

Molecular surface comparison. 2. Similarity of electrostatic vector fields in drug design

Frank E. Blaney,* Colin Edge,* and Rob W. Phippen†

*SmithKline Beecham Pharmaceuticals, Harlow, Essex, England

†IBM UK Laboratories, Ltd., Winchester, Hants, England

In the first article of this series¹ a real-time graphics method was described for molecular similarity of scalar properties. This has now been extended for the comparison of molecular vector properties, most notably electrostatic field. A comparison of the various techniques of calculating fields is presented that includes a new method based on natural orbital fitted point charges. In the two examples described, namely, a series of benzodiazepine agonists and a set of serotonin 5-HT₃ antagonists, the program has been shown to produce useful pharmacophoric overlaps that can be used in the design of novel therapeutic agents.

Keywords: molecular similarity, molecular comparison, molecular graphics, electrostatic fields, vector properties, gnömonic projection, benzodiazepine agonists, 5-HT₃ antagonists

INTRODUCTION

Of the several forces (electrostatic, steric, hydrophobic, dispersive, hydrogen bonding, etc.) that are involved in drug-receptor interactions, it has long been recognized that electrostatic forces are the most important in terms of long-range recognition. A classic description of these forces shows that they vary according to a $1/r$ to $1/r^3$ where r is the distance between the two interacting species. These longer range interactions are in contrast to a variance of $1/r^6$ for dispersive forces or $1/r^{12}$ for steric repulsion. Indeed, the recognition phase of serotonin (5-HT) receptor agonism has been studied in great detail² and it has been shown that this

can be fully described in terms of an electrostatic "orientation vector." Electrostatic interactions are often subdivided into ion-ion, ion-dipole, and dipole-dipole terms. The ion-ion term (which has a $1/r$ distance relationship) has no anisotropic dependence and is frequently described by studies of electrostatic potentials, this being a scalar property. The description of interactions with molecular dipoles is by nature, however, a vector term and is therefore related in turn to the electric field of the molecule. A study of the electrostatic field produced by a potential drug molecule should therefore be especially useful in the prediction of its long-range molecular recognition and, as it approaches its final site of action, of its interacting orientation. When comparing a series of molecules the use of the electrostatic field rather than potential should reflect, to a greater extent, the subtleties around putative binding regions. Indeed, field values tend to be small except around highly electronegative or electropositive parts of a molecule. Furthermore, being a vector quantity, electrostatic fields should be of more use in comparing the orientation of molecules binding to a receptor.

In the absence of any firm structural knowledge of enzyme or receptor binding sites (which is normally the case), most molecular modeling is based on the pharmacophore approach. Simply stated, this assumes that a series of active molecules binds via a number of specific sites common to those molecules. These may be electrostatic, hydrogen bonding, and so on. When these molecules are overlapped (via their specific binding sites) they form an "active" volume within the receptor. Molecules that have the specific binding sites but that are not active often occupy areas in space not accessible to the receptor. This is frequently known as the *excluded volume*. By relaxing torsional freedom, and so on, such approaches can allow for conformational flexibility within the molecules. What is crucial to the whole exercise, however, is how exactly the active (and inactive) molecules are overlapped. We and others have shown that molecular similarity is a powerful technique in

Color Plates for this article are on pp 194-197.

Address reprint requests to Dr. Blaney at SmithKline Beecham Pharmaceuticals, New Frontiers Science Park (North), Third Avenue, Harlow, Essex CM19 5AW, England.

Received 10 February 1995; accepted 23 February 1995.

drug design.^{1,3} This allows for a quantitative comparison of molecules on the basis of properties other than the straight fitting of atoms. As interactions between molecules (e.g., drug-receptor interactions) occur at the electronic rather than the nuclear level, the calculation of similarity on the basis of properties such as electrostatic potentials or fields is more appropriate and often suggests overlaps that are not obvious to the user.

MOLECULAR SIMILARITY AND GNOMONIC PROJECTIONS

Probably the most useful way of displaying the electrostatic potential or field is to calculate its value on a suitable surface around the molecule. The most meaningful surface for these purposes is an electron density surface contoured at a suitably low value. Such surfaces, however, require additional quantum mechanical calculations to generate, so it is often more convenient to use a simple surface of the van der Waals or solvent-accessible type. The difficulty in comparing these surfaces is that they are irregular in shape. It was first proposed by Chau and Dean⁴ in 1987 that a simple way around this problem would be to project the values of the property onto a uniform sphere surrounding the surface (in much the same way as the surface of the earth is projected onto a globe of the world). This reduces the problem of comparing the surfaces of two molecules to three rotational degrees of freedom. The concept was used several years ago in the development of a program to study, in real time, molecular similarity based on electrostatic potentials or overall shapes.¹ These are, of course, scalar properties. This program has now been extended to calculate and display real-time similarity of vector properties, of which electrostatic field is the most widely used.

The first task in the generation of a gnomonic projection is to choose a suitable point from which to project (the centre of interest or COI). This is often the center of mass, although the final choice is under the control of the user. With elongated molecules, for example, it is sometimes more appropriate to choose a COI closer to one end of the molecule, particularly if that end is more polar or is believed to be more closely involved in binding. Problems sometimes arise where a line radiating from the COI passes through more than one point on the surface, as is the case where the surface is reentrant. These have been previously discussed elsewhere¹ in detail. In practice, it has been found that careful choice of the COI can minimize this problem.

To ensure that no part of the projection has a greater significance than any other during the similarity calculation, it is important to choose a set of equally spaced points on the spherical surface. This is impossible for an arbitrary number of points. However, Chau and Dean have shown^{4,5} that a near-optimal polygonal representation of a sphere at any given polygon resolution is an icosahedron, further tessellated by triangles on each face. Having chosen a set of points at the desired resolution, a vector is now drawn to the COI. Where it passes through the appropriate molecular surface, the property of interest is calculated. If, as in the case of a reentrant surface, more than one point is found, the one furthest from the COI is chosen.

