Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 1224–1228, February 1994
Biophysics

Thallium counterion distribution in cubic insulin crystals determined from anomalous x-ray diffraction data

(protein electrostatics/macromolecular structure/cation interactions)

JOHN BADGER, YOULI LI, AND D. L. D. CASPAR

Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02254-9110

Communicated by Paul B. Sigler, October 19, 1993

ABSTRACT **To determine the distribution of monovalent cations around a protein we have measured anomalous scattering diffraction data from Tl-containing cubic insulin crystals at pH 8 and pH 10.5. The differences between Bijvoet reflection pairs within each set of data were used to calculate anomalous scattering difference maps. Both maps show the same six Tl^+ sites, which include two well-ordered Tl^+ ions previously identified from isomorphous exchange experiments. The other four sites constitute a second class of cations, which, while much more mobile than the protein atoms, are associated with particular ligating groups. Three of the six Tl^+ sites are created exclusively by protein main and side chain carbonyl dipoles rather than negatively charged groups. All of the Tl^+ ions are positioned so as to interact with both protein atoms and water molecules. The Tl^+ occupancies appear to depend in a complex way on interactions with each other and flexibility in the protein structure. The combined occupancies of these cations are slightly less than is required to neutralize the net protein charge of approximately −2e at pH 8 but account for only about half of the approximately −5e protein charge at pH 10.5. Thus, more disordered counterions, not seen in these Bijvoet anomalous scattering difference maps, are more numerous at higher protein net charge.**

Physiological environments typically contain ≈0.1 M concentrations of monovalent cations (mostly Na^+ or K^+), which act as counterions for biomolecules (1). Despite the prevalence of Na^+ and K^+ in biological systems, the interactions of these cations with protein molecules are not well characterized by x-ray diffraction experiments because the x-ray scattering amplitude from Na^+ (Z = 10e, the number of scattering electrons) is almost identical to the scattering amplitude from a water molecule; x-ray scattering from K^+ (Z = 18e) is less than twice as great as from water. The heavier monovalent cations Rb^+ (Z = 36e) and Tl^+ (Z = 80e), which are expected to interact with proteins in a similar way to Na^+ or K^+, have occasionally been used to obtain derivative crystals for the purpose of structure determination, and a few examples of localized binding sites for these cations have been reported (2). Protein molecules may contain specific binding sites in which monovalent cations are tightly attached. Relatively more mobile cations, with rms displacements of ≈1–2 Å, may still be considered to interact with a few specific atomic groups. The type of counterions described by Debye–Hückel theory (3) should be distributed over a large volume of space with a population density that depends on a smoothly varying electrostatic potential around the protein. A characterization of these classes of protein–cation interactions using x-ray diffraction data collected from protein crystals will help rationalize electrostatic effects in proteins and aid the development of physically appropriate models for representing the neutralizing counterions in computational studies.

Cubic insulin crystals (4) are stable between pH 7 and pH 10.5 in 0.1 M solutions of monovalent cations with radii <1.5 Å (5). The porcine insulin structure has been determined at 1.7 Å resolution at approximately pH 9 (6), and a series of bovine insulin structures have been obtained at 2 Å resolution between pH 7 and pH 11 (7, 8). Based on the amino acid sequence and the locations of the titratable residues in the crystal, the net charge on the insulin molecule may be estimated to be approximately −2e at pH 8 and approximately −5e at pH 10.5. Isomorphous difference maps obtained from data in the resolution range 10–2.8 Å revealed two well-ordered cation-binding sites on the protein (5). One of these sites is located just off a crystal dyad, in a cavity between two insulin molecules. When this site is occupied by Tl^+, the cation is ligated by the main chain carbonyl atoms of Gln-A5 and Val-A10, the side chain of Gln-A5, and (probably) a water molecule. Diffraction data from crystals containing Na^+ and Tl^+ show that this site is partially occupied at pH 7 and approaches maximal occupancy of one cation per dyad at pH > 9.5 (5, 7). Occupation of this site by a cation is correlated with a conformational change in the side chain of Gln-A5, which rotates into the binding cavity when the cation is present. The second localized cation is at a site that is occluded by one of two possible conformations of His-B10. At pH 7 the histidine side chain almost completely occludes the cation binding site, but at higher pH a second conformation, with the side chain rotated out of the cation site, becomes increasingly populated (7). Since steric factors prevent the combined occupancies of these two localized cation sites from exceeding much more than one positive charge per insulin molecule, the remaining negative charge must be neutralized by additional counterions that are too mobile to observe in electron density maps that lack the low-resolution (d > 10 Å) data.

