Display and interpretation of solvent electron density distributions in insulin crystals

John Badger

Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA, USA

In macromolecular crystallography, three-dimensional contour surfaces are useful for interactive computer graphics displays of the protein electron density but are less effective for presenting static images of large volumes of solvent density. A raster-based computer graphics program which displays depth-cued projections of continuous density distributions has been developed to analyze the distribution of solvent atoms in macromolecular crystals. Maps of the water distribution in the cubic insulin crystal show some well-ordered waters, which are bound to surrounding protein atoms by multiple hydrogen bonds, and an ill-defined solvent structure at a greater distance from the protein surface. Molecular dynamics calculations were used to assist in the interpretation of the time-varying solvent structure within two enclosed cavities in the crystal. Two water molecules that ligate a sodium ion were almost immobile during the simulation but the majority of water molecules were found to move rapidly between the density maxima identified from the crystallographic refinement.

Keywords: electron density map, raster graphics, protein hydration

INTRODUCTION

High-resolution macromolecular crystallography involves fitting atomic models of protein and nucleic acids to three-dimensional electron density maps. For interactive computer graphics displays, "chicken-wire" contour surfaces of the electron density provide a representation that may be manipulated in real time and which does not obscure the atomic model.[1] These contour surfaces provide a useful way of displaying the steeply rising electron density around the well-ordered protein atoms in crystal structures. In contrast, the solvent surrounding the protein molecule contains some well-ordered water molecules that are interacting with the protein surface (identifiable as sharp density features) but also many ill-defined features that are difficult to model in an unambiguous way. Contour surfaces provide a satisfactory representation of the three-dimensional solvent density distribution provided attention is restricted to small portions of the map. Alternative graphical representations[2] that capture the relative significance of the varying pixel density values but which do not reduce the map to an unintelligible set of contour lines are needed for the visual analysis of larger solvent volumes.

Information about protein hydration is required to develop an understanding of protein conformations, interactions and dynamics.[3] Faced with the problem of analyzing small solvent density fluctuations in the large solvent spaces within cubic insulin crystals,[4] I have developed a raster-based computer graphics program for the presentation of static images of continuous electron density distributions. In addition, this paper describes the use of molecular dynamics simulations to help verify atomic models of the solvent structure in two enclosed cavities within the crystal.

GENERAL STRATEGY AND PROGRAM OVERVIEW

The computer program was designed to produce images using a few minutes of CPU time on readily available hardware (VAX 3100 workstations or similar) without utilizing any machine-specific graphics capabilities. The images are computed as 500 × 500 pixel arrays. This array size is large enough to minimize perception of jagged line edges but fits comfortably on most raster graphics terminal screens.

To focus on density features in the solvent space of the crystal it is more convenient to display difference electron density maps, from which the electron density due to the protein structure is removed, than to display electron density maps of the complete structure. The dynamic range of a difference map is smaller than the dynamic range of the complete map, and the problem of distinguishing protein from solvent density is largely avoided. A schematic ball-and-stick atomic model is used to represent the protein model. Depth-cueing and edge lines[5] are added to aid three-dimensional comprehension. Main chain covalent bonds are distinguishable from side chain bonds by using thicker lines,

Color Plates for this article are on page 233.

Address reprint requests to J. Badger, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02254, USA.
Received 11 December 1992; accepted 23 March 1993

oxygen atoms are drawn as enlarged spheres, and the Cα atoms are marked by dark dots.

This graphics program is designed to use a single 256-point color table which is split into two equal parts that correspond to separate colors for the depth-cued protein model and electron density. A more sophisticated computer program[2] employs additional color coding to distinguish strong electron density at the rear of the map from weaker features in the foreground. However, splitting images into separate RGB color components leads to a significant increase in computing time, and many graphics terminals and printers are restricted to 256-point color tables. The goal of the current graphics program is to provide a simple and practical tool for the display and publication of electron density maps using readily available computer equipment at Brandeis University. The common experience that published images suffer a loss of clarity and differ in color balance when compared to the original images[6] suggests that the practical benefit of more advanced rendering methods could be disappointingly small. A minor adaption of the current program produces gray-scale image arrays which may be output in the PostScript graphics language. The PostScript language is a useful tool for molecular graphics programs[7] because shaded drawings may be easily output on a laser printer and because it is straightforward to add lettering and extra artwork to the figures.

