Mapping the atomic environment of functional groups: turning 3D scatter plots into pseudo-density contours

Richard E Rosenfield, Jr, Stanley M Swanson, Edgar F Meyer, Jr, Horace L Carrell* and Peter Murray-Rust*

Biographics Laboratory, Department of Biochemistry and Biophysics, Texas A & M University, TX 77843, USA

*Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA, USA.

A 3D map of the atomic environment about a pharmacophore, or any other functional group, can be constructed by combining data from crystal structures containing that group. 3D scatter plots show the positions where nonbonded atoms interact with the group. By placing a 'smearing' function at each of these points, we convert the scatter plot into a density map. Contours at various densities reveal non-uniformities in the distribution, which may indicate preferred directions of chemical interaction. Quantitative comparisons between density maps are easier to make than between scatter plots. We illustrate the method by examining the distribution of anions about the trimethylammonium group, a substructure of the neurotransmitter acetylcholine.

Keywords: intermolecular contacts, pharmacophoric environment, 3D contouring, cluster analysis, interactive graphics

received 20 February 1984, accepted 28 March 1984

Crystal structure determinations provide 3D information on the atomic environments of molecules and their constituent functional groups. We[1-4] and others[5-9,16] have surveyed the solid state environment of some common groups, and have drawn conclusions about preferred directions and distances of nonbonded atomic interactions through the use of 2D and 3D scatter plots. Interactive computer graphics can have an important role to play in such surveys, helping the examination of 3D information.

Rosenfield and Murray-Rust[3] used interactive graphics to study the environment of the pharmacophore $RCH_2NMe_3^+$, a substructure of the neurotransmitter acetylcholine. 3D scatter plots showing where halide counter ions interact with this pharmacophore were constructed by superimposing data from 34 different crystal structures, retrieved via the Cambridge Crystallographic Database[10]. We found pronounced clustering in certain directions (Figure 1), suggesting preferred directions of ionic binding between acetylcholine and its anionic binding site in receptors and enzymes. In data from nine other crystal structures where oxyanions surround the pharmacophore, however, we could not discern a clear pattern of oxygen clustering (Figure 2). In order to elucidate, quantify, and compare patterns of clustering in

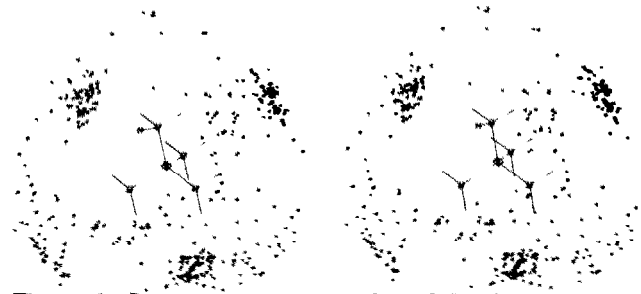

Figure 1. Stereoscopic scatter plot of the distribution of Cl^- (tetrahedra), Br^- (3 intersecting squares) and I^- (3 intersecting lines) anions about a reference $RCH_2NMe_3^+$ cationic group. (Only one carbon atom of the alkyl group is shown.) 188 unique halide positions are shown together with symmetry-related points across the mirror plane bisecting the cation. Clustering occurs along the middle of three of the cation's tetrahedral faces; halide approach to the fourth face is hindered by the alkyl group

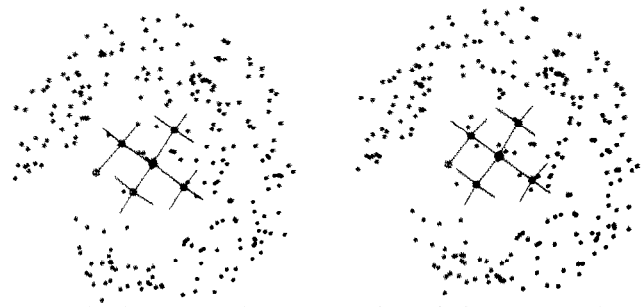

Figure 2. Stereoscopic scatter plot of the distribution of oxygen atoms from oxyanions about a reference $RCH_2NMe_3^+$ group. 119 unique oxygen positions are shown; as in Figure 1, the distribution was mirror-symmetrized. A slight amount of clustering occurs about the cation's tetrahedral faces, but this is difficult to see

 0263-7785/84/020043-04 $03.00

three dimensions, we have been pursuing graphical techniques that recast the scatter data as 'pseudo-density' maps. Here we will show how density maps can help us to clarify the information in these two strikingly different distributions.

