Acta Cryst. (1988). A44, 309-320

Analysis of Solvent Structure and Hydrogen Exchange in Proteins on the Basis of Neutron Diffraction Data from Deuterated and Hydrogenous Crystals

BY ROBERT W. HARRISON* AND ALEXANDER WLODAWER*

Center for Chemical Physics, National Bureau of Standards, Gaithersburg, MD 20899, USA

AND LENNART SJÖLIN

Department of Inorganic Chemistry, Chalmers Institute of Technology, Göteborg, Sweden

(*Received* 4 *August* 1987; *accepted* 10 *December* 1987)

Abstract

A method has been developed to determine the structure of bound solvent and the positions of exchanged hydrogens in proteins, on the basis of neutron diffraction from hydrogenous and deuterated crystals. In this method phases for the hydrogenous and for the deuterated model are refined simultaneously, and an average model is imposed in the volume occupied by non-hydrogen atoms. The densities in the areas of bulk solvent are replaced by their average values, while no modifications are performed in the vicinity of ordered solvents and potentially exchangeable hydrogens. The method was tested on 1·8 Å neutron diffraction data collected from two crystals of bovine pancreatic trypsin inhibitor, one of them deuterated and the other hydrogenous. Significant improvement was observed for the densities corresponding to many partially occupied solvent sites, as well as to partially exchanged hydrogens. The algorithm presented here has been compared with a different approach published recently by Shpungin & Kossiakoff [*Methods Enzymol.*(1986), **127**, 329-342].

Introduction

Protein crystals contain large numbers of water molecules in fully or partially occupied sites, usually in close proximity to the protein surface. Water positions are always assigned during refinement of protein structures at high resolution, but their number varies considerably in different structural investigations. For example, 40 waters were assigned in a very conservative neutron study of carbon monoxymyoglobin (Hanson & Schoenborn, 1981), while 316 solvents were placed in the course of an X-ray refinement of oxymyoglobin (Phillips, 1980). Detailed comparisons of solvent positions have been performed for the same structures refined independently in different laboratories (Wlodawer, Borkakoti, Moss & Howlin, 1986), for the same proteins in different crystal forms (Wlodawer, Deisenhofer & Huber, 1987; Wlodawer, Nachman, Gilliland, Gallagher & Woodward, 1987), or for two crystallographically independent molecules in the same unit cell (Blevins & Tulinsky, 1985). All of these investigations were in agreement that while the majority of solvent positions present in these models were real, some were probably artefacts. In the absence of confirming information, positions of partially occupied waters, or of waters not forming clear hydrogen bonds with the protein, must be treated with a degree of skepticism.

It has been suggested (Schoenborn, 1969; Wlodawer, 1982; Kossiakoff, 1983) that neutron diffraction should be a method of choice for crystallographic investigations of water structure in macromolecular crystals, especially if D_2O is used to replace H_2O. This is due to the much higher relative contribution of deuterium atoms to the overall scattering, since the deuterium scattering length is almost the same as that of oxygen. In practice, however, our results of neutron refinements of solvent structure have been disappointing. Neutron data are usually of poorer quality and extend to lower resolution than the X-ray data from the same crystals (Savage & Wlodawer, 1986; Wlodawer, Walter, Huber & Sjölin, 1984). This is caused by two unfavorable properties of the neutron diffraction technique. The available fluxes of neutrons are five to six orders of magnitude lower than the fluxes of X-rays, and the incoherent scattering of unexchanged hydrogens makes the intrinsic signal-to-noise ratio much lower in neutron diffraction. The latter problem could be circumvented only by the use of completely deuterated proteins (in which even the non-exchangeable hydrogens have been substituted by deuteriums), but no such studies have been reported so far. While the increase of resolution beyond about 2Å does not necessarily improve solvent maps, high-resolution data help in obtaining better phases and thus lead indirectly to easier interpretation of Fourier maps.

Neutron diffraction has been utilized in a number of hydrogen-exchange studies of protein single

* Current address: Crystallography Laboratory, NCI - Frederick Cancer Research Facility, BRI - Basic Research Program, PO Box B, Frederick, Maryland 21701, USA.

0108-7673/88/030309-12$03.00

crystals (for reviews, see Wlodawer, 1982; Kossiakoff, 1985). The computational techniques used to assess the hydrogen or deuterium occupancies of potentially exchangeable sites ranged from the analyses of density in Fourier maps resulting from real-space refinement (Hanson & Schoenborn, 1981), to automated constrained Fourier refinement (Kossiakoff & Spencer, 1981), and to least-squares refinement of occupancies or scattering lengths (Wlodawer & Sjölin, 1982; Mason, Bentley & McIntyre, 1984; Raghavan & Schoenborn, 1984). Wlodawer & Sjölin (1984) analyzed the errors resulting from the application of least-squares refinement to 2Å neutron diffraction data for ribonuclease and concluded that the errors in the estimation of occupancy were of the order of $\pm 0{\cdot}2$ occupancy units. In particular, the average numerical values of occupancies were found to be strongly dependent on the assumed occupancies of the starting models (Wlodawer, Walter, Huber & Sjölin, 1984; Raghavan & Schoenborn, 1984). Since the use of diffraction data for both the hydrogenous and deuterated crystals increases the number of observations without significantly increasing the number of parameters to be determined (since almost all hydrogen occupancies in a hydrogenous crystal are known to be 1·0), we expected to improve the estimates of the amount of exchange by the analysis of differences between refined Fourier maps based on hydrogenous and deuterated data.

A method combining the results of X-ray and neutron diffraction in the same experiment was introduced by Wlodawer & Hendrickson (1982). In this joint refinement procedure both X-ray and neutron structure factors are utilized simultaneously in a process in which a single model is adjusted for the best fit to both sets of data. This method was found to be helpful in refining water sites in ribonuclease (Wlodawer & Sjölin, 1983) and in bovine pancreatic trypsin inhibitor (Wlodawer, Walter, Huber & Sjölin, 1984), but did not work well for insulin, where only 2·2 Å neutron data of questionable quality were available (Savage & Wlodawer, 1986; Wlodawer, Savage & Dodson, 1988). In any case, this method is expected to work only for those waters which are clearly seen in both the electron and neutron density maps and thus is not very useful in the search for marginally occupied solvent sites.

One of the primary reasons for performing neutron diffraction experiments on biological molecules is the ability of this technique to detect hydrogen and deuterium positions. Much of the scattering from hydrogen atoms comes from the solvent areas where a conventional atomic model is unworkable. This means that the phases derived from a refined atomic structure do not include direct information about the solvent, and, therefore, the calculated scattering density in the solvent region can be erroneous. The fundamental difficulty is that in the absence of an atomic model a large number of independent linear factors are refined. The problem is not sufficiently 'overdetermined' for this approach to be reliable (if the grid spacing on the map were greater than or equal to the Nyquist spacing, and all of the data were observed, then this problem would be just 'determined').

