Molecular surface comparison: Application to drug design

Frank Blaney

SmithKline Beecham Pharmaceuticals Research Division, Medicinal Research Centre, Harlow, Essex CM19 5AD, UK

P. Finn

Pfizer Central Research, Sandwich, Kent, UK

Rob Phippen and Mike Wyatt

IBM UK Scientific Centre, Winchester SO23 9DR, UK

An interactive system for the display and manipulation of molecular surface properties is presented. The property at the molecular surface is mapped onto the sphere by gnomonic projection. This representation allows direct comparison of the surface properties of pairs of molecules.

The system allows the user to explore the similarities between a pair of molecules in an interactive manner, and provides extensive visual (color-coded field and field difference maps) and numerical (rms difference value) aids to complement the user's chemical intuition. Examples of the use of the system to study beta-lactam compounds and phosphodiesterase inhibitors are presented.

Keywords: molecular comparison, molecular graphics, gnomonic projection, drug design, molecular surface, surface comparison

INTRODUCTION

The rationalization for using molecular surface property comparison as an aid to drug design has been discussed elsewhere.[7] It would appear to rely on the pharmacophore hypothesis; namely, that a particular spatial distribution of electrostatic potential in the ligand causes recognition at the receptor site.[3] This definition, taken at face value, implies that we should compare volumetric EP data for a pair of molecules to assess the likelihood that they have similar pharmacophoric regions. The use of surface property comparison can be regarded as a simplification of the full volumetric comparison. The molecular surface, however defined, can be regarded as the portion of the volumetric field which is considered to be of most importance. Formally, we could define a weighting function

$$W(\bar{r}) = \delta(|\bar{r} - \bar{r}_{surface}|) \quad (1)$$

where

$$\delta(x) = 1 \quad \text{when } x = 0$$
$$\delta(x) = 0 \quad \text{when } x \neq 0 \quad (2)$$

This would allow us to define a volumetric field $f'(\bar{r})$

$$f'(\bar{r}) = W(\bar{r})f(\bar{r}) \quad (3)$$

whose value was zero everywhere except at the molecular surface.

Once the surface has been defined, there remains the question of how to compare two different surfaces; a number of techniques have been proposed.[4,5] We have taken, as the basis for our work, the gnomonic projection technique described by Chau and Dean *et al.*[6–9] This technique involves projecting the surface property of a molecule onto the surface of a sphere; to do this it is necessary to choose a center of interest (COI) from which the projection can be made. This approach solves the question of how to compare two dissimilar surfaces, although only a single value can be projected onto a point on the spherical surface, as shown in Figure 1. The projection may therefore involve the loss of information about portions of the surface if they are reentrant with respect to a line radiating from the COI. In practice, careful choice of the COI can minimize this problem.

Sampling the surface of a sphere

In order to perform the projection, we need to choose an appropriate set of points with which to sample the spherical

Color Plates for this article are on page 124.

Address reprint requests to Dr. Frank Blaney at SmithKline Beecham Pharmaceuticals Research Division, Medicinal Research Centre, Coldharbour Road, The Pinnacles, Harlow, Essex CM19 5AD, UK.

Received 28 April 1992; accepted 16 June 1992

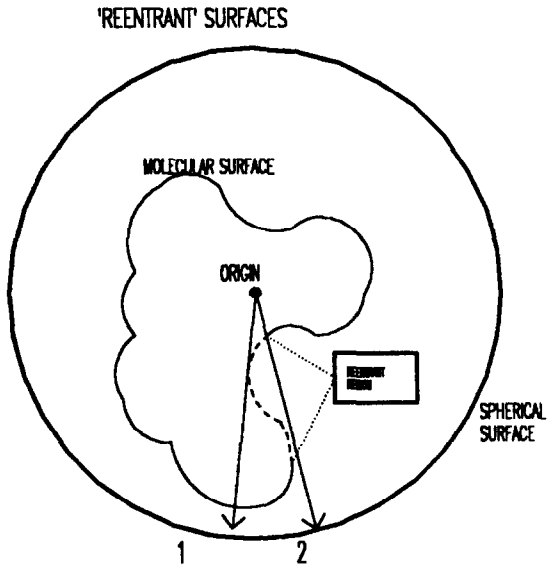


Figure 1. Gnomonic projection, molecular surface which is reentrant with respect to a vector radiating from the COI.

surface. It is *impossible* to generate an arbitrary number of precisely equally spaced points on a sphere. Only the regular polyhedra (tetrahedron, cube, octahedron, dodecahedron, and icosahedron, also known as the Platonic solids) have vertices which are equally spaced on the sphere. From any one of the Platonic solids, a set of points can be chosen which approximates equal spacing. The largest number of points that *can* be spaced equally across the surface of a sphere is 20; they form the vertices of an icosahedron. An icosahedron has 20 equilateral triangular faces which can be used as a crude polygon representation of a sphere. Each face of the icosahedron can be subdivided into n^2 equilateral triangles.[8] An improved representation of a sphere can then be produced by normalizing the distance of each vertex from the center of the icosahedron to be equal to the desired radius. This technique provides the *optimum* polygonal representation of a sphere at a given polygon resolution.