To locate these mobile cations, we have collected 5 Å resolution diffraction data from individual Tl–insulin crystals at pH 8 and pH 10.5 and calculated anomalous scattering difference maps using the Bijvoet differences within these data sets. Compared to isomorphous difference data, the Bijvoet anomalous difference data have the advantage that it is not necessary to measure and scale data collected from different crystals. The significant features in the resulting maps are due only to the anomalous scatterers. We show here that useful information on the distribution of counterions in protein crystals may be obtained from readily detectable anomalous scattering differences ($\Delta f''$ = 8e for Tl^+ at a wavelength of 1.54 Å) on a laboratory x-ray source.

MATERIALS AND METHODS

Bovine cubic insulin crystals (space group $I2_13$, a = 78.9 Å) were grown with 0.5 M Na^+ and then dialyzed overnight in 0.1 M Tl^+ at pH 8 and pH 10.5 as described (5). To prevent crystal slippage during data collection, the crystals were

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

fastened to the capillary tubes with a Formvar film cast from dichloroethane solution (9). Diffraction data to 4.5 Å resolution were measured using an Elliot rotating anode on the Brandeis area detector (10, 11) in 0.1° frames over a total rotation range of 180°. Data collection and initial processing were carried out with the MADNES software (12). Two sets of measurements were made from each crystal at different gain settings to accommodate the full dynamic range of the intensity data. The two sets of measurements made from each crystal were scaled together using a subset of reflections common to both data sets. Intensity measurements flagged as unreliable and measurements beyond 5 Å resolution (data near the edge of the detector) were discarded. After applying an empirical absorption correction according to the crystal rotation angle and eliminating a few discrepant measurements (≈5% of the data collected), the intensity data were merged in groups containing Bijvoet-equivalent reflections. Since the distribution of structure factor amplitudes as a function of resolution was very similar for both the pH 8 and pH 10.5 data, a linear scale factor was used to put both sets of data on the same relative scale.

The anomalous scattering component that is observable at a particular x-ray wavelength is caused by a phase shift of $\pi/2$ in the radiation diffracted by the anomalous scatterers. An unbiased map of the anomalous scatterers in the crystal, which correlates well with their true distribution, is the Bijvoet difference map $[F(+) - F(-)]\exp(i\phi + \pi/2)$ in which the phases, ϕ, are computed from a model that lacks the anomalous scatterers (13). Bijvoet difference maps were computed using all paired Tl–insulin data to 5 Å resolution with phases calculated from refined Na–insulin models. Bijvoet differences are only present for acentric reflections, thus these difference maps do not contain any data at a resolution lower than 32 Å (the 112 reflection) because these low-order reflections are all centric.

RESULTS

Reliability of Diffraction Data. The merging R factors of 8–9% in the four sets of equivalent data reflect a relatively large spread of the independent measurements about the group means (Table 1). The effect of these measurement errors was mitigated by recording large numbers of duplicate reflections over a wide rotation angle. It was possible to measure these data without incurring excessive radiation damage to the crystal because the small number of low-resolution data have relatively large intensities, which could be rapidly measured on the electronic area detector. Because of the high space group symmetry of these crystals, there is an unusually high redundancy of ≈10 measurements to each unique reflection. Thus, the merged data are accurately determined, and nonisotropic absorption effects are minimized. The standard errors in the merged intensities are 2–3% of the average intensities, which are ≈3 times smaller than the R factors of 6% and 10% obtained for the Bijvoet intensity differences.

