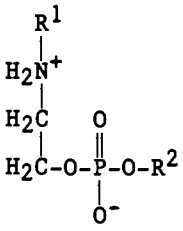
Molecular Similarity of Anti-HIV Phospholipids

David L. Cooper,*,† Kath A. Mort,† Neil L. Allan,‡
Derek Kinchington,§ and Christopher McGuigan‖

Department of Chemistry, University of Liverpool
P.O. Box 147, Liverpool L69 3BX, U.K.
School of Chemistry, University of Bristol
Cantocks Close, Bristol BS8 1TS, U.K.
Department of Virology
Medical College of St. Bartholomew's Hospital
West Smithfield, London EC1A 7BE, U.K.
Department of Chemistry, University of Southampton
Southampton SO9 5NH, U.K.

Received September 24, 1993
Revised Manuscript Received November 1, 1993

Virology data for a particular group of phospholipid show no obvious correlation between HIV inhibition and the identity of the lipid tail or, indeed, of other alkyl substituents on the amine group. It proves possible to rationalize these activity data, and to make successful predictions, using a technique taken from the field of molecular similarity. The families of phospholipids considered in this communication have the general structural formula



in which R^2 represents the lipid tail and R^1 is a relatively small alkyl substituent on the amine group. Molecules with different choices of R^2 and R^1 give 50% inhibition of HIV1 in C8166 T-lymphoblastoid cells with ED_{50} (μM) values[1,2] listed in Table I. It is difficult to make much sense of the various trends by consideration of the chemical structure alone. For example, the replacement by *tert*-butyl of the methyl group at R^1 results in greatly reduced ED_{50} values in the most inactive compounds (HX1 and DD1), and in some enhancement in the activity of two others (OD1 and OL1), whereas it is clear from the data in Table I that EG2 is *less* active than EG1. The series of compounds in which the amine group is free (R^1 = H) tends to have very low ED_{50} values, except that HX3 is *inactive*, unlike HX2. No experimental data are 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. The concept of molecular similarity (see, for example, refs 3 and 4) is particularly useful when the molecular processes involved are extremely complex, as in the present examples, or even unknown. In this communication, we obtain a quantitative answer to the question "How similar are two phospholipids?" by comparing the long-range valence electron densities. A wide variety of techniques of molecular similarity have been developed: these include *inter alia* graph theoretic methods and database searches,[3,4] topological analysis of the three-dimensional shapes of charge densities,[5] and quantitative comparisons of position-space electron densities and density matrices,[6,7] and of electrostatic potentials.[8]

Table I. ED_{50} Values for Inhibition of HIV1 in C8166 T-Lymphoblastoid Cells

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

A central quantity in our approach is the generalized overlap[9-12]

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

in which $\rho_A(\mathbf{p})$ and $\rho_B(\mathbf{p})$ are the electron densities of molecules A and B, expressed as a function of the *momenta* of all the electrons. The inclusion of powers of p into the integrand, via the p^n term, allows us to emphasize different regions of the electron density. Typical values of n are −1, 0, 1, and 2. Putting n equal to −1 (i.e., evaluating $I_{AB}(-1)$) highlights the slowest moving electrons and thus corresponds, in the more familiar position-space representation, to highlighting the long-range slowly-varying valence electron density. The value of $I_{AB}(n)$ depends on the relative orientation of the two molecules, but not on the separation between them in position space, and can take any non-negative value. It is more convenient to generate from $I_{AB}(n)$ a quantity which lies in the range 0–100%, with higher values implying greater similarity. For this purpose, we employ here a definition inspired by the Tanimoto index used in statistical analysis, namely,[12]

$$T_{AB}(n) = 100 I_{AB}(n)/(I_{AA}(n) + I_{BB}(n) - I_{AB}(n)) \qquad (2)$$

There are, of course, alternative ways of "normalizing" $I_{AB}(n)$ (see also refs 6 and 8) which we have used in previous work on structure–activity relations.[12] For most applications, $T_{AB}(-1)$ turns out to be the most useful quantity, with the largest range of values, as described in detail in ref 12. Even so, it is in the nature of quantities derived from momentum-space electron densities that values of $T_{AB}(-1)$ tend to be fairly high.

In view of the size of the phospholipids, we chose to obtain computationally inexpensive position-space wave functions from semiempirical (MNDO) geometry optimizations using the well-

† University of Liverpool.
‡ University of Bristol.
§ St. Bartholomew's Hospital.
‖ University of Southampton.

