PROTEINS: Structure, Function, and Genetics 12:203–222 (1992)

Solvent Structure in Crystals of Trypsin Determined by X-Ray and Neutron Diffraction

Janet S. Finer-Moore,[1] Anthony A. Kossiakoff,[2,3] James H. Hurley,[1] Thomas Earnest,[1] and Robert M. Stroud[1,3]

[1]Department of Biochemistry and Biophysics and Graduate Group in Biophysics, University of California, San Francisco, California 94143-0448; [2]Department of Protein Engineering, Genentech, Inc., So. San Francisco, California 94080; [3]Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

ABSTRACT **The solvent structure in orthorhombic crystals of bovine trypsin has been independently determined by X-ray diffraction to 1.35 Å resolution and by neutron diffraction to 2.1 Å resolution. A consensus model of the water molecule positions was obtained using oxygen positions identified in the electron density map determined by X-ray diffraction, which were verified by comparison to D_2O–H_2O difference neutron scattering density. Six of 184 water molecules in the X-ray structure, all with B-factors greater than 50 Å^2, were found to be spurious after comparison with neutron results. Roughly two-thirds of the water of hydration expected from thermodynamic data for proteins was localized by neutron diffraction; approximately one-half of the water of hydration was located by X-ray diffraction. Polar regions of the protein are well hydrated, and significant D_2O–H_2O difference density is seen for a small number of water molecules in a second shell of hydration. Hydrogen bond lengths and angles calculated from unconstrained refinement of water positions are distributed about values typically seen in small molecule structures.**

Solvent models found in seven other bovine trypsin and trypsinogen and rat trypsin structures determined by X-ray diffraction were compared. Internal water molecules are well conserved in all trypsin structures including anionic rat trypsin, which is 65% homologous to bovine trypsin. Of the 22 conserved waters in trypsin, 19 were also found in trypsinogen, suggesting that they are located in regions of the apoprotein that are structurally conserved in the transition to the mature protein. Seven waters were displaced upon activation of trypsinogen. Water structure at crystal contacts is not generally conserved in different crystal forms. Three groups of integral structural water molecules are highly conserved in all solvent structures, including a spline of water molecules inserted between two β-strands, which may resemble an intermediate in the formation of β sheets during the folding of a protein.

Key words: protein folding, protein structure, hydrogen bond, serine protease

INTRODUCTION

The nature of interactions at the interface between protein and solvent is essential to an understanding of protein stability, folding, and function.[1] A good model of this interaction is required for ligand or drug design, any accurate macromolecular modeling, and for energy calculations. However, the changing structure of liquid water, and water at interfaces has been exceptionally difficult to analyze experimentally or model theoretically. Current models for liquid water are rooted in the proposal by Bernal and Fowler[2] that liquid water is an extended and irregular four-coordinated state. Highly distorted hydrogen bonds undoubtedly occur in water, owing to the relative lack of directionality of the oxygen lone-pair orbitals, and the dominance of entropy over enthalpy above the melting temperature of ice.[3] The current picture of water is one in which there are substantial deformations of hydrogen bonds and in which electrostatic and steric interactions serve to orient water both in bulk and at protein interfaces.

The water at the interface between macromolecules and bulk solvent has thermodynamic properties distinct from bulk water. It is more localized than bulk water and its structure and dynamics

Received March 18, 1991; revision accepted May 20, 1991.

Address reprint requests to Dr. Robert M. Stroud, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143-0448.

Jim Hurley is now at Institute of Molecular Biology, University of Oregon, Eugene, OR 97403.

Thomas Earnest is now at Donner Laboratory, Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720.

Abbreviations used: (D–H), (deuterium–hydrogen); BA, benzamidine; MIP, monoisopropylphosphoryl; DIP, diisopropylphosphoryl; GBZ, guanidinobenzoyl; BPTI, bovine pancreatic trypsin inhibitor.

have been analyzed by spectroscopy, theoretical calculations, and diffraction. Calorimetric and spectroscopic analyses[1,4,5] and computer simulations[6] of the hydration of globular proteins lead to a picture of a partially immobilized monolayer of water at the protein surface. The dependence of the heat capacity and vapor pressure of protein-bound water on relative hydration shows that globular proteins are closely associated with approximately 0.3–0.5 g water of hydration per g protein,[4] enough for monolayer coverage of the protein.[7] This water does not freeze in the usual sense, since protein–solvent geometry may dominate over solvent–solvent interactions. It has a reduced heat capacity, an increased rotational correlation time, and a reduced diffusion coefficient since it is partially immobilized. The water present in a protein crystal ranges from 1 to 10 times this amount[8] and is typically twice this amount. Yet the protein–protein interfaces occupy a significant part of the protein surface. Thus, the protein crystal contains significantly more water than that associated with surface hydration alone, present as bulk solvent in the crystal.

Hydration is a stepwise process, with hydration of ionizable groups occurring first, followed by solvation of polar groups and finally apolar groups.[5] During dehydration, enzymatic activity is lost when approximately one-half of the total water of hydration is removed,[5] presumably implying a structural change in the protein when there is less than full solvation of hydrogen bonding groups. The water molecules of hydration are presumably more tightly associated with the protein, and are the ones most likely to be observed by diffraction.

Diffraction methods provide the highest resolution model for a time-averaged structure of interfacial water. X-ray diffraction provides the most accurate locations of the centers of electron density at oxygen atoms in water. Neutron diffraction studies of proteins, usually carried out in D_2O to reduce inelastic scattering of hydrogen atoms and generally not at atomic resolution, give peaks centered between the three atoms that all scatter neutrons with similar magnitudes (D = 6.7 F, O = 5.8 F). It provides a threefold higher density for water than for an oxygen atom alone. In D_2O–H_2O difference maps, the scattering by the water oxygen cancels giving (D–H) peaks at the deuterium positions in a very high resolution map, or a peak centered between the two deuterium atoms in a moderate resolution (1.5–3.5 Å) map.[9,10] Because of the negative scattering length of hydrogen (−3.85 Fermi) the D–H density remains about twice that of an oxygen atom. The difference in position of X-ray and neutron water peaks gives information on the disposition of the oxygen atom relative to the deuterium. The identities of the donor and acceptor in a hydrogen bond, which are often ambiguous in an X-ray or neutron structure alone, can thus be resolved. In concert, X-ray and neutron diffraction methods also provide an independent assessment of the criteria for reliability of water loci determined by either procedure alone. By comparing X-ray with neutron density maps, we use two independent observations that together give the more complete description of protein-solvent interaction around a protein of 25,000 daltons in size. The positions of water oxygens, and the orientation of water hydrogens obtained from comparison with neutron diffraction densities in trypsin provides an opportunity to assess the relative importance of steric interactions and hydrogen bonding in the hydration of proteins.

