Journal of Mathematical Chemistry 23 (1998) 51–60 

# Quantum molecular similarity via momentum-space indices

Neil L. Allan [a] and David L. Cooper [b]

[a] *School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, UK*
[b] *Department of Chemistry, University of Liverpool, P.O. Box 147, Liverpool L69 7ZD, UK*

A method for assessing quantum molecular similarity has been developed based on the momentum-space representation. The principles of the method are summarized and the results of two applications are presented. The approach emphasizes the variation of the outer valence electron density much more than the bonding topology, it avoids several problems associated with more conventional approaches based on the position-space representation, and it can now be applied to extended series of large molecules. The first of the applications involves different sulfonylurea inhibitors of the enzyme acyl-CoA : cholesterol acyltransferase (ACAT). The second is concerned with the relative toxicity of a number of anti-HIV phospholipids. Further experimental work is suggested in both cases.

## 1. Introduction

Over the past seven years, we have developed various ways of using momentum-space concepts to evaluate quantum similarity measures. The indices employed are of the type suggested for the comparison of the *position-space* ($r$-space) densities of molecules [3,15], but in our novel approach we input *momentum-space* ($p$-space) electron densities. Many of the problems associated with the conventional position-space procedures are avoided and particular emphasis is placed on the variation of the long-range position-space electron density. The momentum-space approach is particularly suited to problems for which the molecular activity depends more on features of this long-range slowly-varying valence electron density and less on precise details of the bonding topology.

Of course, the motivation for classifying molecules according to how similar they are to one another is the premise that similar molecules may exhibit similar behaviour. The overall aim is to detect patterns of activity, to make useful predictions, and to identify rogue data. In our own work, starting with studies of model systems such as $CH_3XCH_3$ (X=O, S, $CH_2$) [5] and hydrofluoromethanes [1,6], we have reached the stage where it is relatively straightforward to consider extensive series of large molecules. Momentum-space similarity concepts have, for example, now been applied to a number of problems relating to HIV inhibition, including anti-HIV1 phospholipids [8,13], derivatives of the nucleoside analogue d4T [12], and non-nucleoside reverse transcriptase inhibitors [14]. Problems from materials chemistry have also been

studied, and useful structure–activity relationships established [12,13] for the molecular hyperpolarizabilities of conjugated systems, such as disubstituted benzenes, styrenes, stilbenes and diphenylacetylenes; the accurate direct calculation of such quantities is far from straightforward.

The momentum-space approach is not restricted to comparisons of *total* electron densities, useful though this has proved to be. Our studies have also involved comparisons of *total valence* electron densities and of the densities associated with a particular atom or group of atoms common to the two molecules. It is also possible to compare the densities associated with particular orbitals of two different molecules, such as their respective HOMOs or localized orbitals describing a particular type of bond. A comparison of two orbitals in the *same* molecule, such as the HOMO and the LUMO, has also proved useful.

In this paper, following a brief summary of the essential methodology and a brief example illustrating the form of the momentum-space density for large molecules, we present the results of two recent similarity studies. The first consists of a number of inhibitors of ACAT (acyl-CoA : cholesterol acyltransferase) and in the second we extend our earlier work on anti-HIV phospholipids. A common link in the present work is that the molecules in each series are all *very* similar to one another, and so we focus attention on momentum-space *dissimilarity* indices, which we have previously found to be especially useful in such cases [1,6].

## 2. Momentum-space electron densities and molecular similarity

We start with the familiar molecular orbitals, $\psi(\mathbf{r})$, of the form

$$\psi(\mathbf{r}) = \sum_j c_j \phi_j^\alpha(\mathbf{r} - \mathbf{R}_\alpha), \tag{1}$$

where the sum is over the $j$ atomic basis functions, $\phi_j^\alpha$, centered on nuclei with position vectors $\mathbf{R}_\alpha$. The momentum-space wavefunction, $\Psi(\mathbf{p})$, is obtained by a Fourier transform of $\psi(\mathbf{r})$:

$$\Psi(\mathbf{p}) = \sum_j c_j \Phi_j^\alpha(\mathbf{p}) \exp(-\mathrm{i}\mathbf{p}\cdot\mathbf{R}_\alpha), \tag{2}$$

in which the $\Phi_j^\alpha(\mathbf{p})$ are the Fourier transforms of the respective $\phi_j^\alpha(\mathbf{r})$.