3D SCATTER PLOTS

3D scatter plots are easily constructed and viewed using interactive computer graphics (Vector General Series 3 controllers are used at both of our places of work). Each atomic position is displayed as a single point (or fiducial mark), and the distribution of points is treated as a single object in the picture, like a giant molecule, that can be manipulated in real time to give the viewer an undistorted and continuously adjustable 3D view of the distribution. When points tend to concentrate within a few regions of space, as in Figure 1, the centres of clustering are easily recognized.

Figure 2, on the other hand, illustrates potential problems and limitations with visual methods. When the variation in density is slight, local maxima and minima cannot easily be perceived. Our visual processes tend to exaggerate the importance of a few points lying outside a cluster. Stereopsis tends to be harder to achieve with point images than with lines or solids. All these effects limit our ability to interpret and communicate our results graphically.

3D PSEUDO-DENSITY CONTOUR MAPS

Two principal types of analysis exist to tackle such difficulties. Cluster analysis based on Euclidean interpoint distances is one[11,12]. However, it may not work well for rod- or ribbon-like distributions, or for highly populated regions linked weakly by a few points. The second type of analysis involves the construction of a 3D density map[11]. Interpoint distances need not be calculated explicitly or compared.

The scatter plot is transformed into a density map by replacing each sample point with a 3D 'smearing' function: the superposition of these functions is the density map. Peak regions in the density map, where smearing functions overlap, will correspod to regions of clustering in the scatter plot. The procedure is straightforward once a suitable smearing function has been chosen. Although there is no generally accepted method of optimizing this choice, we expect the density map to reflect the gross features of the scatter plot for a wide range of choices.

The map is contoured at one or more levels of constant density, and individual levels may be varied dynamically in intensity or colour. It is useful to examine more than one density level, because no single level can convey the density's slope or curvature.

METHODS

We have applied this technique to the distributions in Figures 1 and 2, using an isotropic Gaussian smearing function. The density ρ at (x, y, z) is given as follows. Let:

$$d_i^2(x,y,z) = (x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2$$

where x_i, y_i, z_i are the coordinates of the ith point. We then express a truncated Gaussian function $g_i(x, y, z)$ as

$$g_i(x,y,z) = \begin{cases} (\mathrm{k}/w^3)\exp\left[-(d_i/w)^2\, 1n\ 2\right], d_i \leqslant 2.55w \\ 0, d_i > 2.55w \end{cases}$$

where w is the radius at which the function is half its maximum value, and k is a constant. We truncate the function at three standard deviations from the mean for computational convenience, and choose k so that the total integrated density of g_i is exactly one. (The truncation contains 97 per cent of the density within an infinitely extended Gaussian.) Then,

$$\rho(x,y,z) = (1/N)\sum_{i=1}^{N} g_i(x,y,z),$$

where N is the total number of points in the scatter plot. Thus, the integral of ρ over all space is also exactly one. Density values were calculated on a 3D Cartesian lattice at x, y, and z intervals approximately proportional to the extent of the map along these directions. (To avoid skipping over local peaks the intervals were kept smaller than about $1.4w$.) The range of density values was divided into 250 intervals, and partial sums were computed for each interval. The density was then integrated by combining the partial sums, taking the contribution from the highest density values first. This procedure integrates the density from the peaks outward to the tails. Contours, within x, y, and z layers of the map, can be drawn at a level that corresponds to a particular upper fraction of the total integrated density.

