A quantum molecular similarity approach to anti-HIV activity

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

[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

Abstract

Our quantum molecular similarity approach, which emphasizes the variation of the outer-valence electron density rather than the bonding topology, is applied to four series of molecules showing anti-HIV activity. Applying our techniques to the virology data of a family of anti-HIV1 phospholipids, bioactive molecules are shown to be similar to one another but dissimilar to those that are inactive, and vice versa. We have been able to make a number of successful predictions. In the case of structurally diverse non-nucleoside reverse transcriptase inhibitors, fragments are identified which are possibly important for recognition by the enzyme. A study of masked phosphate derivatives of the nucleoside analogue d4T leads to the successful prediction of a highly active compound. Lastly, a particularly simple pattern emerges from the toxicity data for a further group of phospholipids. © 1998 Elsevier Science B.V.

Keywords: HIV; Molecule similarity; Phospholipids; Nucleosides; Structure-activity

1. Introduction

Molecular similarity techniques [1–5] can be especially useful in the study of processes with mechanisms that are extremely complicated, partly characterized or even completely unknown. Of course, the desire to classify molecules according to how similar they are to one another is motivated by the premise that similar molecules may exhibit similar behavior. The eventual aim is to detect patterns of activity, so as to allow useful predictions to be made and rogue data to be identified.

Our own approach to quantum molecular similarity [6–13] has now been applied to a number of problems relating to HIV inhibition. In the present article we both review and extend our studies of anti-HIV1 phospholipids [11], of non-nucleoside reverse transcriptase inhibitors [12] and of derivatives of the nucleoside analogue d4T [13]. We also describe some new work that relates to the toxicity of a further family of phospholipids. We refer to these last molecules as "second-generation" phospholipids, so as to distinguish them from the first set.

2. Momentum–space similarity

It is conventional in quantum mechanics to express molecular wavefunctions and the associated electron densities as functions of the positions of the particles. However, because of the complementarity of position and momentum, we can choose instead to work with functions of the momenta of the particles. Momentum–space wavefunctions are most readily obtained as the Fourier transform of their position–space counterparts and the relationship in momentum space between the

wavefunction and the total electron density is exactly the same as in position space. The momentum density $\rho(\mathbf{p})$ falls off so sharply with p that it is dominated by low values of p. These last correspond to the slowly varying outer valence position–space electron density and so, unlike $\rho(\mathbf{r})$, which is determined largely by the core electrons, $\rho(\mathbf{p})$ tends to emphasize some of the chemically most interesting parts of the electron density.

The central quantity in our approach to quantum molecular similarity is the generalized overlap

$$I_{AB}(n) = \int p^n \rho_A(\mathbf{p})\rho_B(\mathbf{p}) \, d\mathbf{p} \tag{1}$$

in which $\rho_A(\mathbf{p})$ and $\rho_B(\mathbf{p})$ are two momentum–space electron densities and the inclusion of powers of p allows different regions of momentum space to be emphasized. n is typically chosen in the range -1 to $+2$, with a value of -1 highlighting the slowest moving electrons and often giving the most discriminating (dis)similarity index. The value of $I_{AB}(n)$ depends on the relative orientation of the two molecules, but not on their separation, because the momentum density is independent of choice of position–space origin. It is straightforward to determine $I_{AB}(n)$ for total electron densities, total valence electron densities or the densities associated with particular orbitals or molecular fragments.

For some applications it proves useful to scale $I_{AB}(n)$ into the range 0–100%, with higher values indicating higher similarity. One of the many ways of achieving this is with the Tanimoto-like index:

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

However, with quantities which show very high similarity, such as the momentum–space electron densities for analogous fragments in closely related molecules, it can be convenient to dispense with the normalization of $I_{AB}(n)$ and to use instead momentum–space *dissimilarity* indices:

$$D_{AB}(n) = 100 \times (I_{AA}(n) + I_{BB}(n) - 2I_{AB}(n)) \tag{3}$$

$$P_{AB}(n) = 100 \times ((I_{AA}(n)I_{BB}(n))^{1/2} - I_{AB}(n))$$

These distance-like measures take values from zero upwards. Larger values imply greater dissimilarity,

without an upper bound. Alternative measures are

$$\mathcal{D}_{AB}(n) = \frac{D_{AB}(n)}{N_A N_B} \text{ and } \mathcal{P}_{AB}(n) = \frac{P_{AB}(n)}{N_A N_B} \tag{4}$$

in which $N_x = \int \rho_x(\mathbf{p}) \, d\mathbf{p}$. The $\mathcal{P}_{AB}(n)$ index can be particularly useful when the shape of the density function is the major consideration, because it is independent of the (non-zero) value of m in $\rho_A(\mathbf{p}) = m \times \rho_B(\mathbf{p})$.