In a normal molecular orbital (MO) calculation, this set of points would be passed back to the program, sometimes reoriented to the principal axes of the molecule. In practice, however, provided the property of interest is quickly calculated, it is often more convenient to generate the surface values by interpolation from a regular grid calculated around the molecule. This has the added advantage that more than one surface may be calculated from the same data set, for example, at multiples of the van der Waals surface. In addition to the coordinates of the molecule and the COI, the main input to the program is therefore a regular grid (typically $51 \times 51 \times 51$ points) of Cartesian and associated property data.

To compare two molecules, one simply calculates the difference in the value of the property of interest at each corresponding point of the two spheres and sums this over all points. A numerical value of the similarity at this orientation can then be calculated using one of the methods previously described elsewhere.⁶ Alternatively, this can be described by some correlation function such as a Spearman's rank correlation.⁷ To find the best overlap it is then necessary to maximize this correlation or similarity index (or minimize the sum of differences in values). The advantage of the projection method now becomes apparent. The problem is reduced to a nine dimensional one, namely, three translational degrees of freedom for each COI and three rotational for the relative orientation of the two spheres. In practice it has been found that movement of the COIs has little effect on the final results and therefore they are fixed at the start of the calculation. If a method can be found to calculate the correlation fast enough, then the overlap of two structures becomes a simple matter of "real-time" x , y , or z rotation of one sphere with respect to the other.

Unfortunately, using the techniques of Chau and Dean,^{4,5} any reorientation of a molecule requires a complete recalculation of its spherical projection. To avoid this, the whole surface property is stored as a two-dimensional map. This is similar to the two-dimensional representation of height on a map of the world. Calculating the value of the property at the surface for a particular space-fixed vector, in this case one of the vectors defining the vertices of the subtessellated icosahedron, is reduced to the simple task of locating its position on the map. The new value of the property is calculated from the redefined position by a straightforward linear interpolation.

The reduced recomputation resulting from this technique allows the difference between two maps to be recalculated 5–10 times per second on modern workstations. This means that the orientation of a molecule can be linked to a suitable input device (e.g., a set of dials), and the difference between the surface properties of the two molecules can be recalculated and displayed in real time.

CALCULATION OF ELECTROSTATIC FIELDS

As with all one-electron properties (density, potential, etc.) the electric field is readily calculated from the first-order density matrix, which can be obtained from any quantum mechanical calculation. The field $E(R)$ at point R is readily evaluated within the Born–Oppenheimer approximation. If

R_α is the position vector of nucleus α and r_i the position vector of electron i , E is given by

$$E(R) = \frac{e}{4\pi\epsilon_0} \sum \frac{Z_\alpha(R - R_\alpha)}{|R - R_\alpha|^3} - \frac{e}{4\pi\epsilon_0} \left\langle \Psi_0 \left| \frac{R - r_i}{|R - r_i|^3} \right| \Psi_0 \right\rangle$$

which can be expanded into atomic orbitals, χ_p , as

$$E(R) = \frac{e}{4\pi\epsilon_0} \sum \frac{Z_\alpha(R - R_\alpha)}{|R - R_\alpha|^3} - \frac{e}{4\pi\epsilon_0} \sum_p \sum_q D_{pq}^{\text{HF}} \left\langle \chi_p \left| \frac{R - r_i}{|R - r_i|^3} \right| \chi_q \right\rangle$$

where D_{pq}^{HF} are the elements of the first-order density matrix. These are straightforward two- or three-center potential integrals that can be readily evaluated as either Slater functions or more commonly as Gaussian expansions.

For a set of evenly spaced points on a typical molecular surface, it is perfectly feasible to calculate all the field values analytically. However, the method used above for the comparison of two irregular surfaces requires the prior evaluation of all points on a three-dimensional grid. Even for a relatively coarse grid of, say, $30 \times 30 \times 30$, (i.e., a total of 27 000 points), the time taken for the calculation becomes unmanageable. It was necessary therefore to utilize a more approximate method.

Previous workers⁸ have evaluated the use of semiempirical MO methods for the calculation of electrostatic potential. The main time saving here, however, is in the initial SCF calculation rather than in the calculation of the one-electron potential integrals. The relative reduction in the total number of these integrals resulting from neglect of the core electrons is small. Strictly speaking, under the ZDO approximation, it should not be necessary to calculate those two- and three-center integrals corresponding to the off-diagonal elements of the density matrix. However, these have been shown⁸ to contribute significantly to the overall magnitude of the potential and it is generally necessary to deorthogonalize the density matrix and then calculate all the corresponding off-diagonal potential integrals. The same

integral forms are used in the calculation of electrostatic fields as in potentials. Therefore, because there is no real saving in time by using an analytical semiempirical method, it was concluded that an "accurate" point charge method was required. Using the anxiolytic molecule diazepam as a test case, several of these have been evaluated and compared with the ab initio field calculated using a 3-21G basis set.

EVALUATION OF APPROXIMATE ELECTROSTATIC FIELD METHODS

The choice of diazepam as a test molecule is appropriate because it contains several features generally difficult to reproduce using approximate methods. As seen below, some point charge methods completely fail to give a field that resembles that calculated by ab initio Hartree-Fock-type computations.

Color Plate 1a shows the field corresponding to regions of negative electrostatic potential only, as calculated from an ab initio wavefunction using a 3-21G basis set (with the Gaussian 90⁹ program). The full field for both positive and negative regions is shown in a different representation in Color Plate 1h. The strongest attractive field is found, not surprisingly, in the region around the carbonyl with a secondary field in the vicinity of the imino nitrogen. Smaller field lines can be observed around the halogen atom, which might be thought of as being grouped around its "lone pairs" rather than lying in a spherically symmetrical distribution. There is also another group of field lines going into the face of the aromatic C-ring of the benzodiazepine on the side opposite to the imino nitrogen, but they are noticeably fewer and smaller around the A-ring. The maximum values in these regions are summarized in Table 1.