Having produced the set of points to be projected onto the sphere, the procedure, following Chau and Dean,[8] is as follows:

(1) A point on the surface of the sphere is chosen.
(2) The points at which the vector pierces the molecular surface are calculated.
(3) Of these, the point farthest from the COI is chosen.
(4) The value of the electrostatic potential at that point is calculated by interpolation of data stored in a three-dimensional field array.

This process is repeated for each of the vertices on the tesselated sphere, to produce a list of surface values corresponding to a particular vertex. The different values can be represented as different colors, allowing the entire projection to be represented as a colored sphere.

The projection is repeated for a second molecule, which can also be represented as a colored sphere. More importantly, it is then possible to calculate the difference in the projected surface value at each vertex. This enables a mean-squared difference for the whole surface to be calculated; this provides an index of similarity for two molecules at a given relative orientation. It also enables the graphical representation of the difference in the surface property as a colored sphere.

Of course, the two molecules will probably not initially be oriented in such a way that the difference in their properties is minimal, and some reorientation will be necessary. Using the techniques presented by Chau and Dean,[8] any reorientation would have required a complete recalculation of the spherical projection for each molecule. This is unnecessary since, although the *difference* in the projections changes as their relative orientation is altered, the *individual* projections obviously remain unchanged in their respective molecule-fixed axis systems. We have developed a technique to take advantage of this.

Storing the whole surface property

The basic concept is to create a two-dimensional (2D) map of the values of the property at the chosen surface, similar to the 2D representation of height on a map of the world. Thus the entire surface of the object is represented as a single entity.

The projection of the field-value at the surface to the sphere is carried out for a regular grid of values of θ and ϕ, the spherical polar coordinates of a point on a sphere. The surface field is thus represented as

$$F[(\theta_i), (\phi_j)] \tag{4}$$

where (θ_i) and (ϕ_j) represent the set of discrete values of θ_i and ϕ_j of constant separation.

This set of data points is *not* suitable for direct comparison between two surfaces, because it is not evenly distributed over the surface of a sphere; the points are closer together at the poles ($\theta \approx 0$ and $\theta \approx \pi$). Thus field values in locations close to the poles would be given disproportionate weighting. The set is used simply because it enables the field at any point on the surface to be calculated by interpolation between the values calculated for the grid.

Calculating the field value at an arbitrary point

Calculating the value of the field at the surface for a particular space-fixed vector, in this case one of the vectors defining the vertices of the subtesselated icosahedron, is reduced to the simple task of locating its position on the map.

The field value at an arbitrary point on the surface defined by $(\theta_{arbitrary}, \phi_{arbitrary})$ is found by linear interpolation between the four values of the field at the nearest stored θ_i and ϕ_j map positions.

An alternative way of defining a point on the surface is as a vector,

$$\mathbf{v}_{molecule} = \begin{pmatrix} x_{arbitrary} \\ y_{arbitrary} \\ z_{arbitrary} \end{pmatrix} \tag{5}$$

The field at this point can be found by calculating the corresponding (θ, ϕ) values, explicitly

$$\begin{aligned} \theta &= a\cos(z_{arbitrary}/L) \\ \sin\phi &= x_{arbitrary}/(L\sin\theta) \\ \cos\phi &= y_{arbitrary}/(L\sin\theta) \end{aligned} \tag{6}$$

where L is the length of $\mathbf{v}_{molecule}$. The value of ϕ can simply be computed from sin ϕ and cos ϕ.

Reorienting an object

The particular value of this technique is that when the orientation of the object is altered, the map remains useful. As the object is rotated, the map describing its surface property is also implicitly rotated; thus a particular space-fixed vector will refer to a different point on the map.

If the rotation matrix describing the orientation of the molecule-fixed axis system with respect to the world-fixed axis-system is $\mathbf{r}$, then the (θ, ϕ) location of a world-fixed vector $\mathbf{v}_{world}$ can be calculated as follows:

(1) In molecule-fixed coordinates the vector becomes $\mathbf{v}_{molecule}$, which is defined by $\mathbf{v}_{molecule} = \mathbf{r}\mathbf{v}_{world}$.
(2) The molecule-fixed values of (θ, ϕ) can then be calculated using the procedure described in Equations 5 and 6.
(3) The value of the surface property at the point on the surface corresponding to $\mathbf{v}_{world}$ is recalculated, as before, by interpolation of the values stored in

$$F(\{\theta_i\}, \{\phi_j\}) \quad (7)$$

Rapid remapping allows interactivity

After the initial mapping stage for each molecule, the list of operations required to recalculate the difference between the properties of two surfaces has been reduced to

(1) The value of the surface field at each space-fixed vector is calculated by linear interpolation.
(2) This process is repeated for another object.
(3) The difference between the surface field for the two objects can then be calculated by direct subtraction of the values at each of the predefined points.
(4) The mean squared difference between the values at each of the points then gives an indication of the *overall* difference between the two surface properties.