GRAPHICAL REPRESENTATION OF ELECTRON DENSITY MAPS

Three-dimensional electron density difference maps are computed on grids with a spacing of ~1/4 the maximum resolution of the observed data (e.g., a 0.5-Å grid for a 2-Å resolution map). This grid spacing is slightly smaller than is ordinarily used in protein crystallography, but the finer grid avoids visible artifacts caused by the linear interpolation operations carried out by the program. First, the map is rotated into the desired orientation and is transferred onto a new three-dimensional grid which has its XY-axes in the plane of the output image and its Z-axis (the new map section axis) pointing into the image. The electron density values at positions corresponding to the 500 × 500 image pixels are then obtained by bilinear interpolations within each section. The electron density seen in each pixel of the final image is obtained by a summation down the Z-axis of depth-cued density values above a preset threshold.

Thus,

$$\text{image pixel}\,(X, Y) = \Sigma\,\text{density}\,(X, Y, Z) \times DC(Z)$$

where

$$\text{density}\,(X, Y, Z) = \text{map density}\,(X, Y, Z) - \text{threshold}$$

for

$$\text{map density}\,(X, Y, Z) > \text{threshold}$$

and

$$\text{density}\,(X, Y, Z) = 0$$

for

$$\text{map density}\,(X, Y, Z) < \text{threshold}$$

The depth cue $DC(Z)$ attenuates the contribution of density from each section to the final image according to

$$DC(Z) = Z - Z_{min}/(Z_{max} - Z_{min})$$

The electron density was considered opaque to the atomic model, and the model prevents density beneath it from contributing to the image.

Since the Z-axis spacing is significantly smaller than the characteristic size of density features in the image, this two-dimensional interpolation scheme provides a smoothly varying projected density distribution at low computational cost (typically ~6 minutes CPU time on a VAX/VMS 3100 workstation).

APPLICATION

Water and monovalent cation electron density distributions in cubic insulin crystals have been displayed using this graphics program. Cubic insulin crystals[8] (space group $I2_13$, a = 78.9 Å) contain a solvent volume that is 65% of the unit cell and are stable in 0.1 M solutions of monovalent cations.[9] Therefore, these crystals provide a useful system for the analysis of solvent structure under physiological conditions. Diffraction data from porcine insulin crystals collected by Dodson and coworkers extend to 1.7-Å resolution,[8] and the resulting atomic model[10] is now refined to R = 0.18%. The solvent volume in these crystals should contain ~440 water molecules; 81 ordered waters are identified in the current model. At Brandeis University we have determined the bovine cubic insulin crystal structure as a function of pH to 2-Å resolution,[11] and we are using isomorphous and anomalous scattering effects to determine the distribution of monovalent cations in thallium-containing crystals.[9,12]

ELECTRON DENSITY MAPS OF THE SOLVENT

An electron density map of the solvent distribution in cubic porcine insulin crystals was computed from the Fourier difference terms

$$F_{obs}\exp(i\alpha_{ord}) - F_{prt}\exp(i\alpha_{prt})$$

The structure factor amplitudes F_{obs} were determined from diffraction data.[8] Protein electron density was eliminated by subtraction of the terms with amplitude F_{prt} and phase angle α_{prt}, calculated from the Fourier transform of the protein atomic model. In these illustrations (Color Plates 1–3), the phase angles α_{ord} were obtained from the Fourier transform of the protein and solvent atomic model.[10] Since the structure factor amplitudes F_{obs} contain a scattering contribution from all atoms in the crystal, solvent structure not included in this model is also visible in the map. In specific crystallographic applications other phasing models or types of difference map might be used. Most importantly, all structure factor data, including the very low resolution terms, were used in the calculation of the difference map. By including these data, low-resolution Fourier series truncation artifacts are avoided and the description of the poorly ordered solvent structure is improved.

Color Plate 1 shows the solvent electron density in a box with side length of ~25 Å that covers approximately one

insulin molecule in the cubic crystal. In this map, the most highly ordered water molecules (bright gray-white spheres) are found in spaces between insulin molecules, where they make multiple hydrogen bonds with the protein. With this type of graphical representation it is possible to make large-scale overviews of the distribution of both localized and mobile solvent molecules in the crystal. Comparisons of the experimentally determined solvent density maps in different crystals may prove more objective than comparisons of atomic models that differ in completeness and reliability.

Compared to surface rendering methods, volumetric representations of electron density emphasize the general distribution of density rather than the sharp features in the map. In addition, only a small fraction of the solvent in most macromolecular crystals is sufficiently well ordered to provide a significant scattering contribution to the highest resolution structure factor data. For these reasons, most of the solvent density features observed in difference maps computed to 2.5-Å resolution (more typical of diffraction data obtained in protein crystallography) appear similar to features in the 1.7-Å resolution map. This graphics program is also being used to illustrate anomalous scattering difference density maps at 5-Å resolution, which reveal the distribution of very mobile thallium counterions in cubic insulin crystals.[12]

