Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 6014–6018, October 1984
Biophysics

Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin

(x-ray diffraction/clathrate hydrate/liquid-like)

M. M. TEETER

Department of Chemistry, Boston University, 685 Commonwealth Avenue, Boston, MA 02215

Communicated by Martin Karplus, May 14, 1984

ABSTRACT **The water structure has been analyzed for a model of the protein crambin refined against 0.945-Å x-ray diffraction data. Crystals contain 32% solvent by volume, and 77% of the solvent molecules have been located—i.e., 2 ethanol molecules and 64 water molecules with 10–14 alternate positions. Many water oxygen atoms found form chains between polar groups on the surface of the protein. However, a cluster of pentagonal arrays made up of 16 water molecules sits at a hydrophobic, intermolecular cleft and forms a cap around the methyl group of leucine-18. Several waters in the cluster are hydrogen-bonded directly to the protein. Additional closed circular arrays, which include both protein atoms and other water oxygen atoms, form next to the central cluster. This water array stretches in the *b* lattice direction between groups of three ionic side chains.**

Water is the most abundant molecule in living systems and plays an important role in their structure and function. From crystal structures of proteins (1–3), the following picture of tightly bound internal and surface water molecules has emerged (4, 5). First, internal water is found singly (6), in clusters (7), or bound to metal ions (8, 9). It either may stabilize the protein structure by connecting charged or polar groups, or both, or may serve a catalytic function (10). Second, surface water links polar groups at the intra- and intermolecular protein surface; 2–3 times more water hydrogen bonds are made to main-chain -CO groups than to -NH groups (4, 5) in both crystal structures and simulations, reflecting in part the greater capacity of -CO to form hydrogen bonds. Many surface waters bind at turn regions (11). Here, fewer secondary-structure, backbone hydrogen bonds are formed, and the side chains are often hydrophilic (12). Little if any ordered surface water has been reported at nonpolar side chains, perhaps because such water may be mobile as inferred from computer simulations (13). Finally, less ordered water (about 30% of that in the first shell) is seen in the second shell (14).

Water–protein interactions may be important for protein folding (4). Klotz proposed that hydrophobic residues would organize surface water into five-membered ring arrays analogous to the water clathrate (cage) hydrate structures (15). However, protein crystal structures reveal that many hydrophobic groups fold instead into the protein interior away from water, as suggested by Kauzmann (16). Nevertheless, 40–50% are found on the protein surface in contact with solvent. The folded structure may be a balance between removal of hydrophobic groups from water to form van der Waals attractive contacts and the penalty for burial of polar groups which might not be able to form hydrogen bonds. The arrangement of water at surface hydrophobic sites is not known. If five-membered ring structures were formed (15) and were not hydrogen-bonded to the protein, they might become disordered or vibrate strongly. Such waters would contribute little scattering to an x-ray experiment.

Crystals of the hydrophobic protein crambin diffract at least to 0.88 Å (17) and provide an excellent opportunity to study both protein and water structure at atomic resolution. The volume of solvent (32%) is low but not atypical for proteins. [Values of 30–78% solvent (18) are found.] Crambin is small (M_r 4720), and most residues are solvent accessible. Although crystals were grown from 60% ethanol, only 7% (vol/vol) ethanol has been located. Of the 85 water molecules expected (based on the density and the presence of 2 ethanol molecules), the model includes 64; 10–14 of the waters located have alternate positions. Most waters link polar groups, as is typical for surface water (see above). However, some form pentagonal rings at a hydrophobic, intermolecular contact in crambin crystals.

Diffraction Studies

X-ray data on crambin have been collected to 0.945 Å. Crystals have $P2_1$ space group symmetry, and the unit cell is a = 40.96, b = 18.65, c = 22.52 Å, with β = 90.77°. The model (19) has been refined by restrained least-squares (20) to an R value of 12.9% for the 10.0- to 0.945-Å data (unpublished results). ($R = \Sigma|(|F_o| - |F_c|)|/\Sigma|F_o|$.) There are 21,045 observations (Fs) above 2 σ, and four parameters (x, y, z, and a vibrational factor) are needed to describe each of the 440 nonhydrogen atoms in the model. Thus, the structure is 12 times overdetermined. Based on small-molecule crystals at this ratio of overdetermination and at this resolution, the corefinement of vibrational factors and occupancy is justifiable. Although the occupancy was allowed to vary for the water molecules, many retained values close to 1.0 (full occupancy). Water molecules with clear alternates were fixed at fractions that added to 1.