The reduced recomputation implied by this technique allows the difference between two maps to be recalculated 5–10 times per second on a powerful workstation (e.g., IBM RISC System/6000). This means that the orientation of an object can be linked to a suitable input device (e.g., a set of dials), and the difference between the surface properties of the two objects can be recalculated and displayed in real time.

This real time behavior has allowed us to develop an interactive user interface, which we describe below.

The user interface

We have developed a graphical interface, based on the GL libraries, which is designed to maximize the interactivity of our system; when the program is started, the user is presented with five windows, Figure 2, each of which contains an image representing different information:

(1) Top: shows a graduated bar. The length of the bar represents the value of the mean-squared difference in the surface property between the two surfaces.

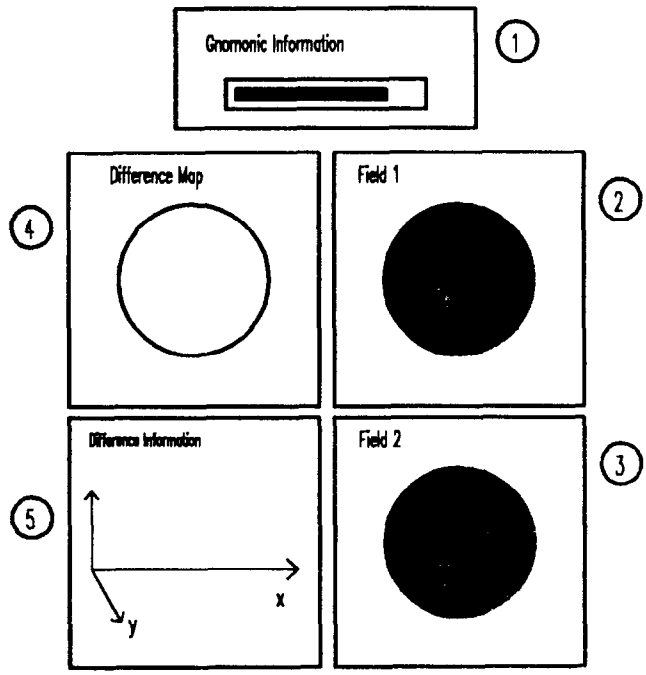


Figure 2. WINMATCH, user interface to surface comparison system.

(2) Bottom right: spherical projection of the surface of molecule 1.
(3) Top right: spherical projection of the surface of molecule 2.
(4) Top left: graphical representation of the difference in the two projections at each point on the spherical surface. An example of the main windows of the user interface is shown in the Color Plate 1.
(5) Bottom left: map of mean-squared differences (see below).

Manipulating the windows

Any of the windows can be chosen using a mouse-driven cursor; when windows 2, 3, or 4 are chosen, the orientation of the chosen projection can be altered using a set of dials. This causes the affected information in each window to be updated in real time; as this is done, the associated information affected by the change is updated in each window. This allows the user to compare the two surfaces in a completely interactive way, and provides a range of visual and quantitative cues for the exploration of local minima in the mean-squared difference.

Representing projections by shape and color

Windows 2, 3, and 4 are effectively spherical projections of a surface property. They are, by default, represented as color-coded spheres. However, it is also possible to use *shape* to represent the projections. The value of the projected property can be used to define a length; this length is used to define the distance of the surface of a star-like shape from a central point. The direction of the vector from the

center to the point on the surface is the same as that used in the spherical representation.

An example of a shape-mapped projection is shown in the Color Plate 2.

The representation of the surface property by shape is particularly appropriate when we choose to compare the shape of two objects. The first step is to create, for each surface to be compared, a projection map containing values which correspond to the distance of the surface from the COI for each orientation of the vector radiating from the COI.

Map of mean-squared differences

The mean-squared difference between the surface values for two surfaces is a good measure of how dissimilar they are. The value of the mean-squared difference alters as the relative orientation of the two surfaces is changed. In order to investigate relative orientations of two surfaces, which maximizes their apparent similarity, it is therefore important to be able to locate, and explore, the minima in the mean-squared difference as a function of the relative orientation of the two surfaces.