The relationship in momentum space between the wavefunction and the electron density is exactly as in position space, i.e., the momentum-space density, $\rho(\mathbf{p})$, for this molecular orbital is given by the product $\Psi^*(\mathbf{p})\Psi(\mathbf{p})$. The $p$-space basis functions, $\Phi_j^\alpha(\mathbf{p})$, fall off sharply with increasing $|\mathbf{p}|$, and so the $p$-space electron density emphasizes the slowest moving valence electrons. In contrast, $r$-space electron densities tend to be dominated by the regions close to the nuclei.

We have discussed in detail elsewhere [2] the form of momentum-space densities for molecules and solids. The momentum-space electron densities of molecules consist

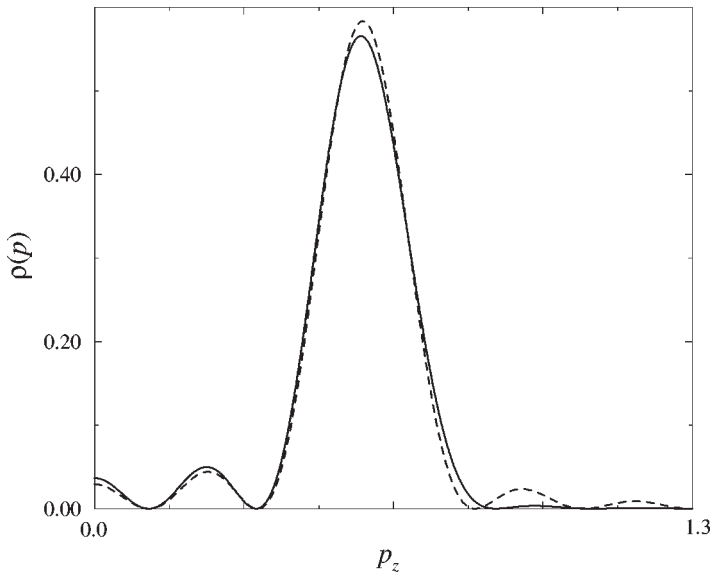



Figure 1. Momentum-space electron density of the highest-occupied molecular orbital (HOMO) of H–(C≡C)$_{10}$–H vs. $p_z$, where $p_z$ is the component of the total momentum along the carbon chain. The component of the momentum perpendicular to the chain is 1 a.u. The solid line denotes the density calculated from SCF wavefunctions (3-21G basis set) and the dotted line the density calculated using a simple Hückel approach [11].

of atomic orbital terms multiplied by oscillatory factors which contain information about the molecular geometry. These oscillatory terms can play a striking role when an orbital is delocalized over a large number of centres. For an infinitely long chain [2], for example, the momentum density is non-zero only for discrete values of $p = |\mathbf{p}|$. This is of course entirely consistent with the Heisenberg uncertainty principle: as we know less about the position of an electron when it is delocalized over more centres in position space, our knowledge of its momentum increases and there are sharper peaks in the momentum density.

We present here just one example. Figure 1 shows the momentum-space electron density associated with the HOMO of the polyyne H–(C≡C)$_{10}$–H, generated using SCF wavefunctions with a 3-21G basis set. For comparison, we also give the density calculated from a simple Hückel treatment [2,11], analogous to that of Duncanson and Coulson [10], using one 2$p$-basis function on each C atom. $p_z$ denotes the component of the total momentum along the carbon chain. In $p$-space, the density for the HOMO is clearly peaked around $p_z \approx 0.6$ a.u., whereas this orbital is, of course, delocalized in $r$-space over all the nuclei. The momentum-space formalism also proves useful when examining the effect of an applied electric field on the electron density of such a system [11].

