Multiple Hydration Layers in Cubic Insulin Crystals

John Badger
Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254 USA

ABSTRACT Cubic insulin crystals contain ~30-Å-diameter channels filled with aqueous solvent, providing a useful system in which to analyze hydration structure at a variety of distances from protein surfaces. Beginning with an atomic model for the protein and ordered water molecules, the density distribution in the solvent volume of the phasing model was iteratively refined to improve the fit of calculated structure factors with x-ray diffraction data. The free *R* value, which compares calculated structure factors with a subset of observed structure factors deliberately omitted from the refinement, was used to provide an objective confirmation of the effectiveness of the refinement procedure. Electron density maps of the solvent, computed using the solvent-refined phases and complete low-resolution diffraction data, reveal multiple hydration layers around the protein.

INTRODUCTION

Most diffraction studies of the distribution of water molecules in protein crystals have concentrated on analyzing the interactions between protein atoms and waters specifically bound to the protein surface (Thanki et al., 1988). Outside the primary hydration layer, electron density maps of the solvent contain few resolvable peaks and are usually represented by a featureless continuum in which the electron density is set to a constant level (Blake et al., 1983). However, the radial density distribution for pure liquid water (Narten and Levy, 1969) and the hydration force phenomena (LeNeveau et al., 1976) suggest that relatively fixed protein surfaces should impose some degree of order in the surrounding solvent over a range of ~10 Å.

This paper describes the application of a new statistical test to x-ray diffraction data (Brünger, 1992) to measure the extent of the hydration structure in cubic insulin crystals (Dodson et al., 1978; Badger et al., 1991). Cubic insulin crystals (a = 78.9 Å, space group $I2_13$) contain solvent channels ~30 Å in diameter and diffract to 1.7 Å resolution (Dodson et al., 1978). About two-thirds of the crystal volume is filled with solvent. Because the insulin molecules are packed in mutually orthogonal rows running parallel to the unit cell edges, ~40% of this solvent is more than 8 Å from any protein atom. The crystals used for this analysis were grown in 0.4 M Na^+; similar crystals remain stable in 0.1 M concentrations of small monovalent cations at pH 7–10 (Gursky et al., 1992a,b). Thus, unlike protein crystals that contain only restricted solvent spaces or crystals obtained at high ionic strength, the cubic insulin crystals provide information on the distribution of water and monovalent cations at various distances from the protein under approximately physiological conditions.

In earlier work we used a solvent density modification scheme to include a contribution from mobile solvent, not represented by atomic coordinates, into the phasing model (Badger and Caspar, 1991). Since solvent density fluctuations in the resulting maps were significantly larger than our estimates for the map noise, we concluded that significant nonuniformities in the solvent density extended across the solvent channels in the crystal. Similar model-free electron density refinements have been applied in small-molecule crystallography (Ito and Shibuya, 1977; Van der Suis and Spek, 1990) and used to adjust portions of solvent density in macromolecular neutron crystallography (Kossiakoff et al., 1992). However, these procedures have not been widely adopted because, until recently (Brünger, 1992), no critical test was available to confirm that the refinement is fitting the scattering from poorly ordered solvent structure rather than absorbing noise in the structure factor data. In this paper I describe a controlled and verifiable approach to improving maps of the solvent electron density in macromolecular crystals. In addition, I argue that observed periodicities in the solvent density outside the primary hydration layer are consistent with the expected distribution of water molecules restrained by a protein surface.

MATERIALS AND METHODS

Calculations were carried out using 1.7 Å resolution structure factor data for cubic porcine insulin crystals (Dodson et al., 1978) and the resulting atomic model (Badger et al., 1991, and entry 9INS in the Brookhaven Data Bank). This set of atomic coordinates contains 81 ordered water sites of an estimated ~440 waters per asymmetric unit. An initial phasing model was obtained by computing an electron density map from the protein and ordered solvent coordinates (Agarwal, 1978) with a constant solvent electron density filling unoccupied grid points. The calculated electron density around each atom extended to 2.6 Å from the atom center, and the grid spacing in the model map was 0.66 Å.

Iterative refinement of the solvent density distribution in this model map consisted of the following steps: (*i*) calculation of a conventional difference Fourier map from the Fourier transform of $(F_{obs} - F_{calc}) \exp(i\alpha_{calc})$ using the observed structure factor amplitude, F_{obs}, and structure factor amplitudes and phases, F_{calc} and α_{calc}, obtained from this phasing model; (*ii*) addition of this difference Fourier map to the phasing model at grid points only within a specified region of the solvent; (*iii*) calculation of a new set of structure factor amplitudes and phases from this solvent-modified map.