Since water molecules interact with proteins through noncovalent forces, there are large thermal fluctuations and statistical disorder of the water structure. While thermal B-factors for water molecules are generally large, and water molecules do not generally contribute to diffraction beyond ~2.0 Å, high resolution data are still required to refine the protein structure adequately, thus minimizing the noise level and giving an accurate electron-density calculation in the solvent region of the crystal. To avoid an unreasonable ratio of parameters to observations, large numbers of water molecules can be included in refinement only at high resolution.

The X-ray structure of trypsin (MIP-inhibited) is one of the most completely refined for a protein of its size, 24,000 daltons.[11–13] At 1.34 Å resolution, the residual is R = 15% for a highly restrained structure. The neutron structures in D_2O at 1.8 Å resolution, R = 19.1%,[14] and in H_2O at 2.1 Å, R = 19.3%[10] have also been described. We have determined structures for three different crystal forms of bovine trypsin,[13,15,16] for bovine trypsinogen,[15,17] and for rat trypsin,[18] and many related serine protease structures have been determined elsewhere.[19] The solvent occupies between 40 and 60% of the unit cell in the crystal forms of trypsin and trypsinogen we compare here.

Although in favorable cases, modified substrates or replacement of amino acid residues can provide information on the role of specific water molecules,[20] no general method exists for the removal or replacement of an individual bound water molecule. Any water that is structurally or functionally relevant is expected to be found in all well-determined structures of a particular protein in different crystal forms. The availability of three different structures of bovine trypsin in different packing arrangements, of rat trypsin, and of the homologous enzymes such as chymotrypsin[21,22] and elastase[23–25] provide a test for hypotheses concerning any conserved function of specific water molecules. Thus, the solvent structures for the serine proteases place limits on hypotheses on the function, or structural roles of bound water molecules. An understanding of the precise role of individual water molecules in protein stability and function can thus be approached.

METHODS

Further Refinement of Trypsin and Water Structure Determination by X-Ray Diffraction at 1.34 Å Resolution

The structure of MIP–bovine trypsin in the first orthorhombic crystal form had been solved and refined by difference Fourier methods to a residual of R = 15% at 1.5 Å by Chambers and Stroud.[11–13] The 1.5 Å structure included 164 refined water molecules initially identified from (Fo–Fc) α^{calc} maps. Subsequent to this analysis, the 3,371 strongest reflections between 1.5 and 1.34 Å resolution, corresponding to 24% of the total theoretical data in this resolution shell, were measured as previously described[11] and included in the refinement at 1.34 Å. The residual for the 1.5 Å model before refinement against the 1.34 Å data was 18.2% for the 30,040 independent reflections with resolution >7 Å and Fo > 2σ(Fo). After cycles of difference Fourier refinement,[12] and restrained least-squares refinement[26] the set of high-B-factor water molecules was edited by examination of (Fo–Fc) α^{calc} "omit" maps using FRODO.[27] Water molecules that refined to temperature factors of greater than 35 Å^2 or occupancies of less than 6 electrons were retained only if density for them reappeared in subsequent maps phased without inclusion of these molecules both in the refinement cycles, and in the map calculation. This serves to limit bias in selection of the less well-ordered water molecules. After removal of the less reliable waters a new (Fo–Fc) α^{calc} map was computed from which additional peaks were chosen and refined. This procedure was repeated several times until a set of consistent water molecules was obtained.

Density Modification to Remove Noise in the Solvent Region by Flattening Under the Protein

In an attempt to refine the disordered water structure in the X-ray crystal structure, we effectively flattened the residual error density under the protein. Procedurally this was done by holding **Fc**'s for the protein and the water molecules that had B < 40 Å^2, fixed, and iteratively refining the solvent structure factors $\mathbf{f}_{solvent}$, to see whether all the remaining discrepancy between | **F'c** | 's and the Fo's could be accommodated by density, even though it might lack discrete peaks, in the solvent region alone. The iterative refinement procedure was as follows:

1. **Fc**'s were calculated from coordinates and temperature factors for the protein structure, water molecules, and bulk solvent represented by a constant density of 0.35 *e*/Å^3. A temperature factor of 40 Å^2 was applied to the bulk solvent density to minimize discontinuities in density at the protein surface.
2. A difference Fourier synthesis was calculated from terms (Fo- | **F'c** |)α_{calc}. The density outside a radius of 1.75 Å from external protein atoms and ordered water molecules and outside a radius of 2.5 Å from internal protein atoms was back-transformed to generate "residual solvent density" structure factors, $\mathbf{f}_{solvent}$.
3. A new set of $\mathbf{F'c} = \mathbf{Fc} + \mathbf{f}_{solvent}$ were calculated and the | **F'c** | 's scaled to the Fo's prior to calculation of another difference Fourier map as in no. 2. The procedure was repeated until convergence.

This procedure, which is not biased from selection of water molecules for inclusion in the phasing model, provides a test of whether the discrepancy between | **F'c** | and Fo is due to inadequate representations of solvent or protein. The results of this test do not contribute to the water structure we discuss here.

Consensus X-Ray/Neutron Water Structure Model

The neutron solvent difference map method is described in the previous paper.[10] Evaluation of the sensitivity of the method indicated that positions of water molecules having as low as 10% occupancy can be identified accurately.[10] A consensus X-ray/neutron water structure model was developed by comparing the X-ray determined sites to the density in the D_2O–H_2O difference density maps. X-ray water positions whose density in the X-ray map intersected neutron solvent density above 3 σ were retained in the water structure model. Peaks that lay outside D_2O–H_2O difference density were retained in the consensus structure if they met at least two of the following criteria: (1) they appeared in other independently refined trypsin structures; (2) (Fo–Fc) α^{calc} X-ray difference maps phased without them had strong, well-resolved, discrete spherical density at their refined positions; and (3) they could make stereochemically reasonable hydrogen bonds to the protein or to well-ordered water molecules. (Fo–Fc) α^{calc} maps were then examined for additional peaks that lay in D_2O–H_2O difference density. These water molecules were refined and checked in "omit" maps in which the waters in question were not included in the phase calculation, thereby reducing bias in their favor.

