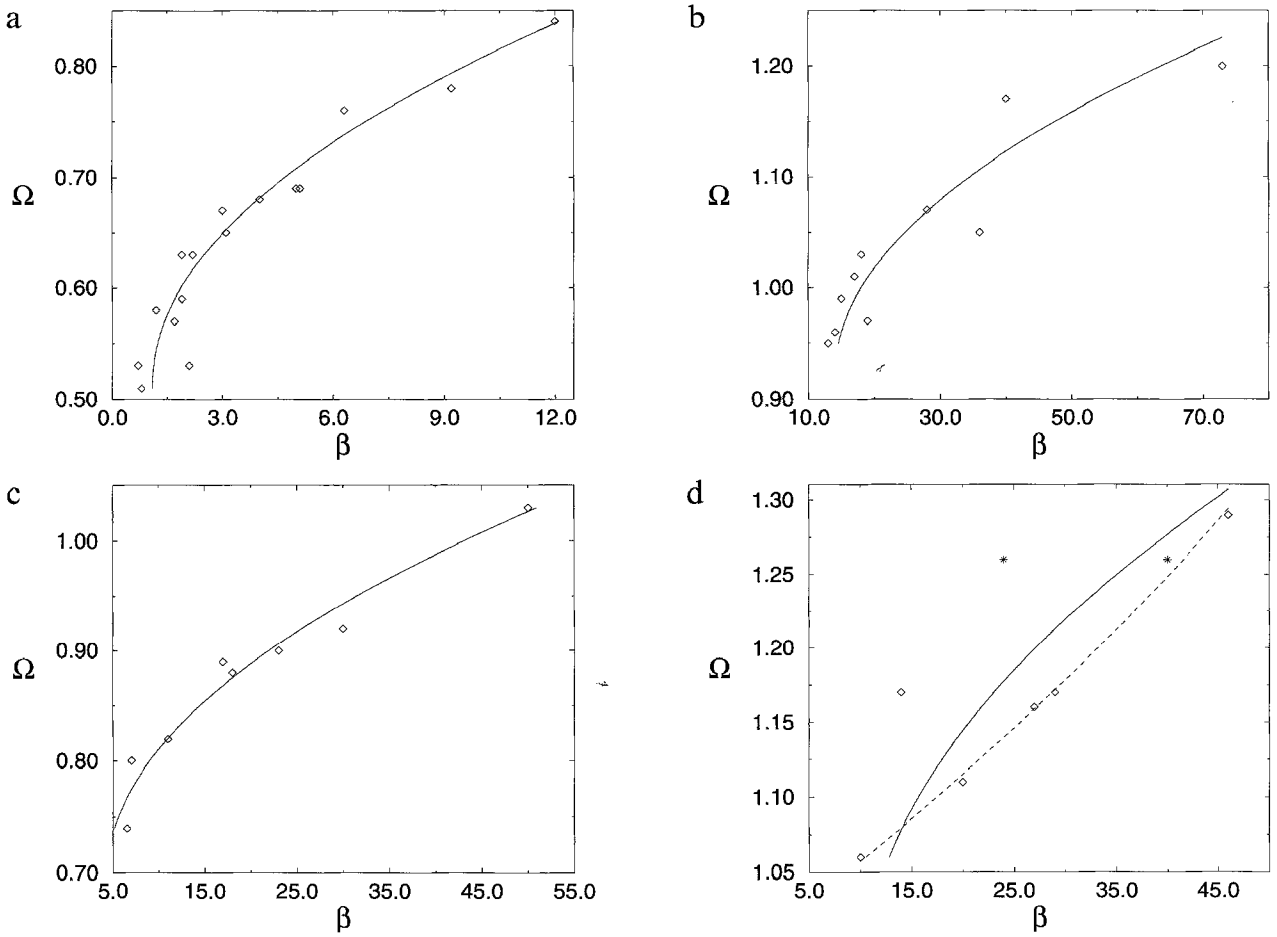


Fig. 3. Calculated values of Ω (defined in Eq. 9) versus experimental values of β (in 10^{-30} esu) taken from Ref. 23. (a) 1,4-Disubstituted benzenes; (b) 4,4'-disubstituted stilbenes; (c) 4,β-substituted styrenes; (d) 4,4'-disubstituted diphenylacetylenes. The two points marked * in Fig. 3d are for $A=NO_2$ and $D=NH_2$ in different solvents. Details of the fitted curves are given in the text.