Various tools, taken from the field of cluster analysis [14], are available for interpreting matrices of similarity indices. These include mapping methods, the construction of similarity trees, hierarchical clustering methods, partitioning schemes, density search techniques and clumping procedures. In much of our own work, it has proved useful simply to replace the plethora of numbers by different degrees of shading and then to rely on visual inspection. We have also made some use of two purely numerical procedures — our implementations of a clumping technique [15] and of a density search scheme [16,17] for analyzing normalized similarity indices, S_{IJ}, are described in detail elsewhere [12]. In essence, the clumping technique assigns molecules to one or both of two clusters, A and B, by minimizing

$$\frac{\sum\limits_{I \in A} \sum\limits_{J \in B} S_{IJ}}{\left[\left(\sum\limits_{I \in A} \sum\limits_{J \in A} S_{IJ} \right) \left(\sum\limits_{I \in B} \sum\limits_{J \in B} S_{IJ} \right) \right]^{\kappa}} \tag{5}$$

with $A \neq B$. The input parameter κ (in the range 1/2 to 1) controls the size of the intersection between the two clusters.

The density search technique assigns molecules to a number of non-overlapping clusters, as follows. We start with a cluster containing only the most similar pair of molecules. At each stage, we locate max(S_{IJ}), such that "I" belongs to the most recently created cluster and molecule "J" is as yet unassigned. To see if "J" should be added to the same cluster as "I", we compute the mean similarity of the proposed cluster — if the change in the mean on including "J" is too large, then this species is rejected from the present cluster and we start a new one by locating the highest similarity between unassigned molecules. The number of clusters is not fixed beforehand, but it is instead
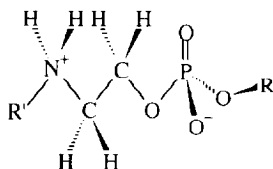
Scheme 1. General formula of phospholipids.

a function of the tolerance τ used for the acceptance/rejection of molecules from clusters.[1]

3. Phospholipids

We have examined the virology data for a family of phospholipids [11,18] with the general structural formula shown in Scheme 1, in which R represents the lipid tail and R′ is a relatively small alkyl substituent on the amine group. Molecules with different R and R′ give 50% inhibition of HIV1 in C8166 T-lymphoblastoid cells with the ED_{50} (μM) values listed in Table 1—a low ED_{50} value indicates high activity. There appears to be no consistent correlation between the HIV1 inhibition and the choice of R and R′. For example, replacing a methyl group at the R′ position by t-butyl results in much greater activity (smaller ED_{50} values) for the most inactive compounds (DD and HX), and some increase in the activity for OD and OL, but a decrease in the activity for EG. The compounds tend to have very low ED_{50} values when R′ is simply a hydrogen atom, except that HX3, unlike HX2, is inactive. Experimental data are not currently available for EG3.

The mechanism by which these phospholipids inhibit the virus is not completely understood, although it is thought that they first insert into the membranes of the virus. 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, rather than the densities for individual orbitals or for molecular fragments.

We generated computationally inexpensive wavefunctions from semi-empirical MNDO geometry

Table 1
Experimental ED_{50} values for the inhibition of HIV1 in C8166 T-lymphoblastoid cells for a series of phospholipids (see Scheme 1)

Mnemonic	R′	R	$ED_{50}(\mu M)$
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

optimizations using the MOPAC code ([19] and references therein). No search for the global minimum conformation was carried out. The conformation used for the framework common to all the systems studied is as shown in Scheme 1.

We started by comparing all the molecules with HX1 and OL2. We found that active compounds have a high similarity to the highly *active* compound OL2 and a low similarity to the *inactive* compound HX1, and vice versa. We were initially informed that OD2 was insoluble/inactive, whereas we expected from our $T_{AB}(-1)$ values that this phospholipid should have a very low ED_{50} value. The revised experimental data (Table 1) suggest $ED_{50} = 3\ \mu$M. The $T_{AB}(-1)$ values for the series with a free amine group were calculated *before* virology data became available to us, and we were able to *predict* correctly from the comparisons with HX1 and OL2 that OL3 and OD3 would have small ED_{50} values, but that HX3 would be inactive. For DD3 we observed moderately high T_{AB} values relative to *both* HX1 and OL2, and so it was impossible to make a definitive prediction in this case.