One solution to this problem is the utilization of D_2O–H_2O neutron difference maps and successive density modifications based on physically reasonable constraints. The use of D_2O–H_2O difference neutron density maps has been described by Shpungin & Kossiakoff (1986). The basic constraints used in that study were based on the assumption that the protein region in the two density maps should be the same. This constraint is effective in two ways. Obviously, there is a contribution from the scattering of the protein non-hydrogen atoms, which, within error limits, must be the same in each map. However, if this were the only effect then there would be an infinite number of solutions for the solvent components. A second effect of this constraint partially resolves this ambiguity since it also requires that the Fourier transform of the solvent component be collinear for both data sets. This limits the number of potential solutions and allows the refinement of the closure between F(obs), F(protein), and F(solvent).

In this paper we describe the use of neutron diffraction data from hydrogenous and deuterated crystals of bovine pancreatic trypsin inhibitor to search for bound solvent. The combined X-ray and neutron structure for this protein has been published earlier (Wlodawer, Walter, Huber & Sjölin, 1984). The method presented here starts from similar assumptions to that of Shpungin & Kossiakoff (1986), but utilizes different algorithms for calculations.

Data collection

Neutron difffraction data collection from a deuterated BPTI crystal was described in detail by Wlodawer, Walter, Huber & Sjölin (1984). The data set contained 4044 observed reflections [$I > 1{\cdot}5\sigma(I)$] extending to 1·8 Å resolution (78% of the theoretical maximum). Over half of the reflections in the outermost shell (2–1·8Å) were observed.

The crystal of form II of BPTI used in this work to collect data in H_2O was grown as described previously (Walter & Huber, 1983). It was soaked in a 30% solution of polyethylene glycol, $M_r = 20\,000$, for several days before use. The size of the crystal was approximately $3{\cdot}5 \times 3{\cdot}5 \times 0{\cdot}5$ mm (volume 4 mm^3) and it was mounted in a quartz tube with its shortest crystallographic axis roughly parallel to the tube axis. The crystal was completely surrounded with quartz wool and only a small amount of mother liquor was present at the end of the tube (Wlodawer, 1980).

Neutron diffraction data were collected using the flat-cone diffractometer (Prince, Wlodawer & Santoro, 1978) at the National Bureau of Standards Reactor. Neutrons were monochromatized by reflection from a 50×100 mm graphite crystal, and the chosen wavelength was 1·68 Å. Beam conditions and the methods employed to measure the reflections were essentially the same as those utilized previously to collect data from a deuterated crystal (Wlodawer, Walter, Huber & Sjölin, 1984). Data were collected in two parts. First, reflections falling on 11 levels of k (0–10) were measured using flat-cone geometry, by rotating the crystal around the shortest (b) real axis by 300° in increments of 0·09°. Each frame was counted for approximately 4 min, with the actual time depending on the accumulated counts on an incident-beam monitor. Almost every reflection was observed more than once. Reflection intensities were integrated using the dynamic mask procedure (Sjölin & Wlodawer, 1981). Symmetry R factors for each level, expressed on F's, were between 0·057 and 0·104. Reflections missing in the blind region of reciprocal space, as well as several hundred reflections needed for scaling, were measured in equatorial geometry. Each peak of width 1·5° was scanned in 64 steps, with total counting time 10 min per reflection. The equatorial data scaled to the flat-cone data with $R =$ 0·098. The total length of time spent collecting data was about 20 weeks. The final data set contained 3280 observed reflections, 63% of the theoretical maximum. Only 25% of the data in the outermost shell (2–1·8Å) were stronger than $1{\cdot}5\sigma(I)$. Smaller crystal size and worse signal-to-noise ratio were responsible for the overall poorer quality of the hydrogenous data set compared with the data from the deuterated crystal.

Computational procedures

Neutron and X-ray scattering

Neutron diffraction and X-ray diffraction can be treated in a similar manner. There is, however, one important difference between them which will manifest itself in several ways throughout this work. Neutron scattering lengths of deuterium and hydrogen are very different; in fact, the hydrogen scattering length is negative. Therefore, in theory, a subtraction of a Fourier $F_o\alpha_c$ map based on hydrogenous neutron diffraction data from an analogous map based on deuterated data ought to produce a map where hydrogens would be enhanced and the other protein atoms would vanish. A map calculated in this manner is not the same as the one resulting from a difference Fourier synthesis. In the latter case, if one considers X-ray diffraction data as an example, the differences used in the calculation, $(F_o - F_c)\alpha_c$, are small perturbations of the native scattering. In general, the native phases are good approximations to the true phases for the difference coefficients. In neutron diffraction analysis, when deuterated and hydrogenous data are being used, this is no longer the case. The differences in structure factors due to different cross sections of hydrogen and deuterium are quite large (typical scaling R factor for such data is 35%) and because of this the phases for the two native data sets are no longer near the phases for the differences. There is also a significant shift in the $F(000)$ term, which is simply the sum of the scattering lengths in the unit cell.

This effect is very important when determining a scale factor between the two data sets. It is reasonable to assume that the scattering density at non-exchanged atoms should be the same in both hydrogenous (H) and deuterated (D) crystal data sets. If this were an X-ray diffraction experiment, a multiplicative scale factor of the form

$$\rho_H = A\rho_D \tag{1}$$

would be correct. With neutron diffraction, because of the difference in the $F(000)$ term, this would lead to systematic errors. Instead, a function of the form

$$\rho_H = A\rho_D + B \tag{2}$$

must be used.

Similarly, the difference in the $F(000)$ has another effect in the calculation of the difference maps. It is usual practice to calculate a standard deviation of a difference map and to examine peaks or holes in the map that are larger than several standard deviations. If the $F(000)$ term is left out [as is required by several popular Fourier programs, such as the one in the *PROTEIN* package of Steigemann (1974)], then while the standard deviation is correctly estimated it is estimated around a mean value lower than zero. (Strictly speaking the mean must be zero, but the density will be skewed as though the mean were non-zero.) If the standard deviation is used relative to a mean value of zero to select a contour level (as is the case with X-ray diffraction), the density will appear as though it were systematically underestimated. This means that significant and interesting features of the map will be ignored as they will appear to be too small.

Babinet's principle

It is common practice when refining crystallographic coordinates without the removal of very-low-resolution data (7–10 Å and lower) to include a correction for the solvent scattering. Often 'Babinet's principle' is invoked to justify such a procedure (Fraser, MacRue & Suzuki, 1978). In its simplest form this principle states that the phase of the scattering from the solvent should be nearly opposite to the phase of the scattering of the molecule. In practice

this principle is used as justification for the refinement of a special scale factor for low-resolution reflections (Chambers & Stroud, 1977). Since we are interested in refining the solvent scattering density we cannot invoke this principle in its simplest form. With general reflections the statement that the phase of the solvent scattering is approximately opposite to the molecular scattering is not sufficient. Further analysis is necessary.

