Fuzzy molecular fragments in drug research

Paul G. Mezey and P. Duane Walker

Pharmaceutical research, chemistry and biochemistry suffer from a special handicap – researchers are unable to see individual molecules directly. Current advanced experimental techniques can generate, at best, only blurry pictures of molecules. However, using a new 'computational microscope' method of quantum chemical, *ab initio* molecular imaging, based on an 'additive fuzzy electron density fragment principle' and a molecular electron density 'Lego' assembler density construction method, realistic detailed images of molecules can be generated. The authors describe how this method can be applied to both small and large molecules and indicate the features of the technique that are relevant to the process of drug discovery.

The drug discovery process still involves a great deal of intuition and serendipity. However, during the last decade, systematic, rational, scientific approaches have also become important and often dominant components of the discovery process[1–7]. One systematic approach that provides a framework for selecting potential drug molecules for further testing relies on a computer-based molecular shape analysis and similarity evaluation. A new advance in this field is based on fuzzy molecular electron density fragments, some of which can be identified with chemical functional groups. This technique can also be used to build detailed and reliable shape representations of potential drug molecules and large biomolecules such as proteins.

Molecular shape analysis in drug discovery

Whereas it is customary to interpret the concept of molecular shape in terms of stereochemical bond diagrams, or fused-sphere van der Waals surfaces, the shape of a molecule is, in fact, the shape of its electronic charge cloud. For example, a fused-sphere van der Waals surface is only a crude representation of the molecular shape, primarily for two reasons:

- Whereas fused-sphere van der Waals surfaces depict a sharp molecular boundary, the actual charge cloud has no sharp boundary – it fades gradually into the vacuum surrounding an isolated molecule. It is impossible to provide an accurate model of a gradual change from high density to zero density using a single fused-sphere van der Waals surface.
- Within a molecule, the electron density is not that of a superposition of spherical atoms. There are drastic deviations from sphericity, manifested most prominently in molecules with π-bonds, where it is impossible to describe π-electron density by spheres centered on the nuclei. For example, any such atomic sphere of a typical carbon atom is bisected by a plane where the π-electron density is actually zero. In particular, aromatic rings are very poorly represented by fused spheres because the electron density is certainly neither spherical nor cylindrical in any aromatic π-bond. For a molecule, π-bonds are among the most important chemical characteristics because they are, in general, more reactive than σ-bonds,

Paul G. Mezey* and P. Duane Walker, Mathematical Chemistry Research Unit, Department of Chemistry and Department of Mathematics and Statistics, University of Saskatchewan, 110 Science Place, Saskatoon, SK, Canada S7N 5C9. *tel: +1 306 966 4661, fax: +1 306 966 4730, e-mail: mezey@sask.usask.ca

1359-6446/97/$17.00. PII:S1359-6446(97)01009-X

yet these very π-bonds are rather poorly described by the traditional van der Waals surface approach.

Electron density

It is electron density that makes the molecules what they are; without electron density the nuclear arrangement would fall apart. There is nothing else in a molecule but a set of atomic nuclei surrounded by an electron density cloud. The nuclear locations coincide with local accumulations of the electron density in this charge cloud. Consequently, the electron density cloud, by itself, fully describes all properties of the molecule because there is simply no other material present that could contain additional information. Hence, the shape of the molecule is, indeed, the shape of the electronic charge cloud, the very material that 'holds' the positively charged nuclei together.

It makes sense then to describe molecular shapes using a shape analysis of the electronic charge clouds.

Because the electron density cloud contains all molecular information, differences in reactivities and, specifically, differences in drug actions are ultimately dependent on the differences in the shapes of the electron density charge clouds. Consider two molecules and the shapes of local charge clouds in their reactive regions. If these shapes are identical, then the reactivities of the two molecules are also identical. If these shapes are very similar, we can expect similar reactivities; whereas, if these shapes are very different, then similar reactivities are unlikely. This simple principle has motivated the introduction of the method of quantitative shape–activity relations (QShAR)[8,9], which has been applied successfully for several molecular families[10–14]. Since the shape differences of the electronic charge clouds are responsible for the differences in the activities of molecules, by correlating shape features of drug molecules with their known activities, these correlations can be used for estimating the activities of potential new drug molecules based on the analysis of their shapes. A related technique, the shape group method[9,15–19], provides an automated computational tool for the shape analysis of molecular charge densities. This method generates a family of numerical shape codes that are used in an automated computational algorithm for the evaluation of molecular similarities in terms of the electron densities. These similarity measures are then correlated with drug activities, forming the basis of the QShAR approach[8,9]. The shape group method, the QShAR approach and related methods have been described in the literature. A detailed introduction can be found in a recent monograph[9].