The most common approximate method for the calculation of electrostatic properties is the use of partial atomic point charges. These can come from a variety of computational methods, including ab initio and semiempirical MO techniques. A number of other totally empirical charging schemes such as those of Gasteiger and Marsili¹⁰ or Abrahams and Grant¹¹ have also been described. Because of its simplicity, the use of partial charges is the most frequently chosen technique, for example, in the calculation of electrostatic potentials. In a similar manner the field of a molecule can be described in terms of point charges. In its simplest form, the electric field $E(R)$ of a molecule at a

Table 1. Values of the maximum field lines around each of the electronegative regions associated with diazepam

Method	Field (kcal Å ⁻¹) around:				
	Oxygen	Imino nitrogen	Chlorine	Phenyl A-ring	Phenyl C-ring
Ab initio 3-21G wavefunction	-47.642	-22.605	-14.386	-4.791	-9.748
CNDO Mulliken charges	-19.730	-3.673	-8.992	—	—
CNDO-VSS-Mulliken charges	-21.092	-3.769	-9.100	—	—
Mopac ESP charges	-25.381	-21.589	—	-16.870	-19.693
NAO-PC AM1 charges	-40.747	-23.098	-28.227	-4.996	-7.230
3-21G Mulliken charges	-42.950	-31.758	-9.649	-17.025	-33.115
3-21G ESP charges	-35.391	-25.013	-10.265	-7.351	-14.516

point with position vector R can be defined in terms of the partial atomic point charges Q_α of the molecule as follows:

$$E(R) = \sum \frac{Q_\alpha(R - R_\alpha)}{|R - R_\alpha|^3}$$

One of the most common sources of point charges is Mulliken population analysis (MPA), which is carried out at the end of almost every MO calculation. Using MPA charges derived from a CNDO semiempirical computation, the field associated with diazepam is shown in Color Plate 1b. This is characterized essentially by field lines distributed spherically around the oxygen and the halogen. The magnitude of these vectors is much lower around the oxygen when compared to ab initio fields. However, those around the chlorine are relatively large, this being a result of the unusually large partial negative charge computed for the chlorine by the CNDO method. There is only a very small attractive field associated with the imino nitrogen and none at all with either of the phenyl rings.

A detailed study of the calculation of electrostatic potentials using the CNDO semiempirical method was carried out by Gessner-Prettre and Pullman.⁸ From this it was concluded that the only way of coming anywhere near ab initio results was to deorthogonalize the one-electron density matrix and then compute all the associated potential integrals. Although the absolute values were smaller, they concluded that a better description of the relationship between potential and distance from the molecule could be calculated by making use of the radial components of the s orbitals, that is, by using the Slater exponents. As only the s orbital terms were used, the resultant atomic charge distribution was still spherically symmetrical. This so-called VSS method has become popular largely because of its availability from Quantum Chemistry Program Exchange (QCPE).¹² Using the VSS approximation, the field associated with atomic point charges is given by the equation

$$E(R) = \sum \frac{Z_\alpha(R - R_\alpha)}{|R - R_\alpha|^3} - \sum_{\alpha \neq H} \frac{V_\alpha^{\text{heavy}}(R - R_\alpha)Q_\alpha^{\text{heavy}}}{|R - R_\alpha|^3} - \sum_{\alpha = H} \frac{V_\alpha^H(R - R_\alpha)Q_\alpha^H}{|R - R_\alpha|^3}$$

where

$$V_\alpha^{\text{heavy}} = \frac{1}{r} \left\{ 1 + \zeta r \left[\frac{3}{2} + \zeta r \left(\frac{1 + \zeta r}{3} \right) \right] e^{-2\zeta r} \right\}$$

$$r = |R - R_\alpha|$$

and

$$V_\alpha^H = \frac{1}{r} [1 - (1 + \zeta r)e^{-2\zeta r}]$$

Q_α^{heavy} and Q_α^H are the total electronic charges associated with each heavy atom and hydrogen, respectively, and ζ is the Slater s exponent for that atom.

The diazepam field calculated using the above described method with CNDO MPA charges is shown in Color Plate 1c. As can be seen from this and the data in Table 1, there is little improvement in the field description except that the absolute values of the field are slightly larger.

The choice of method in an MO calculation can have a profound effect on the final field. Color Plate 1f shows the field produced using MPA charges derived from an ab initio Hartree-Fock calculation with a 3-21G* basis. The magnitude of the field around the heteroatoms is much closer to the analytical ab initio values, with the nitrogen being somewhat higher and the halogen lower. The magnitude of the field associated with the aromatic rings is much too high, however, the region around the C-ring being computed as larger than that associated with the imino nitrogen!

Another commonly used source of atomic point charges is by least-squares fitting to precomputed electrostatic potentials (ESP charges).¹³ Typically, a series of surfaces is constructed around the molecule and the potential is calculated for points on each of these. Charges are then generated that best reproduce the potential. These charges are commonly used in molecular mechanics force fields such as CHARMM¹⁴ or AMBER.¹⁵ Again, the choice of MO program can affect the results dramatically. Color Plate 1d shows the field calculated with one widely used program, MOPAC-ESP.¹⁶ The maximum value around the oxygen, the nitrogen, and the two phenyl rings varies by only 8.5 kcal Å⁻¹, with the magnitude around the region of the carbonyl being half that calculated by ab initio SCF. In addition to these discrepancies, there is, oddly enough, no negative potential around the halogen. Instead there is a large negative charge associated with the adjacent C-7 carbon of ring-A. In Color Plate 1g, the field shown was calculated with 3-21G* ab initio potential fitted charges. This is much closer to the ab initio analytical field than the semiempirical example in Color Plate 1d. The field around the two phenyl rings is still somewhat high. Of the atom-centered point charge methods discussed so far, namely, MPA charges, ESP charges, and the VSS correction, the use of ab initio-derived ESP charges is probably the best. The initial calculations, however, are still computationally expensive and represent no real saving in time for coarse grids.