The map can be divided into two non-overlapping regions - the solvent and the molecule. We define the solvent density as a function that is equal to the scattering density in the solvent region and zero everywhere else, and the molecular density is defined in a similar manner. The integral of the product of these two functions over the unit cell is clearly equal to zero. If the solvent density and the molecular density are represented as Fourier series then this integral can be used to find the phase relationships between the scattering from the molecule and the solvent.

If one starts with the assumption that the solvent and molecular regions do not overlap, one may write

$$\int dV \rho_s^* \rho_m = 0, \tag{3}$$

where ρ_s^* = complex conjugate of solvent density, and ρ_m = molecule density. ρ_s^* is being used primarily as a notational convenience for the next step. ρ_s represents the real value of the density and as such $\rho_s^* = \rho_s$. As the densities ρ_s and ρ_m are defined as the result of a Fourier transform, the substitution of that transform into this integral gives the relationship between the Fourier coefficients of ρ_s and ρ_m:

$$\begin{aligned} \rho_s^* &= \textstyle\sum_-^+ f_s^* \exp(-2\pi s), \\ \rho_m &= \textstyle\sum_-^+ f_m \exp(+2\pi s), \end{aligned} \tag{4}$$

where f_s and f_m are the scattering factors of the solvent and of the molecular densities, respectively. Note that ρ_s^* is defined as a series in which the coefficients are replaced by their Friedel mates.

Placing these series into the integral (3) yields

$$\int dV \rho_s^* \rho_m = V \textstyle\sum_-^+ f_m = 0 \tag{5}$$

or

$$\begin{aligned} f_{s000} f_{m000} &= -\textstyle\sum_1^+ (f_s^* f_m + f_s f_m^*) \\ &= -\textstyle\sum_1^+ \{2|f_s||f_m|\cos[\Delta(\varphi)]\}. \end{aligned}$$

Clearly, f_s^* and f_m are linearly related. For any given reflection the phase of the solvent component is determined by the slope of the line $f_s f_m^*$. This slope is just the tangent of the phase of f_m^* and is independent of the contents of the solvent region.

What has been derived so far is not sufficient to determine any solutions for the value of the solvent scattering. An auxiliary condition is needed. The simplest one is that the magnitude of the sum of the solvent and the molecular scattering be the same as for the observed data,

$$|f_s + f_m| = |F_{obs}|. \tag{6}$$

A geometric construction can be drawn to show the nature of the solutions to the solvent scattering. This construction is shown in Fig. 1. A line through the origin with slope defined by the tangent of the phase of f_m^* gives all the possible values for the solvent scattering. The auxiliary condition is applied by drawing a circle of radius F_{obs} and a center at $-f_m$. The intersections of the line and the circle are then the allowed values for the solvent scattering. In general there are no, one or two solutions. For non-centrosymmetric reflections there will be two solutions, if any are present.

Babinet's principle does not suffice

Babinet's principle does not provide a unique solution if no other conditions are applied to the solvent density. With only one observation per reflection, as is the case with X-ray diffraction, the only reasonable condition is that the solvent be relatively flat. Since solvent flattening became a standard tool of macromolecular crystallography, this result has some important consequences. In solvent flattening it is assumed that within some region (the 'solvent') there is little or no variation in density. By iteratively forcing the condition of featureless density in this region it is hoped that the phases of the reflections will improve. The analysis in the last section shows that the phases derived *via* this method are critically dependent on the molecular transform, which in turn is dependent on the choice of the boundary between the solvent and the molecule. It is therefore essential that an accurate mask be determined when using a

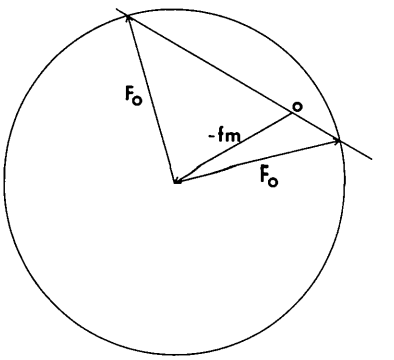

Fig. 1. The closure used to determine the solvent components. The invariant molecular component f_m was chosen to have a magnitude of 2 arbitrary units and a phase of 30°. The circle with radius F_o has a radius of 2·5 units. The origin is marked by ○. The two possible solutions for the phase of F_o are shown.

solvent-flattening method, such as the one developed by Wang (1985).

Even with two observations there is still no unique solution. The reason for this is that while the number of observations is doubled, the number of possible solutions is doubled as well. However, if a relationship exists between either the solvent densities or the molecular densities under the conditions of both observations, then a stronger external condition can be applied to the solution. For example, in the simplest case, where the solvent contains no features other than a constant density, then the transforms of the solvent regions from both the D and H maps are collinear and the degrees of freedom are reduced, so that a unique solution is possible.

Unfortunately, such a simple model does not suffice, because the interesting features of the solvent come from ordered water molecules which contain oxygen as well as D or H. So, while a simple flat solvent model for the bulk regions is extremely useful, further constraints on the phases are necessary. One such constraint is that the solvent density in the H map must always be less than or equal to that in the D map. This is only approximate, since there could be differences in the ordering of the solvent.

Implementation of the algorithm

The external conditions mentioned in the last section have been used to formulate the algorithm. A set of computer programs was written to interface with the *SIRSAS* program system of Wang (1985). As the operations performed on a density grid point depend on whether it is assigned as solvent or protein, the manner in which these assignments were made is critical and will be described in detail. The assignment is done in three passes. In the first pass, all the points around every protein atom (including exchangeable hydrogens) are flagged. In the next pass, the remaining unconnected regions that are too small to contain a water molecule are marked. Finally, the points around any exchangeable site are removed from the masked volume. The van der Waals radii for the atoms were used in the first pass. The second pass was accomplished by moving a sphere the size of a water molecule (1·8 Å diameter) through all points near protein atoms and counting the number of marked grid points. If more than half of the volume of this sphere was already marked then the center of the sphere was added to the marked list (a different flag was used for points marked in this manner and for those near protein atoms, otherwise the approach would have been unstable). Then a 2 Å sphere centered about every exchangeable site was removed from the marked list.

The manner in which the maps are scaled is also critical. An incorrect scale will prevent proper convergence of phase refinement. The scale factor was determined by examining points within 0·8 Å of C_α atoms. This distance was chosen for a 0·5 Å grid as it is slightly larger than the diagonal distance between grid points. Both a simple multiplicative scale ($\sum H/\sum D$) and a linear scale were examined. Better convergence and more-interpretable maps resulted when the linear scale was used. The scaling coefficients A and B of (2) were determined by least-squares refinement.

A density modification cycle then proceeded in the following manner. First the initial maps were calculated with the phases which resulted from separate least-squares refinement of the hydrogenous and deuterated data. Then the scale factor between the two sets of data was determined. Those points within the non-exchanged protein were scaled, averaged, and then restored to the original scales. The original scale was restored in order both to avoid a discontinuity and to maintain the proper scale with respect to the reflections. The density near the protein was then forced to have $\rho_D \geq \rho_H$. The density located farther than 4Å from the protein was averaged independently in each map and was replaced by the average value. The modified maps were then Fourier transformed and the phases were updated.