Table 1. Anomalous scattering data collected from Tl–insulin crystals at pH 8 and pH 10.5 to a resolution of 5 Å

Parameter	pH 8		pH 10.5	
	+	−	+	−
No. of reflections measured	2731	2695	1801	1704
No. of unique reflections	261	261	213	210
Merging R factor on intensities	0.084	0.094	0.087	0.084
Standard error of intensity as a fraction of mean intensity	0.023	0.031	0.029	0.034
R factor on Bijvoet intensity differences	0.063		0.104	
R factor on Bijvoet structure factor differences	0.044		0.063	
R factor on F after refinement of the Tl^+ atomic parameters	0.029		0.039	

The columns labeled + and − refer to sets of reflections in Bijvoet-equivalent groups corresponding to h, k, l and $-h, -k, -l$.

Anomalous Scattering Difference Maps. The anomalous scattering difference map obtained from the Tl–insulin data at pH 8 showed maxima at the two localized Tl^+ sites previously identified (sites 1 and 2 of ref. 5) and four other prominent but more diffuse features (described below) that we identify as more mobile Tl^+ ions. At a density threshold, where the map was almost free from noise, smaller peaks at the sites of two of the three disulfide bridges in the insulin molecule were also visible. With the CuKα radiation used for these experiments, the anomalous scattering amplitude is 0.6e for a sulfur atom and 8e for a Tl^+ ion. Density maxima corresponding to all of these six Tl^+ sites were also seen in the somewhat noisier anomalous scattering Tl–insulin map obtained from the crystal at pH 10.5.

The features we identify as Tl^+ in these maps (Figs. 1 and 2) are

Site 1. Oblate density near the crystal dyad that is due to a Tl^+ bound in either one of two symmetry-related binding sites separated by ≈2 Å (site 1 of ref. 5). The bound Tl^+ is ligated by the main chain carbonyl atoms of Gln-A5, Val-A10, and the side chain of Gln-A5 (Fig. 2*A*). Analysis of analogous structures with K^+ (7) and NH_4^+ at this site (23) indicates that there is a ligating water molecule close to the unoccupied symmetry-related Tl^+ site across the crystal dyad.

Site 2A. Spherical density in a position that is occluded by one conformation of the side chain of His-B10 (site 2 of ref. 5). The bound Tl^+ is ligated by the main chain carbonyl of Tyr-A14 and a side chain oxygen from Asn-A18 (Fig. 2*B*).

Site 2B. Weak and irregular density, with an ill-defined maximum, in a region that would allow ligation by the side chains of Asn-B3 and Glu-B13. This site is probably occluded when the side chain of His-B10 is in a position that allows cation binding at site 2A (Fig. 2*B*).

Site 3A. A strong, diffusely spherical density feature positioned close to the main chain carbonyl atoms of Asn-A18 and Cys-A20 at a site created at the end of the second A-chain helix (Fig. 2*C*).

Site 3B. A weak and somewhat oblate feature close to the two C-terminal main chain atoms of Asn-A21 (Fig. 2*C*).

Site 4. A triangular feature that is centered on the crystallographic three-fold axis between three copies of the poorly ordered side chain Glu-B21 (Fig. 2*D*).

Estimated Tl Site Occupancies. Atomic parameters for the six Tl^+ atoms were estimated by a reciprocal-space least-squares refinement procedure in which model atoms were refined against the Bijvoet Fourier coefficients in a manner analogous to the phased heavy-atom refinement method used for isomorphous difference data (16). The low-resolution data used for this refinement are very insensitive to the Tl^+ thermal factors ($B = 8\pi^2\langle u^2\rangle$), and the assumed values have very little effect on the occupancies obtained. Trial refinements showed that the rms positional fluctuations for Tl^+ at sites 1 and 2A are $u < 1$ Å, and the rms positional fluctuations for Tl^+ at sites 2B–4 are probably $u \approx 1$–2 Å. Electron density maps computed without the low-resolution data ($d > 10$ Å) show very little density at sites 2B–4, confirming that Tl^+ ions at sites 2B–4 are more mobile than the Tl^+ at sites 1 and 2A. Since the positional parameters for Tl^+ at sites 1 and 2 have been determined from isomorphous difference data at 2.8 Å resolution (5), these were not refined against the lower resolution anomalous scattering data. The combined occupancies for the sulfur atoms in the three disulfide bridges in the insulin molecule at pH 8 were used to obtain a self-consistent absolute scale factor for the Tl^+ occupancies.

