Brookhaven Symp. Biol 32
BNL, NEUTRONS IN BIOLOGY, B. Schoenborn, Plenum Press NY
June 1-4, 1982

BIO 4083
BNL - 34680
Conf Art
Bk Chpt

CONF-8206240--9

BNL--34680

DE84 012914

THE NEUTRON STRUCTURE OF THE HYDROPHOBIC PLANT PROTEIN CRAMBIN

Martha M. Teeter[a] and Anthony A. Kossiakoff[b]

[a]Dept. of Chemistry, Boston University, Boston, MA 02215

[b]Biology Dept., Brookhaven National Laboratory
Upton, NY 11973

INTRODUCTION

Crystals of the small hydrophobic protein crambin have been shown to diffract to a resolution of at least 0.88 Å (11). This means that crambin presents a rare opportunity to study a protein structure at virtually atomic resolution. The high resolution of the diffraction pattern coupled with the assets of neutron diffraction present the distinct possibility that crambin's analysis may surpass that of any other protein system in degree and accuracy of detail.

The crambin molecule has 46 amino acid residues and has a sequence which shows strong homology to a class of hydrophobic plant toxins (12). This protein has remarkable temperature stability (7) and has been shown by x-ray crystallographic studies to have a large number of intramolecular backbone-backbone and backbone-side-chain hydrogen bonds (3). The main-chain conformation of the molecule is shown in Figure 1. A packing analysis shows

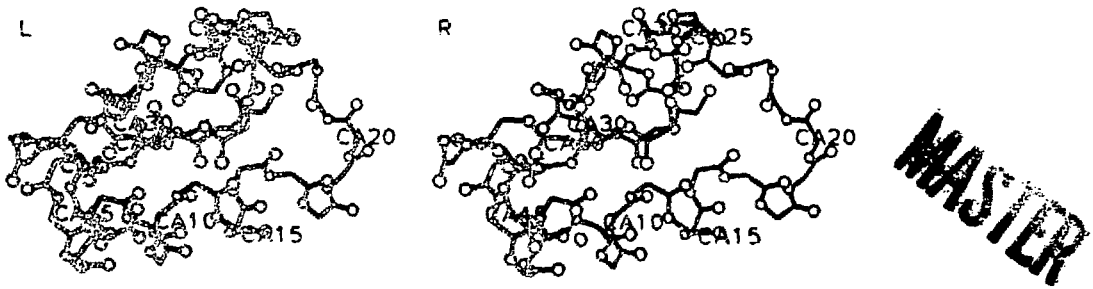

Fig. 1. Stereo pair of the main-chain conformation of crambin.

*Research carried out under the auspices of the U.S.D.O.E.

DISCLAIMER

that the hydrophobic surface of the protein is shielded from crystal solvent by the interaction with the hydrophobic surface of a symmetry-related molecule. There are also three direct hydrogen bonds and one salt bridge formed between adjacent molecules in the crystal. The close intermolecular associations result in such efficient packing of the crystal that the solvent content is only 32% of the unit cell volume and most of this is ordered. The combined effects of the low solvent content, the strong intermolecular associations, and the intramolecular hydrogen bonds may be responsible for the observed high diffractability of this crystal system.

An ultimate goal of the crambin neutron analysis will be to provide pertinent experimental evidence to establish the degree of deviation from "ideal" bonding geometry which actually exists in a protein molecule. Currently, amino acid and peptide crystal structures are used as standards of protein geometry. The theoretical potential functions believed to describe covalent interactions permit some degree of deviation from ideal geometry. However, the details of the extent of these deviations and which structural parameters are most flexible are not known. The unequivocal location of the hydrogen atoms in crambin by neutron diffraction together with the high resolution of the data should provide the degree of accuracy necessary to give direct evidence on the extent and nature of the protein group deviations from normal bonding geometry.

The crambin x-ray analysis revealed an unusually large number of highly ordered water molecules at the protein surface; in fact, in the crystal, only about 10 to 15% of the water can be classified as bulk solvent. Hence, this system is an excellent candidate for a detailed study of the ordered water structure. It has been shown (9,10) that neutron diffraction is potentially better suited to the study of the water environment of proteins than its x-ray counterpart. This advantage is in the threefold greater relative scattering of ordered water (D_2O) molecules in the neutron case, providing a correspondingly greater confidence level in determining their locations, given phasing models of equivalent quality. Another important advantage is that the orientation of the deuteriums on the well-ordered waters should be assignable directly in the neutron Fourier map.