Two further points require clarification. In Wang's (1985) system of programs the phases are updated by phase combination. Such phase combination is unnecessary with the neutron data, and ties the phases to their initial values. The phase combination was disabled by generating Lattman coefficients (Hendrickson & Lattman, 1970) that were small compared with their correct values. These coefficients are normally proportional to $F_oF_c/\sigma(\text{error})$. Phase combination could have been eliminated by using zero probability instead, but an artificially high standard deviation of the error was used to avoid any problems with having all of the coefficients zero for all of the reflections. In addition, it is very important that the two sets of reflections used in this refinement have no members that only exist in one set. Allowing, for example, the D set to contain reflections which are not present in the H set causes generation of a very noisy and incorrect map. The phases are shifted to try to minimize a real difference between the maps.

Results

Phase refinement and calculation of maps

Two independently refined structures of BPTI were used to calculate the initial maps. One of them was the result of a joint X-ray/neutron study of Wlodawer, Walter, Huber & Sjölin (1984), utilizing neutron data from a deuterated crystal. Crystallographic R factors for that model were 0·200 (X-ray) and 0·197 (neutron). This structure, after renaming all deuterium atoms as hydrogens, was used to initiate

Table 1. *Progress of phase refinement for hydrogenous and deuterated BPTI*

The following values are shown for each refinement cycle:
R_H: crystallographic R factor for the hydrogenous model.
R_D: R factor for the deuterated model.
σ: standard deviation of the difference between the hydrogenous and deuterated maps in the invariant molecular region.
Scale: the scale factor between the two maps [A in equation (2)].
dF(000): the estimated difference in F(000) between the two maps [B in equation (2)].

Cycle	R_H	R_D	σ	Scale	dF(000)
1	0·191	0·172	22·3	0·938	2·74
2	0·180	0·148	19·2	0·966	−4·95
3	0·177	0·148	17·9	0·962	−5·64
4	0·171	0·143	17·5	0·965	−6·67
5	0·165	0·139	17·0	0·967	−7·28
6	0·160	0·135	16·7	0·967	−7·88
7	0·155	0·132	16·4	0·968	−8·24
8	0·151	0·129	16·1	0·968	−8·58
9	0·148	0·126	15·8	0·969	−8·95
10	0·145	0·123	15·6	0·968	−9·17
11	0·142	0·120	15·5	0·970	−9·53
12	0·140	0·117	15·3	0·969	−9·65
13	0·139	0·117	15·6	0·981	−9·40
14	0·136	0·114	15·1	0·976	−9·75
15	0·134	0·113	14·9	0·976	−9·82
16	0·132	0·111	14·8	0·977	−10·12
17	0·131	0·110	14·6	0·977	−10·13
18	0·129	0·108	14·6	0·977	−10·32
19	0·127	0·107	14·4	0·976	−10·29
20	0·125	0·105	14·3	0·977	−10·52

12 refinement cycles with the same X-ray data, but with neutron data collected from the hydrogenous crystal. Final R factors for the resulting model were 0·199 and 0·211 for the X-ray and neutron data, respectively. For map calculations, the initial phases were derived from these two models, using only potentially non-exchangeable protein atoms in the calculation. These phases were entered into the *SIRSAS* system and were given Lattman coefficients which were nearly zero. Earlier runs where the correct Lattman coefficients were used restrained the phases to the initial input values too strongly. Lowering the magnitude of the Lattman coefficients releases this restraint.

Initially, when the maps were calculated the standard deviation of the difference between the density around the non-exchanged atoms was 22·3 units (arbitrary scale). The scale factor was 0·938 and a 2·78 unit dF(000) term was added. The R factor for map inversion (Wang, 1985) was 0·191 for the hydrogenous data and was 0·172 for the deuterated data. After 20 cycles of density modification the R factors were lowered to 0·125 and 0·105 respectively. The total phase shifts were approximately 17° for each phase set. The standard deviation of the non-exchanged data was reduced to 14·33 units. The scale factor and dF(000) term had changed to 0·977 and −10·52. Refinement was stopped at 20 cycles as it appeared that all of the derived measures had converged. The process of convergence is shown in Table 1.

Three types of maps were examined to see whether the process had improved the density. The first map was calculated using the hydrogenous data with their phases and the second utilized the deuterated data with their phases. We calculated an 'exchange map', using as coefficients the scaled complex differences between the D and H data. These coefficients must be properly scaled to make a meaningful difference Fourier map. This scaling was accomplished in the same manner as for the maps which were scaled for density modification. However, while the dF(000) term [B in (2)] is applied in scaling maps, only the multiplicative term is used when scaling structure amplitudes. Then when the difference maps are displayed, the contour levels are adjusted to reflect the difference in F(000) terms. This approach is necessitated by the limitation in the Fourier program in the package *PROTEIN* (Steigemann, 1974), which does not use F(000) in the calculations. The maps were carefully examined by computer graphics. We noticed some improvement in the areas corresponding to the protein, but we did not attempt to apply modifications to the areas which were not potentially exchangeable. The observations below are based on visual analyses of the maps calculated as described above, the exchange map in particular.

Hydrogen exchange

Information about hydrogen exchange is not available in the X-ray structure determination, but it becomes accessible during analysis of neutron diffraction data. Neutron diffraction has been used successfully to establish exchange patterns for several proteins (Wlodawer & Sjölin, 1982, 1984; Kossiakoff, 1982; Bentley, Delepierre, Dobson, Mason, Poulsen & Wendin, 1983; Mason, Bentley & McIntyre, 1984; Raghavan & Schoenborn, 1984; Teeter & Kossiakoff, 1984). The comparison of the protection from exchange of individual main-chain hydrogens with the rates established by NMR has been published by us for BPTI (Wlodawer, Walter, Huber & Sjölin, 1984) and for lysozyme by Bentley *et al.* (1983). These investigations showed unambiguously that β sheets provide the highest degree of protection, followed by α helices. Our previous investigations of exchange utilized an algorithm in which occupancies of all atoms in the protein were refined. The errors involved in such a procedure are significant (Wlodawer & Sjölin, 1984) and are very much influenced by the uncertainties of the protein model.

Exchangeable protons are found on both the main chain and the side chains of a protein. The latter are usually completely exchanged under the conditions of a neutron diffraction experiment, but the exact placement of the deuterium atoms is sometimes problematic. Thus, during the analysis of maps resulting from this work, we paid attention to the exchange

status of both the side-chain and the main-chain protons.

Side chains. While some side chains have a similar chemistry to the peptide amides (Asn, Gln), the structural considerations of the environment are different. The peptide amides occur throughout the protein, in both hydrophobic and hydrophilic environments, while the polar side chains occur almost exclusively on the surface of the protein. Because of this structural difference the interesting questions about the two classes of exchangeable groups are somewhat different. With peptide amides the major question is whether the group is fully exchanged with solvent and what role the neighboring residues play in this process, while with side chains the question is not the exchange [which is essentially universal, with HE1 of Trp 28 in lysozyme being the only side-chain hydrogen shown by neutron diffraction to be protected after long exchange (Mason, Bentley & McIntyre, 1984)], but rather how this exchange can be used to further our understanding of the interaction of the side chain with the solvent which surrounds the protein.