We subsequently 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 numerical results by different degrees of shading. The most active molecules are located in the top left-hand corner of the figure.

---

[1] In practice, the mean similarity of the original cluster is subtracted from twice the mean similarity of the proposed cluster and the new molecule is accepted into the cluster if this value exceeds the tolerance τ.
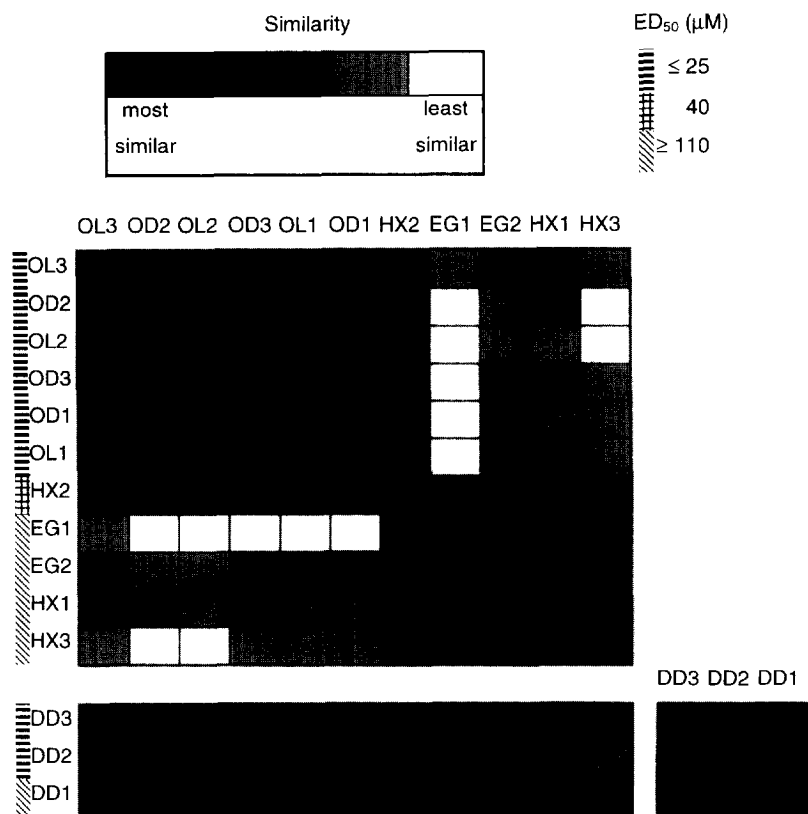
Fig. 1. Pictorial representation of $T_{AB}(-1)$ values for various phospholipids. The molecules are identified in Table 1.

Clearly, the active molecules are very similar to each other and they are dissimilar to the inactive molecules, and vice versa. HX2 shows intermediate similarity and activity. However, the DD molecules (DD1 inactive, DD2 and DD3 active), shown separately in Fig. 1, appear to be very similar to all of the active species *and* to some of the inactive species.

For certain input parameters (κ for the clumping technique and τ for the density search technique), both of our numerical clustering procedures give results in broad agreement with the visual approach. DD2 is correctly predicted to be active in these cases. However, there is no consistency in the results for DD1 and DD3. For example, a slightly modified clumping technique, in which we ignore the self-similarity terms in Eq. (5), leads to:

$$\kappa = 0.54$$

Cluster A   DD1, DD2, DD3, OD1, OD2, OD3, OL1, OL2, OL3
Cluster B   DD1, DD3, HX1, HX2, HX3, EG1, EG2

with DD1 and DD3 in both clusters. Applying this procedure with $\kappa = 0.53$ has the same outcome, except that DD1 is now excluded from cluster B. It is clear that definitive experimental values for the DD molecules would be very useful for assessing the merits of our different approaches.