An independent confirmation of the positions of the water oxygen atoms was obtained from neutron diffraction (21). Crystals grown as for the x-ray experiments were soaked in $^2H_2O/C_2H_5O^2H$ solutions for several weeks; 73% of the data to 1.5 Å and 45% of the data to 1.2 Å were collected (6300 reflections in all). The model refined against the neutron data has an R value of 14.4%. It was not refined to convergence, since only half of the water deuterium atoms had been included. The detailed arrangement of water oxygens is very close to the x-ray results, and no contradictions exist. Figs. 3 and 5 use the neutron coordinates and illustrate features not likely to change with further refinement. The distances reported are from the x-ray structure.

Overall Water Arrangement

Crambin is a neutral molecule with only four charged side chains. These groups lie near the inner bend of this Γ-shaped molecule (the darker molecule left of "2" in Fig. 1), and hydrophobic residues are on the opposite side ("1" in Fig. 1), which gives crambin an amphipathic character (19). Two in-

---

teractions bring three of the four charges close. First, a guanidinium N of arginine-17 binds to the carbonyl O of aspartic acid-43 from a *c*-axis-related molecule. Second, water links the side chains of residues arginine-17 and glutamic acid-23 part of the time. [The bridging molecule (water 90) is an alternative to water 98.]

Fig. 1*a* shows the packing diagram for crambin and the space available for water molecules. In Fig. 1*b*, the water locations in relation to the protein are shown. It is clear that the majority of these waters lie between protein molecules in the *ac* plane (the plane of the figure) and not on top of protein molecules in the *b* direction, where the protein molecules pack tightly.

The average water oxygen–oxygen distance is 2.81 ± 0.16 Å. The ratio of water–backbone carbonyl:water–backbone amide hydrogen bonds is approximately 2:1. The limit assumed for a hydrogen bond (3.3 Å) is close to the limit of 3.2 Å found for small molecule hydrate crystals (22).

Pentagonal Rings of Waters Are Localized

The *c* axis protein contact for the top α helix is quite hydrophobic. Here a cluster of pentagonal water rings is formed (Fig. 2*a*). It is surrounded by leucine-18 and asparagine-14, by isoleucine-7 and valine-8 (from a *c*-axis-related molecule), and by tyrosine-29 (from a screw-axis-related molecule). This cluster is adjacent to the pseudo-dimer contact between screw-axis-related molecules where residues leucine-18, valine-15, isoleucine-25, proline-19, and tyrosine-29 form a water-excluding protein contact (see "1" in Fig. 1). Above and below the cluster in the *b* direction, at the bottom helix contact along *c*, are the charged amino acids residues (arginine-17, glutamic acid-23, and aspartic acid-43) described above.

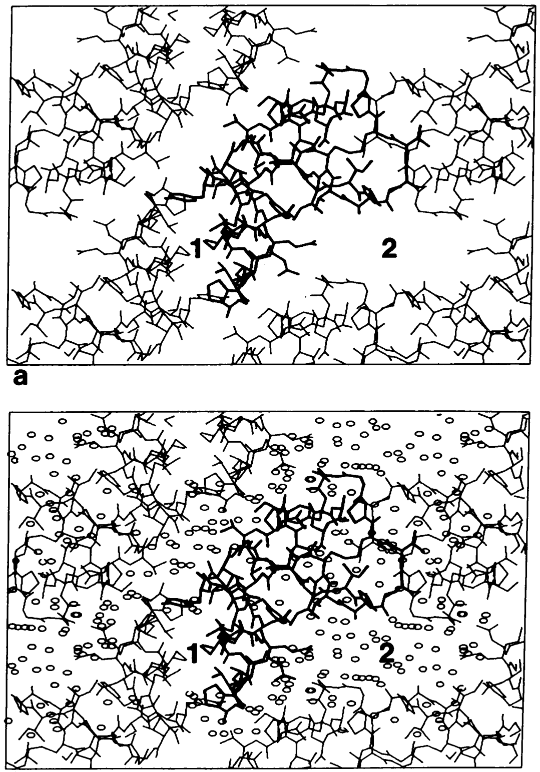

FIG. 1. Packing diagram of crambin in the *ac* plane of the crystals. The *a* axis is horizontal, and the *c* axis is vertical. The area shown is 33 × 47 Å. (*a*) The unique protein molecule is darker than the symmetry-related ones that surround it. The hydrophobic pseudo-dimer contact is indicated by 1, and the "hole," by 2. (*b*) All waters including alternates are shown. Because of limits imposed by the viewing distance of the figure, some screw-axis-related waters are omitted.