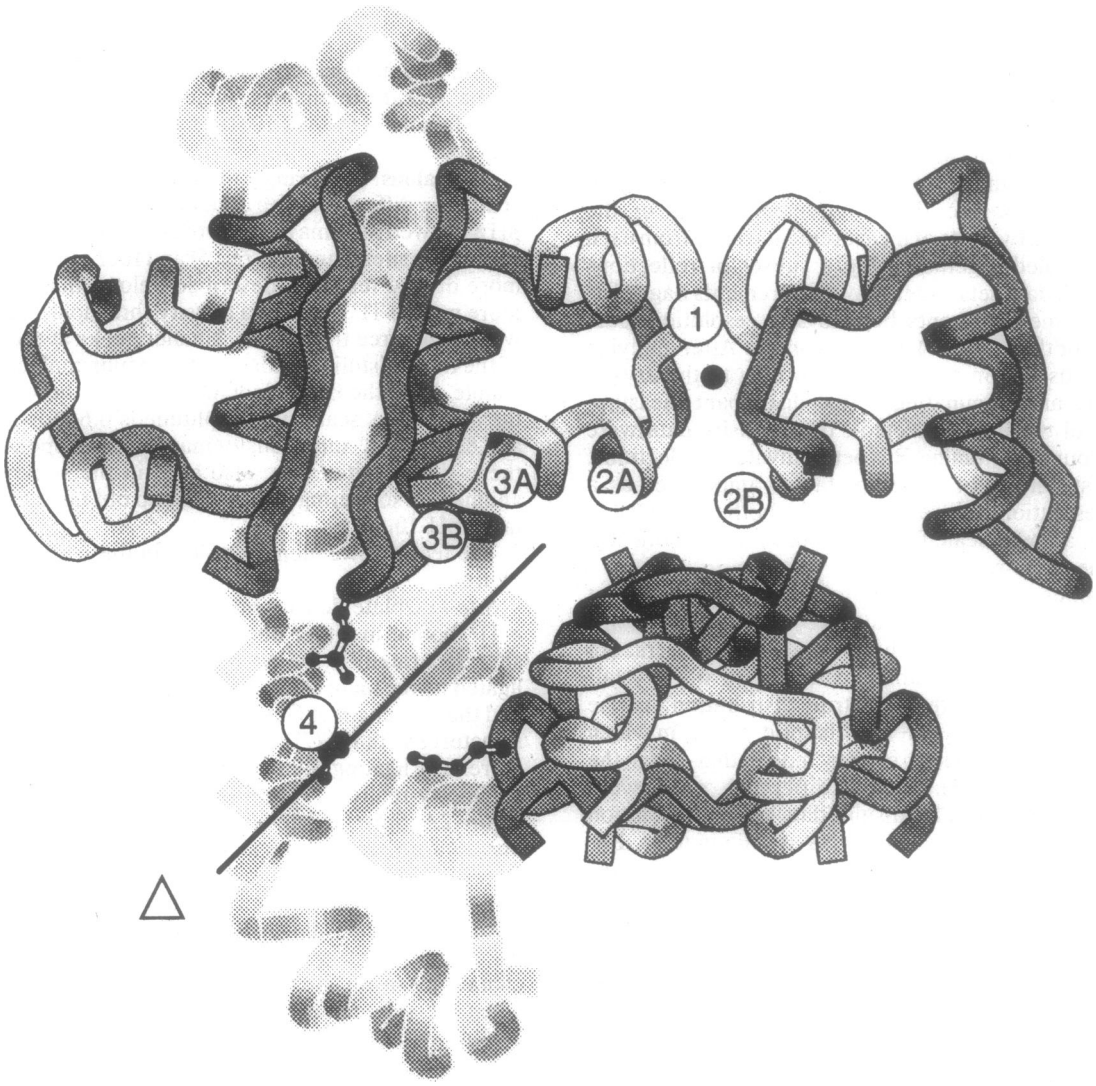

FIG. 1. Schematic illustration drawn with the MOLSCRIPT program (14) showing the positions of the six Tl^+ binding sites identified from the Bijvoet anomalous scattering difference maps. Insulin molecules in the cubic crystal are packed in orthogonal rows. The C^α trace for several molecules in the crystal is shown, with the A chains marked in light grey and the B chains marked in dark grey. A crystal twofold axis is indicated by a solid circle, and a threefold axis is indicated by an open triangle. Sites 1 and 4 lie close to these twofold and threefold symmetry axes, respectively.

In agreement with our previous studies, the refined fractional occupancy of Tl^+ at site 1 is less than the maximum possible value of 0.5 at pH 8 but attains maximum occupancy at pH 10.5. Tl^+ occupancy of site 2A is prevented when the side chain of His-B10 occupies a conformation that occludes the site. According to our high-resolution analysis of pH-dependent conformational changes in cubic insulin crystals (7), this side chain is present in two conformations with almost equal weights at pH 8, but the conformation that occludes site 2A is less visible at pH 10. Both isomorphous replacement data and anomalous scattering data show that site occupancy is incomplete at pH < 9 but exceeds 0.5 at pH > 9.5. Cation occupation of site 2B may only be possible when His-B10 is in the position that occludes site 2A. At pH 10.5 the combined occupancy of sites 2A and 2B is approximately unity.