4. Non-nucleoside reverse transcriptase inhibitors of HIV

Our second example concerns the molecules shown in Fig. 2, which inhibit the action of the reverse transcriptase enzyme in spite of very different chemical structures ([20] and references therein). It is worthwhile to inquire whether our measures of momentum–space similarity can identify a feature which is common to these various non-nucleoside reverse transcriptase inhibitors. Initial energy minimizations
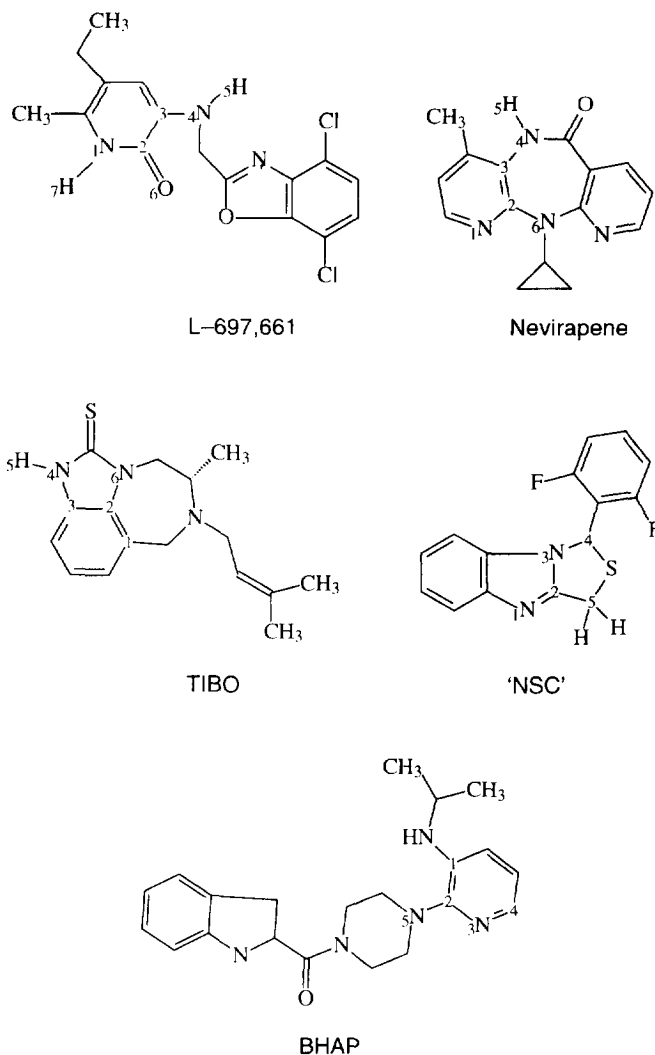
Fig. 2. Non-nucleoside reverse transcriptase inhibitors, with the selected fragments identified by atom numbering.

using the DTMM program [21] were followed by semi-empirical MNDO geometry optimizations using the MOPAC code ([19] and references therein), without searching for global minimum conformations.

All of the molecules feature at least two nitrogen atoms close to each other and, accordingly, we have examined various groupings of atoms which include two or three N atoms. In calculating the indices, we considered the contribution to the total density only from those basis functions associated with the atoms of interest. Since the fragments are *very* similar to one another, the values of normalized indices such as

$T_{AB}(-1)$ are very high indeed and so it proved more convenient to work with the dissimilarity index $D_{AB}(-1)$. In those cases for which heavy atoms are listed, any directly attached hydrogen atoms not explicitly shown in the figure were also included in the comparisons. In the cases of BHAP and TIBO, we examined various possible fragments, each containing two or three nitrogen atoms, but we found that all of the comparisons *very* strongly support one particular choice for each molecule [12], with no ambiguity whatsoever. The chosen groups are all very similar to one another,

Table 2
Values of $D_{AB}(-1)$ for selected fragments of non-nucleoside reverse transcriptase inhibitors (see Fig. 2)

	L-697,661	Nevirapene	"NSC"	TIBO
Nevirapene	14			
"NSC"	16	11		
TIBO	23	8	3	
BHAP	23	13	1	16

with values of $D_{AB}(-1)$ less than 25 (see Table 2), and it is tempting to believe that this has 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 non-competitive, in that they do not bind to the substrate binding site of the enzyme but elsewhere, probably altering the secondary structure 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 might be purely steric. It may well be that the common feature we have identified is a *recognition site* for binding to the enzyme—it is rather unlikely to determine the steric interference subsequently required for deactivation.