Four pentagonal rings (A, C, D, and E) lie along a hydrophobic surface at an intermolecular contact (Fig. 2 *a* and *b*). Rings A, C, and E, which share an apex, create a cap around the $C^{\delta 2}$ of leucine-18 (Fig. 2*b*). Ring A forms a single layer between the *c* axis-related molecules (Fig. 2*a*). It is surrounded on four sides by protein but connected to water molecules in the other two directions. A fourth ring (D) extends this sheet of pentagons (Fig. 2*a*) and lies along the hydrophobic ethylenes of arginine-17 (Fig. 2*b*). Rings C, D, and E have one face to the protein and one face to the solvent. In contrast, ring B extends into the solvent (Fig. 2*b*). This ring has been omitted in Figs. 2*a* and 3 for clarity. Fig. 3 is viewed down the *a* axis, approximately perpendicular to the plane of rings C, D, and E. Note that the methyl group of leucine-18 (darker atom) is visible through the center of ring C. Table 1 lists short contacts (less than 3.5 Å) between water oxygens in the rings and carbon atoms in the protein. Although leucine-18 is not in this table, it has five water contacts less than 4.0 Å, more than any other carbon atom at that limit.

In addition to the hydrophobic "contacts," the rings are extensively hydrogen bonded to the protein. Four carbonyl oxygens at the COO^--terminal end of the top helix (leucine-18, arginine-17, asparagine-14, and the $O^{\delta 1}$ of asparagine-14 in Fig. 2*a*) are attached directly to the cluster. On the other side of the cluster, two amide nitrogens (isoleucine-7 and valine-8) as well as the carbonyl oxygen of proline-5, all from a *c*-axis-related molecule, are bound. The charged residues arginine-17 and glutamic acid-23 (Fig. 4) connect pentagonal clusters in the *b* direction. Six additional five- and six-membered water–protein rings form adjacent to the cluster. Protein atoms from the -CO group of proline-5, from $O^{\delta 1}$ of asparagine-14, from the $-COO^-$ of aspartic acid (2 rings), and from the -OH group of tyrosine-44 are considered part of the ring, in analogy with the semiclathrate structures.

The average O–O distance for the water rings (A–E) is 2.80 ± 0.08 Å. Detailed hydrogen-bonding information for the water pentagons is given in Table 2. The average vibrational factor for all of the pentagon rings is 15.8 (rms vibration of 0.45 Å), and the average occupancy is 0.76. (For all first-shell waters, the average vibrational factor is 19.) Rings A, C, and E have the respective averages for vibrational factor and occupancy of 13.7/0.9, 13.5/0.9, and 13.6/0.85. In the remaining two rings (B and D), at least one atom has a clear alternate position (partial occupancy). For rings B and D, the average vibrational factors and occupancies are 19/0.7 and 15/0.7. Rings A, C, and E are more well-ordered than are B and D, and rings A, C, D, and E form a sheet of water rings. Three of the five rings (A, C, and D) are nearly planar (Fig. 2 *a* and *b*).

The Hole: Water Between the Charged Surfaces in the *ac* Plane

In the crystal, the charged hydrophilic surfaces from four symmetry-related molecules face one another. This "hole" (Fig. 1*a* at "2" and Fig. 5*a*) is about 13 × 8 Å in the *ac* plane and extends continuously in the *b* direction. The cluster of water-ring pentagons is at the side of this hole (Fig. 5*b*). When the rest of the located waters, which are in the hole (40% of the total found), are added (Fig. 5*c*), they appear to fill this hole.