The fourth and final point charge approximation to be evaluated, the natural atomic orbital (NAO) fitted charge method, comes from the recent work of Rauhut and Clark.¹⁷ This differs significantly from the others in that it is able to describe the anisotropic nature of the density around atoms resulting from differences in hybridization. To do this, it makes use of natural atomic orbitals. In a semiempirical (usually AM1) sp basis, for example, these are readily obtained by diagonalization of the 4×4 blocks associated with each nonhydrogen atom in the density matrix (this is illustrated schematically in Figure 1). Charges are then least-squares fitted to each lobe of the NAOs to reproduce the density. This means that each nonhydrogen atom has a total of nine point charges (one nuclear and eight electronic) associated with it. The hydrogens, which have only one valence $1s$ orbital, are still represented by a single point charge centered at the nucleus.

Using the NAO-PC method, the field at a point of position vector R is given by the equation

$$E(R) = \sum_{\alpha \neq H} \frac{Z_{\alpha}(R - R_{\alpha})}{|R - R_{\alpha}|^3} - \sum_{\alpha \neq H} \sum_{NAO=1}^8 \frac{Q_{\alpha}^{NAO}(R - R_{\alpha}^{NAO})}{|R - R_{\alpha}^{NAO}|^3} - \sum_{\alpha=H} \frac{Q_{\alpha}^H(R - R_{\alpha})}{|R - R_{\alpha}|^3}$$

where Q_{α}^{NAO} are the eight orbital point charges for each heavy atom, α , and Q_{α}^H is the NAO point charge for each hydrogen. Using this method, the field for diazepam is shown in Color Plate 1e. This shows a remarkable resemblance to the full ab initio field in Color Plate 1a. The vectors in the regions of the phenyl A- and C-rings are almost identical in magnitude and direction. The same is true of the oxygen and nitrogen atom areas. The magnitudes of the field lines around the chlorine are calculated to be somewhat higher than their corresponding ab initio values but otherwise the similarity between the two methods is high. The full calculation of NAO-PC charges using the AM1 Hamiltonian and a previously minimized structure takes only a few seconds on a workstation, as does the calculation of the field on the surface. It was our conclusion, therefore, that this was the method of choice for the calculation of the grids used in the subsequent similarity calculations.

VISUALIZATION OF ELECTROSTATIC FIELDS

Vector fields have traditionally been displayed with three-dimensional (3D) arrows and this method has also been adopted here. To aid in the study of these, a separate program was written that allowed interactive manipulation of the display. Thus fields could be read in from a variety of different sources, including ab initio and semiempirical analytical calculations and all the approximate methods de-

scribed above. Displays could be generated either from 2D planes through the molecule or on the usual types of molecular surfaces. The field lines themselves were interchangeably depicted either as vector line arrows or as Gouraud-shaded raster objects (see Color Plate 1a-h). The former were used for real-time manipulation of the displays. The magnitudes of the arrows, both length, and (in the case of raster displays) width were scaled according to the absolute value of the field at that point. This could be interactively adjusted using a scaling factor set by the user. It was found useful to color-code the field corresponding to areas of negative and positive potential differently. As was subsequently found (see below), it was often more appropriate to consider only those fields associated with areas of negative potential and the program therefore allowed for the interactive inclusion or exclusion of fields of any user-defined range of values or of user-defined areas of the plane or surface. All these options form part of a general MO package used internally. In the context of this article they were especially useful in the analysis of the field calculation methods described above, but they do not form part of the field comparison program.

THE FIELD COMPARISON PROGRAM

The visual layout of the program (known internally as Winmatch-V) is essentially identical to that of its predecessor,¹ which, as already mentioned, was developed for the comparison of scalar properties. This is illustrated in actual use in Color Plate 2. On normal startup, seven windows appear on the screen. The two molecules themselves are displayed in the separate windows to the left of the display, with their relative orientations linked to the respective projections. Instead of the color-coded surfaces of the scalar program, as indicated above, three-dimensional arrows have been used to represent the field, both for real-time comparisons and for final visualization. Having chosen the "center of projection" (which is under user control), a set of vectors is allowed to radiate outward. When they hit the molecular surface, the components of the field are recorded. These vectors are then normalized to define a sphere. The molecular surface used in the program is an "expanded van der Waals" surface, in which the radius of each atom is expanded by a constant term. Two different representations of the vectors are allowed. For real-time rotations, the fields are represented as simple lines. However, a more detailed representation allows the vectors to be displayed as arrows composed of a cylinder joined to a cone. The cylinder and cone primitives are constructed using nonuniform rational B-splines (NURBS) surfaces.¹⁸ With high-performance workstations such as the Silicon Graphics VGX or IBM RS6000 series on which this program was developed, an adjustable lighting model may be included for the raster displays. The length of the vector represents its magnitude. Also, like its predecessor, Winmatch-V is designed to be highly interactive. The vector field projection for each molecule is displayed in the two separate graphical windows to the right, as the appropriate set of arrows placed at the surface of the sphere. As a molecule is rotated, the vector field associated with it also rotates across the sphere. The vector difference between the projections is displayed