The crystal volume fixed by the protein density model was defined by the protein Van der Waals envelope. Solvent density fluctuations in the difference Fourier map that were outside the refinable solvent region were averaged to a constant value and added to the constant density solvent space.

Received for publication 5 May 1993 and in final form 14 June 1993.
Address reprint requests to Dr. John Badger.

0006-3495/93/10/1656/04 $2.00

Scale factors between observed and calculated structure factor amplitudes were redetermined after each refinement cycle using all observed data (except data deliberately omitted for free R value calculations) to 1.7 Å resolution. A least-squares fit of a multiplier and "temperature factor" was used to scale F_{obs} with F_{calc}.

Test refinement cycles were carried out in which the size of the refinable solvent volume (step *ii*) was varied. The free R value statistic ($R_{free} = \Sigma \mid F_{obs} - F_{calc} \mid / \Sigma F_{obs}$), which compares calculated structure factors, F_{calc}, with a subset of observed structure factors, F_{obs}, omitted from refinement calculations, was used as an objective indicator of improvements in the solvent-modified phasing model (Brünger, 1992). These refinements used structure factor data extending to only 2.6 Å resolution because preliminary tests showed that the refinement procedure had no useful effect on the higher-resolution data. To determine R_{free}, 5% of the observed structure factor data (113 reflections to 2.6 Å resolution) were omitted from the refinements. A test calculation in which the protein portion of the map was allowed to change was carried out to show that fixing the protein density at the values given by the protein model reduces error. The improvement in R_{free} ceased after two or three refinement cycles.

After completing the solvent-density refinement of the phasing model, a 1.7-Å resolution electron density map of the solvent was computed using the observed structure factor amplitudes, F_{obs}, and solvent-modified phases, α_{ref}. To remove the image of the protein from this map, structure factor terms, F_{prt} and α_{prt}, were computed from the protein model alone and subtracted from the complete structure factors. Thus, the solvent density map of the solvent, ρ_{sol}, was obtained from the Fourier transform of the set of structure factor differences

$$\rho_{sol} = FT[F_{obs}\exp(i\alpha_{ref}) - F_{prt}\exp(i\alpha_{prt})].$$

All low-resolution structure factor data were included in these maps to avoid series truncation artifacts and to improve the map of the poorly ordered solvent. The values of the unmeasured F(000) and F(011) terms were chosen so that the average solvent density throughout the crystal was approximately equal to the density for liquid water.

RESULTS AND DISCUSSION

Verification of solvent density refinement

The free R value, which compares a small subset of observed structure factors omitted from refinement calculations with structure factors calculated from the refined phasing model, should effectively distinguish between genuine improvements in modeling the data from meaningless overfitting (Brünger, 1992). The solvent density refinement was initiated from a model that included the 81 ordered solvent sites identified from the conventional crystallographic refinement. Therefore changes in R_{free} are due to alterations and additions to this structure. The value of R_{free} was reduced by ~3% by the refinement. Although significant, this improvement is relatively small because solvent density fluctuations absent from the initial atomic model make only a small scattering contribution to the diffraction data. According to the R_{free} statistic, hydration structure beyond the protein-solvent interface is required to improve agreement with x-ray diffraction data in cubic insulin crystals (Table 1). For the refinement of the three-dimensional phasing model, an adjustable volume of solvent density extending ~8 Å from the protein surface, incorporating two or three hydration layers, was found to be optimal. Since this refinement is based on differences between observed and calculated structure factors that include experimental errors, the true extent of the solvent density fluctuations should be at least as large as the volume that could be refined.

TABLE 1 Free *R* value statistic for solvent density phasing models

Phasing model	R_{free}
Atomic model and flat bulk solvent	0.161
Refined solvent (6 Å) after one cycle	0.141
Refined solvent (8 Å) after one cycle	0.140
Refined solvent (10 Å) after one cycle	0.147
Refined protein and solvent (8 Å) after one cycle	0.154
Refined solvent (8 Å) after two cycles	0.133

This statistic compares a set of observed structure factors omitted from each refinement with structure factors calculated from the refined phasing models. The R value for the protein and ordered solvent model with flat bulk solvent density (0.161) was obtained after conventional model refinement with all structure factors included. This value serves as a baseline for assessing the effectiveness of the various solvent-density refinements. R_{free} is given for phasing models that include different refinable solvent volumes after one refinement cycle. Additional entries show R_{free} after refining both the protein and partial solvent volume, and the solvent-refined phasing model after two iterations. The maximum distance from the protein into which solvent density differences were introduced (given in parenthesis) is defined as the distance to the nearest protein atom.