5. Derivatives of the anti-HIV nucleoside analogue d4T

We have also studied a series of nine d4T derivatives with the general formula shown in Scheme 2 [13]. The parent nucleoside analogue (d4T) and various "masked" phosphate derivatives have been tested for their ability to inhibit the replication of virus in HIV1$_{IIIB}$-infected C8166 cells and of HIV1 in CEM cells [22–24]. These phosphate pro-drugs, designed to act as intracellular sources of the free 5'-monophosphate, tend to be significantly less toxic to uninfected cells than d4T itself. Although they may be less potent, the reduced toxicity can lead to improved antiviral selectivity. The masking of the phosphate group is necessary to improve penetration through the cell membrane. The different R groups are listed in Table 3, together with the corresponding molecular labels and biological activities. The activities of molecules
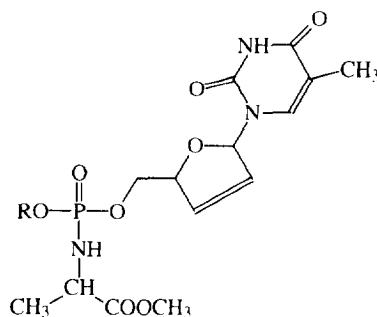
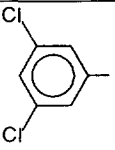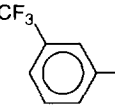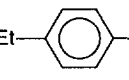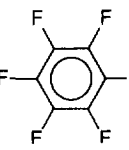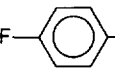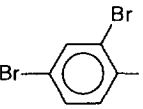Scheme 2. General structure of d4T derivatives.

8 and 9 had not yet been determined when we carried out our calculations.

Given our interest in the effects of the masking on the penetration of the nucleoside analogue derivative through cell membranes, it is logical to concentrate on the R substituents. With this in mind, we carried out MNDO geometry optimizations [19] for the corresponding ROH molecules and calculated dissimilarity indices $\mathcal{P}_{AB}(-3/4)$ for each pair, overlaying as closely as possible the C–O–H groups. This set of $\mathcal{P}_{AB}(-3/4)$ values, reported in Ref. [13], led to the successful prediction of a highly active compound. We denote the alcohols derived from molecules 1 to 9 with the corresponding letters of the alphabet (see Table 3).

We have now extended this work by allowing the relative orientations of the molecules to vary, minimizing $\mathcal{P}_{AB}(-3/4)$ for each pair using a straightforward simplex scheme. The "optimized" values of $\mathcal{P}_{AB}(-3/4)$ are shown in Fig. 3 by varying degrees of shading. Alcohol e seems to be different from most of the other molecules. Furthermore, alcohols c, d, f and g are all rather similar to each other, in spite of very different ED_{50} values for the corresponding d4T derivatives. Nonetheless, we are able to make some predictions for the molecules with unknown ED_{50}. We expect the activity of molecule 8 to be similar to that of molecule 2. Molecule 9 should show comparable activity to molecule 1. These predictions coincide with those we made in Ref. [13].

Distance-like values of $\mathcal{P}_{AB}(-3/4)$ were converted to similarity measures (S_{IJ}) by selecting the maximum dissimilarity, subtracting each dissimilarity index from this value, and then scaling these new quantities so that they lie in the range 0–100%. Applying the clumping procedure, alcohol e is separated from all

Table 3

R groups of nine nucleosides analogues (see Scheme 2), together with the corresponding molecular labels and respective ED_{50} values for HIV1 in CEM cells

Molecule	R-	ED_{50} (μM)	Alcohol
1		0.06	a
2		0.08	b
3		0.085	c
4	Et—	0.2	d
5		2.5	e
6	Et—	20	f
7	Pr—	100	g
8	F—	—	h
9	Br—	—	i

the other alcohols when $\kappa < 0.72$ and only molecules g and e are not in the intersection when $\kappa > 0.72$. To gain further insight, we excluded alcohol e from the clustering procedure and used Eq. (5) without the self-similarity terms. We then found four different domains:

	$\kappa < 0.5997$	$0.5998 < \kappa$ < 0.6115	$0.6116 < \kappa$ < 0.6267	$\kappa > 0.6268$
Cluster A	c, d, f, g, h	c, d, f, g, h	c, d, f, g, h	b, c, d, f, g, h, i
Cluster B	a, b, i	a, b, h, i	a, b, c, d, h, i	a, b, c, d, f, h, i

Rather disappointingly, alcohols b and c do not always cluster together, in spite of the comparable ED_{50} values for molecules 2 and 3. Nevertheless, the results obtained with the third choice of κ are striking: a, b, i (most active) are separated from f, g (least active), with c, d, h (intermediate activity) assigned to both clusters. Furthermore, the scheme also consistently places alcohol i with alcohol a, and it places alcohol h with alcohol b. As such, we predict activities for molecules 8 and 9 that are consistent with those inferred from visual inspection of Fig. 3.