	1 2S	1 2Px	1 2Py	1 2Pz	2 2S	2 2Px	2 2Py	2 2Pz	3 2S	3 2Px	3 2Py	3 2Pz	• • •
1 2S	###	###	###	###	***	***	***	***	***	***	***	***	• • •
1 2Px	###	###	###	###	***	***	***	***	***	***	***	***	• • •
1 2Py	###	###	###	###	***	***	***	***	***	***	***	***	• • •
1 2Pz	###	###	###	###	***	***	***	***	***	***	***	***	• • •
2 2S	***	***	***	***	###	###	###	###	***	***	***	***	• • •
2 2Px	***	***	***	***	###	###	###	###	***	***	***	***	• • •
2 2Py	***	***	***	***	###	###	###	###	***	***	***	***	• • •
2 2Pz	***	***	***	***	###	###	###	###	***	***	***	***	• • •
3 2S	***	***	***	***	***	***	***	***	###	###	###	###	• • •
3 2Px	***	***	***	***	***	***	***	***	###	###	###	###	• • •
3 2Py	***	***	***	***	***	***	***	***	###	###	###	###	• • •
3 2Pz	***	***	***	***	***	***	***	***	###	###	###	###	• • •
•	•	•	•	•	•	•	•	•	•	•	•	•	• • •
•	•	•	•	•	•	•	•	•	•	•	•	•	• • •
•	•	•	•	•	•	•	•	•	•	•	•	•	• • •

Figure 1. A diagrammatic density matrix with a valence SP basis. Values are represented by *** or ### and the row and column continuations by •. The NAOs are determined by diagonalizing the 4×4 blocks indicated.

on a third (upper middle) sphere. In places where the vector fields are similar, the arrows displayed in the difference projection are short and will reduce to a point if the field is the same for each molecule. The vector correlation is defined as

$$\frac{\sum(\mathbf{v}_1 \cdot \mathbf{v}_2)}{\{[\sum(\mathbf{v}_1 \cdot \mathbf{v}_1) + \sum(\mathbf{v}_2 \cdot \mathbf{v}_2)]/2\}}$$

where $(\mathbf{v}_1 \cdot \mathbf{v}_2)$ is the dot product of the vectors \mathbf{v}_1 and \mathbf{v}_2 . This is continuously calculated to give a running estimate of the similarity of the vector projections and is displayed as a sliding scale, ranging from -1.0 to $+1.0$, in its own small window at the top. The fast interpolation method used means that the user can rotate one sphere relative to another in real time while watching for the maximum similarity, both from the color of the third sphere and from the similarity or correlation value in the uppermost window.

It often happens when calculating the similarity of two molecules that more than one possible overlap results from the computation. This is just another example of the classic multiple minima problem.¹⁹ As an aid to this, the similarity index or correlation coefficient is precomputed at 10° intervals for all possible x , y , and z rotations. This produces a three-dimensional block of data that is stored in a compact file format.²⁰ The global minimum can hence be quickly located by searching this data block. Furthermore, other local minima can be quickly located visually. The lower middle window is used to display the precomputed similarity index data as a "terrain map." This is a simple deformed rectangular surface with the values of the similarity or correlation displayed as the height above the surface. The surface is displayed for all possible x and y rotations at a chosen value of z rotation. The map is shown diagrammatically in Figure 2. A three-dimensional marker points to the position on the map that represents the current relative orientation of the two molecules. At any instant, the z rotation represented by the terrain map is fixed. This can be changed by rotating the appropriate dial, which then causes the shape of the terrain map to be altered. Alteration of the x or y rotations results in the marker moving across the terrain map, while simultaneously updating the difference information and orientations in all the other windows.

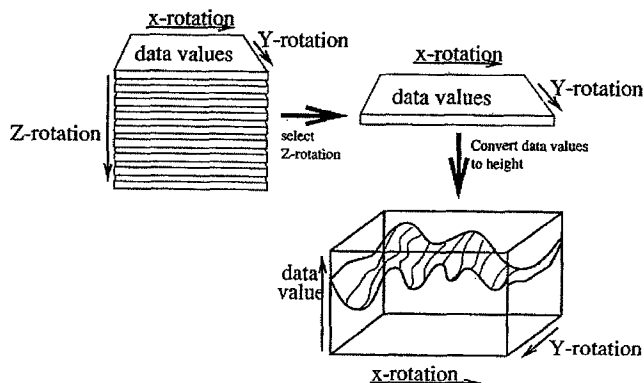


Figure 2. Using precalculated similarity index data in the terrain map display.

USE OF ELECTROSTATIC FIELD COMPARISON

Comparison of a small series of benzodiazepine agonists

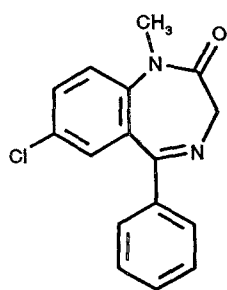
One of the most widespread mental conditions in the western world today is that known as anxiety. A chance discovery by workers at Hoffman-LaRoche²¹ led to the first series of drugs to be used in the treatment of this condition, namely, the benzodiazepines. Many of these have now been used in the treatment of anxiety, including chlordiazepoxide (Librium) and diazepam (Valium). It was later discovered that these compounds bound to specific ligand-binding sites associated with the gamma-aminobutyric acid (GABA) receptor^{22,23} and that a wide variety of structurally diverse ligands could also bind to this site, some of which are shown in Figure 3. Within the computational chemistry community we and many others²⁵ have attempted to define a pharmacophore that would explain how such diverse ligands bind to the same receptor site.

It has already been noted that diazepam has three electronegative regions associated with its heteroatoms. Using electrostatic potential similarity methods, it was discovered that many of the other compounds also had these common features and that two of these electronegative regions were separated by a distance of approximately 8.0 – 10.4 Å. In Color Plate 3, these regions of electronegative potential are shown for diazepam, CL218872, and triazolam.

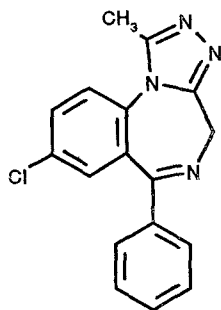
One particularly potent agonist, which at first glance did not appear to fit the pharmacophore pattern, was CGS9896. This compound, however, can in theory exist as a mixture of three different tautomers (Figure 4).