A computational method based on additive, fuzzy molecular fragments extends the scope of most of the QShAR approaches to large molecules, important in drug design. Whereas this is a rather new technique, it has already been applied successfully for the study of toxicity of large polycyclic aromatic hydrocarbon molecules[13,14], identifying local and global shape features relevant to their toxicity.

The AFDF principle

Until recently, rigorous electron density shape analysis and the application of the QShAR methodology have been limited to relatively small molecules. However, a new technique, based on an additive fuzzy density fragmentation approach (the AFDF principle), has opened the way for the construction of high resolution shape representations of large molecules, potential drug molecules and proteins. In this section, a brief description and examples of this methodology are presented, with potential implications in drug discovery.

The AFDF approach is the basis of the molecular electron density 'Lego' assembler (MEDLA) density construction method[20–24]. The two key features of the scheme are:

- Additivity, which leads to a physically meaningful recombination of 'custom-made' molecular fragments into new molecules.
- The use of fuzzy fragments (i.e. electron density clouds without boundaries), which eliminates the large 'mismatch' errors typical of fragments with boundaries.

These two conditions are fulfilled by the following AFDF scheme, motivated by a technique known as Mulliken's population analysis, but involving no integration. The computations are based on high quality (in quantum chemistry terminology: *ab initio* 6-31G** quality) electron densities.

The same set of equations applies if fuzzy fragments are used to rebuild a molecule, or if custom-made fuzzy fragments are used to build a large molecule. In the latter case, the fuzzy fragments are taken from smaller molecules where the local surroundings of the fragments are the same as in the large molecule. These equations involve only the elementary quantum chemistry concepts of atomic orbitals and density matrices, where the latter contain information equivalent to an LCAO (linear combination of atomic orbital) molecular wavefunction. One may think of these equations as the procedure for a synthetic project, except that the synthesis is now accomplished by the computer.

If n is the number of atomic orbitals $\varphi_i(\mathbf{r})$ ($i = 1,2,\ldots,n$) of an *ab initio* wavefunction of a molecule, vector $\mathbf{r}$ is the position variable, and the $n \times n$ density matrix of elements P_{ij} is denoted by $\mathbf{P}$, then the electronic density $\rho(\mathbf{r})$ of the molecule is given by:

$$\rho(\mathbf{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij}\varphi_i(\mathbf{r})\varphi_j(\mathbf{r}) \tag{1}$$

The nuclei of the molecule are partitioned into m mutually exclusive groups to generate m electron density fragments. Any set of nuclei from the molecule can be selected as 'anchor points' for the kth fragment $\rho^k(\mathbf{r})$ of the electron density $\rho(\mathbf{r})$, using the following definition for the $n \times n$ fragment density matrix $P^k{}_{ij}$:

$P^k{}_{ij} = P_{ij}$ if both $\varphi_i(\mathbf{r})$ and $\varphi_j(\mathbf{r})$ are atomic orbitals centered on nuclei of the fragment

$P^k{}_{ij} = f_{ij}P_{ij}$ if precisely one of $\varphi_i(\mathbf{r})$ and $\varphi_j(\mathbf{r})$ is centered on a nucleus of the fragment

$P^k{}_{ij} = 0$ otherwise (2)

In the simplest scheme, following Mulliken, $f_{ij} = f_{ji} = 0.5$ is chosen; however, alternative, more general schemes fulfilling the condition $f_{ij} + f_{ji} = 1$ have also been proposed[25]. The electron density of the kth fuzzy density fragment is defined as:

$$\rho^k(\mathbf{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} P^k{}_{ij}\varphi_i(\mathbf{r})\varphi_j(\mathbf{r}) \tag{3}$$