Tyrosine residues. Four tyrosine residues are present in BPTI. They are all well ordered and are useful for checking the results of the method. While it is possible to explain poor density for the exchangeable protons on a long and flexible amino acid such as a lysine as a result of poor ordering of the side chain, no such explanation is possible with the tyrosines.

Initial examination of the map showed that three of the four tyrosines had no density in the position occupied by the hydroxyl proton (deuteron) of the model. However, in each of these cases there was density nearby, and it was clear that the hydrogen had simply been misplaced. In the one case where the hydrogen resided in density (Tyr 35) there was only one reasonable choice of a hydrogen bond (to the sulfate ion) and that is where the atom was found. In the other cases there were several choices for the hydrogen bonding and presumably the wrong one was taken. Typically, a water was added to the model in the location near where the proton should be and the geometry of the hydroxyl was strained.

Tyr 10 shows this problem clearly (Fig. 2). In the original conformation the hydroxyl is pointed at the symmetry equivalent of water 157. The distance is rather long for a hydrogen bond. Rotating the hydroxyl to point at Wat 203 both fits the density and results in a better hydrogen bond. Wat 157 can be rearranged as a hydrogen-bond donor to the hydroxyl oxygen of the tyrosine.

Basic residues. The arginine and lysine residues show differing amounts of density in the exchange map. The critical factor with these kinds of residues is not whether there is exchange, but rather whether the residue is well ordered or not.

The lysine residues are all disordered to a certain extent and very little density near the terminal amine is visible. With Lys 26, for example, there is no density for the side chain in either the H map or the D map and, not surprisingly, there is no density in the exchange map.

The side chains of arginine residues, on the other hand, very often make good hydrogen bonds to other groups in the protein (for example the amides of glutamines or asparagines), and so are well ordered. In these cases the exchange density is very clear, as shown for Arg 20 in Fig. 3. This residue is near the sulfate and makes hydrogen bonds to Asp 44 and Wat 127. A nearby water molecule appears to have been misplaced in the original refinement.

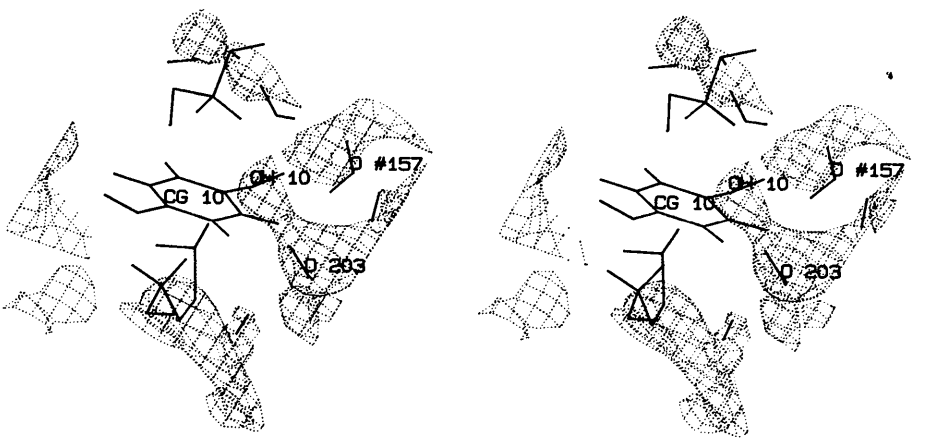

Fig. 2. The exchange density in the vicinity of Tyr 10. The coordinates are those of the refined model of Wlodawer, Walter, Huber & Sjölin (1984). Rotation of the hydroxyl atom down towards Wat 203 will result in a better fit to the density and shows less strain. This map, as well as all of the following ones, is contoured at the 1σ level.

Neutral polar residues. These residues are almost all on the surface but, unlike the basic residues, tend to be well ordered. The amide residues, with the one exception of Asn 43, tend to make hydrogen bonds with other side chains of the protein, while the hydroxyls form hydrogen bonds exclusively with waters or with main-chain carbonyl groups.

Asn 43 is particularly interesting (Fig. 4). It is located within the interior of the protein and shows anomalous exchange behavior. Unlike what was reported in the refinement of Wlodawer, Walter, Huber & Sjölin (1984), current analysis of the H and D maps indicates only a little exchange for the side-chain amide hydrogen of this residue. What is interesting is that it can hydrogen bond to the peptide amide of Tyr 23 which is one of the several amides that did not exchange under the conditions of the experiment (Wlodawer, Walter, Huber & Sjölin, 1984). It is possible that this lack of exchange occurs for similar reasons.

Thr 54 shows a typical result for the vicinity of side-chain hydroxyls. Its side chain hydrogen bonds to the main-chain carbonyl of Asp 50, which helps to stabilize an α helix, and it has two waters, 310 and 311, which may hydrogen bond to it (Fig. 5).

Backbone amides. Wlodawer, Walter, Huber & Sjölin (1984) refined the occupancies of peptide amide hydrogens in an attempt to quantify the hydrogen exchange with bulk solvent. It was found that 11 of the internal amides showed very little exchange, in good agreement with NMR results of Wagner & Wüthrich (1982). Since our current approach does not produce atomic occupancies, the results obtained here are not directly comparable to those reported previously. There is a general correlation, though, since the exchange density is weak where the deuterium occupancy was expected to be low and the exchange density was strong where the occupancy data predict exchange. Complete lack of exchange is

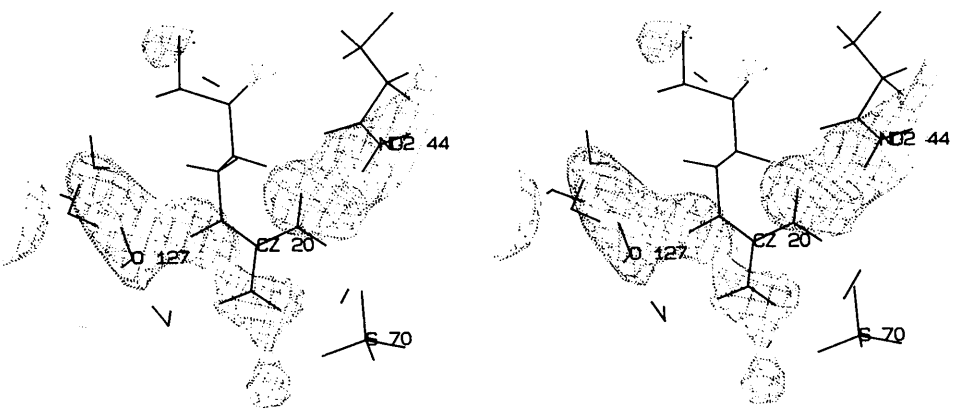

Fig. 3. The exchange density near Arg 20. The exchanged hydrogens are clearly indicated in this map.