The basic approach used here to quantify the momentum-space similarity is the analogue of the scheme first proposed for position-space densities by Carbó et al. [4].

A generalized overlap between momentum-space densities $\rho_A$ and $\rho_B$ takes the form

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

As indicated above, the momentum-space densities can be total electron densities, total valence densities, or those associated with one or more orbitals of interest or with particular molecular fragments. The function $p^n$ is included in the integrand to emphasize particular regions of the density [12]. For example, the value of $n$ of $-1$, which we use throughout this paper, focuses on the slowest moving electron density. This corresponds in turn to emphasizing the long-range valence density in position space.

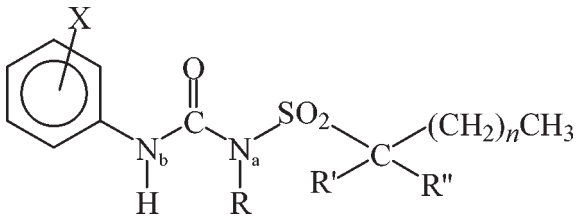
It is often useful to scale $I_{AB}(n)$ so that it takes values in the range 0–100%. This can, of course, be achieved in many ways and a number of these have been employed in our previous work. However, a distance-like dissimilarity index can be more informative when similarities are extremely high. In the present work, we employ the dissimilarity index $\mathcal{D}_{AB}(n)$, which takes the form

$$\mathcal{D}_{AB}(n) = 100\frac{I_{AA}(n) + I_{BB}(n) - 2I_{AB}(n)}{\left(\int \rho_A(\mathbf{p})\,\mathrm{d}\mathbf{p}\right)\left(\int \rho_B(\mathbf{p})\,\mathrm{d}\mathbf{p}\right)} \tag{4}$$

with a minimum value of zero (total similarity) and no upper limit. Full details of the calculation of the momentum-space wavefunctions (including the explicit forms of the $\Phi_j^\alpha$) and discussions of further similarity measures based on $I_{AB}$ and of the methods used for numerical integration, are to be found in [7]. A particular advantage of the $p$-space approach is that $I_{AB}(n)$ is completely independent of the separation between molecules A and B in position-space, and so depends only on their relative orientation.

## 3. Sulfonylurea inhibitors of ACAT

The importance of inhibition of acyl-CoA : cholesterol acyltransferase (ACAT) [16,17] is linked to the significant role played by this enzyme in the absorption of dietary cholesterol from the intestine, in the secretion of very low density lipoproteins by the liver, and in the esterification and storage of cholesteryl esters in the arterial wall. Roth et al. [16] have prepared and tested a series of sulfonylureas with the general formula shown in scheme 1. Their most complete set of data, reproduced in table 1, relates to *in vitro* measurements made by incubating the test compounds with [1-$^{14}$C]-oleoyl-CoA and microsomes isolated from the livers of cholesterol-fed rats. The $IC_{50}$ values are the concentrations of sulfonylurea required to achieve 50% inhibition – the most potent compounds have the smallest $IC_{50}$ values. Roth et al. [16] reported relatively poor agreement between the inhibition of ACAT *in vitro* with rat liver and rabbit intestinal microsomes and the inhibition of lower plasma cholesterol *in vivo* for acute and chronic rat models of hypocholesterolemia. This could be due to variations in the bioavailability of the different compounds and to a wide range of




Scheme 1. General formula of the sulfonylureas considered in this work. The various combinations of X, R, R′, R″ and $n$ are listed in table 1.