The sum of the fragment density matrices is the density matrix of the molecule, and the sum of the fragment densities is the density of the molecule:

$$P_{ij} = \sum_{k=1}^{m} P^k{}_{ij} \tag{4}$$

and

$$\rho(\mathbf{r}) = \sum_{k=1}^{m} \rho^k(\mathbf{r}) \tag{5}$$

The simple fragment additivity rules specified by Eqns (4) and (5) are exact within the given *ab initio* framework.

Such fragment densities can also be combined to form approximate electron density for a different molecule, by selecting and arranging fragments so that the nuclear positions match those in the target molecule. The 'fuzzy' density fragments also account for the interfragment interactions occurring within their molecular neighborhoods.

In the first computational step, an electron density fragment databank is generated, based on accurate, high quality *ab initio* quantum chemical calculations for small molecules, specified with a (possibly distorted) nuclear arrangement matching that needed in the target molecule. *Ab initio* quality electron densities can be constructed for large molecules, for any nuclear arrangement, using experimental or theoretically determined nuclear coordinates, or distorted arrangements assumed to occur along reaction paths or in protein-folding processes. The method has been shown to be highly accurate, faithfully reproducing the shapes of 6-31G** *ab initio* quality molecular electron densities, including nonbonding interactions, conjugated and aromatic π-electron distributions, and hydrogen bonds. Note that many features of these important interactions are poorly described or simply ignored by conventional van der Waals, fused-sphere or 'space-filling' representations.

Examples of electron density images

In Figure 1 some electron density fragments obtained from phenanthrene are shown, each at three density thresholds, at 0.1 atomic unit (au), 0.01 au and 0.001 au, where the atomic unit (electronic charge per cubic bohr) of electron density is used. For the two carbon fragments, as well as for the two CH fragments, the location of the carbon atoms along the molecular skeleton is specified using the standard IUPAC numbering convention. The subtle shape differences between these local density clouds are evident, especially at low densities; these differences are caused by the different local surroundings within the phenanthrene molecule.

The phenanthrene system, as well as the benzophenanthrene molecule shown in Figures 2 and 3, have been used to model the role of deviations of aromatic systems from planarity in the interactions of potential drug molecules and toxicants with aromatic systems[13,14]. As Figures 2 and 3 show, in benzophenanthrene the nonplanarity is very pronounced, caused by the high degree of steric hindrance between two hydrogens on the two terminal rings. Overall planarity would require a planar ring-like arrangement of seven nuclei, including the two interfering hydrogen nuclei and five carbon nuclei; such a hypothetical arrangement would be extremely strained. As shown in Figure 3, the molecule avoids this strain at a price of smaller strains in each aromatic ring, and the nonplanarity is approximately evenly distributed along the four actual rings of benzophenanthrene.

Conformational studies, molecular similarity and complementarity analysis, relying on the actual shape of the electron