However, analysis of the space available for additional waters from a Lee and Richards surface construction (23) for the protein plus solvent reveals that at least 12–15 more waters can be packed into this hole. The total obtained by combining these with the 64 water molecules already found approaches 85, the number calculated from the density of the

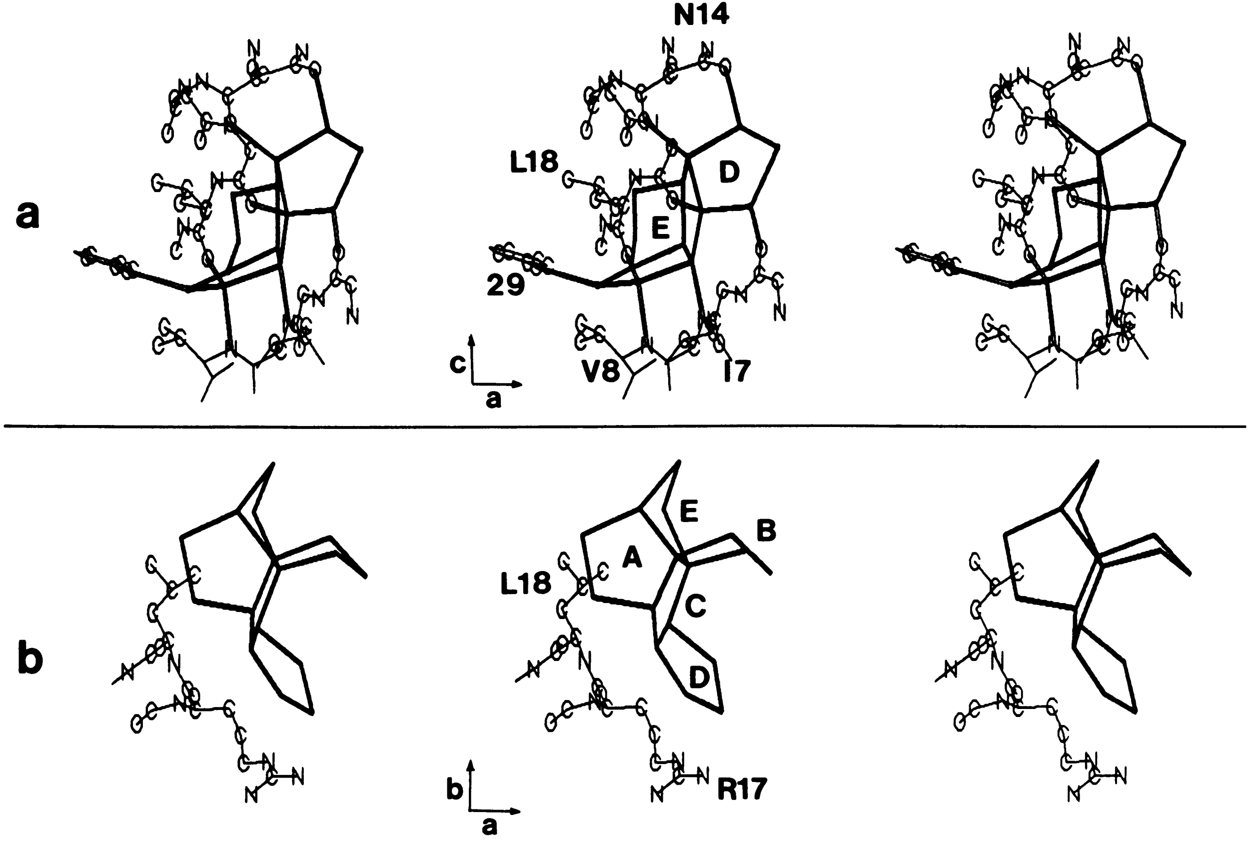

FIG. 2. Stereo diagrams showing the relation of the pentagon rings to crambin's side chains. (*a*) At the hydrophobic contact between symmetry-related crambin molecules, hydrogen bonds are formed between the water cluster and carbonyl groups at the COO$^-$-terminal end of the top helix and amide groups at the NH$_3^+$-terminal end of the same helix on a *c*-axis, translationally related molecule. The figure is viewed along the *b* axis. (*b*) Rings A, C, and E surround C$^{\delta 2}$ of leucine-18, and ring D lies above the hydrophobic methylenes of arginine-17. Ring B extends into solvent. The figure is viewed along the *c* axis. Single-letter codes are used for amino acids: N, asparagine; L, leucine; Y, tyrosine; V, valine; I, isoleucine; and R, arginine. The left-hand pair of diagrams can be viewed with a stereo viewer. The right-hand pair should be viewed if ones eyes cross.

crystals (see above). It should be noted that water is not well-localized around several methyl groups at the edge of this hole (threonine residues 1 and 39).