Simple MO calculations suggested that the enol form (tautomer III) was the most stable and that, in this form, the expected geometric relationship between the two electronegative sites was once again achieved (Color Plate 3 [bottom]) shows the calculated electronegative regions for the three tautomeric forms of CGS9896). In their elegant work on the study of molecular determinants for benzodiazepine receptor affinities, Villar *et al.*²⁶ came to a similar conclusion, that the enol form of CGS9896 was the bioactive species. Using the techniques of molecular similarity, therefore, a common overlap of benzodiazepine receptor ligands could be defined that had two lipophilic and three electronegative regions as shown in Color Plate 4.

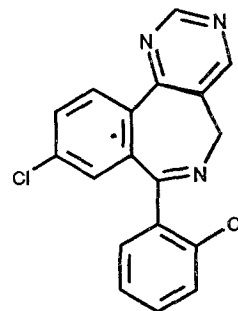
As an initial test for the field similarity program, a small set of benzodiazepine agonists was chosen to see if the above-described pharmacophore could be reproduced. This set of compounds had the added advantage from a similarity index calculation viewpoint that they are all conformationally rigid, each having only one significant rotatable bond (i.e., not including methyl groups). The structures were optimized with the semiempirical MO program VAMP,²⁷ using the AM1 Hamiltonian, and NAO fitted point charges were calculated for these structures. The fields were calculated for a $51 \times 51 \times 51$ cubic grid using the method described above. Precomputed full terrain maps of field similarity were used to examine all local maxima and these were manually fine-tuned using the dials. The correlations obtained from the full grid of field values appeared at first glance to be rather low, although it should be remembered that the range for these is from -1.0 to $+1.0$. One way by



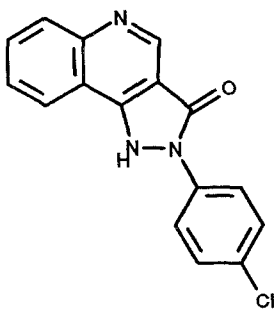
Diazepam



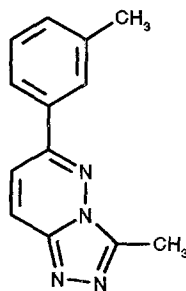
Triazolam



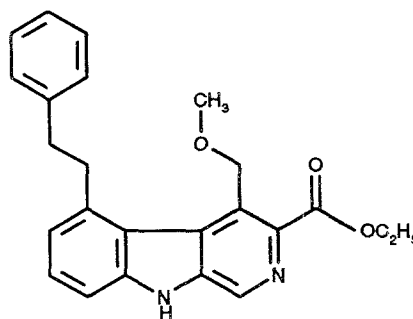
RO-223245



CGS9896



CL218872



ZK91296

Figure 3. Six diverse benzodiazepine agonists.

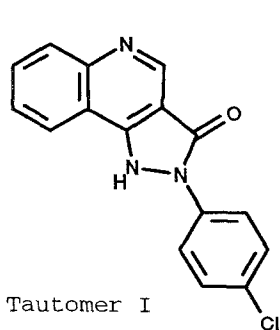
which they could be improved was to ignore the influence of the field associated with areas of positive electrostatic potential. These are generally of less interest in drug design because they do not represent potential ionic binding sites to positively charged lysine or arginine residues, nor do they represent hydrogen bond acceptor sites to donor residues such as serine, threonine, tyrosine, and histidine. The problem is particularly noticeable with molecules such as the benzodiazepine agonists, which are largely electropositive across the surface except in those regions noted above. By flagging those regions of positive potential in the projection map and then ignoring them in the summation of the similarity index or correlation, the magnitude was increased generally by a value of between 0.2 and 0.4. The correlation matrix for the set of maximum similarities, ignoring the positive areas, is shown in Figure 5.

The final calculated overlap of diazepam and triazolam was the same as that previously derived from the potential similarity calculations given above. The overlap of these

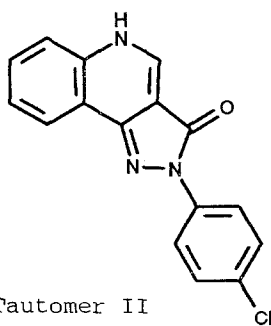
with CGS9896 and CL218872 was, however, different. Instead of the two chloro-substituted phenyl rings of diazepam and CGS9896 being overlaid as was predicted from potential calculations, this ring of CGS9896 now overlapped with the C phenyl ring of the benzodiazepines as shown in Figure 6. With CL218872, the trifluoromethyl-substituted phenyl ring also overlaps with the C-ring of diazepam. In this overlay, the regions of maximum field strength, that is, the carbonyl of diazepam and the basic ring nitrogens of CL218872 and CGS9896, are all perfectly overlapped. The new overlap, which is shown in Color Plate 5, has not been previously considered in our generation of benzodiazepine agonist pharmacophores and could therefore lead, in theory, to new directions in drug design.

A study of 5-HT₃ antagonists

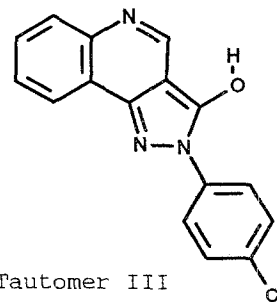
Of all the neurotransmitter systems in the body, 5-HT (serotonin) is one of the most widely studied. It is now well



Tautomer I



Tautomer II



Tautomer III

Figure 4. The three possible tautomers of CGS9896.

Diazepam			
CGS9896	0.4486		
Triazolam	0.2894	0.2477	
CL218872	0.2689	0.1780	0.3779
	Diazepam	CGS9896	Triazolam

Figure 5. Maximum correlation indices calculated for overlaps of the four benzodiazepine agonists diazepam, CGS9896, triazolam, and CL218872.

established that a large number of 5-HT receptor subtypes exist, and these have been broadly categorized into "5-HT₁-like," 5-HT₂, and 5-HT₃. The first two classes are members of the G protein-coupled receptor superfamily and, as such, there is now a considerable amount of knowledge about their receptor sequences and structure. The 5-HT₃ receptor, on the other hand, is an ion channel and little is known about its structure. Despite this, interest continues to increase, particularly in antagonists, because many of these compounds have been linked therapeutically with the treatment of emesis, anxiety, and schizophrenia. Several previous modeling studies have been carried out^{28,29} to determine possible pharmacophore models. These have all been of the atom alignment type mentioned above. This seemed therefore to be another ideal case to use in evaluating the use of the electrostatic field alignment software. Five compounds were chosen for this study. These were BRL24682, BRL43694 (granisetron), BRL46470, ICS205930 (tropisetron), and GR38032 (ondansetron). The structures are illustrated in Figure 6.