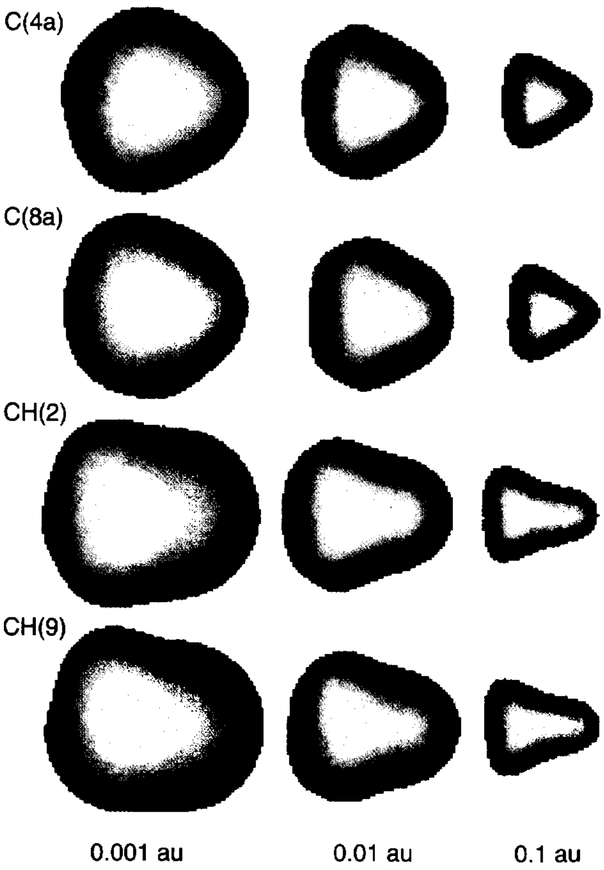

Figure 1. *Typical electron density fragments taken from phenanthrene are shown at the density thresholds of 0.1 atomic units (au), 0.01 au and 0.001 au. Note the subtle differences in the shapes of the two CH fragments, especially at the 0.001 au density level, and the somewhat less noticeable differences in the two C fragments. The carbon atoms are numbered using the standard IUPAC numbering convention.*

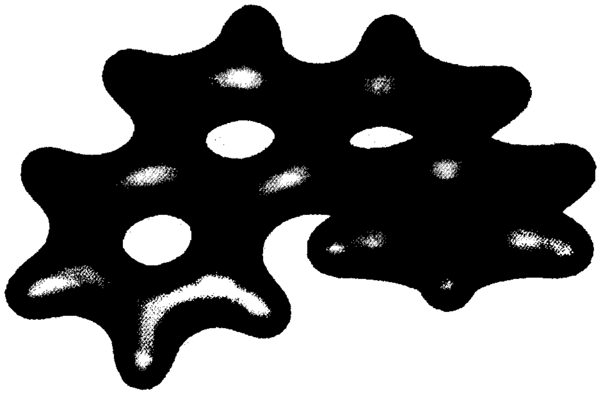

Figure 2. *The 0.1 au electron density contour of the benzophenanthrene molecule showing a marked nonplanarity.*

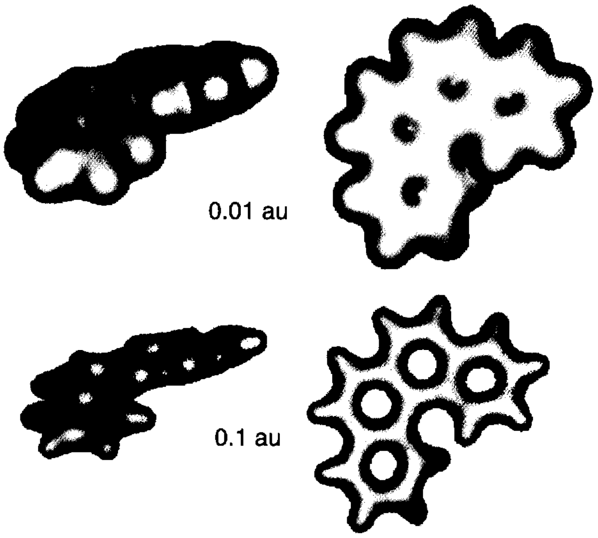

Figure 3. *Two additional views of the benzophenanthrene molecule indicating nonbonded interactions between two interfering H atom moieties at the lower density threshold.*

density, can now be extended from small systems, such as β-alanine[26] shown in Figure 4, to much larger molecules such as taxol[23] (Figure 5) and Leu-enkephalin[21] (Figure 6), as well as to macromolecules such as the protooncogene tyrosine kinase protein 1ABL, which contains 873 atoms (Figure 7). In Figure 5, one low-energy conformation of taxol[23] is shown. It is not currently known whether this geometrical arrangement is indeed the one with the most pronounced bioactivity. The skeletal bond diagram (also shown in Figure 5) underlines the contrast between the electron density images of molecular bodies and the conventional stereodiagrams. Based on the stereodiagram alone, it is hard, if not impossible, to envisage the actual shape of any conformation of the taxol molecule.