Table 1
Experimental $IC_{50}$ values [16] for *in vitro* inhibition of ACAT by a series of sulfonylureas (see scheme 1).

| Compound | X | R | R′ | R″ | $n$ | $IC_{50}$ ($\mu$M) |
|---|---|---|---|---|---|---|
| 8 | 2,6-$[CH(CH_3)_2]_2$ | H | H | H | 6 | 22.6 |
| 9 | 2,6-$[CH(CH_3)_2]_2$ | H | $CH_3$ | H | 7 | 1.3 |
| 10 | 2,6-$[CH(CH_3)_2]_2$ | H | Ph | H | 7 | 9.2 |
| 11 | 2,6-$[CH(CH_3)_2]_2$ | H | H | H | 8 | 13.6 |
| 12 | 2,6-$[CH(CH_3)_2]_2$ | H | H | H | 10 | 3.3 |
| 13 | 2,6-$[CH(CH_3)_2]_2$ | $CH_3$ | H | H | 10 | 0.675 |
| 14 | 2,6-$[CH(CH_3)_2]_2$ | H | H | H | 11 | 2.0 |
| 15 | 2,6-$[CH(CH_3)_2]_2$ | H | H | H | 12 | 0.94 |
| 16 | 2,4-$[F]_2$ | H | H | H | 12 | $\gg 10$[a] |
| 17 | 2,4,6-$[OCH_3]_3$ | H | H | H | 12 | $>5$[b] |
| 18 | 2,6-$[CH(CH_3)_2]_2$ | $CH_3$ | H | H | 12 | 8.83 |
| 19 | 2,6-$[CH(CH_3)_2]_2$ | $CH_2Ph$ | H | H | 12 | $>5$[c] |
| 20 | 2,6-$[CH(CH_3)_2]_2$ | H | $CH_3$ | H | 12 | 0.83 |
| 21 | 2,6-$[CH(CH_3)_2]_2$ | H | $CH_3$ | $CH_3$ | 12 | 0.682 |
| 22 | 2,4,6-$[OCH_3]_3$ | H | $CH_3$ | $CH_3$ | 12 | 8.1 |
| 23 | 2,6-$[CH(CH_3)_2]_2$ | H | Ph | H | 12 | 2.3 |
| 24 | 2,6-$[CH(CH_3)_2]_2$ | H | H | H | 14 | 1.27 |
| 25 | 2,6-$[CH(CH_3)_2]_2$ | H | $CH_3$ | H | 15 | 1.39 |
| 26 | 2,4,6-$[OCH_3]_3$ | H | $CH_3$ | H | 15 | 0.085 |

[a]No inhibition at 10 $\mu$M; [b]40% inhibition at 5 $\mu$M; [c]38% inhibition at 5 $\mu$M.

pharmokinetic factors [16]. Nevertheless, it is interesting to look for patterns in the *in vitro* rat liver data.

On the whole, the 2,6-$[CH(CH_3)_2]_2$ derivatives tend to be the most potent (see table 1), except that compound 26 has a lower $IC_{50}$ than does compound 25. With $R' = R'' = H$, the potency generally increases with increasing $n$, except that compound 24 is less active than is compound 15. Replacement of R′, R″ with phenyl groups tends to increase $IC_{50}$ whereas methyl groups tend to decrease $IC_{50}$, but there are some exceptions. There is no consistent pattern in the change in potency on replacing R. It seems worthwhile to examine these various compounds using our momentum-space quantum molecular similarity techniques.