Correspondence of Water Positions From Different Sets of Coordinates

For evaluation of the correlation between X-ray and neutron waters, it is necessary to estimate the errors in X-ray coordinates. The expected random errors in atomic positions determined from an X-ray structure is a function of B-factor according to

$$\sigma_{err}(B) = 3/4\ R^*r^*\{0.00107B^2 - 0.01493B + 0.264\} \quad (1)$$

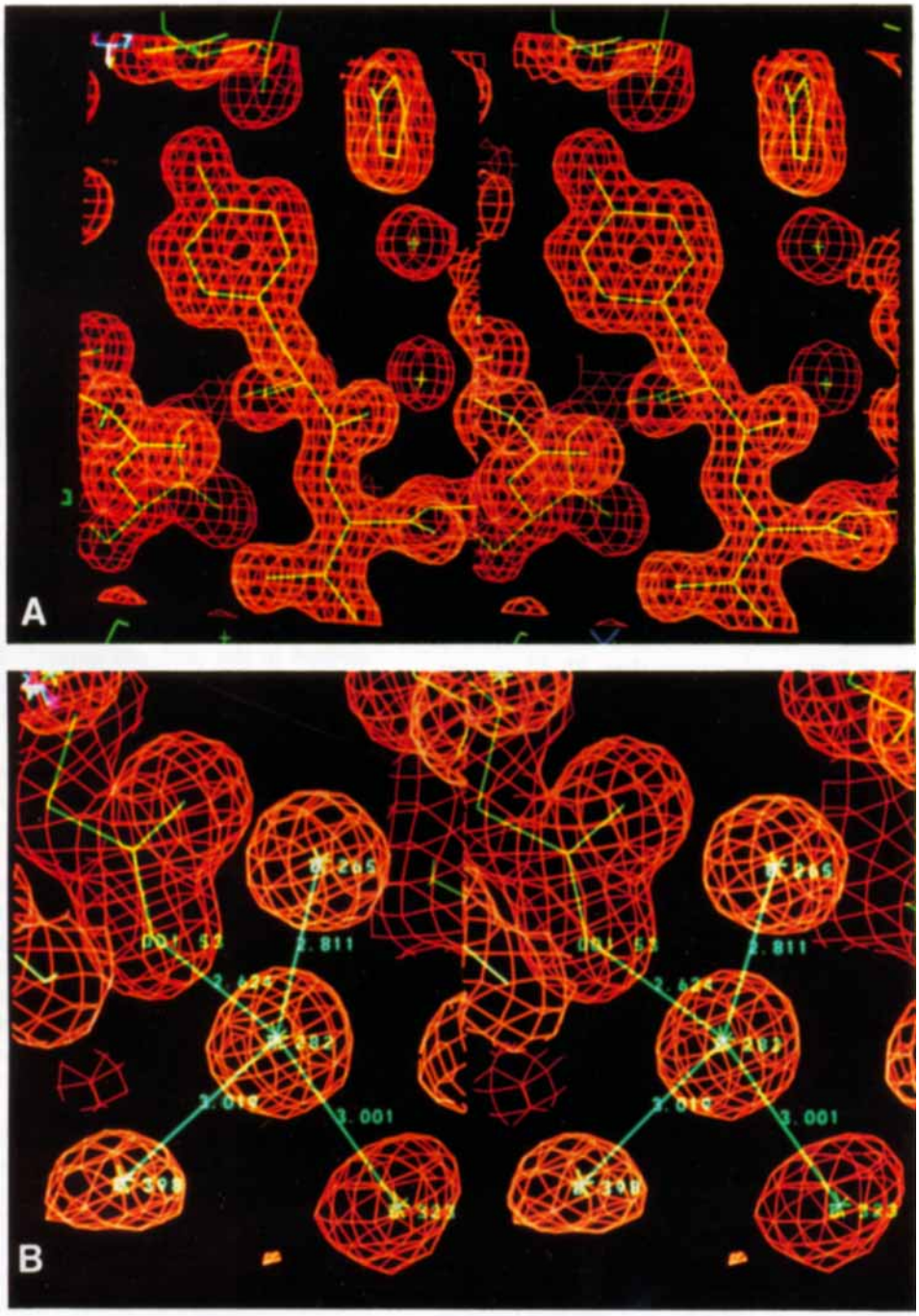

Fig. 1. Crossed-eye stereo plots of 1.35 Å (2Fo–Fc) α^{calc} maps for MIP–trypsin. Waters stand out as resolved spherical density and generally had good hydrogen bonding geometry.

where R = (Σ | Fo | – | **Fc** |)/Σ | Fo | for Fo's > 2 σ(Fo), r = resolution in Å, and B is the thermal factor in Å^2.[11,28]

To evaluate whether we could use displacement of X-ray waters from D_2O–H_2O density peaks to orient hydrogens on the waters, we did the following test. Hydrogen positions were calculated for X-ray waters in arbitrary orientations. The waters were then rotated by hand using FRODO so that the center of a line connecting the hydrogens was as close to the center of the D_2O–H_2O difference peak as possible. Based on the orientation of the water hydrogens relative to the side-chain heavy atoms, the water was classified as a hydrogen bond donor or acceptor to the side chain. The hydrogen positions on the side chain as determined by neutron diffraction

were then checked for consistency with this assignment.

In order to identify buried waters, solvent accessible surfaces[29] for the hydrated structures were calculated by the algorithm of Connolly[30] using a 1.4 Å radius water probe. A water was considered internal if it had no solvent accessible surface area. A water was considered a crystal contact water if it was within 4.0 Å of at least two symmetry-related protein molecules.

An overlap coefficient between two sets of water molecules was defined by

$$C = (1/N_2)\,(\Sigma_{i=1,N1}\Sigma_{j=1,N2} \quad e^{-|x_{i1}-x_{j2}|^2/s^2}) \quad (2)$$

where s = 1.5 Å, **x** denotes a water position vector, and N_1 and N_2 are the number of water molecules in the larger and smaller sets of water molecules, respectively. C is proportional to the overlap integral of the two water structures when each water is represented by a Gaussian envelope. The overlap coefficients were used to evaluate similarities of the X-ray water structures we compared.