The combined occupancy of sites 3A and 3B is close to unity at pH 10.5, and the occupancies for both sites are smaller at pH 8. Simultaneous occupation of sites 3A and 3B in the protein structure at pH 8–10.5 may be energetically unfavorable. The cubic insulin crystal structure at pH 11 in 1 M Na^+ shows a conformational change involving Asn-A21 (7), which could be due to oversaturation of these sites.

Site 4 consists of one lobe of the trigonal feature near three symmetry-related Glu-B21 carboxyl groups. The refined occupancy of ≈33% for this site corresponds to one cation shared between the three closely spaced binding sites about the crystal threefold axis. (The slightly lower refined occupancy at pH 10.5 is within the error margin of occupancy determination.) The occurrence of discrete disorder in Glu-B21 and the nearby Lys-B29 (6, 7) is consistent with the idea of slightly different cation-binding and cation-free conformations for these side chains.

The combined contents of sites 1–4 accounts for 90% of the expected charge of approximately $-2e$ on the insulin molecule at pH 8 but only about one-half of the expected charge of approximately -5 at pH 10.5 (Table 2). The total increase in the Tl^+ occupancies at sites 1–4 when the pH is raised from pH 8 to pH 10.5 is only approximately $+1$, which is much less than the expected increase of approximately -3 in the net negative charge on the insulin molecule. The additional counterions needed to neutralize these crystals must be too disordered to visualize in these Bijvoet anomalous difference maps. Furthermore, most of the increase in the protein net charge between pH 8 and pH 10.5 appears to be compensated for by an increase in the number of these very mobile Tl^+.