These and other issues will be discussed in the following sections. The neutron crambin structure is currently being refined at 1.50 Å (44.5% of the data to 1.2 Å has also been included). It is expected that a nominal resolution of 1.0 Å can be achieved. Although the quality of the current model is excellent, as judged by the refinement statistics and the clarity of the resulting Fourier maps, we expect that an even finer detailed

picture of the structure will be obtained as the model is further refined and the resolution is extended.

EXPERIMENTAL

Crambin is isolated from the defatted seeds of Crambe abyssinica, a plant in the Crucifer family (13). The protein is soluble in a number of aqueous organic solvents but is insoluble in water. Crystals suitable for diffraction analysis were grown by vapor diffusion of a 60% ethanol/water mixture into a protein solution (30 mg/ml) containing 80% ethanol/water. The resulting crystals are monoclinic, space group $P2_1$, with unit-cell dimensions a = 40.96 Å, b = 18.65 Å, c = 22.52 Å, and $\beta = 90.77^{\circ}$. Under these conditions, some few large crystals can be grown, and a well-formed crystal of about 1.4 mm^3 (0.8 x 1.1 x 1.6 mm) was chosen for the neutron data collection. The crystal was soaked in deuterated mother liquor in two stages. First, the crystal was placed in 50% deuterated solvent for several days, and then it was soaked in fully deuterated solvent for a week.

The data were collected with a two-dimensional position-sensitive detector (1,2) using normal beam rotation geometry and λ of 1.55 Å. The majority of these data were collected at the instrument χ_0 position by rotating the crystal around the ϕ axis (parallel to b*) over a range of $180^{\circ} + 2\theta_{max}$. Those reflections falling in the "blind region" were collected by moving the crystal to χ_{90} and rotating around the instrument ω axis. Certain limitations in the present instrument setup prevented collection of the complete blind region data. The data were collected at 17 unique settings of the detector. An absorption curve measured, as a function of rotation angle, was closely isotropic; therefore, no absorption correction was applied to the data.

The data were reduced by the processing technique of Spencer and Kossiakoff (10). A total of 3871 reflections were in the 1.5-Å data set (73.3% of the theoretically obtainable data), and, on the basis of F values, 97.5% of these were above 3σ (98.6% above 2σ). The 2432 additional reflections collected to 1.2 Å represented 44.9% of the data. Analysis of the symmetry pairs gave a 6.2% R-factor between >3σ data and an R-factor of 8.3% for all pairs (both based on intensity).

Initial Phasing Model

The model used to calculate the first neutron map was derived from the 0.945-Å refined x-ray structure (Teeter and Hendrickson, unpublished). The accuracy of these coordinates is of extremely high quality as judged by the low R-factor for the 21,000 observations to 0.945 Å: 14%.

The only hydrogen atoms included in the initial model were those unexchangeable ones whose positions could be fixed by their parent atoms (i.e., phenyl ring or C_α hydrogen atoms, etc.). Those hydrogens which had rotational ambiguity and were thus unconstrained (i.e., methyl hydrogens) were omitted, together with the exchangeable protons. The calculated hydrogen positions were assigned temperature factors of $B + B^{1/2}$, where B is the temperature factor of the parent atom. No solvent molecules were included at this stage in the phasing model. A structure factor calculation using this partial model (79.0% of the total possible protein atoms) gave an R-factor of 30%.

To facilitate the location of the unconstrained and exchangeable hydrogen atom positions and to identify bound water molecules, a difference Fourier synthesis was computed and examined. From inspection of this map, most of the methyl rotor orientations could be assigned by the method of Kossiakoff and Spencer (4). All but two of the 26 rotors appeared to be in the staggered conformation. In this map most of the hydroxyl deuterium orientations (70%) were also apparent. Interestingly, a significant number of the largest peaks in this map were located at the sites of the well-ordered water molecules as determined from the x-ray analysis. The refinement of the structure was initiated with the positions assigned for 99% of the total protein atoms and 70 "water" molecules (i.e., oxygens, see below).

Refinement

Refinement of the atomic coordinates was carried out by a version of the restrained least-squares refinement procedure of Hendrickson and Konnert (15), modified to incorporate negative neutron scattering lengths of hydrogen (Wlodawer and Hendrickson, personal communication). In this method, the model is fitted to both the diffraction data and certain stereochemical restraints. The stereochemical restraints used were for bond lengths and angles, for planarity, for chirality, for thermal variation, for close contacts (van der Waals repulsions), and for excessive shifts. These can be weighted individually or as a group against the diffraction data, and their importance can be changed during the refinement.