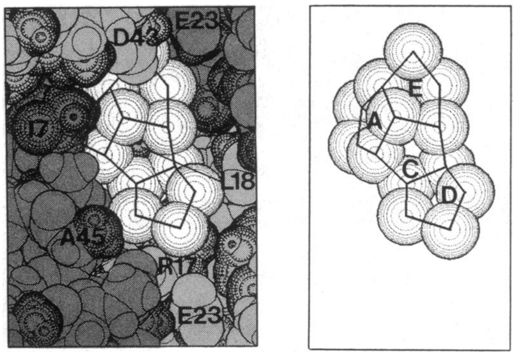

FIG. 3. van der Waals surface diagram of the water pentagons A, C, D, and E viewed in the negative *a* direction. Rings A, C, and E form a cap around leucine-18. Hydrophobic atoms have dark circles, and water oxygens have light circles. The methyl group of leucine-18 (shown in Fig. 2*b*) can be seen through the C ring. Adjacent translationally related molecules are shaded. The van der Waals radii used for the protein C, N, and O atoms are 1.7, 1.4, and 1.4 Å, respectively, and for water oxygen, 1.8 Å. The larger radius is used for the water oxygens because hydrogen atoms have been omitted. Atomic coordinates are from the neutron model without the water deuterium atoms.

Discussion

Ordered five-membered water arrays apparently have not been described before in protein structures. However, such arrangements have been seen in the crystal structure of d(CpG)-proflavine (24) and may be of general importance. A sheet of five pentagonal water rings are formed, four with only water molecules and a fifth with four water molecules and one phosphate oxygen in one area of this complex. Oth-

Table 1. Hydrophobic water contacts of <3.5 Å

Protein carbon atom	Ring water oxygen	Distance, Å	Ring
Pro-5 C$^{\beta}$	67	3.50	Tyr-44*
	72	3.49	Tyr-44*
Ile-7 C$^{\delta 2}$	79	3.49	B, E
Val-8 C$^{\gamma 2}$	50	3.33	A
Pro-22 C$^{\alpha}$	67	3.49	Tyr-44*
Ile-34 C$^{\delta 1}$	55	3.42	—
Pro-41 C$^{\gamma}$	115	3.40	Asp-43*
Tyr-44 C$^{\zeta}$	72	3.41	Tyr-44*

Contacts have been excluded to a carbon atom adjacent to the protein doner or acceptor atom to which the water is bound—e.g., an α-carbon atom next to a backbone -NH.

*The ring consists of water molecules bonded to the cluster with a polar atom from the designated residue.