RESULTS AND DISCUSSION

Comparison of X-Ray and Neutron Results

Two hundred ninety-one peaks in the neutron solvent difference density were assigned as water molecules.[10] Approximately one-half the peaks (140) found were discrete and resolved water molecules in well-defined neutron density. In many cases, the other peaks were in extended neutron density, and served effectively as markers of local maxima in extended partially ordered water networks. The geometric relationships of the water molecules in these diffuse water channels contain valuable information for an assessment of protein hydration.

A total of 184 water molecules were chosen from X-ray diffraction electron density using (Fo–Fc) α^{calc} maps (Fig. 1). The neutron (D–H) peaks centered on the center of mass of deuterium positions are expected to be about 0.6 Å from the X-ray peak corresponding to the oxygen positions. Figure 2 illustrates the correspondence of the discrete X-ray water molecules (orange density) with D_2O–H_2O neutron difference density. Maxima in the D–H density fall in between deuterons and are systematically displaced from the water oxygen positions; 86 X-ray water molecules matched peaks in the D_2O–H_2O difference density to within $\{0.6 + 2^*\sigma_{err}(B)\}$ Å, where 0.6 Å is the expected shift between corresponding peaks in (Fo–Fc) α^{calc} X-ray maps and D_2O–H_2O difference maps and σ_{err} is the estimate of error in the X-ray water oxygen position (see Methods). This number represents a minimum estimate of the number of matching water molecules because estimated errors for the neutron water positions were not considered. Peaks in the more diffuse and extended D_2O–H_2O difference density (Fig. 2B) in particular, did not necessarily represent discrete water molecules and were therefore not strictly comparable to discrete positions for X-ray water molecules; 72 additional water molecules were in density that intersected or bordered D_2O–H_2O difference density, even though they were greater than $\{0.6 + 2^*\sigma_{err}(B)\}$ Å from a D_2O–H_2O difference density peak. These were judged consistent with the neutron diffraction results giving a total of 158 "common" water molecules (86 + 72).

Six water molecules that did not lie in D_2O–H_2O difference density and had B-factors greater than 50 Å^2 and either made no contacts with the protein or were ≤3.0 Å from a carbon atom were judged spurious and were eliminated from the list. Twenty water molecules were not consistent with D_2O–H_2O difference density but had strong electron density in the X-ray maps and reasonable hydrogen bonding geometry. All these waters had refined temperature factors of less than 47 Å^2 and occupancies of at least four electrons. It is possible that the density for some of these water molecules was inadvertently masked out during the solvent refinement procedure used for the neutron diffraction analysis. Because they met our reliability criteria (see Methods), these water molecules were retained in the consensus water structure.

Thirty-four peaks in the electron density map that were not included in the water structure determined by X-ray diffraction overlapped D_2O–H_2O difference density and were added to the consensus water list. Six of these water molecules had occupancies of less than four electrons (occupancy <0.5) and would probably not have been interpreted as water during X-ray refinement without neutron diffraction maps available for comparison. B-factors for the final set of 211 water molecules ranged from 8 to 60 Å^2 and averaged 28 Å^2. Refined occupancies for the water molecules averaged 7 electrons out of a total of 8 for a fully occupied oxygen, and only seven were less than 4 electrons. Since it is not possible to reliably co-refine the occupancies and temperature factors of waters in very weak density, we fixed the occupancies of these seven waters at 4 electrons and refined only their B-factors. As shown in the histogram of B-factor and occupancy distributions for the water molecules (Fig. 3) almost one-half the water molecules refined to full occupancy. All but one of the 19 waters that were completely conserved in the eight compared trypsin X-ray structures refined to full occupancy. The occupancy of 7.2 electrons for the exception was fixed at 8 electrons, since it seemed unlikely that any of the conserved solvation sites would be less than fully occupied. There was little correlation between occupancies and B-factors for the water molecules, except that most waters with B-factors <20 Å^2 were fully occupied (Fig. 3C).

The estimated average noise level in difference maps,[31] which is the root mean square (rms) electron density difference between one-half the true