With the 5-HT₃ antagonists described above, a greater degree of conformational flexibility over the benzodiaze-

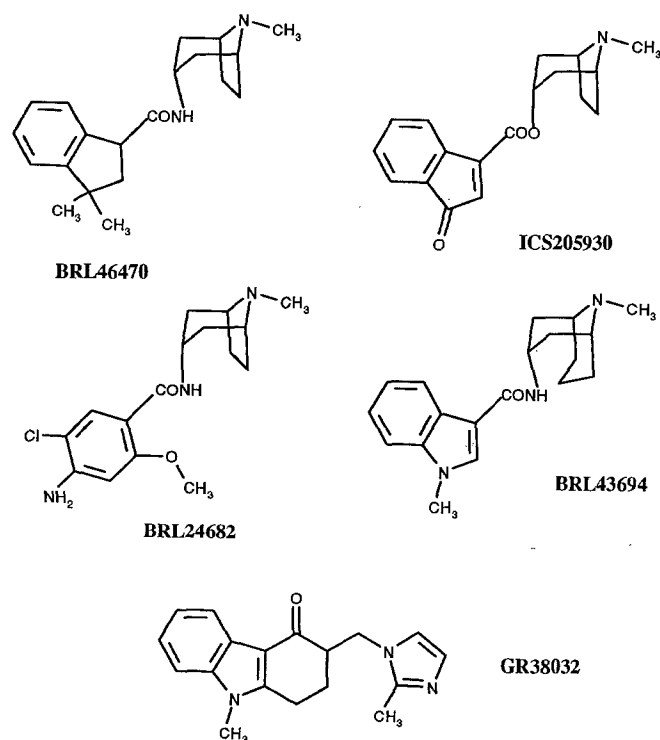


Figure 6. Five typical 5-HT₃ antagonists.

pine agonists would be expected. Before performing the field similarity calculations, therefore, it was necessary to examine this in detail. Previous studies^{28,29} on these or similar compounds have concentrated on the torsional freedom about the rotatable bonds between the "rigid" amine group and the aromatic ring system. However, no consideration was given in these studies to the conformations of the various ring systems. All the compounds, with the exception of ondansetron, have the aromatic moiety in an axial configuration, as drawn in Figure 6. In the case of the bicyclo[3.2.1] system (tropane), it is possible for the piperidine ring to flip into a boat conformation, thus relieving the steric interactions with the ethylene bridge. With the bicyclo[3.3.1] system (granatane), either or both piperidine rings can flip into boat conformations. In addition to these ring flips, it is possible for the nitrogen to invert, usually with little difference in energy. To account for these possibilities with each molecule, all possible ring conformations were generated and minimized using the COSMIC force field.³⁰ The rotatable bonds around the amide group were relaxed to suitable low-energy conformations. The results of these calculations are summarized in Table 2.

The calculations show that in general, for the tropane ring compounds, the chair conformation is preferred. The piperidine ring is able to pucker sufficiently to relieve the strain from the axial amide substituent. With granisetron, there is little difference in energy between the chair-chair and the boat-chair forms. In all cases, it is interesting to note that there is only a small energy barrier to inversion at the *N*-methyl nitrogen. This may have some significance when these compounds are compared with ondansetron. No alternative ring conformations were chosen for ondansetron; the only conformational freedom that was considered in the minimization was rotation about the two exocyclic single bonds. In all cases the final lowest energy conformation was chosen to perform the quantum mechanical steps.

To calculate the fields, each final conformer was further optimized, again using the semiempirical molecular orbital program VAMP²⁷ with the AM1 Hamiltonian. NAO point fitted charges were calculated as previously described and subsequently used for the field calculations.

Not surprisingly, the final overlapping conformations resulting from the field similarity calculations were dominated by the overriding effect of the electrostatic field around the carbonyl oxygen. This will almost certainly align all compounds in a similar manner during the molecular recognition phase. Interestingly, this alignment places the aromatic rings of all compounds in an almost coplanar arrangement, allowing for slight conformational flexibility around the amide or ester bonds. There has been some concern expressed in the past about the overlap of the basic nitrogens between the tropane/granatane compounds and ondansetron. Hibert²⁸ explained this by assuming that the positive charge in the imidazole ring was delocalized and this could place it in some proximity to the protonated N-H group of the bicyclic structures. We have already noted that the N-CH₃ group can flip readily between equatorial and axial forms with little loss in energy. Using the predicted overlap from the program and allowing the *N*-methyl to be axial, this allows the protonated N-H of ondansetron to overlap almost perfectly with the N-H of granisetron and the other compounds (see Color Plate 6). Thus from this

Table 2. Conformational study of 5-HT₃ antagonists^a

Compound	Ring conformation	N-CH ₃ conformation	Energy (kcal mol ⁻¹)
BRL24682	Boat	Axial	24.99
BRL24682	Boat	Equatorial	22.36
BRL24682	Chair	Equatorial	17.52
BRL24682	Chair	Axial	18.48
BRL46470	Chair	Axial	26.57
BRL46470	Chair	Equatorial	26.22
BRL46470	Boat	Axial	33.14
BRL46470	Boat	Equatorial	30.42
ICS205930	Boat	Equatorial	15.90
ICS205930	Boat	Axial	18.79
ICS205930	Chair	Equatorial	11.68
ICS205930	Chair	Axial	13.25
Granisetron	Rng1-chair Rng2-chair	Axial	28.13
Granisetron	Rng1-chair Rng2-chair	Equatorial	28.15
Granisetron	Rng1-chair Rng2-boat	Equatorial	31.18
Granisetron	Rng1-chair Rng2-boat	Axial	29.89
Granisetron	Rng1-boat Rng2-chair	Axial	31.96
Granisetron	Rng1-boat Rng2-chair	Equatorial	27.65

^aIn granisetron, Rng-1 is the piperidine ring with the indole substituent. Axial or equatorial refers to the *N*-methyl conformation in this ring.

work it would appear that there is a high degree of similarity between all five compounds, both sterically and in their electrostatic fields. In cases in which recognition and binding are largely electrostatic in nature, this approach promises to be highly useful in the future.