(1) McGuigan, C.; O'Connor, T. J.; Swords, B.; Kinchington, D. *AIDS* **1991**, *5*, 1536–1537.
(2) Kinchington, D.; McGuigan, C. To be published.
(3) See, for example: *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990.
(4) *Proceedings of the First Girona Seminar on Molecular Similarity*; Carbó, R., Mezey, P. G., Eds., to be published.
(5) Duane-Walker, P.; Artera, G. A.; Mezey, P. G. *J. Comput. Chem.* **1991**, *12*, 220.
(6) Carbó, R.; Leyda, L.; Arnau, M. *Int. J. Quantum Chem.* **1980**, *17*, 1185. Carbó, R.; Domingo, Ll. *Int. J. Quantum Chem.* **1987**, *32*, 517.
(7) Ponec, R.; Strnad, M. *Int. J. Quantum. Chem.* **1992**, *42*, 501. Ponec, R., in ref 4.
(8) (a) Hodgkin, E. E.; Richards, W. G. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *No. 14*, 105. (b) Richards, W. G.; Hodgkin, E. E. *Chem. Br.* **1988**, *24*, 1141 and references therein. (c) Burt, C.; Richards, W. G. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 231.
(9) Cooper, D. L.; Allan, N. L. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 253–259.
(10) Cooper, D. L.; Allan, N. L. *J. Am. Chem. Soc.* **1992**, *114*, 4773–4776.
(11) Allan, N. L.; Cooper, D. L. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 587–590.
(12) Cooper, D. L.; Allan, N. L., in ref 4.

0002-7863/93/1515-12615$04.00/0

known MOPAC program.[13] Bearing in mind that it is likely that the phospholipids must be able to mimic the membrane lipids, we compare total densities for the complete molecules, rather than for individual orbitals or for molecular fragments. The numerical evaluation of $T_{AB}(-1)$ was straightforward and relatively inexpensive. Values of $T_{AB}(-1)$ were computed for each pair of molecules, with the phospholipids aligned so as to match as closely as possible the positions of those atoms common to all systems. Results are presented here only for the zwitterions, although all of the key features were also exhibited in comparisons of the neutral molecules.

We display in Figure 1 a representative subset of our results, namely, comparisons relative to the *inactive* compound HX1 and comparisons relative to the highly *active* compound OL2. The corresponding ED_{50} values from Table I are also shown, with an arbitrary value of 200 μM assigned to the most inactive compounds (HX1, DD1, and HX3). It is clear from Figure 1 that active compounds have a high similarity to OL2 and a low similarity to HX1, and *vice versa*. The way in which the various maxima and minima coincide is particularly striking in both cases. It is important to point out that the $T_{AB}(-1)$ values for the series with a free amine group were computed *before* the virology data were available. In this way we were able to *predict* that OD3 and OL3 would both have low ED_{50} values, but that HX3 would be inactive.

(13) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1 and references therein.

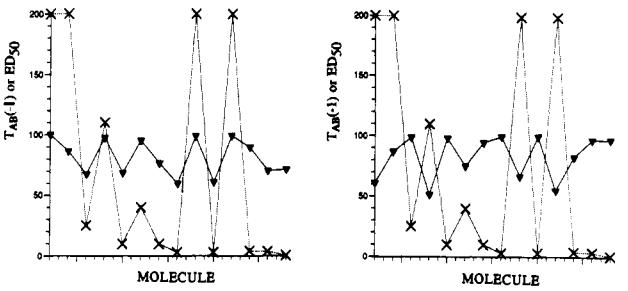


Figure 1. Values of the molecular similarity index $T_{AB}(-1)$ (–▼–) for various phospholipids relative to (a) the *inactive* compound HX1 and (b) the highly *active* compound OL2. The molecules are in the same order as in Table I. Also shown in each frame are the corresponding ED_{50} (μM) values (···×···), with an arbitrary value of 200 assigned to the most inactive compounds (HX1, DD1, and HX3).

For DD3 we observed relatively high $T_{AB}(-1)$ values relative to *both* HX1 and OL2, and so it was not possible to make a definitive prediction in this case.

It is clear that the particular approach to molecular similarity employed here can be used to rationalize HIV virology data for the families of phospholipids considered and even to make some successful predictions of active compounds. Preliminary results for various reverse transcriptase inhibitors are also very encouraging, and these will be reported in due course.