In order to aid this process, we use the spherical projection technique described above to calculate the mean-squared difference for all possible x-, y-, and z-rotations to a resolution of 10 degrees. This produces a three-dimensional block of data, which is stored in the familiar WINSOM field format.[10]

Window 5 uses this precomputed mean-squared difference data. It is presented to the user as a 2D surface, or *terrain map*. The horizontal axes represent the possible x- and y-rotations of one projection with respect to the other; the height of the surface above a particular point in the horizontal plane represents the mean-squared difference for a particular relative orientation of the two surfaces. A marker points to the position on the map which represents the *current* relative orientation of the two surfaces. At any instant, the z-rotation represented by the terrain map is fixed; it can be changed by rotating a dial; this causes the shape of the terrain map to be altered. Alteration of the x- or y-rotations simply causes the marker to move across the terrain map. Any movement of the orientation marker on the terrain map causes a corresponding change in the orientation of the projection of surface 2, and causes the difference information in windows 1 and 5 to be updated.

Additional facilities

The graphical facilities described above provide a powerful basis for quantitative surface comparison; the window-driven system is particularly flexible, enabling us to provide a very large number of options at every stage of the comparison process. For instance, the mean-squared difference data, used in the terrain map, can be searched automatically to locate the relative orientation which gives the global minimum.

Among many other facilities, the user can specify the desired polygon resolution of the graphical representation of the spherical projections, and change the color coding used to represent the values projected onto the spheres to suit the application.

RESULTS

Application to beta-lactam antibiotics

Beta-lactam antibiotics provide one of the major therapeutic approaches to the treatment of bacterial infections. These structures consist of a nucleus (usually a bicyclic ring system with an acidic functional group) and a side chain (usually attached via an amide). While a large number of variations in the chemical nature of the side chain is known, the design of active variants of the nucleus has proven to be far more difficult, so that only a relatively small number of chemical classes is known. We have applied the gnomonic projection methodology to four nuclei (penam, penem, lactivicin, and pyrazolidinone) to investigate its behavior where the structural modifications are relatively minor. The four structures are shown in Figure 3. The COI was specified as the position of the nitrogen of the cyclic amide. Compound 1 was used as a reference; the other three structures were compared to it. Properties compared were the electrostatic potential (EP) calculated from the CNDO point charges using the VSS

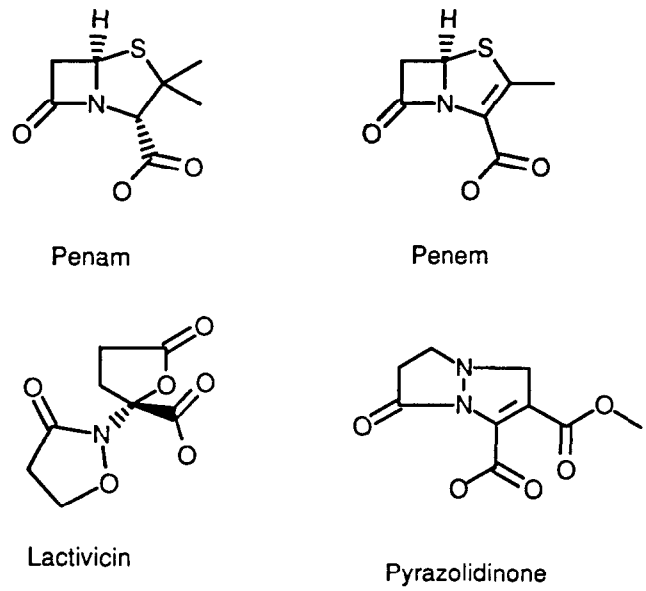


Figure 3. Beta-lactam antibiotics, molecular structures of the nuclei of penam, penem, pyrazolidinone, and lactivicin.

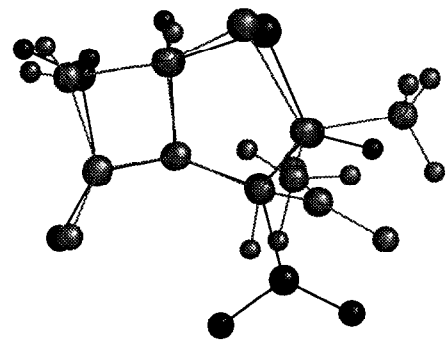


Figure 4. Beta-lactam antibiotics, optimal overlap for penam and penem.

method,[11] and a shape descriptor defined as the distance from the COI to the van der Waals surface.

Figure 4 shows the optimal orientation between penam and penem for electrostatic potential. The potential is dominated by the large negative region associated with the carboxylate functional group. The difference information shows a deep valley running parallel to the *y*-rotation axis with steep walls on either side. Because the carboxylate negative potential is roughly at the "south pole," *y*-rotations cause small movements of this feature and are therefore relatively well tolerated, whereas *x*-rotations cause rapid rises in the rms difference as the negative region is moved away.