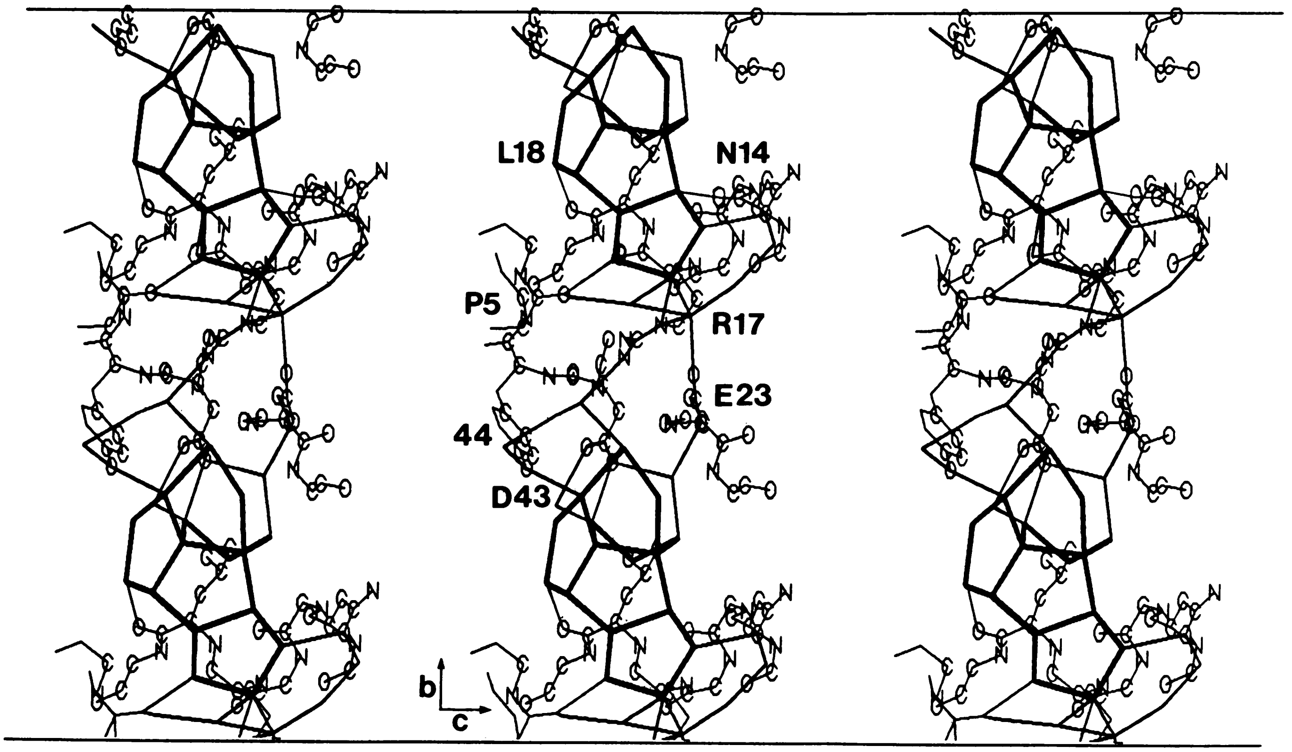

FIG. 4. Stereo diagram of water pentagons and adjacent rings, which include hydrophilic protein atoms, viewed as in Fig. 3. The pentagon cluster has the darkest lines. The adjacent ring bonds are lighter, and the protein-to-cluster bonds are the lightest. Single-letter codes are used for amino acids: L, leucine; P, proline; Y, tyrosine; D, aspartic acid; N, asparagine; R, arginine; and E, glutamic acid. Stereo pairs are arranged as in Fig. 2.

er waters in the structure form a polygonal water disc. The pentagons are approximately planar and bridge the charged nitrogen atoms in heterocyclic rings stacked 3.4 Å apart. The pentagons fill a cavity where water–water interactions may predominate over water–nucleic acid or water–drug contacts (24).

The waters ordered in pentagons, which are found in crambin and in the d(CpG)-proflavine structure, are reminiscent of the semiclathrate hydrate structures. Crystals of piperazine hydrate (25) and pinacol hydrate (26) have layers of host water pentagons joined by hydrogen bonds to the N or O of the guest. In *n*-propylamine hydrate (27), four nonequivalent guest molecules form hydrogen bonds to each other and to waters in concave regions, while the propyl groups sit in "open bowls" made by the convex polyhedra.

Table 2. Hydrogen-bonding contacts in the pentagon cluster

Oxygen atom			
i	*j*	Pentagon	D_{ij}, Å
47	74	A	2.91
74	78	A, E	2.75
78	63	A, C	2.70
63	50	A	2.77
50	47	A	2.85
79	78	B	2.75
78	64	B, C, E	2.80
64	80	B	2.77
80	81*	B	2.75
81*	79	B	2.93
63	62	C	2.65
62	61	C, D	2.79
61	64	C	2.83
69	61	D	2.72
62	84	D	2.50
84	98*	D	2.92
98*	69	D	2.65
64	60	E	2.77
60	59	E	2.92
59	74	E	3.18

D_{ij}, distance between oxygen atoms *i* and *j*.
*These water molecules have an alternate position.

The questions of the generality of the pentagonal cluster in crambin crystals and the nature of the forces involved are difficult to answer. Ring A is surrounded by hydrophobic contacts on four of its six sides. Do these neighbors limit hydrogen bonding geometry of ring A to a planar pentagonal ring and, thus, is the geometry of rings C and E also restricted? Could ring A form a six-membered ring with waters in the remaining two directions? Do the charged groups play a role here? Aspartic acid-43 is close to ring A. If rings C and E are formed because of ring A's hydrogen-bonding restriction, are rings B and D also so generated? Ring B has one atom in common with ring A, but ring D has none. An analysis of the energies involved in these arrays for the crystal could be made with empirical potential energy functions, for instance, those in CHARMM (28) or AMBER (29). Monte Carlo methods might be applied to predict the water locations (30), but so far they have had limited success for the CpG-proflavine structure (31).