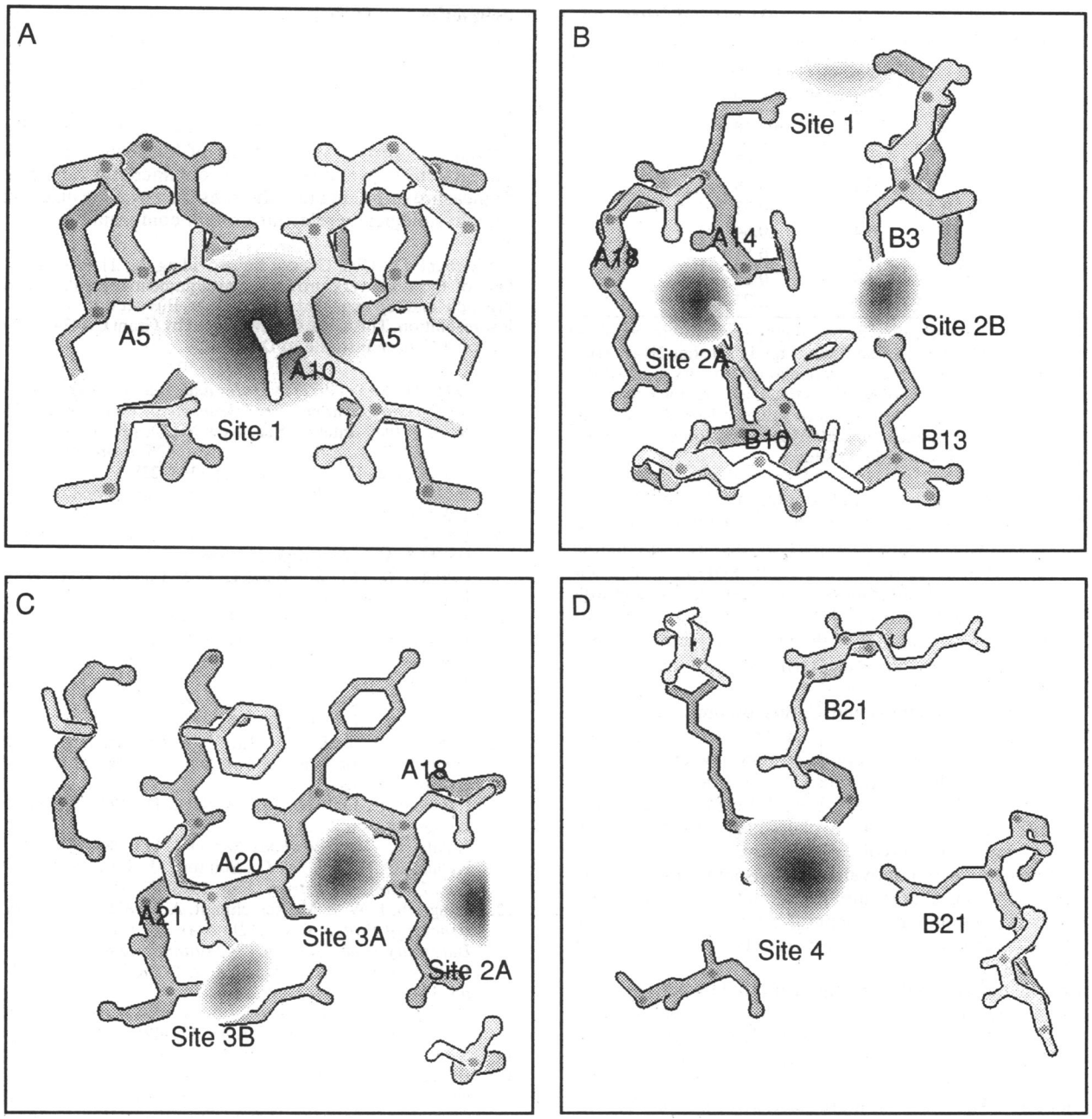

FIG. 2. Anomalous scattering difference electron density maps showing the six Tl^+ ions. The main chain covalent bonds are drawn thicker than the side chain bonds. Oxygen atoms are depicted as enlarged spheres and the C^α positions are marked by dark dots. The density threshold used for displaying significant features in these illustrations was chosen so that two of the three disulfide bridges were just visible. The grey level in each pixel was then determined by summing depth-cued density values along the view direction. With this graphical representation (15) the volume occupied by the electron density is the same as for a "cage" density contour surface, but the shading emphasizes "low-resolution" features for which significant density covers a large volume of space.