RESULTS

Density maps of the distributions of halides and oxygens about $RCH_2NMe_3^+$ are contoured at the 50 per cent level in Figures 3 and 4. The enclosed volume of above average density in the halide map is more than three times smaller than that in the oxygen map, clearly indicating that the halides cluster more than the oxygens. The halide contour map shows five major regions of higher density, well-separated from each other except for one dumbell-shaped region (straddling the plane of symmetry) where two modes of clustering partly overlap.

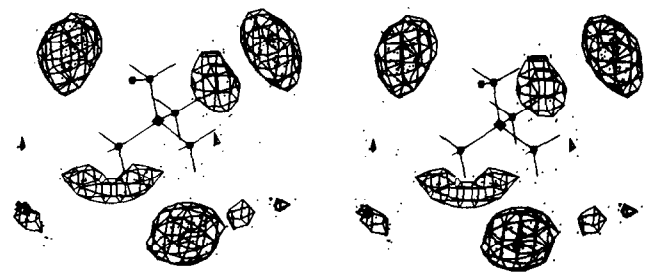

Figure 3. Stereoscopic pseudo-density contours of the halide distribution shown in Figure 1. The density map was made with $w = 0.417$ Å. The contour level, 0.0091 Å^{-3}, encloses 50 per cent of the total integrated density in a volume of 22 Å^3. The halide distribution is superimposed for comparison

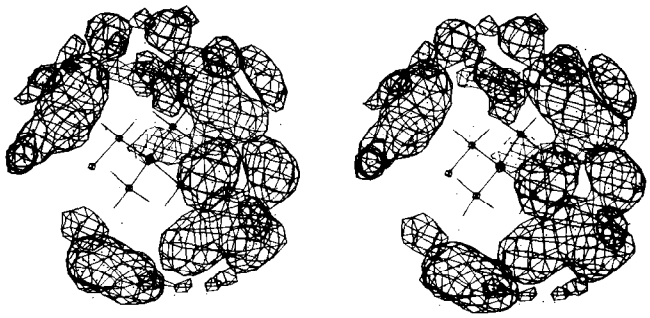

Figure 4. Stereoscopic pseudo-density contours of the oxygen distribution shown in Figure 2. The density map was made with w = 0.521 Å. The contour level, 0.0037 Å^{-3}, encloses 50 per cent of the total integrated density in a volume of 75 Å^3. The oxygen distribution is superimposed, for comparison

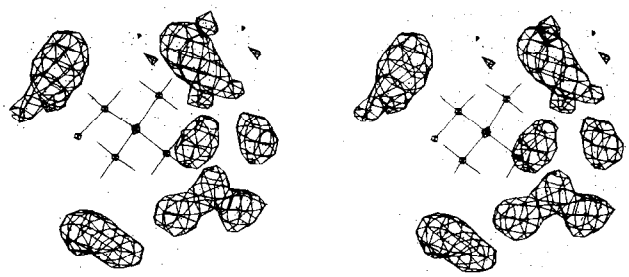

Figure 5. The oxygen density map in Figure 4 is recontoured at a higher level, 0.0068 Å^{-3}. The contours enclose 25 per cent of the total itegrated density in a volume of 25 Å^3

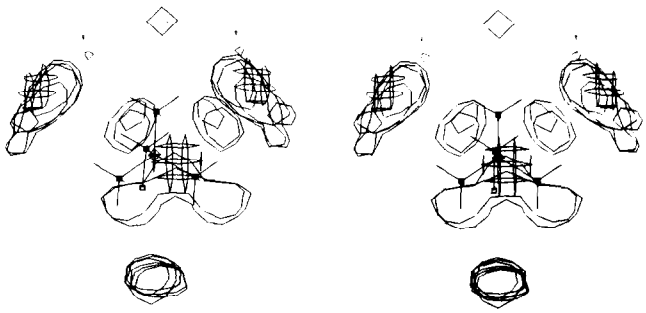

Figure 6. Stereoscopic pseudo-density contours of the halide and oxygen distributions are superimposed for comparison. Halide contours are drawn perpendicular to the plane of the page; oxygen contours are parallel to this plane. Density maps, made as in Figures 3 and 4, were contoured at levels that enclose 25 per cent of each map's total integrated density. For the halides, the contour level is 0.028 Å^{-3} and the contours enclose a volume of 5.6 Å^3. Corresponding oxygen values are given in Figure 5