CONCLUSIONS

The use of molecular similarity calculations in drug design has been well documented and has been shown to be a valuable tool in the generation of pharmacophore hypotheses. By extending our earlier similarity program¹ to include electrostatic vector fields, we have been able to demonstrate that this method can give results complementary to those obtained from electrostatic potential studies. These should more accurately represent the similarity of interaction between a series of drug molecules and the dipole moments associated with their binding sites. During this work a new fast electric field calculation was developed and used. This was based on the natural orbital fitted charge method of Rauhut and Clark¹⁷ and was shown to give results closer to ab initio calculations than previous approximate methods. In the examples described here, some novel potential binding modes were suggested and because of this, the field similarity method would appear to offer much promise in its use with future drug design studies.

REFERENCES

- 1 Phippen, R.W., Wyatt, M.J., Finn, P., and Blaney, F.E. *J. Mol. Graphics*, 1993, **11**, 98–105
- 2 Weinstein, H., Osman, R., and Green, J.P. In *Computer Assisted Drug Design. ACS Symposium Series* (E.C. Olsen and R.E. Christoffersen, eds.), Vol. 112. American Chemical Society, Washington, D.C., 1979, pp. 161–187; and references cited therein
- 3 Johnson, M.A. and Maggiora, G.M. (eds.). *Concepts and Applications of Molecular Similarity*. Wiley, New York, 1990
- 4 Chau, P.L. and Dean, P.M. *J. Mol. Graphics* 1987, **5**, 152–158
- 5 Chau, P.L. and Dean, P.M. *J. Mol. Graphics* 1987, **5**, 97–100
- 6 Hodgkin, E.E. and Richards, W.G. *Chem. Br.* 1988, **24**, 1141
- 7 Namasivayam, S. and Dean, P.M. *J. Mol. Graphics* 1986, **4**, 46–50
- 8 Giessner-Prettre, C. and Pullman, A. *Theor. Chim. Acta* 1972, **25**, 83–88
- 9 Frisch, M.J., Head-Gordon, M., Trucks, G.W., Foresman, J.B., Schlegel, H.B., Raghavachari, K., Robb, M.A., Binkley, J.S., Gonzalez, C., Defrees, D.J., Fox, D.J., Whiteside, R.A., Seeger, R., Melius, C.F., Baker, J., Martin, R.L., Kahn, L.R., Stewart, J.J., Topiol, S., and Pople, J.A. *Gaussian 90*. Gaussian, Inc., Pittsburgh, Pennsylvania, 1990
- 10 Gasteiger, J. and Marsili, M. *Tetrahedron* 1980, **36**, 3219–3228
- 11 Abraham, R.J. and Grant, G.H. *J. Comput. Aided Mol. Design* 1992, **6**, 273–286
- 12 Giessner-Prettre, C. *Quantum Chemistry Program Exchange Bulletin*, 1974, **11**, 249
- 13 Chirlian, L.E. and Francl, M.M. *J. Comput. Chem.* 1987, **8**, 894–905
- 14 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. *J. Comput. Chem.* 1983, **4**, 187–217

- 15 Weiner, P. and Kollman, P. *J. Comput. Chem.* 1981, **2**, 287–303
- 16 Besler, B.H., Merz, K.M., Jr., and Kollman, P.A. *J. Comput. Chem.* 1990, **11**, 431–439
- 17 Rauhut, G. and Clark, T. *J. Comput. Chem.* 1993, **14**, 503–509
- 18 NURBS. Foley, J.D., Van Dam, A., Feiner, S.K. and Hughes, J.F. in *Computer Graphics: Principles and Practice*. Addison-Wesley, 1990.
- 19 Scheraga, H.A. *Prog. Clin. Biol. Res.* 1989, **289**, 3–18
- 20 Watson, D.W. *WINSOM User's Guide and Reference Manual*. IBM UK Scientific Centre Report
- 21 Sternbach, L.H. *J. Med. Chem.* 1979, **22**, 1–7
- 22 Mohler, H. and Okada, T. *Science* 1977, **198**, 849–851
- 23 Squires, R.F. and Braestrup, C. *Nature (London)* 1977 **266**, 732–734
- 24 Skolnick, P. and Paul, S.M. *Med. Res. Rev.* 1981, **1**, 3–22
- 25 Villar, H.O., Davies, M.F., Loew, G.H., and Maguire, P.A. *Life Sci.* 1991, **48**, 593–602
- 26 Villar, H.O., Uyeno, E.T., Toll, L., Polgar, W., Davies, M.F., and Loew, G.H. *Mol. Pharmacol.* 19??, **36**, 589–600
- 27 Clark, T., Rauhut, G., and Chandrasekhar, J. University of Erlangen, Erlangen, Germany, 1991
- 28 Hibert, M.F., Hoffman, R., Miller, R.C., and Carr, A.C. *J. Med. Chem.* 1990, **33**, 1594–1600
- 29 Rosen, T., Nagel, A.A., and Rizzi, J.P. *Synlett.* 1991, 213–221
- 30 Vinter, J.G., Davis, A., and Saunders, M.R. *J. Comput. Aided Mol. Design* 1987, **1**, 31–51