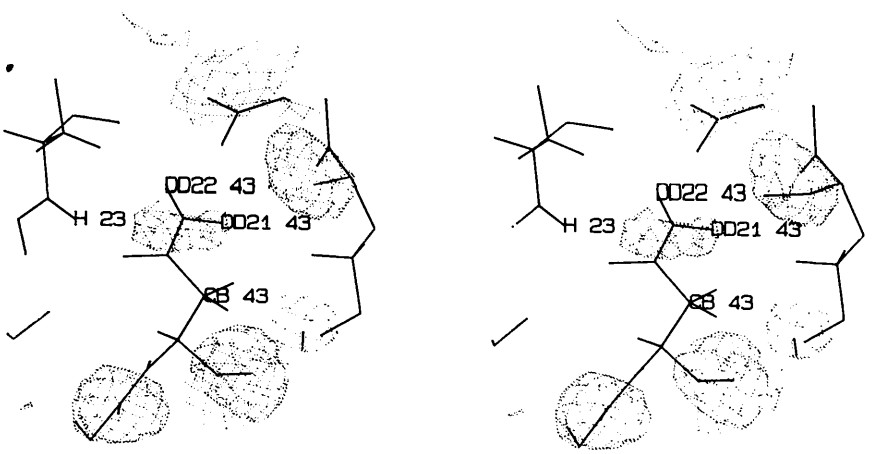

Fig. 4. Asn 43, shown here, does not seem to show full exchange. The amide of Tyr 23 also shows no exchange. The exchange density at the bottom of residues 43 and 44 is typical of fully exchanged amides.

strongly indicated only for two well ordered amides (14 and 23). The C-terminal amides (55–58) show a lack of exchange density, but this is probably an effect of disorder.

Water structure

Water molecules in BPTI can be loosely divided into two classes. A fraction of water molecules (about a fourth) are very tightly bound to the protein and are well ordered; these waters occur in the same positions in all three BPTI crystal structures which have been compared (Wlodawer, Deisenhofer & Huber, 1987; Wlodawer, Nachman, Gilliland, Gallagher & Woodward, 1987). The remaining waters, present mostly on the protein surface, are not as well ordered.

Tightly bound water molecules. The tightly bound water molecules show excellent exchange density and are well defined. Wat 113 is typical of the internal waters (Fig. 6). It interacts with waters 111 and 112, which are also tightly bound and well ordered, as well as with the side chain of Asn 44. Removing this water would break the hydrogen-bonding pattern and clearly is unfavorable.

Loosely bound water molecules. The loosely bound waters more typically occur on the surface and make fewer ordered hydrogen-bonded interactions with either the protein or with other waters. About half (29) of the waters identified in the published model of form II of BPTI (Wlodawer, Walter, Huber & Sjölin, 1984) are in discrete density in the (D–H)

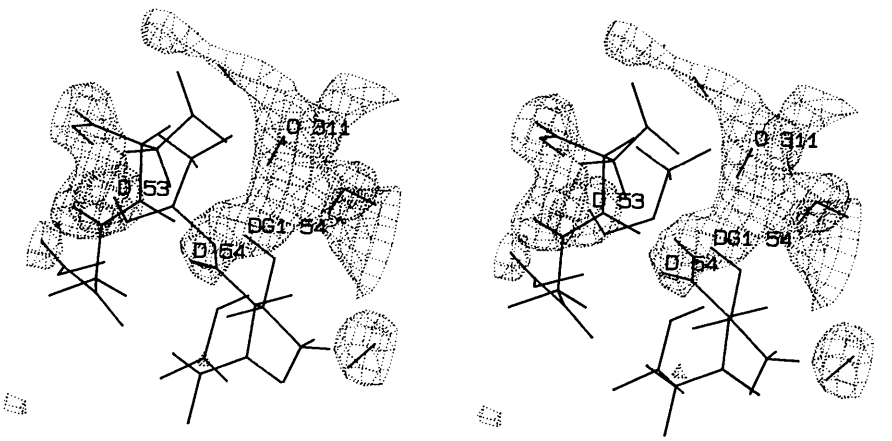

Fig. 5. Thr 54 is typical of neutral polar residues. In this case there is evidence for a hydrogen bond to its own main-chain amide and to waters 311 and 310.

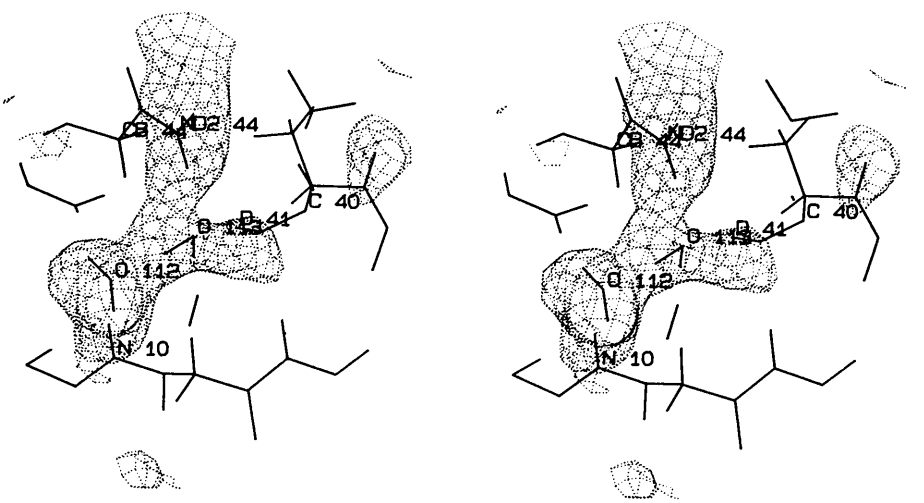

Fig. 6. Wat 113, shown here, is one of the several water molecules whose positions are conserved in the three BPTI structures solved so far (Wlodawer, Nachman, Gilliland, Gallagher & Woodward, 1987). Another conserved water, Wat 112, is also shown. The density is strong and shows components from the hydrogen-bonding groups which help to hold them in place.

map, and the remainder are either in bands of density (14) or show no density at all (14). Those waters found in discrete densities are usually near well ordered hydrogen-bonding side chains of the protein. Bands of density, which are more difficult to interpret, tend to occur in regions where there might be a unique hydrogen-bonding network. Approximately half of the waters with discrete density have neighboring peaks or bands of density, suggesting that more solvent waters could be added to the model.

Wat 138 is typical of a water in discrete density (Fig. 7). It has a well defined hydrogen bond to the peptide amide of Lys 46, and there are several peaks nearby which represent the locations of additional waters. Some of them could bridge and form a network of connections with the symmetry mate of Arg 39. Wat 219 is also quite typical (Fig. 8). This water is bound to the side chains of Gln 31 and Asn 24. Again, there are discrete peaks nearby which could be due to additional waters, not accounted for in the previous models.