With the density search technique, a tolerance of $\tau = 0.82$ gives results most consistent with the experimental ED_{50} data:

Cluster A	Cluster B	Cluster C	Cluster D
c, d, h	f, g	a, b, i	e

Alcohol e is again separated from all the others. Molecule 9 is predicted to behave in a similar fashion to molecule 1. Molecule 8 is predicted to show comparable activity to molecules 3 and 4.

All of our methods of analysis suggest high activity for molecule 9. Subsequent to our work, molecule 9 was shown experimentally to have an extremely high activity ($ED_{50} = 0.04$ μM). This success suggests that our momentum–space approach can be effective even in situations where the data is relatively sparse.

6. Second-generation phospholipids

Virology data have been made available to us [25] for a series of molecules with the general structures shown in Scheme 3, in which R' is methyl or t-butyl and R is one of the groups shown in Fig. 4. Values of CC_{50} (μM) are listed in Table 4: these are the concentrations that inhibit proliferation of uninfected cells by 50%.
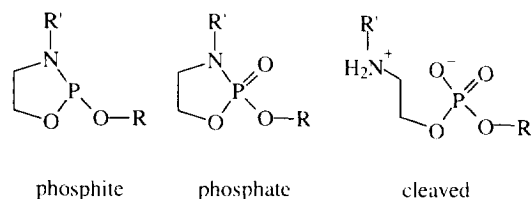
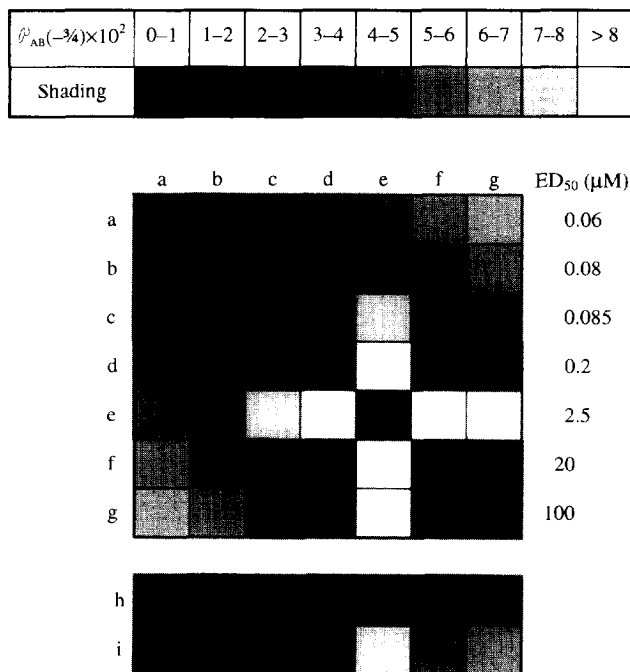Scheme 3. General structures of second-generation phospholipids.

$P_{AB}(-3/4)\times 10^2$	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	> 8
Shading									

Fig. 3. Pictorial representation of $P_{AB}(-3/4)$ values for various alcohols. The ED_{50} values relate to the corresponding d4T derivatives (see Table 3).

Initial energy minimizations using the DTMM program [21] were followed by AM1 calculations using MOPAC [19]. We aligned as closely as possible the C–C–O groups that link the nitrogen and phosphorus atoms, and then calculated values of $D_{AB}(-1)$. These last are represented in Fig. 5 by different shading, with black denoting the most similar and white the least. With the exception of molecule 162, which is shown
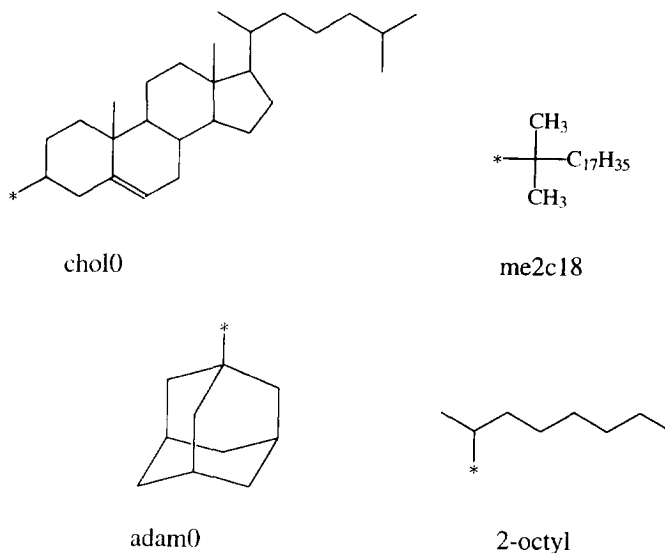