Distance dependence of hydration structure

Solvent density points equidistant from the protein surface in the three-dimensional map, ρ_{sol}, were averaged to give a one-dimensional electron density profile (Fig. 1). For x-ray diffraction data, the density maxima in this plot are caused by nonuniformities in the average positions of the water oxygen atoms. This analysis of distance-dependent solvent density variations obtained from a three-dimensional map is distinctly different from the parameter analysis of shells of pseudoatoms presented by Cheng and Schoenborn (1990). Fig. 1

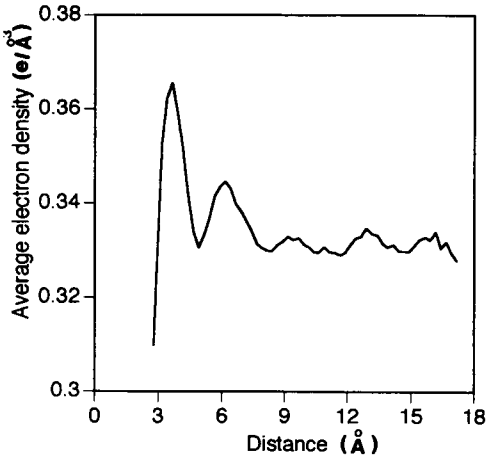

FIGURE 1 Average solvent electron density as a function of distance from the protein. This plot shows density maxima due to multiple hydration layers around the insulin molecule. A strong primary density maximum is observed in contact with the protein surface, and four smaller peaks are found at greater distances from the protein. The number of solvent density pixels in each shell decreases with increasing distance from the protein surface. Thus the proportion of the solvent that is 15–18 Å from any protein atom is about 15% of the amount of solvent 3–6 Å from the protein.