Note that, in all these figures, the brighter color indicates higher density, near the atomic nuclei, whereas the fading color at greater distances from the nuclei is intended to indicate the 'fading' of the electron density cloud, as, eventually, these densities converge to zero at large distances from the nuclei.

The level of quantum chemical accuracy, the demonstrated *ab initio* quality of the computed electron densities, gives sufficient confidence in the results for each of the

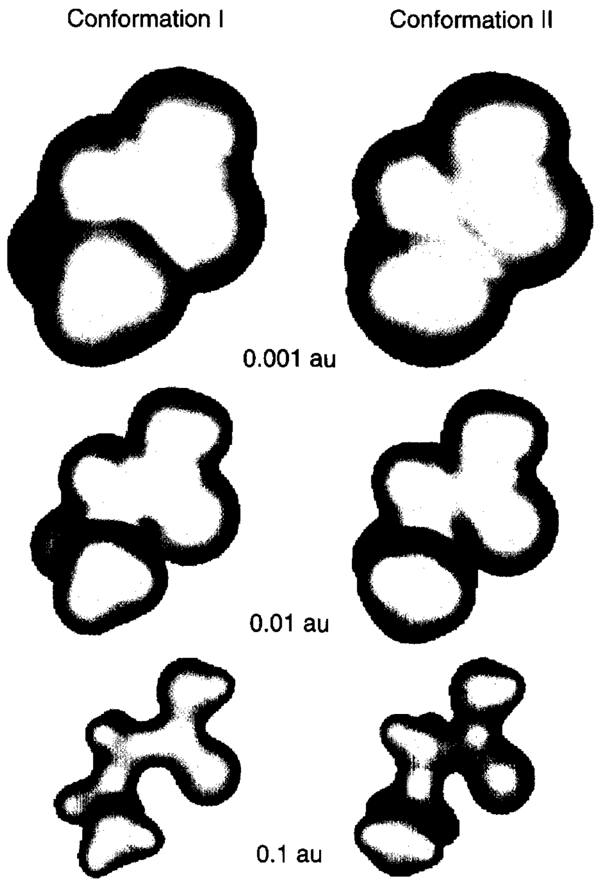

Figure 4. Images of the two most stable conformations of the β-alanine molecule. Numerical shape similarity measures provide a more reliable basis for shape comparisons than simple visual inspection. Although these two conformations are clearly distinct, being separated by an energy barrier, as well as having different shapes (they differ primarily in the orientation of the NH_2 group), an observer relying only on visual cues may find it difficult to recognize and distinguish these two distinct structures.

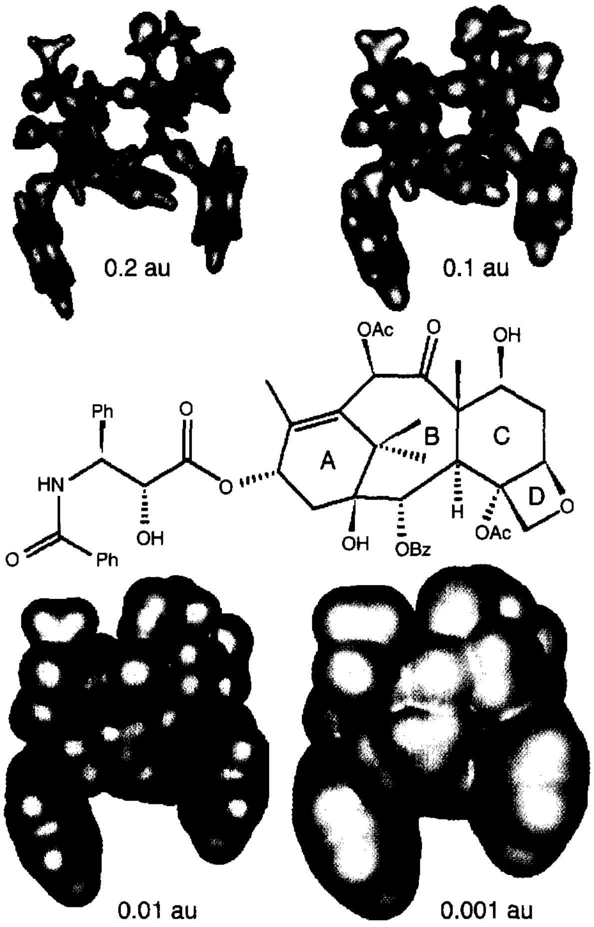

Figure 5. Electron density images of the taxol molecule, an important anticancer drug, built using the additive, fuzzy electron density fragment methodology. The structural formula, also shown, does not provide many clues as to the actual shape of the molecule.