Fig. 4. R substituents in the second-generation phospholipids (see Scheme 3). The point of attachment has been marked with an asterisk.
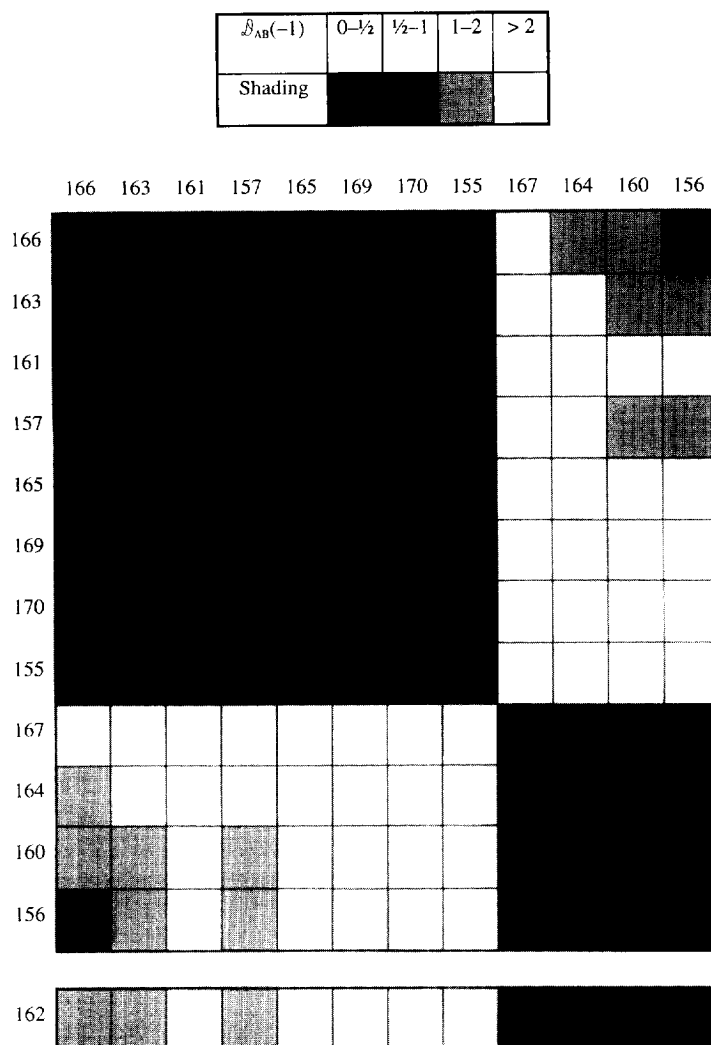
Fig. 5. Pictorial representation of $\mathcal{D}_{AB}(-1)$ values for second-generation phospholipids. The molecules are identified in Table 4.

separately, we have sorted the rows and columns in order of decreasing toxicity. The figure certainly distinguishes between a group of molecules with CC_{50} values in the range 10–400 μM, and a less toxic group with $CC_{50} \geq 1000$ μM. This appears to be another success for our momentum–space methodology. However, we find on closer inspection of the data (Table 4) that the second group comprises molecules with smaller R groups (ℓ-2-octyl and adam0) and the first one consists of molecules with larger R groups (cho10 and me2c18). As such, the calculation of

$\mathcal{D}_{AB}(-1)$ has done little more in this case than could probably have been achieved simply by inspecting the chemical structures. A remaining puzzle, however, is the toxicity of molecule 162. The pattern of shading in Fig. 5 is the same for molecules 160 and 162, and so we should expect molecule 162 to have a CC_{50} value in excess of 1000 μM, rather than the experimentally determined value of only 100 μM. Assuming the experimental data to be correct, it could be very informative to investigate the reasons for the enhanced toxicity of this molecule relative, say, to molecule 160.