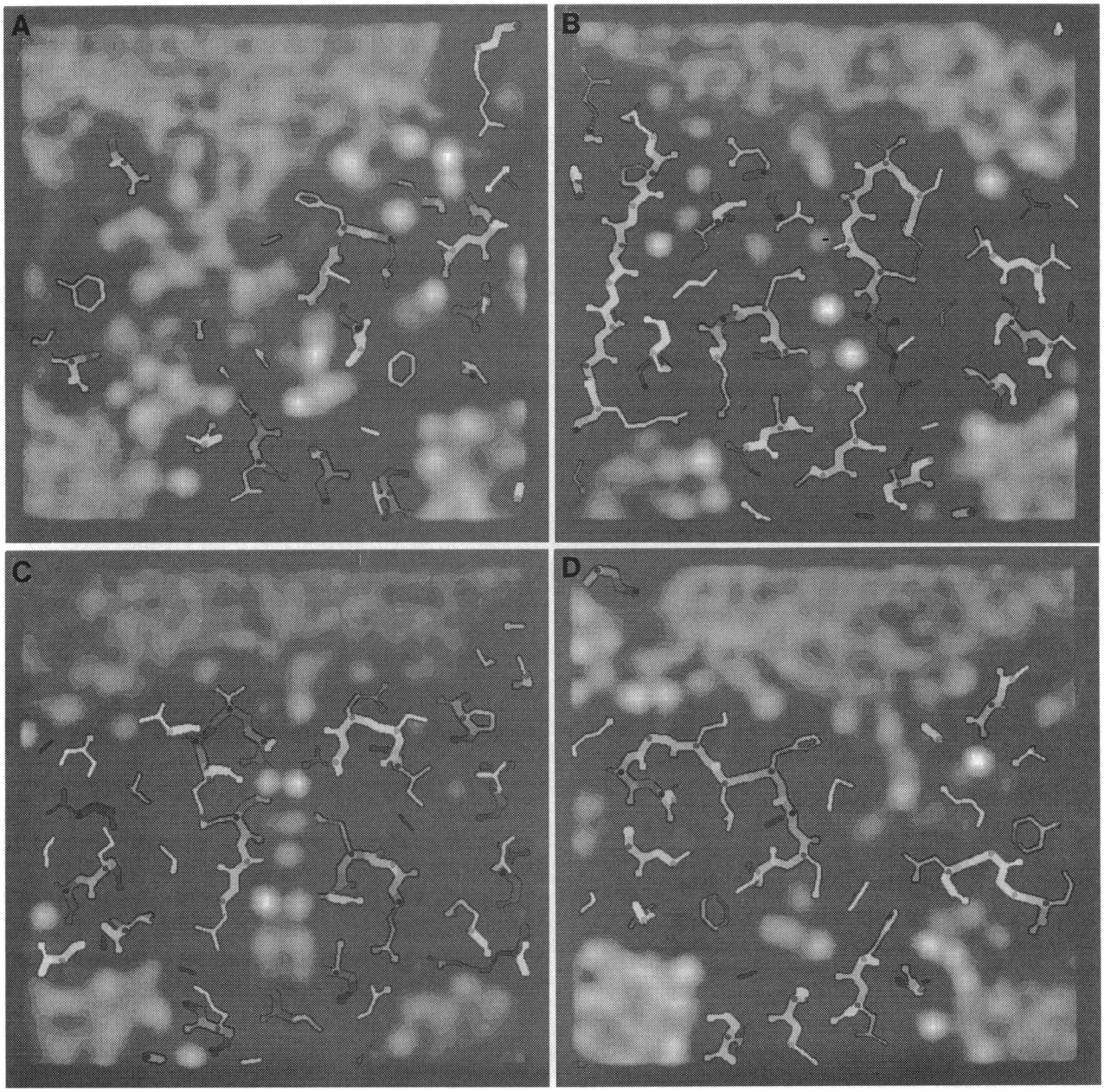

FIGURE 2 (*a–d*) Depth-cued projections of the three-dimensional solvent electron density map showing successive 5-Å-thick slabs of density with a side length of 40 Å. The diameter of one insulin molecule is about one-half the width of each slab of density. The phenylalanine ring at the left-hand side of (*a*) is at the topmost surface of one insulin molecule. Well-ordered solvent atoms (seen as bright isolated features) at the center of (*b*) and (*c*) lie between two insulin molecules that are related by a twofold crystal dyad. Indistinct ringlike features, most visible at the top center of (*a*) and (*d*), correspond to poorly ordered solvent structure that extends from the protein surface. In (*d*), symmetry-related copies of parts of the protein in (*a*) reappear. The protein is represented by a stick model in which bonds between backbone atoms are drawn thicker than bonds between side-chain atoms. The Cα atoms are marked with dark dots, and oxygen atoms are drawn as enlarged spheres.

shows a strong density maximum, corresponding to the primary hydration shell, 3–4 Å from the protein. This density maximum contains contributions from both the ordered solvent, hydrogen-bonded to polar groups, and the mobile solvent in contact with nonpolar surfaces. The distance between primary and secondary hydration maxima is similar to that obtained in theoretical calculations for tetrahedrally coordinated water molecules near fixed plane surfaces (Lee et al., 1984; Valleau and Gardner, 1987). The positions of three weaker density maxima >8 Å from the protein surface are consistent with a continuation of this network. These smaller peaks are in a region of the map where the phasing model is a flat density distribution but are revealed by this averaging of the experimentally determined solvent density over many equidistant gridpoints.

Although the mean phase difference between structure factors calculated from the initial and solvent-refined phasing models is only 6°, larger phase changes occur for some

of the small number of low-resolution data. The importance of these phase differences was verified by a calculation of the one-dimensional density profile with the unrefined phases, which showed an unphysically small primary hydration maximum and severe distortions in the density further from the protein. The refined density profile obtained from a set of bovine insulin crystal data (pH 9 in 0.1 M Na^+; Gursky et al., 1992b) is very similar to the profile in Fig. 1, but a comparable plot from a crystal soaked in a 3 M glucose solution does not show these features (Kapulsky et al., unpublished observations). Thus, the distance-dependent density fluctuations in Fig. 1 are unlikely to be artifacts of this particular set of porcine insulin data or systematic artifacts caused by the solvent density refinement procedure.

Three-dimensional solvent density distributions

Volume rendering methods were used to display the three-dimensional solvent electron density maps to permit a graphical analysis of large volumes of both the well-ordered and diffusely distributed solvent density (Badger, 1993).