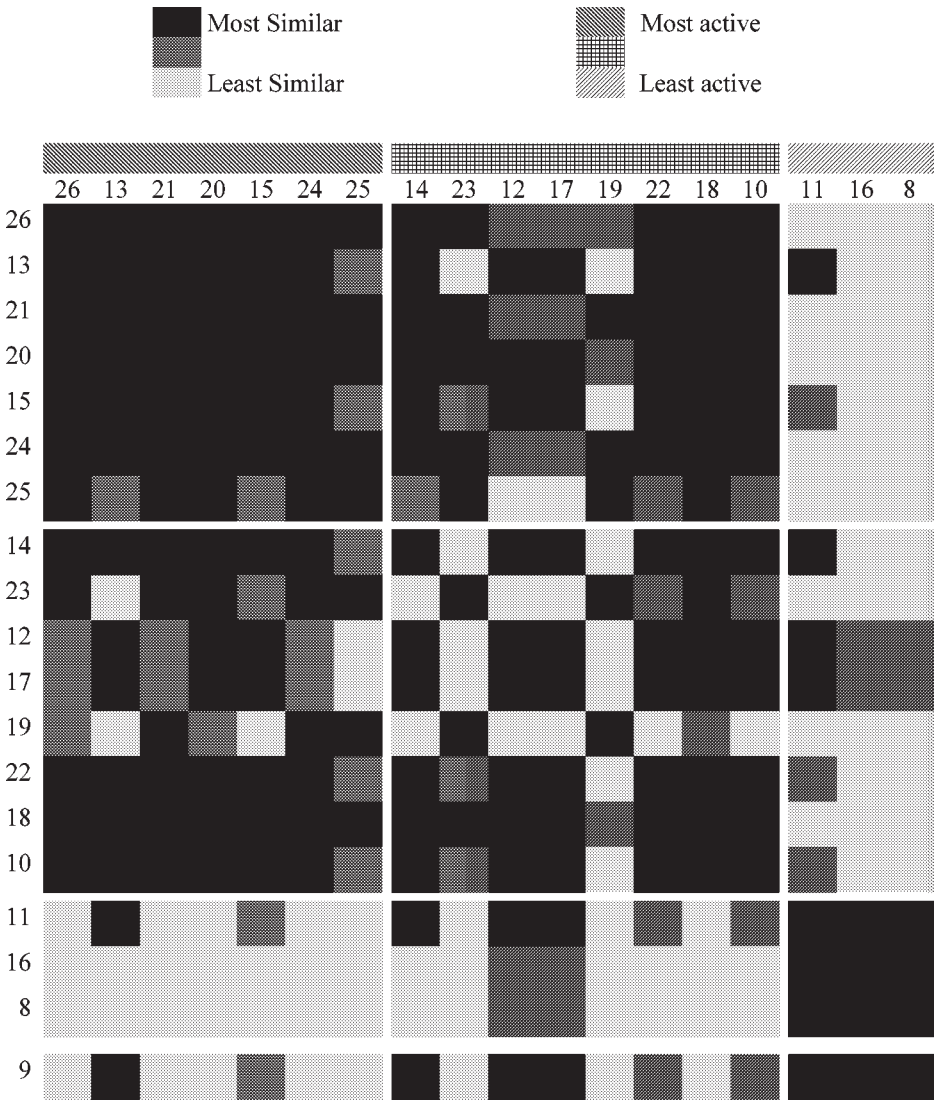



Figure 2. Pictorial representation of $\mathcal{D}_{AB}(-1)$ values for the sulfonylureas identified in scheme 1 and table 1.

Structures for the various molecules were constructed in a consistent fashion using the optimization scheme available in the DTMM program [9] and AM1 wavefunctions were obtained using the MOPAC code [18]. Values of the momentum space dissimilarity index $\mathcal{D}_{AB}(-1)$ were then calculated for each pair of molecules, with a common alignment of the $N_a$–C bonds and of the $N_aCN_b$ planes. The molecules have been ordered in figure 2 according to their $IC_{50}$ values. The three groupings correspond to high ($IC_{50}$ ($\mu$M) $< 1.4$), intermediate ($2.0 \leqslant IC_{50}$ ($\mu$M) $\leqslant 9.2$) and low ($IC_{50}$ ($\mu$M) $> 10$) potency, except that compound 9 has been shown separately. Differ-

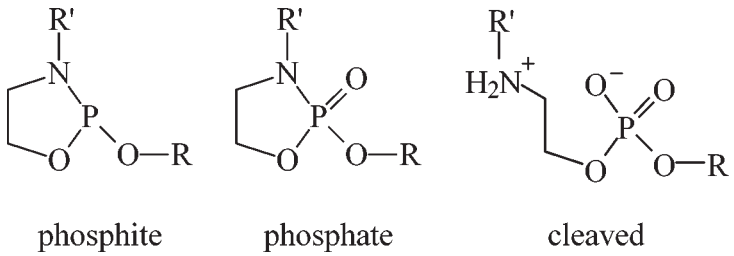
ent degrees of shading have been employed to represent the lowest ($<10$), intermediate (10–20) and highest ($>20$) values for the dissimilarity index.