DISCUSSION

Determination of Thallium Distribution by Anomalous Scattering. We have used anomalous scattering difference maps, in which all differences are obtained from data collected from a single crystal, rather than isomorphous differences to map diffusely localized Tl^+ within the cubic insulin crystal. Although the Bijvoet anomalous scattering differences from Tl–insulin crystals are only approximately one-fifth of the scattering difference between Tl^+ and Na^+ (ΔZ = 70e), eliminating the need to compare crystals that differ in lattice order, x-ray absorption, and unit cell contents is a great practical advantage, which allowed us to obtain readily interpretable Bijvoet difference maps. Earlier attempts to obtain low-resolution isomorphous difference maps (unpublished data) did not give completely satisfactory results because of difficulties in correctly scaling data over a large dynamic range from crystals with different compositions.

A homogenous distribution of highly mobile Tl^+ in the crystal solvent channels will significantly affect only the very low-resolution data and will, according to Babinet's principle, scatter with phase angles that differ by $\approx\pi$ from the protein structure. Since Bijvoet anomalous differences tend to zero when the diffracted amplitudes from the anomalous scatterers are parallel to the diffracted amplitudes from the complete structure (13), Bijvoet anomalous difference maps are not expected to provide much information on a completely disordered counterion density distribution. A difference density map of the distribution of highly mobile Tl^+ will require either an accurate determination of the isomorphous

Table 2. Refined occupancy factors for Tl^+ at the six sites identified from anomalous scattering data at pH 8 and pH 10.5

Site	No. of Tl^+ ion per insulin monomer	
	pH 8	pH 10.5
1	0.35	0.47
2A	0.17	0.58
2B	0.32	0.32
3A	0.42	0.64
3B	0.27	0.51
4	0.34	0.25
Total	1.87	2.77

The values listed were obtained assuming temperature factors of 20 $Å^2$ for Tl^+ at sites 1 and 2A and 100 $Å^2$ for Tl^+ at sites 2B–4. Trials in which the temperature factors were allowed to vary showed that the temperature factors were poorly defined and that the occupancy values were insensitive to their assumed values. The expected errors in the refined Tl^+ occupancies are ±0.05 at the two ordered sites and ±0.1 at the four less well-ordered sites.

differences between the crystals containing Tl^+ and Na^+ or, if sufficiently accurate data could be collected, measurements of changes in the real part of the Tl^+ anomalous scattering factor using multiwavelength synchrotron radiation (17). These measurements would add a considerable number of centric data to the Tl^+ density map (all structure factors to 32 Å and one-half of the structure factors to 15 Å are centric) and should allow a determination of the complete distribution of Tl^+ in the crystal.

Electrostatic Interactions and Cation Binding. All the Tl^+ ions identified in this study are positioned so that they may make multiple contacts with both protein oxygen atoms and water molecules. Three of the six Tl^+ sites (including the two highly localized Tl^+ ions) are bound by neutral carbonyl groups rather than by negatively charged groups. Main chain carbonyl groups are involved in many of the observed interactions between monovalent cations and protein atoms. The prevalence of charge–dipole interactions involving the protein backbone may have general significance since these groups could create arrays of sites for charge conduction in ion channels (18).

Although cation localization appears to depend more on a suitable arrangement of ligating groups than the presence of a nearby negative charge, long-range charge–charge interactions do influence cation binding. Site 1 is a particularly clear example of the importance of these long-range electrostatic interactions. This site is more distant from the protein charged groups than almost any other point on the protein surface. Although there is little cation binding at pH 7, the site is fully occupied at pH > 9. Thus, the change in charge due to titration of relatively distant ionizable groups in the protein induces both cation binding and rotation of the side chain of Gln-A5 into the cation-binding cavity (5, 7).

Long-range electrostatic interactions in proteins are considered to be largely due to the ionizable groups because the strength of interactions with the protein atomic dipoles diminishes rapidly with separation. These charge–charge interactions are implicated in controlling rates of association with charged ligands (19) and play a role in altering the pK values of widely separated titratable groups (20, 21). In the absence of experimental data, the effect of neutralizing counterions on these interactions has usually been described in terms of Debye–Hückel theory. In the cubic insulin crystal, we have shown that a large proportion of the counterions are associated with specific protein atomic groups, with a complex pattern of occupancies that depend on interactions between the cation sites and protein structural changes. Improved descriptions of long-range electrostatic interactions in proteins would be obtained if it were possible to measure or predict (22) the sites for these localized ions in the physiologically relevant protein conformation.