conformations considered. It is possible to explore a whole range of possible conformations. By changing the input nuclear coordinates, the electron density of virtually any conformation of these molecules can be computed. Even for a protein, these images can be computed in less than 30 minutes on a desktop workstation if a suitable fragment databank is available, providing a glimpse into the molecular world.

Prospects and future applications

The detailed view of molecular electron densities obtained using the AFDF methodologies provides a 'true' picture of the fuzzy, three-dimensional 'bodies' of molecules. Besides providing a more accurate description of the expected bonding patterns than the more conventional van der Waals surface models, the method reveals the onset and scope of various nonbonded interactions, steric congestions, hydrogen bonds, as well as the uneven 'swelling' of aromatic and other π-bonded molecular regions, as one switches from the 'skeletal' high density contours to the bulkier low density contours.

It is worth mentioning that the MEDLA method was the first approach that could generate detailed images of electron densities for proteins, at a resolution exceeding current crystallographic techniques by about a factor of 100. This technique, which provides a detailed view of the actual shapes of macromolecules, has many potential applications, especially in biochemistry, drug research, toxicology and biotechnology.

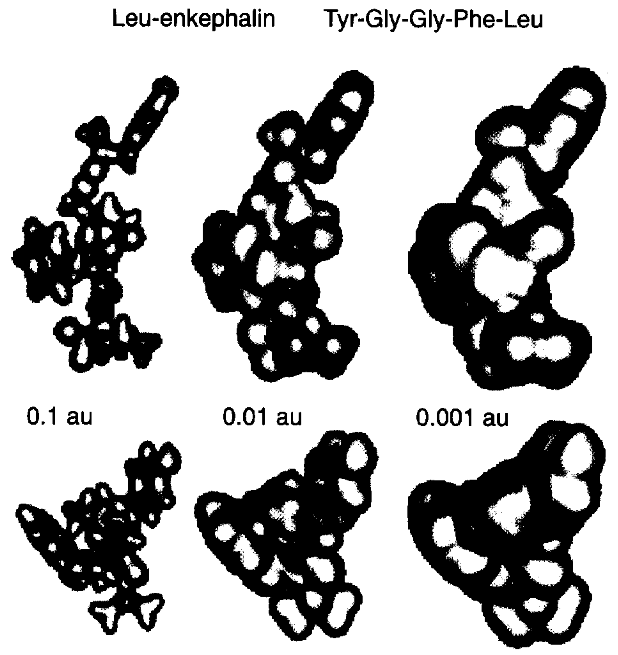

Figure 6. *Electron density images of the Leu-enkephalin molecule showing a whole range of nonbonded interactions not usually revealed by simple van der Waals surface models.*

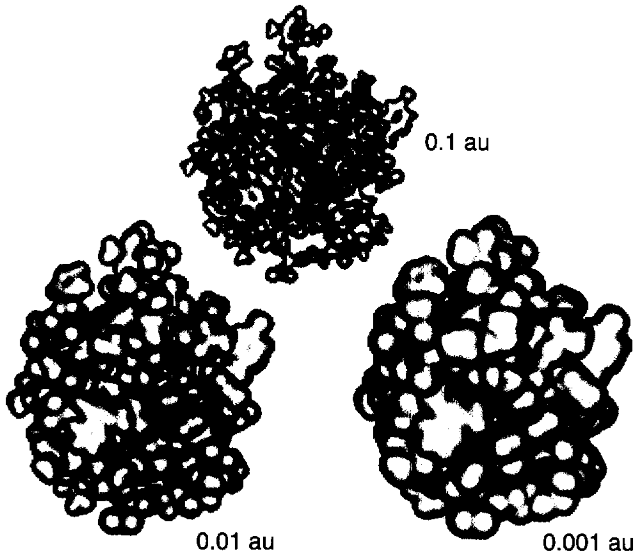

Figure 7. *Electron density images of the protooncogene tyrosine kinase protein 1ABL, a macromolecule containing 873 atoms. The resolution of these MEDLA images compares favorably with images obtained for proteins using crystallographic X-ray diffraction methods.*