A partial cage of pentagonal water rings has been proposed when hydrophobic side chains in an α helix or β sheet are separated by a single layer of water molecules (32). Némethy and Scheraga found that the formation of pentagonal rings had negative free energy when compared with the separately hydrated protein regions as the starting state rather than with pure water. These localized pentagonal clusters might occur not only at intermolecular, hydrophobic contacts, as reported here, but also at adjacent, hydrophobic side-chains in an α helix or β sheet.

Theoretical calculations for liquid water and solutions of

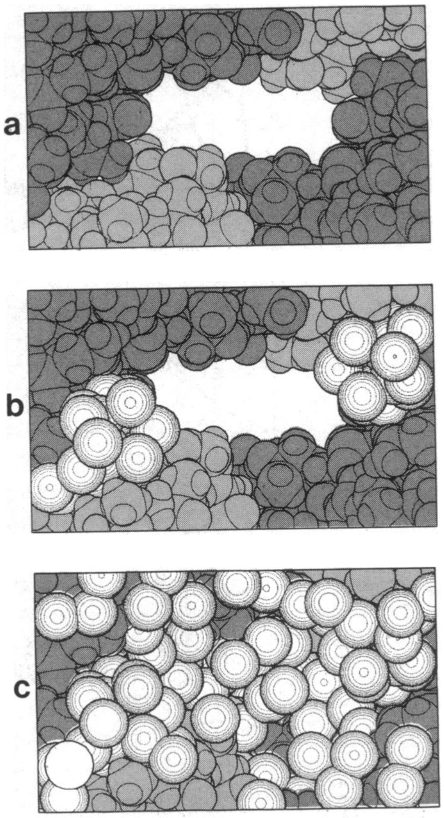

FIG. 5. van der Waals surface representation of the solvent "hole" between adjacent crambin molecules viewed along *b* (see area 2 in Fig. 1). Symmetry-related and translationally related molecules are distinguished by shading. (*a*) No water is included. (*b*) The cluster of pentagons of water molecules has been added. (*c*) The water molecules in the current model appear to fill the "hole" but are loosely packed when compared to the pentagon cluster.

nonpolar molecules show the presence of pentagon arrays of water but do not unequivocally demonstrate that the pentagon is the best way to order water around nonpolar groups. In contrast to ice, which extends as hexagons in a regular lattice, liquid water from simulations organizes into chains and rings of fewer than 10 oxygens that are continually forming and breaking (13). The most abundant polygon formed, assuming a hydrogen-bonding energy of −4.0 kcal/mol, is the pentagon, followed closely by the hexagon (33). Monte Carlo simulations have predicted quasiclathrate structures with nonplanar water pentagons for methane in water (34). Molecular dynamics calculations for the alanine dipeptide in water suggest that the water associated with the hydrophobic groups is less mobile than the bulk or the hydrophilic-associated water and is clathrate-like in geometry (35). However, it is more like bulk water than like a solid in terms of the number and strength of hydrogen bonds, perhaps because hydrogen bonding is maintained despite the geometrical restraints of a nonpolar neighbor (35).

Analysis of the cluster hydrogen bonds reveals that the 16 water molecules making up the cluster form 72% of their hydrogen bonds to water molecules (63% of the total within the cluster and 9% to other water oxygens) and 23% to protein, with 5% (3 bonds) remaining unsatisfied. If ring B, which extends into solvent, is omitted, the percentages are 69% to water and 29% to protein, with 2% unsatisfied bonds. One interpretation of the observed pentagon cluster in crambin from this might be that the hydrophobic surface at the protein interface permits the water there to order as normal, liquid-like water, preferring interactions with itself to those with the protein.

The author thanks Drs. W. A. Hendrickson and A. A. Kossiakoff for their collaborative efforts. The helpful comments of Drs. H. M. Berman, G. J. Quigley, R. Hudson, D. L. Beveridge, K. Wüthrich, and J. T. Edsall are much appreciated. Support of National Science Foundation Grant PCM 8303024 is gratefully acknowledged.