The course of the refinement is summarized in Table 1. The refinement procedure involved several cycles of automated least-squares refinement interspersed with inspection of difference map peaks to identify regions of poor fit and to locate potential water and deuterium positions. Because of the high quality of the starting model, refitting involved essentially only adding deuterium atoms or changing the H/D ratios. No substantive changes in the main-chain atoms were found to be necessary.

Table 1. Course of the Refinement

Model	Cycles	Number of Atoms: Protein	Number of Atoms: Water	RMS Delta[a]	R[b]
1	2	662	W[c]	0.079-0.045	0.286 to 0.221
2	4	664	73 O, 55 D	0.020-0.008	0.200 to 0.170
3	5	664	73 O 75 D	0.007-0.017	0.177 to 0.144

[a]Root mean square deviation from idealiry (Å) for the first and last cycles in this series.
[b]$R = \Sigma (F_o - F_c)/\Sigma F_o$.
[c]Model for water as twice the oxygen scattering (see text).

Because a major objective of the analysis was to assess how well a protein molecule conforms to ideal geometry, the weighting scheme used for structural restraints during refinement is important. The philosophy adopted was to assign high weights to the structural restraints initially (i.e., after refitting) and then to relax them while simultaneously increasing the weight assigned to the diffraction data. When all the deuterium atoms that can be found have been located, we will try to determine whether the restraints are needed and, if they are, to what degree.

Dealing with the protein portion of the unit cell was quite straightforward, but there are inherent problems associated with refining the water molecules at the initial stages of the analysis. (i) Crambin has a large number of well-ordered waters which have to be considered, and (ii) these waters cannot be fully described at high resolution as a single symmetric peak (i.e., an oxygen atom). This is because each deuterium atom in a water molecule (D_2O) has the potential to scatter at a magnitude comparable with that of the parent oxygen. It might therefore be possible to establish the orientation of the well-ordered D_2O molecules from the density features of a Fourier map. The water oxygen atom, known from the x-ray structure, could be subtracted out in a difference synthesis to isolate, in effect, the scattering density for the deuterium atoms alone. However, during the initial stages of the analysis, the maps contain a significant amount of noise in the water region making precise judgments about the location of deuteriums tenuous.

The method initially employed for the D_2O refinement was to refine the well-ordered water molecules as a single atom having the scattering magnitude twice that of an oxygen atom. This treatment establishes the centroid of the water (D_2O) and is necessary to minimize errors in the map which could be propagated by

the substantial scattering of the deuterium atoms. In the first cycle of refinement, which included the protein and these "water" molecules, the R-factor dropped from 28.6 to 22.3%. After a second cycle, "waters" were replaced by the x-ray water oxygen atoms. The locations of the water deuteriums which were hydrogen-bonded to protein carbonyl oxygen atoms were calculated, assuming ideal bond distances. These were also included in the model. In four cycles of refinement, the R-value dropped to 17.0%. Finally, 20 deuterium atoms, which had difference peaks at the calculated positions for water deuteriums, were added to the model. In five cycles, the R-factor was reduced to 14.4%.

The occupancy factors (H/D ratios) for exchangeable protons were not refined. They were varied instead by inspection of difference maps. The starting ratios, for the protein exchangeable hydrogens, were chosen from the first difference map (R = 30%) to be one of five values: 1.0/0.0 (H), 0.75/0.25, 0.50/0.50, 0.25/0.75, and 0.0/1.0 (D). As the fitting proceeded, it became apparent that the largest H/D ratio was actually 0.25/0.75, indicating that probably the soaking solutions were slightly contaminated with H_2O.

DISCUSSION

Quality of the Neutron Maps

The high degree of detail in the crambin maps is evident from the following particular features: the methyl rotors, the ring structures, the scattering density difference between H and D amide protons, and the arginine-carboxyl salt bridge. Figure 2 shows the neutron difference density from several methyl rotors from which the parent carbon has been removed. The location of the hydrogen atoms is indisputable in the first three cases and strongly suggested in the fourth. In each of the first three methyl groups, the temperature factor is about 6, whereas in the fourth it is 10, possibly indicating higher mobility.