Figure 5 shows the superposition of the minimum rms positions with respect to penam for electrostatic potential for penem, pyrazolidinone, and lactivicin. As might be expected, the two dominant electrostatic features align themselves: the negative regions of the carboxylate and beta-lactam carbonyl. The difference information display allows other local minima to be found with ease. The next lowest minimum for EP of penam compared with penem is shown in Figure 6. The two electronegative regions have swapped positions.

Similar alignments are also seen for the shape superpositions, the minimum rms superpositions being shown in Figure 7. In this case, the orientations are dominated by the groups protruding from the roughly spherical bicyclic system on the right-hand side of the molecules as drawn in Figure 8.

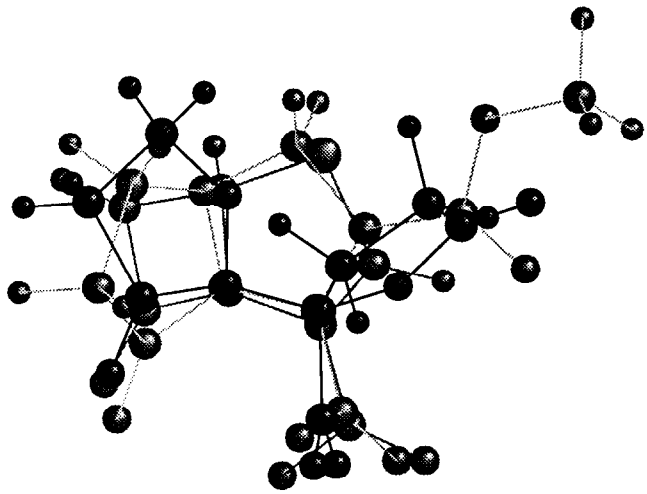


Figure 5. Beta-lactam antibiotics, superimposition of all three molecules in the orientations given by their EP minima with respect to penam.

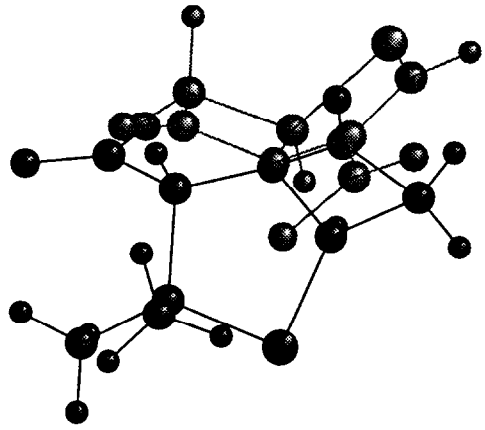


Figure 6. Beta-lactam antibiotics, superimposition of penam and penem in the orientation specified by the second lowest EP minimum.

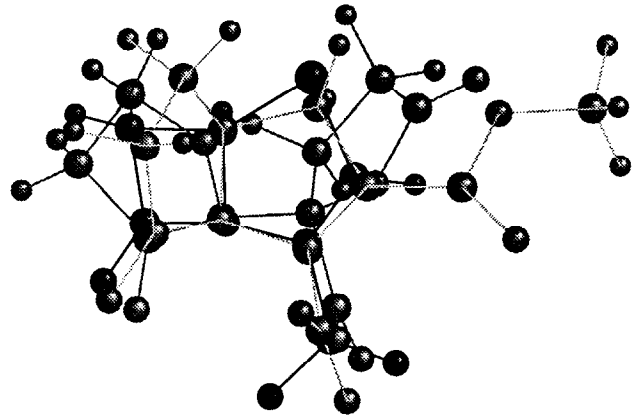


Figure 7. Beta-lactam antibiotics, superimposition of all three molecules in the orientations given by their shape minima with respect to penam.

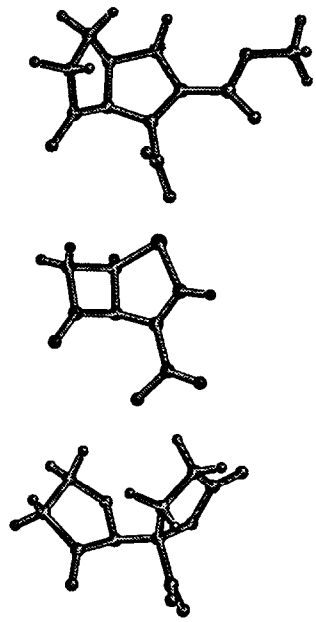
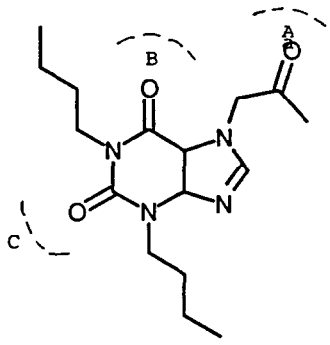


Figure 8. Beta-lactam antibiotics, the main groups affecting the shape overlap.