However, viewing molecules is only one application. The additive, fuzzy electron density fragments also provide new tools for accurate molecular modeling using computer algorithms for the evaluation of numerical measures for molecular similarity and complementarity. For large systems, as well as for smaller drug molecules interacting with them, several new shape analysis techniques have been proposed[27], utilizing in part the fundamental properties of electron densities. Use of additive fuzzy fragments leads to considerable simplifications in the computations, and gives new insight and motivation for a more efficient use of molecular modeling in the process of drug discovery.

ACKNOWLEDGEMENT

Research work leading to this report has been supported by both operating and strategic research grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1 Martin, Y.C. (1978) *Quantitative Drug Design: A Critical Introduction*, Marcel Dekker
2 Richards, W.G. (1983) *Quantum Pharmacology*, Butterworths
3 Franke, R. (1984) *Theoretical Drug Design Methods*, Elsevier
4 Dean, P.M. (1987) *Molecular Foundations of Drug–Receptor Interaction*, Cambridge University Press
5 Johnson, M.A. and Maggiora, G.M., eds (1990) *Concepts and Applications of Molecular Similarity*, Wiley
6 Dean, P.M., ed. (1995) *Molecular Similarity in Drug Design*, Chapman & Hall/Blackie
7 Carbó, R., ed. (1995) *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, Kluwer Academic
8 Mezey, P.G. (1992) *J. Chem. Inf. Comput. Sci.* 32, 650–656
9 Mezey, P.G. (1993) *Shape in Chemistry: An Introduction to Molecular Shape and Topology*, VCH
10 Walker, P.D. and Mezey, P.G. (1995) *J. Comput. Chem.* 16, 1238–1249
11 Walker, P.D. *et al.* (1995) *J. Chem. Inf. Comput. Sci.* 35, 568–578
12 Walker, P.D. *et al.* (1995) *J. Comput. Chem.* 16, 1474–1482
13 Mezey, P.G. *et al.* (1996) *J. Chem. Inf. Comput. Sci.* 36, 602–611
14 Mezey, P.G. *et al.* (1997) *Use of QShAR to model the photoinduced toxicity of PAHs: electron density shape features accurately predict toxicity* (in press)
15 Mezey, P.G. (1986) *Int. J. Quantum Chem. Quant. Biol. Symp.* 12, 113–122
16 Mezey, P.G. (1987) *J. Comput. Chem.* 8, 462–469
17 Mezey, P.G. (1987) *Int. J. Quantum Chem. Quant. Biol. Symp.* 14, 127–132
18 Mezey, P.G. (1988) *J. Math. Chem.* 2, 325–346
19 Mezey, P.G. (1988) *J. Math. Chem.* 2, 299–323
20 Walker, P.D. and Mezey, P.G. (1993) *J. Am. Chem. Soc.* 115, 12423–12430
21 Walker, P.D. and Mezey, P.G. (1994) *J. Am. Chem. Soc.* 116, 12022–12032
22 Walker, P.D. and Mezey, P.G. (1994) *Can. J. Chem.* 72, 2531–2536
23 Walker, P.D. and Mezey, P.G. (1995) *J. Math. Chem.* 17, 203–234
24 Borman, S. (1995) *Chem. Eng. News* 73, 29–29
25 Mezey, P.G. (1995) in *Molecular Similarity* (*Topics in Current Chemistry*, Vol. 173) (Sen, K., ed.), pp. 63–83, Springer-Verlag
26 Heal, J.A. *et al.* (1996) *Can. J. Chem.* 74, 1660–1670
27 Carbó-Dorca, R. and Mezey, P.G., eds (1996) *Advances in Molecular Similarity*, JAI Press