One of the more striking areas of the map is in the vicinity of Pro 5 and Tyr 44, which pack in the structure one on top of the other (see Figure 3). Pro 5 has atomic features, for both the positive and negative scatterers in the ring. A large hole at the ring center and strong O and D peaks are evident for Tyr 44. Hydrogens are well separated from the heavier atoms in both cases.

The clarity with which the exchanged and unexchanged protons can be distinguished is apparent in the difference maps shown in Figure 4. The amide proton in Ile 7 has been fully exchanged and clearly shows the characteristic positive scattering for a deu-

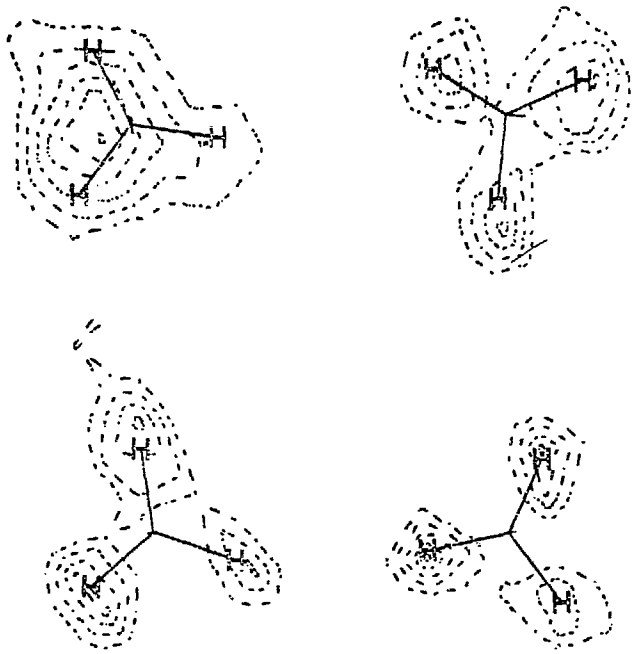

Fig. 2. Neutron densities of methyl rotor hydrogens viewed down the rotor axes. The parent carbon atoms were removed in the difference syntheses. The methyl groups show differing degrees of order as can be seen by comparing the first three plots with the fourth.

terium. It is at the end of the first turn region and is hydrogen-bonded to a solvent molecule. Ser 11 is in the second loop of the α-helix and is protected from exchange through a hydrogen bond to the carbonyl of Ile 7.

Arg 10 forms a salt bridge with the carboxy terminus (Asn 46). Two deuterium atoms on the guanidinium group (Figure 5) point directly to the carboxyl oxygen atoms of the terminus.

Amide Proton Exchange

The observed H/D exchange pattern of crambin is worthy of note because it does not follow the same general trend that has been reported for other proteins studied by neutron diffraction. In the structures of myoglobin (8), ribonuclease (14), and trypsin (5,6), there is a definite exchange pattern observed for the α-helix units. The unexchanged groups are located predominantly on