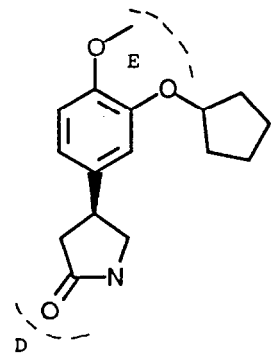
Application to rolipram and denbufylline

We have compared the surface electrostatic and shape properties of the phosphodiesterase inhibitor rolipram with denbufylline (BRL30892), which is a novel phosphodiesterase inhibitor and is chemically unrelated. Structures and important regions for the superpositions discovered are shown in Figure 9.

The EP comparison gave three overlaps with low rms (6.98, 7.39, and 7.87; see Figures 10a, 10b, and 10c).



Denbufylline



Rolipram

Figure 9. Rolipram and denbufylline.

These overlaps represent different ways of superposing the three electronegative areas of denbufylline (the three carbonyl oxygens, labeled *A*, *B*, and *C* in Figure 9) with the two electronegative areas of rolipram (the cyclic amide at one end and the two ether oxygens at the other, labeled *D* and *E*, respectively). The lowest minimum superimposes areas *A* with *E*, and *C* with *D*. Minimum 2 swaps these to match *A* with *D*, and *C* with *E*. For the third minimum a different pattern is seen, with areas *A* and *B* of denbufylline sharing an overlap with area *E*, and *C* again matching with *D*.

Shape superposition gave two low rms orientations (1.16 and 1.27, shown in Figures 11a and 11b). These correspond to either of the butyl groups of denbufylline superposing with the cyclopentyl group of rolipram.

Interestingly, the third EP overlap corresponds quite closely to the second shape overlap (Figure 12). That two such different properties have a common low rms orientation suggests a possible enzyme binding mode. Although overlaps of these structures on the basis of electrostatic potential on the molecular surface had been examined previously,[12] we had not found this orientation.

Sensitivity to the choice of the COI

Clearly it is important that the gnomonic projection methodology is not excessively sensitive to the choice of COI. To test this dependence we have repeated the rolipram/denbufylline EP comparison with varying COIs. The COI used in the preceding section was shifted by 1.5 Å along each of the *x*-, *y*-, and *z*-axes. The resulting coordinates correspond to the centers of the faces of a cube of side 3 Å. The mapping study was repeated for each of these new COIs. Rotation angles and rms figures are given in Table 1.

In five of the six cases, the orientation is very similar to the original. In the sixth case the two dominant regions are retained, but the third is rotated to the opposite side. This shift is also associated with the highest rms value. In this case, the second minimum corresponds to an orientation similar to the others.

Negative shifts in *x* and *y* produce significantly lower minimum rms values for the overlaps than the COI chosen

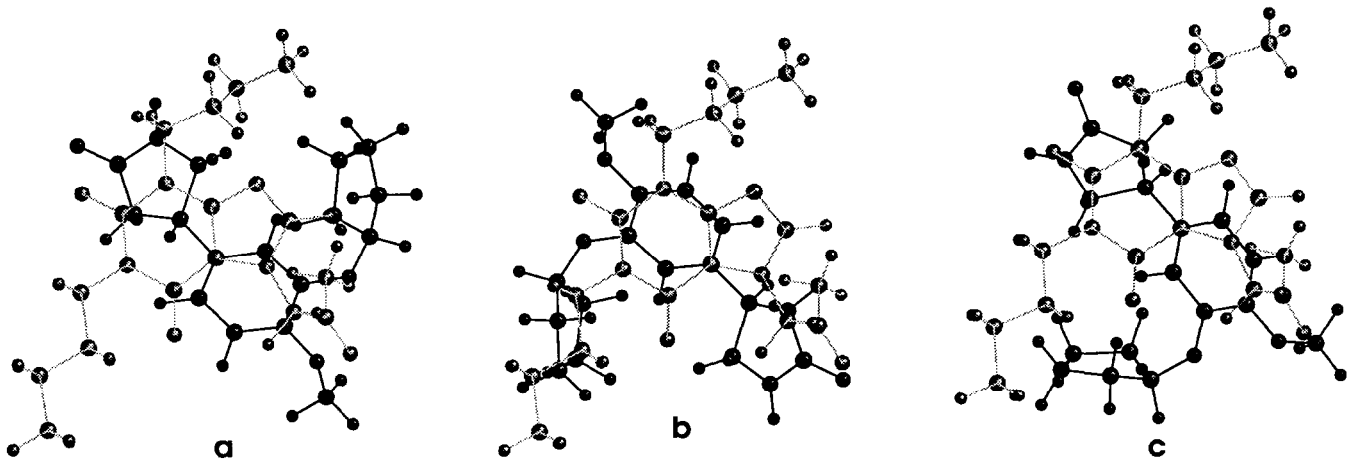


Figure 10. a) Rolipram and denbufylline, superimposition specified by the lowest EP rms minimum. b) Superimposition specified by the second lowest EP rms minimum. c) Superimposition specified by the third lowest EP rms minimum.