Discussion

The aim of the work described in this paper was twofold. On the one hand, we attempted to develop a new algorithm for processing of the neutron diffraction data collected from both hydrogenous and deuterated crystals, and on the other hand, we analyzed the results of such an investigation for a test system, which was bovine pancreatic tryspin inhibitor. The technique described here enabled us to improve the estimates of the extent of hydrogen exchange in BPTI, and also provided a tool for locating water molecules. By combining the neutron diffraction data on the two types of crystals with density modification, we were

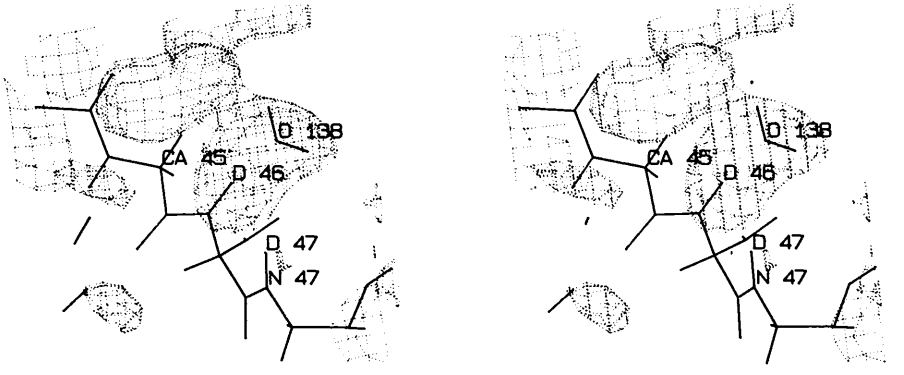

Fig. 7. This water molecule (Wat 138) shows density typical of a well ordered water. There is density for several neighboring water molecules which have not been placed in the model.

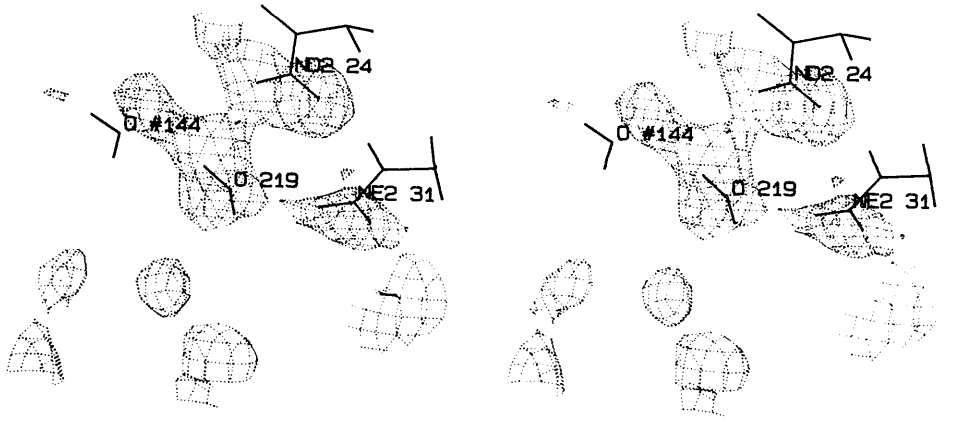

Fig. 8. The density shown in this figure is typical of a well ordered water molecule. Wat 219 is located in density and there is density nearby for several other waters. Wat 144 is the symmetry equivalent for Wat 144 and appears to be slightly misplaced in the model.

able to correct a number of assignments of exchangeable protons which were previously made on the basis of least-squares refinement of atomic occupancies using only neutron data from deuterated crystals (Wlodawer, Walter, Huber & Sjölin, 1984). Wlodawer (1982) and Kossiakoff (1982) have discussed the application of the results of hydrogen exchange as a starting point in the studies of protein dynamics. Such studies have proved so far to be elusive, since owing to the static nature of exchange results and of the relatively low level of confidence in the assignments of individual protons, detailed analysis of the breathing motions has not been done yet. Mason, Bentley & McIntyre (1984) suggested an experimental approach to collecting neutron diffraction data at low temperature from a protein crystal at various stages of hydrogen exchange, in order to follow the true rates of exchange of individual protons. A density-modification procedure of the type described here could be very useful in the interpretation of the results of such an experiment. Such results could also be combined with the analysis of both isotropic and anisotropic temperature factors, since the two types of phenomena are closely related. Taken together, it should be possible to arrive at more precise models of the predominant motions of the structural elements in a protein molecule.

While the use of neutron diffraction data from the hydrogenous and deuterated crystals provided us with a good experimental tool for the studies of the structure of bound solvent in BPTI, we have not seen any unusual waters which would not be otherwise seen in the models based on X-ray diffraction data. We have, nevertheless, been able to locate a number of water molecules in the second hydration layer which would not be easily interpretable as such in the X-ray maps. We have not found any clusters of waters around the hydrophobic residues, like those observed by Teeter (1984) in crambin, but this may be due to true differences between these two proteins, one of which is hydrophobic, and the other hydrophilic. One problem is that the water structure may not be identical in different crystals. While our approach would enhance only those waters which are similar in both crystals, the less-ordered waters may tend to be the most different.

Comparison with another published algorithm

Shpungin & Kossiakoff (1986) have published an algorithm utilizing neutron diffraction data which they have used to analyze the water structure of trypsin. That algorithm and the one described here are very similar in spirit, but differ in the details of implementation. Both algorithms are based on the assumption that some regions of the scattering density should remain the same in hydrogenated and deuterated crystals.

The approach of Shpungin & Kossiakoff (1986) is to refine a model of the difference between H and D maps against the modulus of F_d. F_h is fixed and the exchange and solvent components are added to it [*i.e.* $F_d = F_h + f_{\text{solvent}} + f_{\text{exchange}}$ and minimize $(|F_d| - |F_{d,\text{calc}}|)^2$]. It is assumed in this approach that the estimate of the non-exchanged density from the H case is correct. With this approach a model for the difference in density is refined.

Our approach uses the fact that we have two estimates for the density of the non-exchanged density, contained in both the D and H maps. In each case there are some errors, most of which will be uncorrelated. The appropriately scaled average of the two maps will give a better estimate of the non-exchanged density than either map alone will. Babinet's principle assures us of the ability to use this information and the moduli of the observed reflections to determine the phases reliably for both the D and H cases. Two closures, the closure of the non-exchanged case with F_d and the closure of the exchanged case with F_h are refined [*i.e.* $F_d = f_{\text{invariant}} + f^d_{\text{solvent}} + f_{\text{exchange}}$; $F_h = f_{\text{invariant}} + f^h_{\text{solvent}} - f_{\text{exchange}}$ and the functionals $(|F_d| - |F_{d,\text{calc}}|)^2$ and $(|F_h| - |F_{h,\text{calc}}|)^2$ are minimized]. Here, two models, one for H and one for D, are refined.