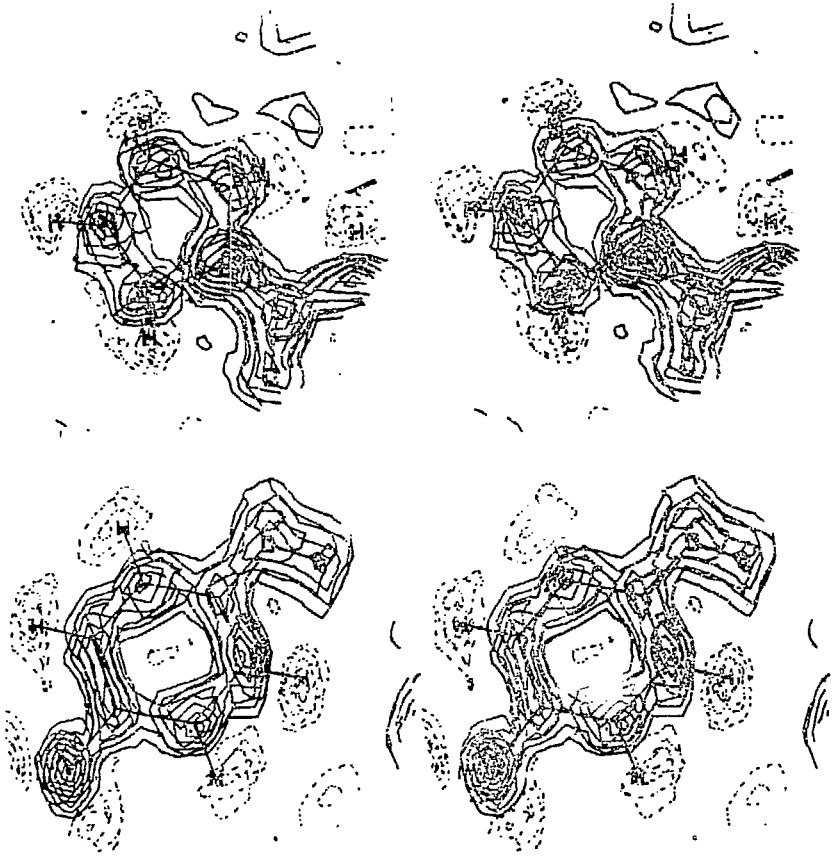

Fig. 3. Stereo view of the neutron Fourier map of side-chain rings Pro 5 (top) and Tyr 44 (bottom). The views are perpendicular to the least-squares plane of the rings.

the inside, the less accessible surface of the helices. The impeding of exchange seemingly only at points of intramolecular contact implies that steric shielding plays an important role in exchange of the helices (5).

In crambin, on the other hand, nearly all the peptide protons in the α-helix regions are found to be unexchanged, whether or not they are located at the intramolecular contact points. In fact, 90% of the amide protons in the helix were calculated to be within 3 Å of the solvent, i.e., they are very solvent accessible. (This is because the crambin molecule is so small that it has very

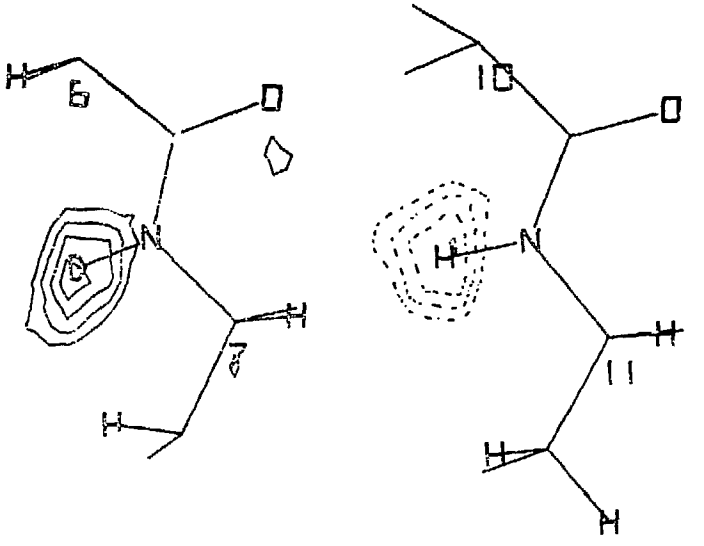

Fig. 4. Difference density for two amide planes with the amide proton or deuteron removed from the phasing mode. Ile 7 hydrogen-bonds to a water molecule at the first bend in the chain and is exchanged. Ser 11 falls in the middle of the top α-helix and is protected from exchange by a hydrogen bond.

Fig. 5. Stereo view of the salt bridge between the carboxy terminus (Asn 46) and Arg 10. The contact is viewed perpendicular to the guanidinium plane.

little protected interior.) With the exception of [illegible] all unexchanged amide protons are hydrogen-bonded to either main-chain carbonyl oxygens or side-chain acceptor groups. The peptide proton of Ser 9 points toward the interior of the molecule, and, although it is not itself hydrogen-bonded, it appears to be protected from exposure by the high degree of secondary structure of the adjacent groups.

Crambin has an unusually large percentage of unexchanged amide groups (Figure 6): 53% unexchanged (mainly those hydrogen-bonded in the α-helix and β-sheet), 24% partially exchanged, and 22% fully exchanged (the amide protons of the latter two categories are located mainly in the turn regions). There are several possible explanations for this behavior. It might be considered that because of its tight molecular packing in the crystal (i.e., only 32% solvent and strong intermolecular contacts), some of the normal fluctuations in structure are impeded by lattice forces. However, NMR solution studies on crambin (7) showed that this molecule is very resistant to exchange even under extreme conditions. After treatment with deuterated glacial acetic acid for several hours at 50°C, about 67% of the amides remained unexchanged. Interestingly, the protons which exchange do so quite rapidly, particularly in aqueous acetone or aqueous ethanol (Llinás, private communication).