HYDRATION OF PROTEIN CAVITIES

Electron density maps provide a time-averaged picture of the crystal structure in which the atomic positions are blurred by thermal motion. Since protein cavities frequently bind drug and inhibitor molecules, the solvation of these enclosed spaces is a consideration for knowledge-based molecular design. Because these water molecules are surrounded by protein atoms, a realistic boundary condition is provided for molecular dynamics simulations of their positions and mobilities. I have carried out 50-ps molecular dynamics simulations of small numbers of water molecules confined within two rigid protein cavities created by packing defects between protein molecules in the cubic insulin crystal. In these calculations only the solvent atoms were allowed to move, and the effects of induced polarization were neglected. These are clearly approximations, but they save a great deal of computer time in exploratory calculations of the type described here. The simulations were carried out using the united-atom protein force field,[13] a SPC water model[14] at 288 K using the AMBER program.[15] Interactions between nonbonded atoms were calculated from all residues within 10 Å of each mobile atom, with a dielectric constant of unity. Energy minimizations of the solvent positions were carried out prior to molecular dynamics calculations.

A sodium binding cavity

Electron density maps show two well-defined streaks of density in a space between two insulin molecules in the cubic crystal (Color Plate 2). The end of the density streak that closely approaches the main chain carbonyl atoms of A5 Gln and A10 Ile contains a sodium ion.[9] Since simultaneous occupation of both of the closely spaced symmetry-equivalent sodium sites is prevented by electrostatic repulsion, any particular cavity in the crystal must contain a sodium ion and ligating waters in an asymmetric structure. Because the asymmetrically positioned sodium and waters are averaged with the crystal symmetry in the experimentally determined electron density map, the detailed interpretation of the density was initially unclear. We now interpret these density streaks as containing a sodium ion, a water molecule forming hydrogen bonds with the main and side chains of A5 Gln, and a second water molecule forming hydrogen bonds with the main chain of A10 Ile and a second copy of the side chain of A5 Gln.[11] The dimensions of the cavity appear too small to allow any additional water molecules to ligate the sodium ion. In Color Plate 2 the one self-consistent set of sodium and water positions is marked with black circles; the sodium is concealed beneath a serine side chain in one half of the stereo pair. Thus, the sodium ion is tetrahedrally coordinated, and all hydrogen bonding possibilities within the cavity are satisfied. This arrangement was found to be very stable throughout a 50-ps molecular dynamics simulation, with no significant departure from the starting model and no large translations or rotations of the sodium and its two ligating waters.

An additional water molecule deeper in the cavity is too distant to directly ligate the sodium and was found to be very mobile in the simulation, in qualitative agreement with the weak electron density. In preliminary calculations, in which only one sodium-ligating water was present (as we suggested earlier by analogy with expectations for larger cations[9]) the water placed deeper in the cavity immediately moved to ligate the sodium, leaving the lower part of the cavity empty and forming the stable tetrahedral structure just described. Thus, these calculations, more qualitative stereochemical consideration, and the most recent crystallographic refinement[11] support the new conclusion that the sodium ion at this site is tetrahedrally coordinated.

A mixed hydrophobic/hydrophilic cavity

A second cavity in these crystals is formed between two copies of B13 Glu with a hydrophobic region between two copies each of B12 Val and B16 Tyr (Color Plate 3). Since the crystals were obtained[8] at ~pH 9, both copies of B13 Glu were assumed to be unprotonated. Two copies each of waters 1, 10, 18, 27 (numbered according to the Protein Data Bank entry 9INS) were placed in the cavity. The crystallographic coordinates for the more ordered waters (numbers 1, 10 and 18) are shown in Color Plate 3 as dark dots inside the bright density maxima; a mobile water (number 27 with $B = 98$ Å^2) occupies the diffuse density in the lower part of the cavity. Two pairs of waters that hydrate B13 Glu were placed outside the cavity to partially offset the charge on the glutamic acid side chains. A preliminary simulation suggested that two copies of an additional water molecule, at positions corresponding to the weak electron density spikes at the left and right sides of the cavity, were needed to prevent the other waters from moving to stable but incorrect positions. Other simulations with selected waters omitted also led to a similarly incorrect set of time-averaged positions. Except for one water, which is almost immobilized by strong interactions with B13 Glu and B3 Asn, rapid exchanges of water molecules between the crystallographically identified sites were observed during the simulation.

The most hydrophobic portions of the cavity remained unoccupied during the simulation, and no density is observed in the corresponding region of the map. This hydrophobic space is filled by a dichloroethane molecule in dichloroethane-treated crystals. The results of both the simulation and the crystallographic analysis imply that dichloroethane binding will not incur a large desolvation penalty.