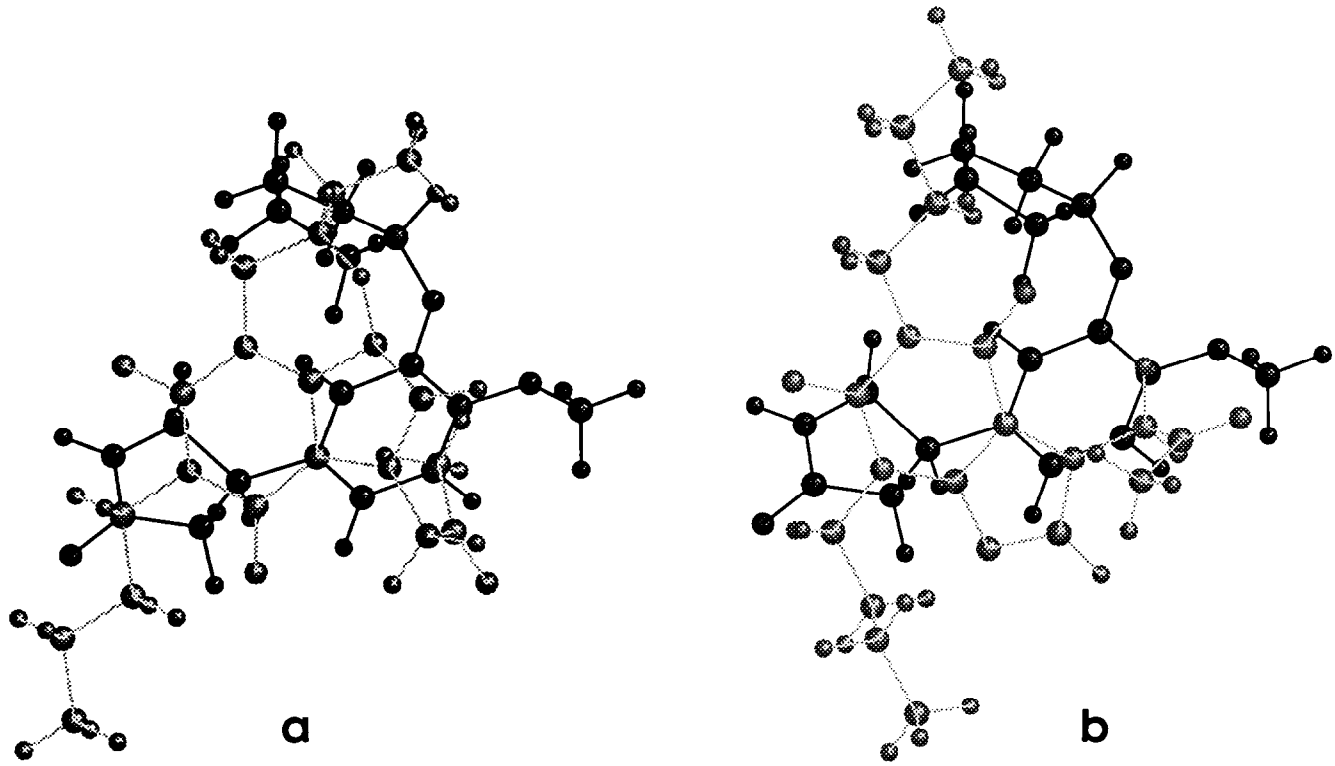


Figure 11. a) Rolipram and denbufylline, lowest rms shape superposition. b) Second lowest rms shape superposition.