It is clear from figure 2 that our values of $\mathcal{D}_{\mathrm{AB}}(-1)$ can be used to distinguish between the most potent and least potent compounds, in that we observe blocks of low dissimilarity for comparisons within the groupings and a block of high dissimilarity between these two groupings. The compounds with intermediate $IC_{50}$ values show higher similarity to the most potent compounds than to the least potent ones. Compounds 19 and 23 appear to be somewhat different from the other intermediate compounds, but the patterns of $\mathcal{D}_{\mathrm{AB}}(-1)$ values for comparisons with the most and least potent grouping are entirely consistent with the observed potency. Indeed, the only anomalous case is compound 9 ($IC_{50} = 1.3\ \mu$M) for which the pattern of $\mathcal{D}_{\mathrm{AB}}(-1)$ values is the same as for compound 11 ($IC_{50} = 13.6\ \mu$M). Assuming the experimental data to be correct, we can offer no explanation for this discrepancy and we suggest that this enhanced *in vitro* potency merits further study.

## 4. Second-generation phospholipids

We have shown previously [8,13] how momentum-space quantum molecular similarity indices may be used to rationalize, and even predict, the relative success of a series of phospholipid molecules for inhibition of HIV1 in C8166 T-lymphoblastoid cells. Experimental data are now available for a further series of molecules, concentrating this time on the *toxicity* of the various compounds. These "second-generation" phospholipids have the general structures shown in scheme 2, in which R′ is methyl or *t*-butyl, and R is one of the groups shown in figure 3. The values denoted $CC_{50}$ ($\mu$M) in table 2 are the concentrations that inhibit proliferation of uninfected cells by 50% [14].

Initial energy minimizations using the DTMM program [9] were followed by AM1 calculations using MOPAC [18]. We calculated $\mathcal{D}_{\mathrm{AB}}(-1)$ for each pair of molecules, aligning as closely as possible the C–C–O link between the N and P atoms. The resulting numerical values are represented in figure 4 by different shading: black denotes the most similar and white the least. The rows and columns are sorted in order of decreasing toxicity, except for molecule 162, which is shown separately. Figure 4 clearly distinguishes a group of molecules with $CC_{50}$ values in the range 10–400 $\mu$M from a less toxic group with $CC_{50} \geqslant 1000\ \mu$M. This appears to be another success




phosphite phosphate cleaved

Scheme 2. General structures of second-generation phospholipids.

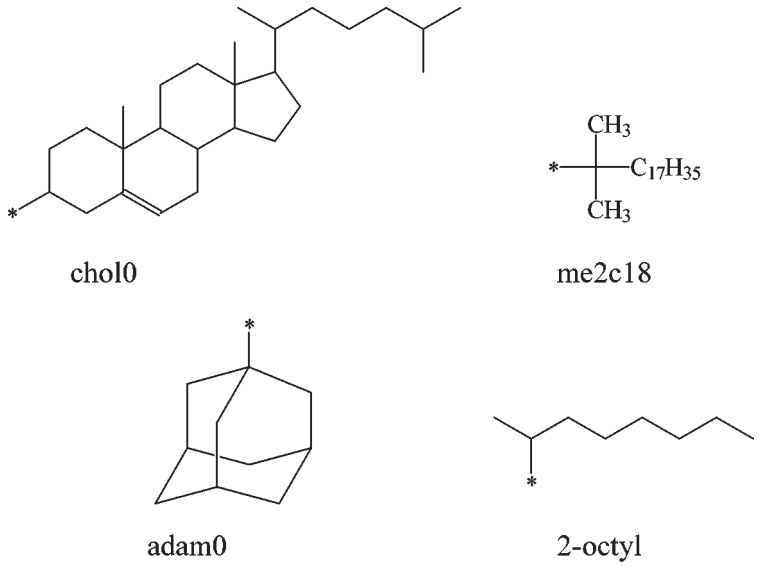



Figure 3. R substituents in the second-generation phospholipids (see scheme 2). The point of attachment has been marked with an asterisk.