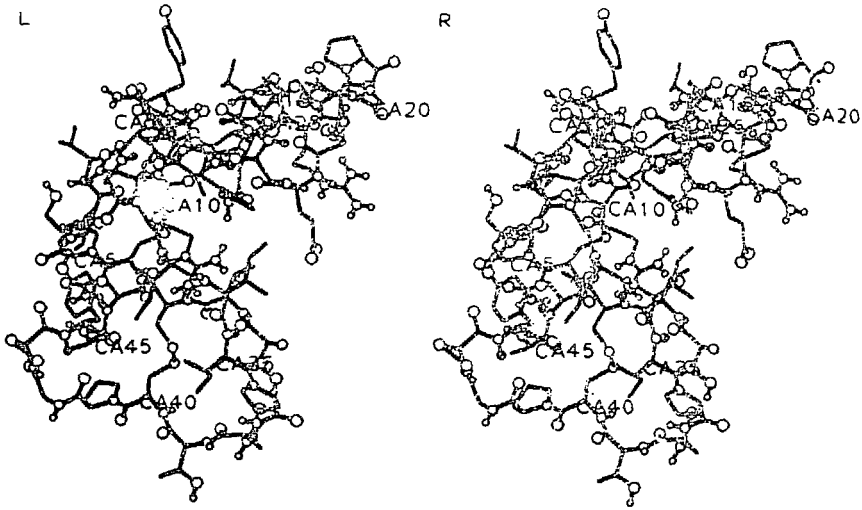

Fig. 6. Exchange character of the amide peptide protons. Hydrogen atoms are represented as filled circles (bonded to the peptide nitrogens) and deuteriums as open circles.

Although effects of crystal packing on exchange cannot be ruled out, the fact that crambin is highly structured, i.e., has an extensive intramolecular hydrogen-bonding network, must play a major role in its resistance to exchange. As a future experiment, we plan to take a set of data on a crystal which has been soaked for about one year. A year is comparable with the soaking times of other reported neutron H/D exchange studies, and these data should provide a more sensitive measure of the relative protection of the α-helical protons.

Protein Geometry

Although the ultimate aim of this work will be a more fundamental understanding of the factors governing protein geometry, such an analysis of the structure at this point would be premature. The regularities and irregularities are quite close to those described earlier for the 1.5-Å x-ray structure (3). After the final cycles of refinement with all constraints removed, the in-depth correlation of bond angles, bond lengths, and dihedral angles will be more meaningful.

Solvent Structure

One of the most interesting features of the crambin structure analysis is the number of well-ordered water molecules associated with the protein surface. Out of ~90 possible waters (calculated from the density), 73 were located from the x-ray experiment at reasonable positions and have been included in the neutron model. Eight of these have alternative positions. Table 2 summarizes the types of water coordination and their vibrational amplitudes.

Table 2. Summary of Water (D_2O) Structure

Type[a]	Number	<B>	Number of Deuteriums
Primary	58(7)[b]	18	52
Secondary	12(1)	32	3
Tertiary	3	35	1

[a]A primary water has its oxygen within 3.2 Å of a donor or acceptor group on the protein. A secondary water has at least one primary water in its coordination shell. A tertiary water has only secondary waters in its first coordination shell.

[b]The number in parenthesis is the number of waters in alternative positions for this type.

In the x-ray determination of crambin, only two ethanol molecules were located. More would be expected since the crystals are grown from 60% ethanol. However, the hydrophobic contact which forms between crambin molecules in the crystal may eliminate the need for some of the hydrophobic solvent that originally solubilized it. In the neutron diffraction experiment, we cannot readily distinguish ethanol from bound water since monodeuterated ethanol was used. Although we do not expect the deuterated crystals used in the neutron analysis to have a different solvent distribution from the x-ray, we plan to soak crystals with perdeuterated ethanol in a future experiment in order to determine the number of bound ethanol molecules more accurately.

ACKNOWLEDGMENTS

We are grateful to Joseph Shpungin (BNL) for considerable technical assistance. We also acknowledge the assistance of Dr. B.P. Schoenborn, who designed and developed the protein crystallography station at the HFBR.