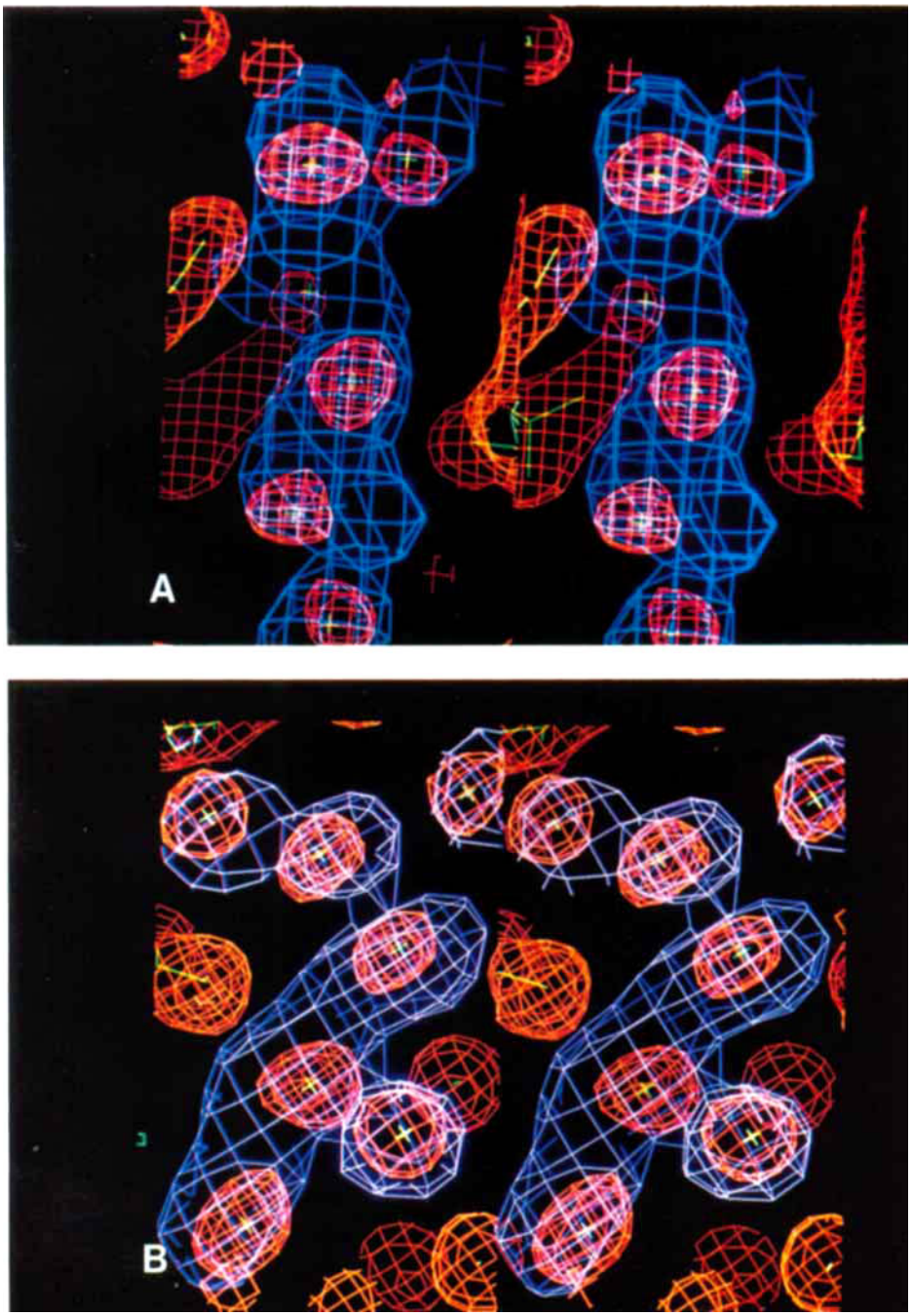

Fig. 2. Crossed-eye stereo plot of the X-ray structure of MIP–trypsin plotted against a (2Fo–Fc) α^{calc} X-ray map (in orange) and D_2O–H_2O difference density (in blue). The weak D_2O–H_2O difference density in **B** is not resolved into discrete peaks but is partially filled with discrete X-ray water molecules.

structure (as would be seen in a difference map) and the experimentally derived structure, calculated by

$$<\Delta\rho^2> = 1/V^2\Sigma_{h=0-\infty}\Sigma_{k=-\infty-\infty}\Sigma_{1=-\infty-\infty}\{\Delta F^2 + \delta^2\} \quad (3)$$

for data between 7.0 and 1.34 Å resolution, where $\Delta F = |Fo| - |Fc|$ and $\delta = \sigma(Fo)$, was 0.028e/Å^3. This is well below the maximum level of $\Delta\rho$ = 0.75 electrons for detecting a 2-electron occupancy water in a 1.35 Å resolution difference map. The rms deviation in (Fo–Fc) α^{calc} electron density was 0.07 e/Å^3, and the maximum density was 0.40 e/Å^3. Thus, it may be possible to identify more low occupancy water molecules from the X-ray structure than we

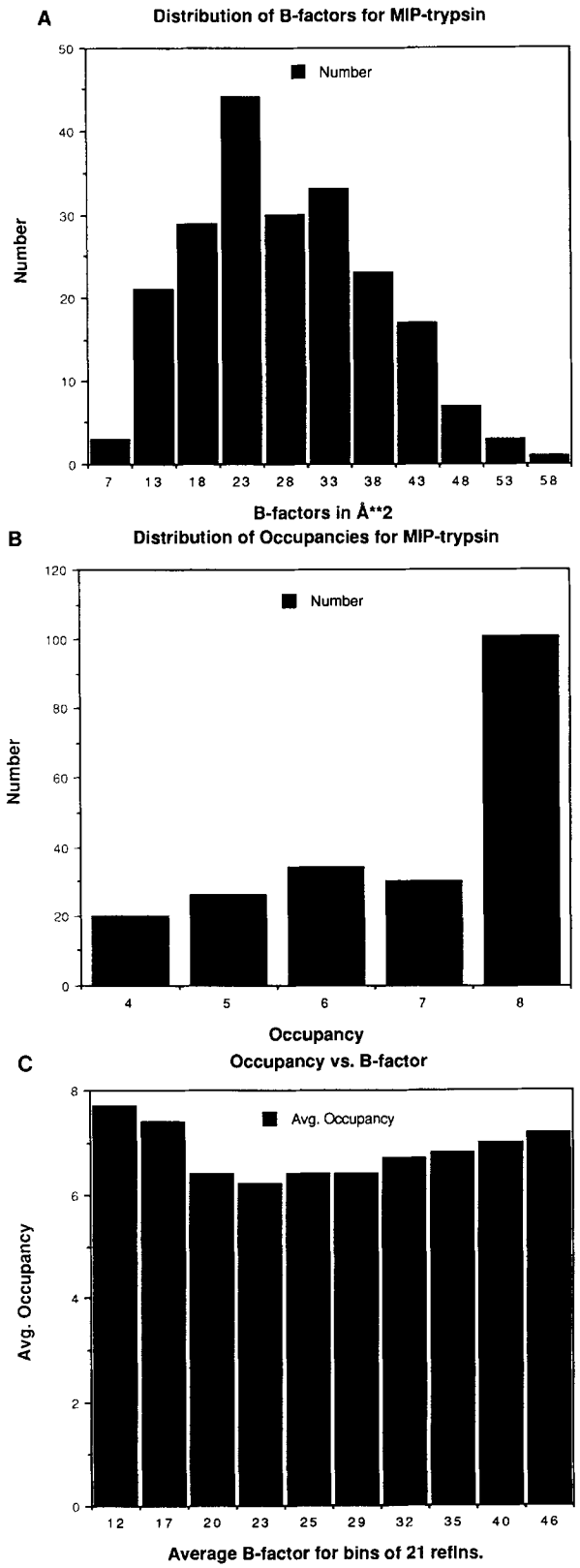

Fig. 3. **A.** Histogram of number of water molecules in the consensus list with B-factors within a given 5 Å^2 range. **B.** Plot of number of water molecules in the consensus list with a given occupancy (in electrons). **C.** Plot of occupancy vs. B-factor for water molecules in the consensus list.

TABLE I. Refinement Statistics for MIP–Trypsin

No. reflections (7–1.34 Å):	30,041
No. restraints:	11,921
No. variables:	7,604
R-factor:	15.2%
Average B-factor:	16 Å^2
rms deviations	
Bond distances:	0.017 Å
Angle distances:	0.052 Å
B-factors, adjacent Main-chain atoms:	1.3 Å^2
B-factors, adjacent Side-chain atoms:	3.6 Å^2
rms deviation in ΔF map:	0.07 electrons
Largest peak in ΔF map:	0.40 electrons
$<\rho 2>^{1/2}$ in ΔF map:	0.028 electrons

accepted. The R-factor for the refined 1.34 Å structure with all 211 water molecules was 15% for reflections with Fo > 2σ(Fo). Table I summarizes the refinement statistics.