Although we can barely see any clustering in the oxygen scatter plot (Figure 2), broad regions of higher density are found in the 50 per cent contour map (Figure 4). Contours at higher density (Figure 5) reveal the modes of clustering more clearly. The upper 25 per cent of the total integrated density separates into four major cluster regions, one of which is composed of five submodes. These regions of peak density overlap with those in the halide map, as seen in Figure 6.

DISCUSSION

The precise results of this method depend on the form of the smearing function. We chose an isotropic Gaussian function, although an ellipsoidal Gaussian or other functional forms (such as an atomic electron density function [4]) could have been used. The breadth of the Gaussian (the radius w at half height) determines the degree of smoothness in the density map. As w is made larger, broad density maxima form as low density regions vanish between Gaussian centres. We wish to avoid the trivial results of extreme w-values, where each point lies within its own peak (w is too small) or where all points lie within a single peak (w is too large). We have attempted[13] to apply statistical methods in choosing an 'optimal' smearing radius, based on the number and distribution of points in the scatter plot[14]. The figures shown here were constructed with optimized values. In other maps, we allowed w to be a function of the local point density; these maps are not substantially different in appearance from the ones shown here.

Pseudo-density maps and contours have several advantages over scatter plots. The location and relative density of maxima can be easily found from high-level contours or from computerized peak searches. For scatter plots with distinct clustering, higher-level contours of the density map encapsulate the cluster regions, clearly revealing their shape. Even when clustering is hard to perceive, density maps offer a sensitive way to characterize quantitatively the non-uniformities in scatter distributions. Finally, differences between distributions (eg halides versus oxyanions) can be made clear by overlaying density contours, or by calculating difference density maps*.

APPLICATIONS

We have applied this technique to several other chemical systems:

- the distribution of OH groups around ketones and ethers, showing directional hydrogen bonding related to the assumed directions of electron lone pairs[4],
- the interaction of C—I bonds with nucleophiles, again emphasizing the importance of lone pairs[15].

When a functional group is symmetric, as are $RCH_2NMe_3^+$ (C_{2v}) and C—I ($C_{\infty v}$), we expect its atomic environment to reflect this symmetry and generate symmetry-related points. In principle, this has a subtle effect on the distribution, since we no longer have totally independent data, but in practice it does not effect the contours. In the particular case of

C–I
 .
 .
 .
 X

* Separate density contours were calculated for Cl-, Br-, and I- distributions about RCH_2NMe_3+, and difference maps were also examined. As in Figure 6, there were three principal peaks in each map. Their distance from the nitrogen clearly increased with halide radius, as expected. But there were other differences as well without an obvious explanation.

(X a nucleophile), symmetry is applied to the data by rotating each X point randomly about the C—I axis; the resulting contour will have almost exact cylindrical symmetry[15].

CONCLUSIONS

This method is suitable for depicting and comparing the environment of chemical groups in crystal structures. It is particularly useful for analyzing hydrogen bonding and the coordination of electron donors to acceptors (eg metals), from the viewpoints of both acceptor and donor. We have argued elsewhere[1-4] that the distribution of atoms around a functional group represents features of the intermolecular potential, and that these features might be used by a pharmacophore when it binds to a macromolecule. To this extent, diagrams such as Figures 3–6 may represent the likely distribution of anionic binding groups in an acetylcholine receptor or in cholinesterase.