In the cubic crystal, the sodium-ligating waters may belong to the very small class of "structural" waters inferred from NMR measurements on other proteins.[16,17] The rapid exchange of other waters between sites for which there are resolved electron density maxima may be compared with observations on cyclodextrin crystals in which the water positions could be modeled by a set of partially occupied oxygen and hydrogen positions in a diffraction analysis[18] but were, nevertheless, found to move rapidly between sites when studied by quasielastic neutron scattering[19] and deuterium magnetic resonance.[20] These combined graphical and computational studies of cavity hydration in cubic insulin crystals show that there is no conflict between these observations if it is recognized that mapping the solvent electron density distribution (rather than identifying and modeling only the solvent peak positions) shows that large volumes are significantly occupied by solvent, and that the transit time between crystallographically refined sites is short compared to the time spent around the density maxima.

ACKNOWLEDGMENTS

This work was supported by grant CA47439 to Prof. D.L.D. Caspar from the National Cancer Institute and by Shared Instrumentation Grant 1-S10-RR04671-01 from the National Institutes of Health. I wish to thank Prof. Caspar for interest and ideas relating to the work described here. I am grateful to Dr. Gina Sosinsky for assistance with using the various graphics display devices in this laboratory.

REFERENCES

1 Jones, T.A. A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* 1978, **11,** 268–272
2 Goodsell, D.S., Main, I.S., and Olsen, A.J. Rendering volumetric data in molecular systems. *J. Mol. Graphics.* 1989, **7,** 41–47
3 Caspar, D.L.D. and Badger, J. Plasticity of crystalline proteins. *Curr. Op. in. Str. Biol.* 1991, **1,** 877–882
4 Badger, J. and Caspar, D.L.D. Water structure in cubic insulin crystals. *Proc. Natl. Acad. Sci.* 1991, **88,** 622–626
5 Namba, K., Caspar, D.L.D., and Stubbs, G. Enhancement and simplification of macromolecular images. *Biophys. J.* 1988, **53,** 469–475
6 Goodsell, D.S. and Olsen, A.J. Molecular illustration in black and white. *J. Mol. Graphics.* 1992, **10,** 235–240
7 Kraulis, P.J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* 1991, **24,** 946–950
8 Dodson, E.J., Dodson, G.G., Lewitova, A., and Sabesan, M. Zinc-free cubic pig insulin: crystallization and structure determination. *J. Mol. Biol.* 1978, **125,** 387–396
9 Gursky, O., Li, Y., Badger, J., and Caspar, D.L.D. Monovalent cation binding to cubic insulin crystals. *Biophys. J.* 1992, **61,** 604–611
10 Badger, J., Harris, M.R., Reynolds, C.D., Evans, A.C., Dodson, E.J., Dodson, G.G., and North, A.C.T. Structure of the pig insulin dimer in the cubic crystal. *Acta Crystallogr.* 1991, **B47,** 127–136
11 Gursky, O., Badger, J., Li, Y., and Caspar, D.L.D. Conformational changes in cubic insulin crystals in the pH range 7–11. *Biophys. J.* 1992, **63,** 1210–1220
12 Badger, J., Li, Y., and Caspar, D.L.D. Thallium counterion distribution in cubic insulin crystals determined from anomalous diffraction data. submitted
13 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Salvatore, P., and Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 1984, **106,** 765–784
14 Hermans J., Berendsen, H.J.C., and van Gunsteren, W.F. A consistent empirical potential for water–protein interactions. *Biopolymers.* 1984, **23,** 1513–1518
15 Singh, U.C., Weiner, P.K., Caldwell, J., and Kollman, P.A. AMBER 3.0. University of California at San Francisco, 1986
16 Otting, G., Liepinish, E., and Wüthrich, K. Protein hydration in aqueous solution. *Science.* 1991, **254,** 974–980
17 Forman-Kay, J.D., Gronenborn, A.M., Wingfield, P.T., and Clore, G.M. Determination of the positions of bound water molecules of reduced human thioredoxin by heteronuclear three-dimensional nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* 1991, **220,** 209–216
18 Steiner, T., Mason, S.A., and Saenger, W. Disordered guest and water molecules. Three-center and flip-flop O–H . . . O hydrogen bonds in crystalline ethanol octahydrate at $T = 295$ K: A neutron and X-ray diffraction study. *J. Am. Chem. Soc.* 1991, **113,** 5676–5678
19 Steiner, T., Saenger, W., and Lechner, R.E. Dynamics of orientationally disordered hydrogen bonds and of water molecules in a molecular cage: A quasielastic neutron scattering study of β-cyclodextrin: 11 H_2O. *Mol. Phys.* 1991, **72,** 1211–1232
20 Usha, M.G. and Wittebort, R.J. Structural and dynamical studies of the hydrate, exchangeable hydrogens and included molecules in β- and γ-cyclodextrins by powder and single-crystal deuterium magnetic resonance. *J. Am. Chem. Soc.* 1992, **114,** 1541–1547