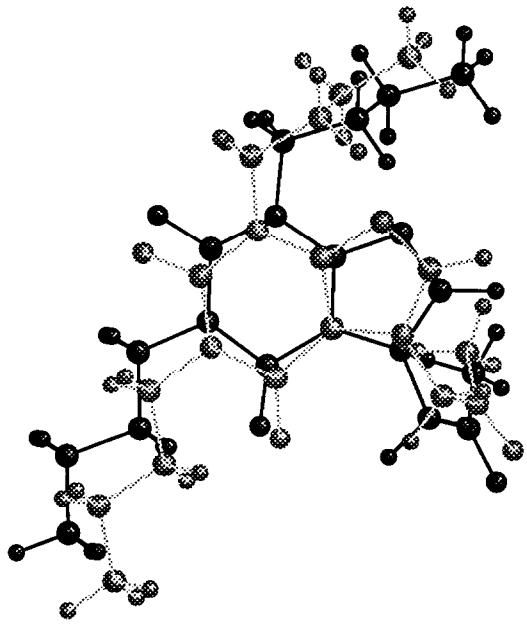


Figure 12. Rolipram and denbufylline, superimposition of the orientations found for the third EP minimum and the second shape minimum.

for the preceding section. Although it has been shown that this does not significantly affect the minimum rms orientation produced (i.e., the *xyz* rotations), it may give information as to the optimum *xyz* translations.

Over the central volume of the molecule, the COI is not crucial to the orientations produced: differences in detail *are* observed. For structurally dissimilar molecules, it is advisable to try 2 or 3 COIs in the early stages of a comparison study to test this sensitivity; it will depend on the molecular systems to be compared.

CONCLUSIONS

We have developed an interactive graphical interface to the gnomonic projection technique, and have tested its usefulness and reliability on two series of known compounds. The technique reproduces molecular overlaps obtained using other methods, and in some cases, produces convincing new overlaps. An initial source of concern, the choice of the COI, has proven to be largely groundless since the results are not drastically affected by changing the location of the COI. Indeed, choice of functions other than the rms difference as the basis of comparison could reduce this effect still further; we intend to investigate the use of Spearman's rank correlation coefficient, successfully tested by Namasivayam and Dean,[7] for this purpose.

Given that the basic technique appears to be sound, the usability of the graphical interface becomes paramount; in this initial study we have given priority to maintaining a strong link between the user's manipulation of the gnomonic projections and the visual cues. We have avoided complete automation of the process of locating minima since the aim of this system is to augment the chemist's intuition rather than to replace it. However, a facility to locate the nearest local minimum once the user has entered the "concave" region surrounding it is being considered as a convenience function.

Table 1. Effect of moving the COI

Coordinate shift in *Angstrom*	RMS difference	*x*-rotation	*y*-rotation	*z*-rotation
(0.0, 0.0, 0.0)	6.53	56.3	−45.0	−45.0
(−1.5, 0.0, 0.0)	6.10	45.0	−22.5	−33.8
(0.0, −1.5, 0.0)	6.05	45.0	−39.4	−22.5
(0.0, 0.0, −1.5)	6.42	56.3	−28.1	−33.8
(1.5, 0.0, 0.0)	6.95	45.0	−28.1	−11.3
(0.0, 1.5, 0.0)	7.04	−168.7	5.6	−112.5
(0.0, 0.0, 1.5)	6.77	33.8	−39.4	−22.5

ACKNOWLEDGMENT

The authors would like to thank Dr. Colin Edge for his comments on the text and for his help in constructing the diagrams.

REFERENCES

1 Berne, B.J. and Kushick, J. Computer Simulation of Anisotropic Fluids. *J. Chem. Phys.* 1976, **64**(4), 1362
2 Frisch, M. *et al.* Gaussian 86. Carnegie-Mellon University, Pittsburgh (1986)
3 Marshall, G.R. in *Computer-aided molecular design.* (W. Graham Richards, Ed.) IBC Technical Services, London (1989)
4 Mezey, P.G. The shape of molecular charge distribution: Group theory without symmetry. *J. Comp. Chem.* 1987, **8**(4), 462–469
5 Arteca, G.A., Gund, T.M., Hermsmeier, M.A., Jammal, V.B., Mezey, P.G., and Yadav, J.S. Shape group studies of molecular similarity. *J. Mol. Graphics* 1988, **6,** 45–53
6 Danziger, D.J. and Dean, P.M. The search for functional correspondences in molecular structure between two dissimilar molecules. *J. Theor. Biol.* 1985, **116,** 215–224
7 Namasivayam, S. and Dean, P.M. Statistical method for surface pattern-matching between dissimilar molecules. *J. Mol. Graphics* 1986, **4**(1), 46–50
8 Chau, P.L. and Dean, P.M. Molecular recognition: 3D surface structure comparison by gnomonic projection. *J. Mol. Graphics* 1987, **5**(2), 97
9 Chau, P.L. and Dean, P.M. Molecular recognition: Optimized searching through rotational 3-space. *J. Mol. Graphics* 1987, **5**(3), 152–158
10 Watson, D.W. WINSOM user's guide and reference manual. IBM UK Scientific Centre Report, in preparation
11 Giessner-Prettre, C. and Pullman, A. *Theor. Chimica Acta* 1972, **25,** 83
12 Blaney, F. private communication