ACKNOWLEDGMENTS

P Murray-Rust and H L Carrell acknowledge financial support from the US National Institutes of Health (grants Nos CA10925, CA22780, and RR05539), the National Science Foundation (PCM-8104532), and the Commonwealth of Pennsylvania. E Meyer, S M Swanson, and R E Rosenfield acknowledge financial support from the Robert A Welch Foundation (A-328) and the Texas Agricultural Experiment Station.

REFERENCES

1 **Rosenfield, R E, Jr, Parthasarathy, R and Dunitz, J D** 'Directional preferences of non-bonded atomic contacts with divalent sulfur. 1. Electrophiles and nucleophiles' *J. Am. Chem. Soc.* Vol 99 (1977) pp 4860–4862

2 **Murray-Rust, P and Motherwell, W D S** 'Computer retrieval and analysis of molecular geometry. 4. Intermolecular interactions' *J. Am. Chem. Soc.* Vol 101 (1979) pp 4374–4376.

3 **Rosenfield, R E, Jr and Murray-Rust, P** 'Analysis of the atomic environment of quarternary ammonium groups in crystal structures, using computerized data retrieval interactions' *J. Am. Chem. Soc.* Vol 104 (1982)pp 5427–5430

4 **Murray-Rust, P and Glusker, J** 'Directional hydrogen bonding to sp^2- and sp^3-hybridized oxygen atoms and its relevance to ligand-macromolecule interactions'*J. Am. Chem. Soc.* Vol 106 (1984) pp 1018–1025

5 **Einspahr, H and Bugg, C E** 'The geometry of calcium–water interactions in crystalline hydrates' *Acta Crystallogr. Section B* Vol 36 (1980) pp 264–271

6 **Lebioda, L**'On the geometry of urea–cation bonding in crystalline urea adducts' *Acta Crystallogr. Section B* Vol 36 (1980) pp 271–275

7 **Einspahr, H and Bugg, C E** 'The geometry of calcium–carboxylate interactions in crystalline complexes' *Acta Crystallogr. Section B* Vol 37 (1981) pp 1044–1052

8 **Britton, D and Dunitz, J D** 'Directional preferences of approach of nucleophiles to sulfonium ions' *Helvetica Chimica Acta* Vol 63 (1980) pp 1068–1073

9 **Guru Row, T N and Parthasarathy, R** 'Directional preferences of nonbonded atomic contacts with divalent sulfur in terms of its orbital orientations. 2. S...S interactions and nonspherical shape of sulfur in crystals' *J. Am. Chem. Soc.* Vol 103 (1981) pp 477–479

10 **Allen, F H, Bellard, S, Brice, M D, Cartwright, B A, Doubleday, A, Higgs, H, Hummelink, T, Hummelink-Peters, B G, Kennard, O, Motherwell, W D S, Rodgers, J R and Watson, D G** 'The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information' *Acta Crystallogr. Section B* Vol 35 (1979) pp 2331–2339

11 **Meisel, W S** 'Computer-oriented approaches to pattern recognition' in **Bellman, R (ed)** *Mathematics in Science and Engineering* Vol 83, Academic Press, NY, USA (1972)

12 **Murray-Rust, P and Raftery, J** unpublished results

13 **Rosenfield, R E, Jr** unpublished results

14 **Habbema, J D F, Hermans, J and van den Broek, K** 'A stepwise discriminant analysis program using density estimation' in **Bruckmann, G, Ferschl, F and Schmetterer, L (eds)** *Compstat 1974, Proc. Computat. Stat.* Physica Verlag, Vienna, Austria 1974) pp 101–110

15 **Cody, V and Murray-Rust, P** 'Iodine . . . X (O, N, S) intermolecular contacts: models of thyroid hormone-protein binding interactions using information from the Cambridge Crystallographic Data Files' *J. Mol. Struct.* Vol 112 (1984) pp 189–199

16 **Chakrabarti, P and Dunitz, J D** 'Directional preferences of ether O-atoms towards alkali and alkaline earth cations' *Helvetica Chimica Acta* Vol 65 (1982) pp 1482–1488