We thank Dr. Olga Gursky for providing the Tl-insulin crystals. One of the reviewers provided the argument regarding Babinet's principle and the expected magnitude of Bijvoet differences at very low resolution. This work was supported by Grant CA47439 from the National Cancer Institute to D.L.D.C.

1. Dick, D. A. T. (1972) in *Water and Aqueous Solutions: Structure, Thermodynamics and Transport Processes*, ed. Horne, R. A. (Wiley–Interscience, New York), pp. 265–294.
2. Blundell, T. L. & Johnson, L. N. (1976) *Protein Crystallography* (Academic, New York), pp. 183–238.
3. McLaughlin, S. (1989) *Annu. Rev. Biophys. Biophys. Chem.* **18,** 113–136.
4. Dodson, E. J., Dodson, G. G., Lewitova, A. & Sabesan, M. (1978) *J. Mol. Biol.* **125,** 387–396.
5. Gursky, O., Li, Y., Badger, J. & Caspar, D. L. D. (1992) *Biophys. J.* **61,** 604–611.
6. Badger, J., Harris, M. R., Reynolds, C. D., Evans, A. C., Dodson, E. J., Dodson, G. G. & North, A. C. T. (1991) *Acta Crystallogr. Sect. B: Struct. Crystallogr. Cryst. Chem.* **47,** 127–136.
7. Gursky, O., Badger, J., Li, Y. & Caspar, D. L. D. (1992) *Biophys. J.* **63,** 1210–1220.
8. Badger, J., Gursky, O., Li, Y. & Caspar, D. L. D. (1993) *Synchrotron Radiation in Life Sciences* (Oxford Univ. Press, Oxford), in press.
9. Rayment, I., Johnson, J. E. & Suck, D. (1977) *J. Appl. Crystallogr.* **10,** 356.
10. Kalata, K. (1985) *Methods Enzymol.* **114,** 486–510.
11. Kalata, K., Phillips, W. C., Stanton, M. & Li, Y. (1990) *Proc. Soc. Photo-Opt. Instrum. Eng.* **1345,** 270–280.
12. Pflugrath, J. W. & Messerschmidt, A. (1987) in *Computational Aspects of Protein Crystal Data Analysis: Proceedings of the Daresbury Study Weekend 23–24 January, 1987*, eds. Helliwell, J. R., Machin, P. A. & Papiz, M. Z. (SERC, Daresbury, U.K.), pp. 149–161.
13. Hendrickson, W. A. & Sheriff, S. (1987) *Acta Crystallogr. Sect. A: Found. Crystallogr.* **43,** 121–125.
14. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
15. Badger, J. (1993) *J. Mol. Graphics*, in press.
16. Dickerson, R. E., Weinzierl, J. E. & Palmer, R. A. (1968) *Acta Crystallogr. Sect. B: Struct. Crystallogr. Cryst. Chem.* **24,** 997–1003.
17. Hendrickson, W. A. (1991) *Science* **254,** 51–58.
18. Eisenman, G. & Dani, J. A. (1987) *Annu. Rev. Biophys. Biophys. Chem.* **16,** 205–226.
19. Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L. & Hellewell, R. A. (1992) *Nature (London)* **358,** 347–351.
20. Gilson, M. K. & Honig, B. H. (1987) *Nature (London)* **330,** 84–86.
21. Sternberg, M. J. E., Hayes, F. R. F., Russell, A. J., Thomas, P. G. & Fersht, A. R. (1987) *Nature (London)* **330,** 86–88.
22. Yamashita, M. M., Wesson, L., Eisenman, G. & Eisenberg, D. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 5648–5652.
23. Badger, J., Kapulsky, A., Gursky, O., Bhyravbhatla, B. & Caspar, D. L. D. (1994) *Biophys. J.*, in press.