1. James, M. N. G. & Sielecki, A. R. (1983) *J. Mol. Biol.* **163**, 299–361.
2. Sakabe, N., Sakabe, K. & Sasaki, K. (1981) in *Structural Studies of Molecules of Biological Interest*, eds. Dodson, G., Glusker, J. P. & Sayre, D. (Clarendon, Oxford), pp. 509–526.
3. Blake, C. C. F., Pulford, W. C. A. & Artymuick, P. J. (1983) *J. Mol. Biol.* **167**, 693–723.
4. Finney, J. L. (1979) in *Water: A Comprehensive Treatise*, ed. Franks, F. (Plenum, New York), Vol. 6, pp. 47–123.
5. Edsall, J. T. & McKenzie, H. A. (1983) *Adv. Biophys.* **16**, 53–183.
6. Diesenhofer, J. & Steigemann, W. (1975) *Acta Crystallogr. Sect. B* **31**, 238–250.
7. Bode, W. & Schwager, P. (1975) *J. Mol. Biol.* **98**, 139–143.
8. Rees, D. C., Lewis, M. & Lipscomb, W. N. (1983) *J. Mol. Biol.* **168**, 367–387.
9. Branden, C. I., Jornvall, H., Eklund, H. & Furugren, B. (1975) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), 3rd Ed., Vol. 11, pp. 103–190.
10. Eklund, H., Nordström, B., Zeppezauer, E., Söderland, G., Ohlsson, I., Boiwe, T. & Bränden, C.-I. (1974) *FEBS Lett.* **44**, 200–204.
11. Carter, C. W., Kraut, J., Freer, S. T. & Alden, R. A. (1974) *J. Biol. Chem.* **249**, 4212–4225.
12. Rose, G. D. (1978) *Nature (London)* **272**, 586–590.
13. Geiger, A., Rahman, A. & Stillinger, F. H. (1979) *J. Chem. Phys.* **70**, 263–276.
14. Watenpaugh, K. D., Margulis, T. N., Sieker, L. C. & Jensen, L. H. (1978) *J. Mol. Biol.* **122**, 175–190.
15. Klotz, I. M. (1959) *Science* **128**, 815–822.
16. Kauzmann, W. (1959) *Adv. Protein Chem.* **14**, 1–63.
17. Teeter, M. M. & Hendrickson, W. A. (1979) *J. Mol. Biol.* **127**, 219–224.
18. Matthews, B. W. (1976) *Annu. Rev. Phys. Chem.* **27**, 493–523.
19. Hendrickson, W. A. & Teeter, M. M. (1981) *Nature (London)* **290**, 107–113.
20. Hendrickson, W. A. & Konnert, J. (1981) in *Biomolecular Structure, Function, Conformation and Evolution*, ed. Srinivasan, R. (Permagon, Oxford), pp. 43–57.
21. Teeter, M. M. & Kossiakoff, A. A. (1983) in *Neurons in Biology*, ed. Schoenborn, B. (Plenum, New York), pp. 335–348.
22. Falk, M. & Knop, O. (1973) in *Water: A Comprehensive Treatise*, ed. Franks, F. (Plenum, New York), Vol. 2, pp. 115–233.
23. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400.
24. Neidle, S., Berman, H. M. & Shieh, H. S. (1980) *Nature (London)* **288**, 129–133.
25. Schwarzenbach, D. (1968) *J. Chem. Phys.* **48**, 4134–4140.
26. Kim, H. S. & Jeffrey, G. A. (1970) *J. Chem. Phys.* **53**, 3610–3615.
27. Brikenkamp, C. S. & Panke, D. (1973) *J. Chem. Phys.* **58**, 5284–5295.
28. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1982) *J. Comp. Chem.* **4**, 187–217.
29. Kollman, P. & Weiner, P. (1981) *J. Comp. Chem.* **12**, 287–303.
30. Madison, V., Osguthorpe, D. J., Dauber, P. & Hagler, A. T. (1983) *Biopolymers* **22**, 27–31.
31. Mezei, M., Beveridge, D. L., Berman, H. M., Goodfellow, J. M., Finney, J. L. & Neidle, S. (1983) *J. Biomol. Struct. Dyn.* **1**, 287–298.
32. Némethy, G. & Scheraga, H. A. (1962) *J. Phys. Chem.* **66**, 1773–1789.
33. Rahman, A. & Stillinger, F. H. (1973) *J. Am. Chem. Soc.* **95**, 7943–7948.
34. Swaminathan, S., Harrison, S. W. & Beveridge, D. L. (1978) *J. Am. Chem. Soc.* **100**, 5705–5712.
35. Rossky, P. J. & Karplus, M. (1979) *J. Am. Chem. Soc.* **101**, 1913–1937.