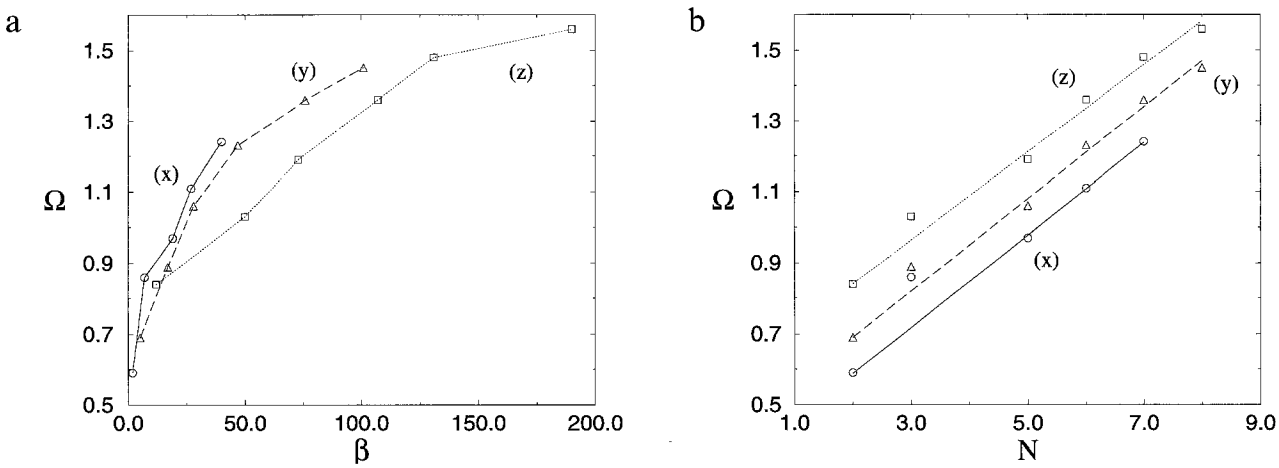


Fig. 4. Effect of increasing conjugation length with fixed donor and acceptor groups. (a) Ω versus β (in 10^{-30} esu, from Ref. 23); (b) Ω versus conjugation length N (see text). The three series are: (x) A=CN, D=OMe; (y) A=NO_2, D=OMe; (z) A=NO_2, D=NMe_2. Ω is defined in Eq. 9.

values of $R_{AB}(-1)$, are collected in Table 3. The variation of Ω with β, shown in Fig. 4a, is similar for each family. Figure 4b presents a plot of Ω versus the conjugation length N, defined as in Ref. 23, according to which a benzene ring contributes two units and an olefinic double bond contributes one. These plots are clearly linear for each family. It is interesting to note that in each family the point furthest from the straight line is the styrene (n=1), as was also observed in the linear plots of log(β) versus log(N) presented in Ref. 23, and we have not included these points in our linear fits.

There has been recent interest in disubstituted diphenylpolyynes, in which the aliphatic carbon chain consists entirely of alternate single and triple bonds [24,29; Kanis, D.R. et al., to be published]. Experimental values of β for two molecules [24] and some theoretical evidence [29; Kanis, D.R. et al., to be published] appear to suggest that, although a single C≡C unit results in a value of β comparable with that for the equivalent stilbene, the increase in β with conjugation length is much smaller than for the polyenes. We note in passing that the values of β calculated for polyynes by Matsuzawa and Dixon [29] appear to have a much more marked increase with carbon chain length than we had expected from the logarithmic relationship quoted by these authors. In view of apparently very limited experimental data, we have studied only one diphenylpolyyne with more than one C≡C unit (see the last entry in Table 3). The value of Ω is considerably larger for this molecule than for the corresponding diphenylacetylene. It will be interesting to consider such molecules further in future work.

Conclusions

We have applied our momentum-space similarity approach to problems such as HIV inhibition and hyperpolarisabilities, where the systems of interest are much larger than those considered previously. Bioactive phospholipids have been shown to be similar to one another, and dissimilar to those that are inactive. Comparisons of the densities associated with fragments of non-nucleoside reverse transcriptase inhibitors may also have provided information useful in the context of their bioactivity. For conjugated organic molecules we have proposed a series of structure–activity relationships relating the hyperpolarisability β to p-space indices.

It is straightforward in our approach to obtain a quantitative answer as to the similarity of two molecules, two orbitals or two molecular fragments. Momentum-space similarity and dissimilarity indices tend to be especially useful when bonding topology and the nature of the atomic backbone are much less important than the long-range valence electron density. In general, the method can be particularly effective for situations where conventional chemical intuition is insufficient.

The choice of relative orientation has been relatively straightforward in the applications described here. Conformational flexibility is more of a problem, particularly when the molecules are *very* similar. The p-space indices for total densities tend to be rather high, even when the atomic backbones are very different. Of course, it is not hard to anticipate situations for which a p-space approach is unlikely to be useful. Obvious examples include those where the activity is largely dependent on the values of bulk physical properties or on well-defined steric effects. A comparative methodology such as ours provides only very indirect information concerning interactions with biological receptors.

In conclusion, our unconventional approach to quantum molecular similarity, based on indices derived from momentum-space electron densities, appears to show considerable promise for a wide range of applications. It is a useful addition to conventional similarity techniques as a means of rationalising variations in physical or biological behaviour, and we anticipate using it extensively in future studies.

References

1 a. Johnson, M.A. and Maggiora, G.M. (Eds.) Concepts and Applications of Molecular Similarity, Wiley, New York, NY, 1990.
b. J. Chem. Inf. Comput. Sci., Vol. 32, 1992, pp. 577–768.