Table 4
Toxicity data for various phospholipids. The structures are shown in Scheme 3 and the mnemonics for the R groups are identified in Fig. 4

Mnemonic	Structure	R'	R	CC_{50} (μM)
155	Cleaved	*t*-butyl	cho10	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	cho10	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	cho10	200
170	Phosphate	methyl	cho10	400

7. Conclusions

Momentum–space similarity and dissimilarity indices concentrate on variations in long-range valence electron densities. They tend to be especially useful when the observed activity appears to have no obvious dependence on the bonding topology of the molecules or on the nature of the atomic backbone. The method can be particularly effective for situations in which conventional chemical intuition is insufficient.

Applying our techniques to the virology data for a family of anti-HIV1 phospholipids, we have demonstrated that bioactive molecules are similar to one another, but dissimilar to those that are inactive, and vice versa [11]. We have been able to make a number of successful predictions. In the case of structurally diverse non-nucleoside reverse transcriptase inhibitors, we identified fragments which may be important for recognition by the enzyme [12]. Our study of masked phosphate derivatives of the nucleoside analogue d4T has already led to the successful prediction of a highly active compound [13]. A particularly simple pattern emerges from the toxicity data for a further group of phospholipids, although the CC_{50} value for one molecule appears to be anomalous.

Of course, any comparative methodology, such as that presented here, can only very indirectly provide information concerning interactions with biological receptors, but it is clear that our techniques hold considerable promise for the study of anti-HIV compounds.

Acknowledgements

References

[1] M.A. Johnson, G.M. Maggiora (Eds.), Concepts and Applications of Molecular Similarity, Wiley, New York, 1990.
[2] R. Carbó (Ed.), Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches, Kluwer Academic, Dordrecht, 1995.
[3] K.D. Sen (Ed.), Molecular Similarity I, Topics in Current Chemistry 173 (1995).
[4] K.D. Sen (Ed.), Molecular Similarity II, Topics in Current Chemistry 174 (1995).
[5] R. Carbó, P.G. Mezey (Eds.), Advances in Molecular Similarity 1 (1996) in press.
[6] D.L. Cooper, N.L. Allan, J. Comput.-Aided Mol. Design 3 (1989) 253.
[7] D.L. Cooper, N.L. Allan, J. Am. Chem. Soc. 114 (1992) 4773.
[8] N.L. Allan, D.L. Cooper, J. Chem. Inf. and Comp. Sci. 32 (1992) 587.
[9] D.L. Cooper, N.L. Allan, in R. Carbó (Ed.), Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches, Kluwer Academic, Dordrecht, 1995, p. 31.
[10] N.L. Allan, D.L. Cooper, in K.D. Sen (Ed.), Molecular Similarity I, Topics in Current Chemistry 173 (1995) p. 85.
[11] D.L. Cooper, K.A. Mort, N.L. Allan, D. Kinchington, C. McGuigan, J. Am. Chem. Soc. 115 (1993) 12615.
[12] P.T. Measures, K.A. Mort, N.L. Allan, D.L. Cooper, J. Comput.-Aided Mol. Design 9 (1995) 331.
[13] P.T. Measures, N.L. Allan, D.L. Cooper, in: R. Carbó, P.G. Mezey (Eds.), Advances in Molecular Similarity 1 (1996) 61.
[14] B. Everitt, Cluster Analysis, Heinemann Educational, London, 1974.
[15] R.M. Needham, The Statistician 19 (1967) 45.
[16] J.W. Carmicheal, J.A. George, R.S. Julius, Syst. Zool. 17 (1968) 144.
[17] J.W. Carmicheal, P.H.A. Sneath, Syst. Zool. 18 (1969) 402.
[18] C. McGuigan, T.J. O'Connor, B. Swords, D. Kinchington, AIDS 5 (1991) 1536.
[19] J.J.P. Stewart, J. Comput.-Aided Mol. Design 4 (1990) 1.
[20] D. Jeffries, Chemistry and Industry (1993) 746.

[21] M.J.C. Crabbe, J.R. Appleyard, C.R. Lay, Desktop Molecular
     Modeller (Version 3.0 for Windows), Oxford University Press,
     Oxford, 1994.
[22] C. McGuigan, H.M. Sheeka, N. Mahmood, A. Hay, Bioorg.
     Med. Chem. Lett. 3 (1993) 1203.

[23] C. McGuigan, D. Cahard, H.M. Sheeka, E. De Clercq, J.
     Balzarini, J. Med. Chem., 39 (1996) 1748.
[24] C. McGuigan, D. Cahard, H.M. Sheeka, E. De Clercq, J.
     Balzarini, Bioorg. Med. Chem. Lett., 6 (1996) 1183.
[25] C. McGuigan, N. Mahmood et al., personal communication.