The consensus water list accounts for approximately two-thirds of the water located in neutron diffraction density and describes the solvation of 3,104 Å^2 of the 6,569 Å^2 solvent accessible surface of trypsin in the crystal. Even though most of the waters were hydrogen bonded to the protein, 1,387 Å^2 (45%) of the surface buried by the water was hydrophobic. In contrast to the water structure determined by neutron diffraction, only 16 water molecules in the consensus water structure, 7.5% of the total, were not in direct hydrogen-bonded contact with protein polar groups. All but two of these water molecules were in contact with nonpolar protein atoms. Fifteen of the 16 second shell waters are hydrogen bonded to other waters, often participating in water networks that bridge hydrophobic surface atoms. One water contacts only hydrophobic atoms in the X-ray structure but is seen to be part of a disordered water network in the D_2O–H_2O difference maps. Most of the water molecules not in hydrogen bonded contact with the protein surface are presumably sufficiently mobile that they remain undetected in the water structure determined by X-ray diffraction.

Does Further Solvent Density Modification Improve the Interpretation of Solvent Density in X-Ray Maps? What Residual Errors in the X-Ray Model Remain?

Implementing the solvent refinement scheme described in the second section under Methods lead to convergence with an R = 4%, approximately the error in the Fo's, and density in the protein region of (2Fo–Fc) α^{calc} maps improved slightly. However, the solvent region contained several relatively strong irregularly shaped density peaks that were not consistent with neutron diffraction solvent density.

Therefore, the scheme was not used to define the disordered solvent structure.

Although a crystallographic residual of 15% has become acceptable for a protein structure, the estimated uncertainty in the measured structure factor amplitudes is usually less than 5%. Modifying solvent density alone to reduce the large discrepency between calculated and observed structure factor amplitudes resulted in a solvent structure which was inconsistent with the neutron diffraction results and therefore probably incorrect. We conclude that inadequate modeling of the protein is partly to blame for the high crystallographic residuals.

Hydrogen Bonding in Water–Protein Interactions

Position and orientation of water molecules in the D_2O–H_2O difference density

Because water locations in the D_2O–H_2O difference map are given by the positions of the hydrogens (D–H) while in the X-ray case water positions are defined by the oxygen alone, the centers of the peaks in the respective maps will not coincide. Figure 2, which shows the electron density of X-ray waters superimposed on D_2O–H_2O difference density, clearly illustrates this. This characteristic peak displacement in the neutron maps away from the oxygen position and toward the hydrogen positions results in apparent calculated hydrogen bond distances that are systematically short where water is a hydrogen bond donor and that appear systematically long where the water acts as the acceptor (Fig. 4). This effect was quantified by comparing X-ray and neutron coordinates for 60 hydrogen bonded water molecules. The average difference between the sets of coordinates was 0.67 Å (±0.24). In the neutron case, hydrogen bonds in which the water is the donor appear systematically short (2.52 ± 0.33 Å), whereas hydrogen bonds in which the water molecules act as acceptors are on average systematically long (3.32 ± 0.23 Å).

In principle, hydrogens on water molecules may be oriented based on the displacement of the X-ray water molecules from the centers of the D_2O–H_2O difference peaks. Savage was able to define water networks in crystals of vitamin B_{12} coenzyme by comparing high resolution neutron and X-ray diffraction density maps.[32] The trypsin D_2O–H_2O difference density, at 2.1 Å resolution, is not sufficiently resolved to orient the water hydrogens precisely, particularly the extended density. No attempt was made, therefore, to define detailed water networks or analyze distortion in hydrogen bonding geometries by defining hydrogen orientations for the X-ray waters.

To assess the ability to orient water molecules, we checked the density around 37 solvated Asn, Gln, Tyr, Thr, and Ser side chains to see whether we could predict the donors and acceptors in the hydrogen bonds based on displacement of X-ray oxygen positions from D_2O–H_2O peaks. We were able to predict correctly whether a water was a hydrogen bond donor or acceptor to a particular side chain in 30 of 49 cases. For three other hydrogen bond interactions, the prediction was not consistent with the side chain hydrogen as determined by neutron diffraction. These hydrogen bonds were to serine hydroxyls. For 16 of the 49 hydrogen bond interactions, we were unable to determine the orientations of the water hydrogens because neutron density was either absent or poorly resolved, or because there was no obvious displacement of the oxygen peaks from the centers of the D_2O–H_2O difference peaks. In general, the donor-acceptor roles of waters in hydrogen bonds were readily assigned for high occupancy waters in resolved D_2O–H_2O difference density (see Fig. 4b).

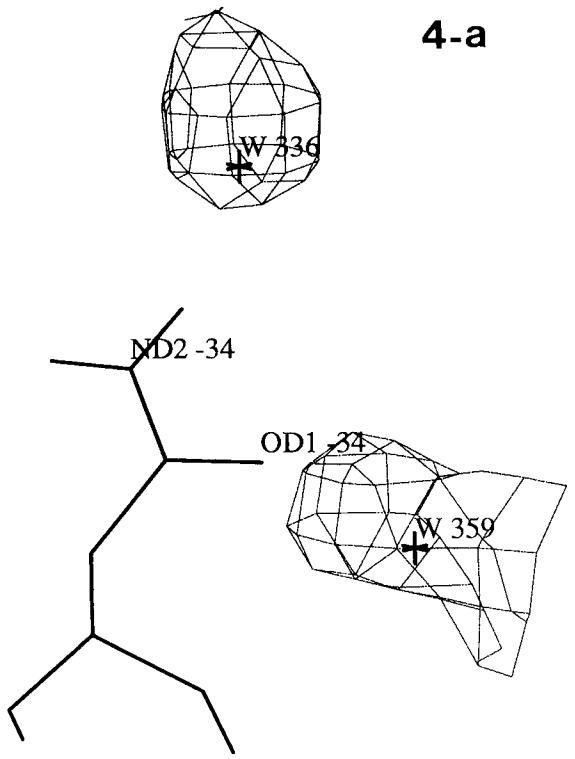

Fig. 4a. Legend appears on page 211.