Table 2
Toxicity data for various phospholipids. The structures are shown in scheme 2 and the mnemonics for the R groups are identified in figure 3.

| Mnemonic | Structure | R′ | R | $CC_{50}$ ($\mu$M) |
|---|---|---|---|---|
| 155 | cleaved | *t*-butyl | chol0 | 400 |
| 156 | cleaved | *t*-butyl | adam0 | >1000 |
| 157 | cleaved | *t*-butyl | me2c18 | 100 |
| 160 | cleaved | *t*-butyl | *l*-2-octyl | >1000 |
| 161 | phosphate | *t*-butyl | chol0 | 80 |
| 162 | phosphate | *t*-butyl | adam0 | 100 |
| 163 | phosphite | *t*-butyl | me2c18 | 40 |
| 164 | cleaved | methyl | adam0 | 1000 |
| 165 | cleaved | methyl | chol0 | >100 |
| 166 | cleaved | methyl | me2c18 | 10 |
| 167 | cleaved | methyl | *l*-2-octyl | 1000 |
| 169 | phosphite | methyl | chol0 | 200 |
| 170 | phosphate | methyl | chol0 | 400 |

for our momentum-space methodology. Closer inspection of the data in table 2 shows that the first group comprises molecules with larger R groups (chol0 and me2c18) and the second one consists of molecules with smaller R groups (*l*-2-octyl and adam0). From a practical point of view, the calculation of $\mathcal{D}_{AB}(-1)$ has done little more in this case than could in fact have been achieved simply by inspecting the chemical structures. A remaining puzzle, however, is the toxicity of molecule 162, for which the experimentally determined $CC_{50}$ is 100 $\mu$M. The pattern of shading in figure 4