2 Duane-Walker, P., Artera, G.A. and Mezey, P.G., J. Comput. Chem., 12 (1991) 220.

3 a. Hodgkin, E.E. and Richards, W.G., Int. J. Quantum Chem., Quantum Biol. Symp., 14 (1987) 105.
b. Richards, W.G. and Hodgkin, E.E., Chem. Br., 24 (1988) 1141. See also: Burt, C. and Richards, W.G., J. Comput.-Aided Mol. Design, 4 (1990) 231.

4 a. Carbó, R., Leyda, L. and Arnau, M., Int. J. Quantum Chem., 17 (1980) 1185.
b. Carbó, R. and Domingo, Ll., Int. J. Quantum Chem., 32 (1987) 517.
c. Carbó, R. and Calabuig, B., Int. J. Quantum Chem., 42 (1992) 1681.
d. Carbó, R. and Calabuig, B., Int. J. Quantum Chem., 42 (1992) 1695.

5 a. Ponec, R. and Strnad, M., J. Phys. Org. Chem., 4 (1991) 701.
b. Ponec, R. and Strnad, M., Int. J. Quantum Chem., 42 (1992) 501.

6 Cooper, D.L. and Allan, N.L., J. Comput.-Aided Mol. Design, 3 (1989) 253.

7 Cooper, D.L. and Allan, N.L., J. Am. Chem. Soc., 114 (1992) 4773.

8 Allan, N.L. and Cooper, D.L., J. Chem. Inf. Comput. Sci., 32 (1992) 587.

9 a. Cooper, D.L. and Allan, N.L., In Carbó, R. (Ed.) Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches, Kluwer, Dordrecht, 1995, pp. 31–55.
b. Allan, N.L. and Cooper, D.L., Top. Curr. Chem., 173 (1995) 85.

10 Cooper, D.L., Mort, K.A., Allan, N.L., Kinchington, D. and McGuigan, C., J. Am. Chem. Soc., 115 (1993) 12615.

11 Li, D., Marks, T.J. and Ratner, M.A., J. Am. Chem. Soc., 110 (1988) 1707.

12 Sekino, H. and Bartlett, R.J., J. Chem. Phys., 85 (1986) 976.

13 See for example: Coulson, C.A. and Duncanson, W.E., Proc. Camb. Philos. Soc., 37 (1941) 55, 67, 74, 397, 406; Epstein, I.R. and Tanner, A.C., In Williams, B.G. (Ed.) Compton Scattering, McGraw-Hill, New York, NY, 1977, p. 209; Rawlings, D.C. and Davidson, E.R., J. Phys. Chem., 89 (1985) 969.

14 Kaijser, P. and Smith Jr., V.H., Adv. Quantum Chem., 10 (1977) 37.

15 Stewart, J.J.P., J. Comput.-Aided Mol. Design, 4 (1990) 1, and references cited therein.

16 An alternative representation of Fig. 1 is available on the WWW with URL http://rs2.ch.liv.ac.uk/dlc/Lipids.html. In this complementary representation, different degrees of similarity have been represented by different colours.

17 Jeffries, D., Chem. Industry, (1993) 746, and references cited therein.

18 A great deal of information is available via the National Library of Medicine gopher service with URL gopher://gopher.nlm.nih.gov:70.

19 Crabbe, M.J.C., Appleyard, J.R. and Lay, C.R., Desktop Molecular Modeller (V. 3.0 for Windows), Oxford University Press, Oxford, 1994.

20 Pauwels, R., Andries, K., Desmyter, J., Schols, D., Kukla, M.J., Breslin, H.J., Raeymaeckers, A., Van Gelder, J., Woestenborghs, R., Heykants, J., Schellekens, K., Janssen, M.A.C., DeClercq, E. and Janssen, P.A.J., Nature, 343 (1990) 470.

21 Li, D., Marks, T.J. and Ratner, M.A., Chem. Phys. Lett., 131 (1986) 370.

22 Dirk, C.W., Tweig, R.J. and Wagniere, G., J. Am. Chem. Soc., 108 (1986) 5387.

23 Cheng, L., Tam, W., Stevenson, S.H., Meredith, G.R., Rikken, G. and Marder, S.R., J. Phys. Chem., 95 (1991) 10631.

24 Stiegman, A.E., Graham, E., Perry, K.J., Khundkar, L.R., Cheng, L.-T. and Perry, J.W., J. Am. Chem. Soc., 113 (1991) 7658.

25 Kanis, D.R., Ratner, M.A. and Marks, T.J., Chem. Rev., 94 (1994) 195, and references cited therein.

26 Prasad, P.N. and Williams, D.J., Non-Linear Optical Effects in Molecules and Polymers, Wiley, New York, NY, 1991.

27 Anulewicz, R., Häfelinger, G., Krygowski, T.M., Regelman, C. and Ritter, G., Z. Naturforsch., 42B (1987) 917.

28 Anderson III, G.M., Kollman, P.A., Domelsmith, L.N. and Houk, K.N., J. Am. Chem. Soc., 101 (1979) 2344.

29 Matsuzawa, N. and Dixon, D.A., Int. J. Quantum Chem., 44 (1992) 497.