Hydrogen bonding geometries in the consensus water structure

Water molecules in the primary hydration sphere of trypsin are extensively hydrogen bonded, with an average of 2.6 potential hydrogen-bonding neighbors per water in the consensus list (Table II). The patterns of hydration are typical of those observed in high-resolution X-ray structures[33–35]; 21% of main-chain carbonyl groups have at least two neigh-

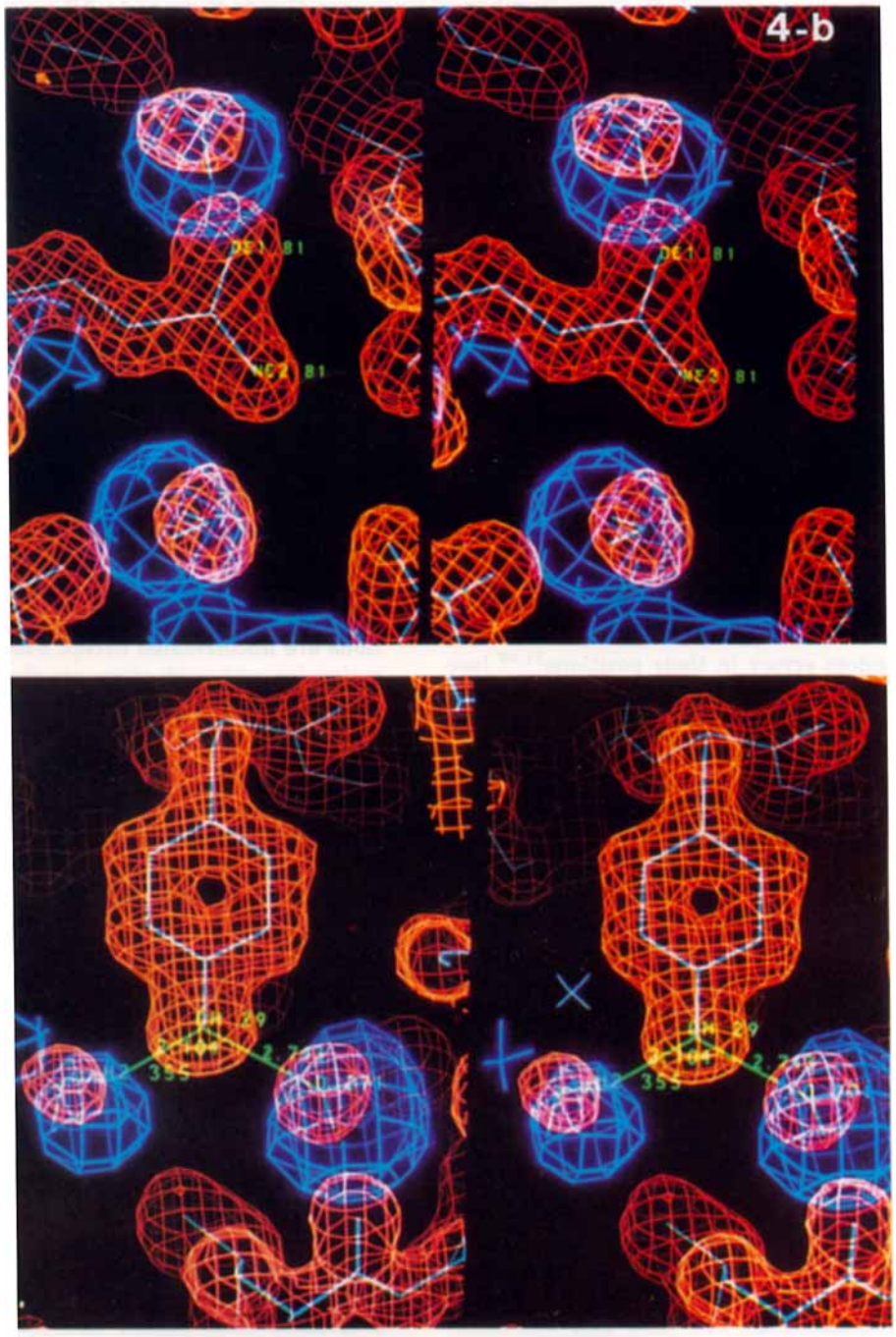

Fig. 4. **a.** Displacement of D_2O–H_2O difference density from water oxygen positions gives rise to systematic errors in hydrogen bond length calculations using neutron water peaks. **b.** This displacement may be used for orienting waters, thus determining whether they are hydrogen bond donors or acceptors. Two crossed-eye stereo plots of X-ray (2Fo–Fc) α^{calc} density superposed on D_2O–H_2O solvent density show examples in which the hydrogen bond donor or acceptor roles of waters could be deduced in this manner.

boring hydrogen-bond donors, reflecting the ability of this moiety to form hydrogen bonds with each oxygen lone pair orbital. By contrast, only 3% of main-chain amide hydrogens participate in more than one hydrogen bond. Among protein side chains, carboxylate oxygens are the most highly hydrogen bonded, with 1.5 hydrogen bonds per oxygen acceptor.

The distributions of hydrogen bond distances and angles between main chain carbonyls and amides and the consensus water molecules (Fig. 5) follow the same trends observed by Baker and Hubbard[36]

TABLE II. Mean Number of Hydrogen Bonds Formed to All Potential Partners and Fraction With More Than One Hydrogen Bond, by Functional Group*

Group	Mean No. of H-bonds	Fraction with N > 0 H-bonds	Fraction with N > 1 H-bonds
Water	2.63	1.00	0.84
Main-chain CO	1.00	0.78	0.21
Side-chain COO^-	1.32	0.69	0.50
Side-chain CO	1.04	0.73	0.35
Side-chain OH (O)	0.69	0.44	0.13
Main-chain NH (H)	0.81	0.78	0.03
Side-chain NH_2 (H)	0.50	0.40	0.00
Side-chain OH (H)	0.30	0.24	0.00

*Hydrogen bonds are defined by an O–O distance 2.3 Å < d < 3.3 Å or an H–O distance of 1.6 Å < d < 2.6 Å. Hydrogens are model-built based on known stereochemistry, except for hydroxyl hydrogens, which were obtained from the protein structure determined by neutron diffraction.

in their analyses of 13 well-refined protein structures. As noted by Baker and Hubbard,[36] the distributions of hydrogen bond distances and angles are broader than those seen for small molecules. The broader distributions may be attributed to the higher B-factors of the protein waters, which result in larger random errors in their positions[11,28] [see equation (1) for estimate of random error from B-factors]. No contraints on hydrogen bond lengths were applied during refinement.