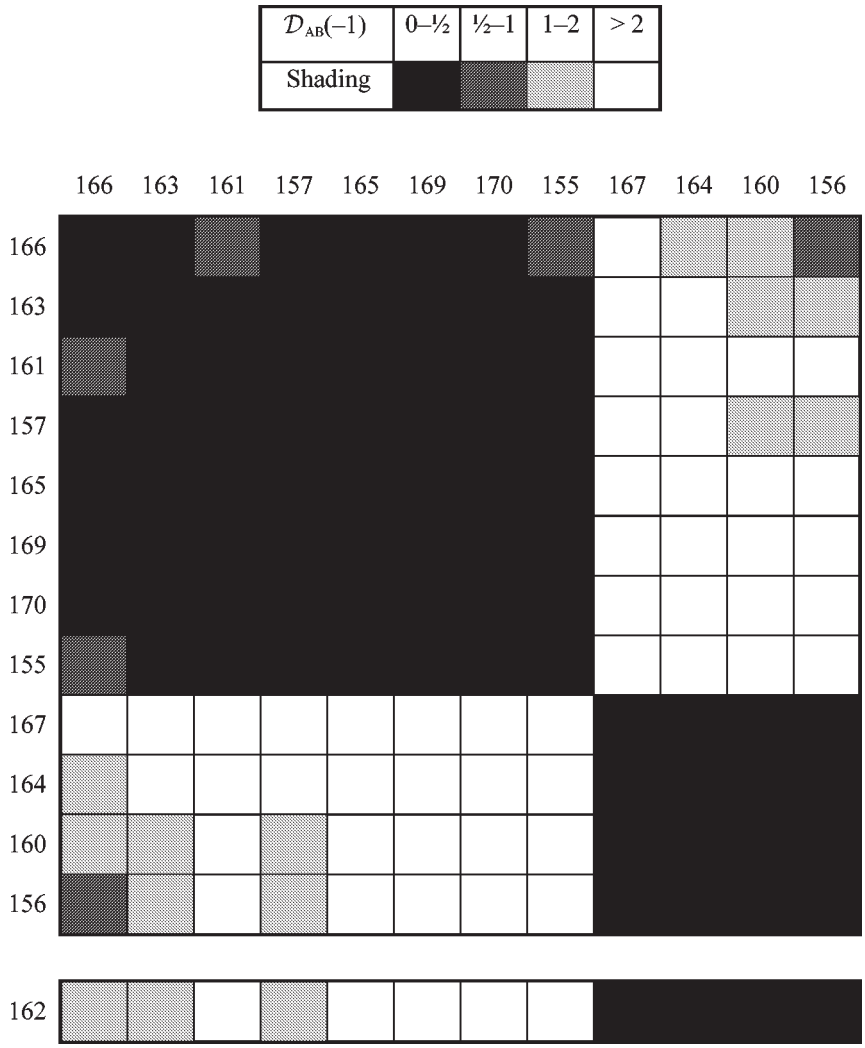



Figure 4. Pictorial representation of $\mathcal{D}_{AB}(-1)$ values for second-generation phospholipids. The molecules are identified in table 2.

is identical for molecules 160 and 162, and so we would have expected a $CC_{50}$ in excess of 1000 $\mu$M. Assuming the experimental data for molecule 162 to be correct, it could be very informative to investigate the reasons for the apparent enhancement of the toxicity of this molecule.

## 5. Conclusions

We have shown in this and in previous work how momentum-space quantum similarity techniques may allow us to rationalize biological activities even when the

details of the processes involved are complex, partially characterized, or simply unknown. The momentum-space techniques are a valuable addition to other molecular similarity tools, particularly when the bonding topology is less important than the variation of the valence electron density. We would be pleased to learn of problems for which it is difficult to rationalize the observed physical or biological behavior with conventional chemical intuition and for which our novel techniques could be a useful aid.

## References

[1] N.L. Allan and D.L. Cooper, J. Chem. Inf. Comput. Sci. 32 (1992) 587.
[2] N.L. Allan and D.L. Cooper, in: *Molecular Similarity*, ed. K.D. Sen, Topics in Current Chemistry, Vol. 173 (1995) p. 85.
[3] R. Carbó-Dorca, E. Besalú, L. Amat and X. Fradera, Adv. Mol. Sim. 1 (1996) 1.
[4] R. Carbó, L. Leyda and M. Arnau, Int. J. Quantum Chem. 17 (1980) 1185.
[5] D.L. Cooper and N.L. Allan, J. Comput. Aid. Mol. Design 3 (1989) 253.
[6] D.L. Cooper and N.L. Allan, J. Am. Chem. Soc. 114 (1992) 4773.
[7] D.L. Cooper and N.L. Allan, in: *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, ed. R. Carbó (Kluwer, Dordrecht, 1995) p. 31.
[8] D.L. Cooper, K.A. Mort, N.L. Allan, D. Kinchington and C. McGuigan, J. Am. Chem. Soc. 115 (1993) 12 615.
[9] M.J.C. Crabbe, J.R. Appleyard and C.R. Lay, *Desktop Molecular Modeller (Version 3.0 for Windows)* (Oxford University Press, Oxford, 1994).
[10] W.E. Duncanson and C.A. Coulson, Proc. Phys. Soc. 65 (1952) 825.
[11] P.T. Measures, Ph.D. thesis, University of Bristol (1996).
[12] P.T. Measures, N.L. Allan and D.L. Cooper, Adv. Mol. Sim. 1 (1996) 61.
[13] P.T. Measures, K.A. Mort, N.L. Allan and D.L. Cooper, J. Comput. Aid. Mol. Design 9 (1995) 331.
[14] P.T. Measures, K.A. Mort, N.L. Allan and D.L. Cooper, Theochem (in press).
[15] R. Ponec, Adv. Mol. Sim. 1 (1996) 121.
[16] B.D. Roth, W.H. Roark, J.A. Picard, R.L. Stanfield, R.F. Bousley, M.K. Anderson, K.L. Hamelehle, R. Homan and B.R. Krause, Bioorg. Med. Chem. Lett. 5 (1995) 2367.
[17] D.R. Sliskovic and B.K. Trivedi, Curr. Med. Chem. 1 (1994) 204.
[18] J.J.P. Stewart, J. Comput. Aid. Mol. Design 4 (1990) 1.