The other major difference between the approaches involves the scaling. Shpungin & Kossiakoff use a simple multiplicative scale factor, which is adequate with difference Fourier approaches, while we use a combination of multiplicative and additive scales, which is more correct and is necessary for our approach.

The approaches also differ in their sensitivity to ripples induced by introducing discontinuities in the map in the process of density modification. (Induced ripple is related to but different from series-termination ripple. For example, induced ripple can occur when series-termination ripple is truncated.) It is impossible to avoid introducing a discontinuity either in the density or in its first derivative when modifying the density. Either case will induce a ripple in a finite-resolution Fourier transform (and as in the Gibbs effect may introduce ripple into an infinite-resolution Fourier transform). Since this effect is a numerical one it is possible and appropriate to discuss the conditioning of the algorithm with respect to induced ripple. The magnitude of the ripple is bounded by a small constant times the magnitude of the discontinuity (Tolstov, 1976). Clearly if the density and its Fourier transform are large with respect to the magnitude of the discontinuity then the induced ripple will be small and the density-modification operation will be well conditioned with respect to it. This is the case when the operations are performed on the complete density. When the operations are performed on difference density this condition is not met and care must be taken to avoid inducing spurious effects.

We are indebted to Dr A. Kossiakoff for providing us with unpublished neutron diffraction data for hydrogenous and deuterated trypsin, as well as for fruitful discussions. Certain commercial equipment, instruments and materials are identified in this paper in order to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

BENTLEY, G. A., DELEPIERRE, M., DOBSON, C. M., MASON, S. A., POULSEN, F. M. & WENDIN, R. E. (1983). *J. Mol. Biol.* **170**, 243-247.

BLEVINS, R. A. & TULINSKY, A. (1985). *J. Biol. Chem.* **260**, 8865-8872.

CHAMBERS, J. L. & STROUD, R. M. (1977). *Acta Cryst.* B**33**, 1824-1827.

FRASER, R. D. B., MACRUE, T. P. & SUZUKI, E. (1978). *J. Appl. Cryst.* **11**, 693-694.

HANSON, J. C. & SCHOENBORN, B. P. (1981). *J. Mol. Biol.* **153**, 117-146.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136-143.

KOSSIAKOFF, A. A. (1982). *Nature (London)*, **296**, 713-721.

KOSSIAKOFF, A. A. (1983). *Annu. Rev. Biophys. Bioeng.* **12**, 159-182.

KOSSIAKOFF, A. A. (1985). *Annu. Rev. Biochem.* **54**, 1195-1227.

KOSSIAKOFF, A. A. & SPENCER, S. A. (1981). *Biochemistry*, **20**, 6462-6474.

MASON, S. A., BENTLEY, G. A. & MCINTYRE, G. J. (1984). In *Neutrons in Biology*, edited by B. SCHOENBORN, pp. 323-334. New York: Plenum Press.

PHILLIPS, S. E. V. (1980). *J. Mol. Biol.* **142**, 531-554.

PRINCE, E., WLODAWER, A. & SANTORO, A. (1978). *J. Appl. Cryst.* **11**, 173-178.

RAGHAVAN, N. V. & SCHOENBORN, B. P. (1984). In *Neutrons in Biology*, edited by B. SCHOENBORN, pp. 247-259. New York: Plenum Press.

SAVAGE, H. & WLODAWER, A. (1986). *Methods Enzymol.* **127**, 162-183.

SCHOENBORN, B. P. (1969). *Nature (London)*, **224**, 143-146.

SHPUNGIN, J. & KOSSIAKOFF, A. A. (1986). *Methods Enzymol.* **127**, 329-342.

SJÖLIN, L. & WLODAWER, A. (1981). *Acta Cryst.* A**37**, 594-604.

STEIGEMANN, W. (1974). PhD Thesis. Technische Univ. München, Federal Republic of Germany.

TEETER, M. M. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6014-6018.

TEETER, M. M. & KOSSIAKOFF, A. A. (1984). In *Neutrons in Biology*, edited by B. SCHOENBORN, pp. 335-348. New York: Plenum Press.

TOLSTOV, G. P. (1976). *Fourier Series*. New York: Dover.

WAGNER, G. & WÜTHRICH, K. (1982). *J. Mol. Biol.* **160**, 343-361.

WALTER, J. & HUBER, R. (1983). *J. Mol. Biol.* **167**, 911-917.

WANG, B. C. (1985). *Methods Enzymol.* **115**, 90-112.

WLODAWER, A. (1980). *Acta Cryst.* B**36**, 1826-1831.

WLODAWER, A. (1982). *Prog. Biophys. Mol. Biol.* **40**, 115-159.

WLODAWER, A., BORKAKOTI, N., MOSS, D. S. & HOWLIN, B. (1986). *Acta Cryst.* B**42**, 379-387.

WLODAWER, A., DEISENHOFER, J. & HUBER, R. (1987). *J. Mol. Biol.* **193**, 145-156.

WLODAWER, A. & HENDRICKSON, W.A. (1982). *Acta Cryst.* A**38**, 239-247.

WLODAWER, A., NACHMAN, J., GILLILAND, G., GALLAGHER, W. & WOODWARD, C. (1987). *J. Mol. Biol.* **198**, 469-480.

WLODAWER, A., SAVAGE, H. & DODSON, G. (1988). *Proteins: Struct. Funct. Genetics.* Submitted.

WLODAWER, A. & SJÖLIN, L. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 1418-1422.

WLODAWER, A. & SJÖLIN, L. (1983). *Biochemistry*, **22**, 2720-2728.

WLODAWER, A. & SJÖLIN, L. (1984). In *Neutrons in Biology*, edited by B. SCHOENBORN, pp. 349-364. New York: Plenum Press.

WLODAWER, A., WALTER, J., HUBER, R. & SJÖLIN, L. (1984). *J. Mol. Biol.* **180**, 301-331.

Acta Cryst. (1988). A**44**, 320-326

Anisotropic Bond Polarizabilities in Birefringent Crystals

BY HANS-J. WEBER

Institut für Physik, Universität Dortmund, 4600 *Dortmund, Federal Republic of Germany*

(*Received* 4 *August* 1987; *accepted* 16 *December* 1987)

Abstract

Based on the concept of bond polarizabilities the optical properties of a crystal are traced back to the orientation of its bonds and to the individual bond parameters $\bar{\beta}$ and $\Delta\beta$. A test with various sulfate compounds shows that the average bond parameters $\bar{\beta}$ describe the average refractive indices with a higher accuracy than the traditional model of ionic polarizabilities. The anisotropic part of a bond polarizability, $\Delta\beta$, is considered to be the origin of optical birefringence. Values for $\Delta\beta$ are evaluated in MCO_3 and MSO_4 (M = Mg, Ca, Sr, Ba). In different crystals the same values for $\Delta\beta$ have been discovered for the C–O and the S–O bonds. $\Delta\beta$ of the M–O bonds depends in a characteristic way on the bond length. A simple method is used to calculate local field effects in accordance with the model of bond

0108-7673/88/030320-07$03.00