REFERENCES

1. Alberi, J., Fisher, J., Radeka, V., Rogers L.C., and Schoenborn, B.P., IEEE Trans. Nucl. Sci. NS-22:255-68 (1975).
2. Cain, J.E., Norvell, J.C., and Schoenborn, B.P, Brookhaven Symp. Biol. 27:VIII 43-50 (1976).
3. Hendrickson, W.A. and Teeter, M.M., Nature (London) 290:107-13 (1981).
4. Kossiakoff, A.A. and Spencer, S.A., Biochemistry 20:6462-74 (1981)
5. Kossiakoff, A.A., Nature (London) 296:713-21 (1982).
6. Kossiakoff, A.A., See paper in this Symposium.
7. Llinás, M., DeMarco, A., and Lecomte, J.T., Biochemistry 19:1140-5 (1980).
8. Phillips, S.E.V., See paper in this Symposium.
9. Schoenborn, B.P. and Hanson, J.C., "Water and Polymers," p. 215, S.P. Rowland, ed., ACS Symp. 127, Am. Chem. Soc., Washington, DC (1980).
10. Spencer, S.A. and Kossiakoff, A.A., J. Appl. Crystallogr. 13:563-71 (1980).
11. Teeter, M.M. and Hendrickson, W.A., J. Mol. Biol. 127:219-23 (1979).
12. Teeter, M.M., Mazer, J.A., and L'Italien, J.J., Biochemistry 20:5437-43 (1981).
13. Van Etten, C.H., Nielsen, H.C., and Peters, J.E., Phytochemistry 4:467-73 (1965).

14. Wlodawer, A. and Sjölin, L., Proc. Natl. Acad. Sci. USA 79:1418 (1982).
15. Hendrickson, W.A. and Konnert, J.H., "Biomolecular Structure, Function, Conformation and Evolution," R. Srinivasan, ed., Vol. I, pp. 43-57, Pergamon, Oxford (1980).

DISCUSSION

KOSSIAKOFF: The intermolecular packing at the hydrophobic interface in crambin seems to involve an uncharacteristically large number of groups almost as though it were a natural point of aggregation. What is the extent of the contact region? Most lattice packing contacts are mediated by solvent; is solvent completely excluded from this particular contact?

TEETER: This hydrophobic contact extends over two turns of the helix, from Pro/Ser 22 to Tyr 29. There is no solvent which mediates the intermolecular contact. The ethanol molecule, which is most certain from the x-ray analysis, lies at the edge of this surface with the methyl toward it and the hydroxyl bound to Asn 12. Thus it acts as a "buffer" between the hydrophobic and hydrophilic regions. There is some evidence that in solution this hydrophobic contact is important. Early molecular weights for crambin were twice that expected from the amino acid sequence which is now known.

MASON: Is there any evidence (at the present stage of the x-ray and neutron analysis) for differences in structure between the deuterated and non-deuterated crystals? Are the cell dimensions the same for the two crystals?

TEETER: At this point, aside from the heterogeneity differences I mentioned, I have not seen other differences. However, I will look in detail as the refinement proceeds. The cell dimensions measured independently in the x-ray and neutron analyses are nearly identical.

MOORE: I am intrigued by the multiple positions seen for aromatics. Does your associate doing NMR on crambin see signs of this heterogeneity?

TEETER: The Tyr 29 positions are generated by an intermolecular contact. Thus, they clearly would not appear in solution. Dr. Llinás at Carnegie-Mellon does, however, see multiple signals because of the heterogeneity at position 25. The Ile and Leu at 25 have different couplings with Tyr 29, which is one turn of the helix away. Llinás does not see effects at Phe 13.

KOEPPE: Would you expect to find ethanol in your crystals?

TEETER: Naively, yes, since the crystals are grown from 60% ethanol. However, in the x-ray structure we find only two, one of which is more certain than the other. In the crystal, crambin has a strong hydrophobic intermolecular contact that may make ethanol -- needed to dissolve the protein, which is insoluble in water -- superfluous in the crystal.

YEAGER: Has a mechanism of action been proposed for the plant toxins based on the crambin structure? For instance, the amphipathic structure may allow formation of a multimeric pore in which the hydrophobic protein surfaces are adjacent to hydrocarbon chains. The hydrophilic surfaces could then form a pore through the membrane. Cell lysis could then result by redistribution of water and electrolytes.

TEETER: No. We have several people studying the toxins to provide biochemical evidence for their action. We have also demonstrated that both crambin and one of the toxins associate to phospholipids. We do not know whether the level of association can quantitatively account for the activity. Formation of pores could be a possible mechanism for lysis, but we want some hard biochemical and physical evidence before we build a model.