The degree of localization observed for individual water molecules is highly dependent on the number and proximity of well-ordered polar atoms in the protein. The most ordered water molecules, recognizable in these maps as sharp, bright, spherical features (Fig. 2), are at sites where the water molecules may make multiple hydrogen bonds with the protein. Most of these sites occur in small spaces between insulin molecules in the crystal lattice. In partially enclosed volumes, 2–3 water molecule diameters across, individual water peaks are generally blurred but are still resolvable. The cubic insulin crystal lattice does not contain any enclosed hydrophobic cavities, but the hydrophobic portion of a mixed hydrophilic/hydrophobic cavity does not appear to contain electron density (Badger, 1993). At exposed protein surfaces, individual water molecules are not resolved, but some diffuse, ringlike features are seen extending away from the protein. The dimensions of these rings suggest a tetrahedrally connected network, with dimensions comparable to those of water rings in various forms of ice.

Although much of the solvent electron density seen in these maps is not readily interpretable in terms of an unambiguous atomic model, it should be generally possible to improve and graphically analyze these maps using methods of the type described here. These electron density maps contain information regarding the solvent "missing matter" that is not present in atomic coordinate files but which is important for the screening of protein electrostatic interactions. Properly validated model-free refinement procedures could also be useful for improving maps of biologically important interfacial solvent regions prior to interpretation in terms of atomic models.

I am grateful to Prof. Caspar for his continued interest and support in these studies.

This work was supported by grant CA-47439 from the National Cancer Institute to Prof. D. L. D. Caspar.

REFERENCES

Agarwal, R. C. 1978. A new least-squares refinement technique based on the fast Fourier transform algorithm. *Acta Crystallogr.* A34:791–809.

Badger, J. 1993. Display and interpretation of solvent density distributions in insulin crystals. *J. Mol. Graphics.* In press.

Badger, J., and D. L. D. Caspar. 1991. Water structure in cubic insulin crystals. *Proc. Natl. Acad. Sci. USA.* 88:622–626.

Badger, J., M. R. Harris, C. D. Reynolds, A. C. Evans, E. J. Dodson, G. G. Dodson, and A. C. T. North. 1991. Structure of the pig insulin dimer in the cubic crystal. *Acta Crystallogr. Sect. B Struct. Sci.* 47:127–136.

Blake, C. C. F., W. C. A. Pulford, and P. J. Artymiuk. 1983. X-ray studies of water in crystals of lysozyme. *J. Mol. Biol.* 167:693–723.

Brünger, A. T. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature (Lond.).* 355:472–475.

Cheng, X., and B. P. Schoenborn. 1990. Hydration in protein crystals. A neutron diffraction analysis of carbon monoxymyoglobin. *Acta Crystallogr. Sect. B Struct. Sci.* 46:195–208.

Dodson, E. J., G. G. Dodson, A. Lewitova, and M. Sabesan. 1978. Zinc-free cubic pig insulin: crystallization, and structure determination. *J. Mol. Biol.* 125:387–396.

Gursky, O., Y. Li, J. Badger, and D. L. D. Caspar. 1992a. Monovalent cation binding to cubic insulin crystals. *Biophys. J.* 61:604–612.

Gursky, O., J. Badger, Y. Li, and D. L. D. Caspar. 1992b. Conformational changes in cubic insulin crystals in the pH range 7–11. *Biophys. J.* 63: 1210–1220.

Ito, T., and I. Shibuya. 1977. Phase refinement by a Fourier transformation of difference electron density peaks. *Acta Crystallogr.* A33:71–74.

Kossiakoff, A. A., M. D. Sintchak, J. Shpungin, and L. G. Presta. 1992. Analysis of solvent structure in proteins using neutron D_2O-H_2O solvent maps: pattern of primary and secondary hydration in trypsin. *Proteins.* 12:223–236.

Lee, C. Y., J. A. McCammon, and P. J. Rossky. 1984. The structure of liquid water at an extended hydrophobic surface. *J. Chem. Phys.* 80:4448–4455.

LeNeveau, D. M., R. P. Rand, and V. A. Parsegian. 1976. Measurement of forces between lecithin bilayers. *Nature (Lond.).* 259:601–603.

Narten, A. H., and H. A. Levy. 1969. Observed diffraction pattern and proposed models of liquid water. *Science (Washington DC).* 165:447–454.

Thanki, N., J. M. Thornton, and J. M. Goodfellow. 1988. Distributions of water around amino acid residues in proteins. *J. Mol. Biol.* 202:637–657.

Valleau, J. P., and A. A. Gardner. 1987. Water-like particles at surfaces. I. The uncharged, unpolarized surface. *J. Chem. Phys.* 86:4162–4170.

Van der Suis, P., and A. L. Spek. 1990. BYPASS: an effective method for the refinement of crystal structures containing disordered solvent regions. *Acta Crystallogr.* A46:194–201.