The average C = O . . . OW hydrogen bond distance was 2.9 ± .21 Å, slightly greater than the average O . . . O hydrogen bond length of 2.76 Å seen for water. However, the asymmetric distribution of hydrogen bond distances (Fig. 5A) peaks at 2.8 Å, and the distributions of bond lengths less than 3.1 Å are fairly symmetrically distributed about this distance.

The average hydrogen bonding angles to protein oxygens (Table III) range from 120° to 133°, as expected based on the sp^2 hybridization of the carbonyl oxygen. Hydrogen bonding angles to amide nitrogens (Table III) are close to 160°, as expected given the ability of this group to form only one linear hydrogen bond. The distributions of main chain hydrogen bond angles, C = O . . . OW (Fig. 5a) and N–H . . . OW (Fig. 5B) were very similar to those reported by Baker and Hubbard,[36] except that the distributions were slightly narrower than those for the thirteen proteins: 88% (compared to 80%) of the C = O . . . OW angles were in the range 110°–150° and 95% (compared to 90%) of the N-H . . . OW angles were in the 140°–180° range. Positions of water molecules forming hydrogen bonds with carbonyl groups are expected to lie near the plane of the carbonyl oxygen. The distribution of angles out of the peptide N-C-O plane (Fig. 5C), analogous to a N–C–O–O(water) torsion angle, but defined between 0° and 90°, show that 84% of water molecules are found within 40° of this plane and 54% are less than 20° out of plane. Therefore, there is some tendency, although not a strong one, to form hydrogen bonds in the plane of the peptide bond.

Clustering of waters

Most water molecules are found in large clusters. Only 41 of the 187 water molecules located on the protein surface hydrate the protein surface at isolated sites. In a simple model in which water positions are uncorrelated except that overlap of water molecules with each other are disallowed, the probability of a water occurring in a cluster is

$$P\,(\text{cluster}) = \Sigma_{i=1,Nwat\text{-}1}\, A_{neighborhood}/(A_{total}\text{-}(i+1)A_{wat}) \quad (4)$$

where N_{wat} is the total number of water molecules, $A_{neighborhood}$ is the area of the zone around a water molecule within which another water must lie to form a hydrogen bond, A_{total} is the total solvent accessible surface area of the protein, and A_{wat} is the area covered by a single water. The summation runs to $N_{wat}{}^{-1}$ to exclude the water under consideration, and A_{wat} is multiplied by i + 1 to take into account the area excluded by all water molecules, including the one under consideration. Based on 6,569 $Å^2$ solvent accessible surface area in the MIP–trypsin crystal form, a 14.5 $Å^2$ ring allowed for hydrogen-bonding neighbors around a given water molecule (2.5–3.3 Å), and a 15 $Å^2$ average area covered by a single water, the calculated probability of finding one or more hydrogen-bonding partners is 54%. Given the simplicity of the model for the probability of random water cluster occurrence, definitive conclusions about cooperativity cannot be drawn. The observation that 78% of the tightly bound surface waters participate in clusters of more than one water, compared to an estimated 54% in the absence of cooperativity, does suggest that cooperative effects may be a driving force in protein hydration.

The type of closed rings found in the structure of the hydrophobic protein crambin[35] was not found in a systematic search of hydrogen-bonded water networks in solvent structure. The presence of a single poorly ordered or partially occupied water in a ring structure, below the threshold of reliable detection

TABLE III. Mean Hydrogen-Bonding Distances and Angles by Protein Functional Group*

	Distance (Å)	Hydrogen bond angle	Out-of-plane angle
Main-chain CO	2.90 ± 0.21	133 ± 16	29 ± 19
Side-chain COO^-	2.77 ± 0.29	120 ± 21	30 ± 23
Side-chain CO	2.85 ± 0.34	129 ± 15	19 ± 15
Side-chain OH (O)	2.92 ± 0.24	120 ± 18	—
Main-chain NH (H)	2.06 ± 0.21	161 ± 14	—
Side-chain NH_2 (H)	2.19 ± 0.22	152 ± 18	

*Peptide and amide hydrogens are placed based on known stereochemistry. Hydroxyl hydrogens have been omitted here because of the conformational heterogeneity described by Kossiakoff et al.[37] Lysine parameters are not shown because this side chain is generally not well ordered, and arginine, histidine, and tryptophan are excluded because of the small number of these residues in trypsin.

in this study, would prevent a ring from appearing as such in a search of water networks. In any case, closed rings of water molecules are apparently not a predominant motif in hydration of trypsin and may be more characteristic of hydrophobic proteins in tightly packed crystals than of globular proteins in general.

Comparison of Solvent Structures Determined by X-Ray Diffraction

To differentiate water molecules that were conserved in trypsin and its close structural homologues, and therefore likely crucial to the structure or function of the protein, from those that are dependent on crystal packing, inhibitor identity, or activation state, we have compared our consensus water list to water structures of seven highly refined trypsin and trypsinogen structures[11,12,16–18,38–41] (Table IV). This structural data base includes three isomorphous crystal structures determined in two different laboratories, to determine reproducibility of water structure; nonisomorphous crystal structures of bovine trypsin to study crystal-packing effects; two bovine trypsinogen structures to evaluate effects of activation state; and a rat trypsin mutant to study effects of side-chain substitutions.

Prior to their comparison, all structures were superposed on the MIP-trypsin model by least-squares fit of backbone and β-carbon positions. We established an overlap coefficient as a means of evaluating the overall similarity between water structures or subsets of water structures. Tables of overlap coefficients used for comparisons of the solvent structures discussed here are in the appendix. The overlap coefficients are highly correlated with percentage of matching waters in the overlapped structures, but also are sensitive to the degree of overlap, or plasticity, of the water structures.

Individual water molecules were judged conserved if they were within 1.5 Å of each other. Water molecules within 1.5 Å in two superimposed structures usually formed hydrogen bonds with the same protein groups, hence can be considered the "same" water. In borderline cases, water molecules were inspected to determine whether they had the same hydrogen-bonding partners. Most of the conserved waters were within 1 Å of each other in the overlapped structures.

The highest correlations between water structures, as determined by the overlap integrals for all possible pairs of water structures are among the three isomorphous trypsin structures: MIP-inhibited[11–13] (MIP), benzamidine-inhibited[38] (BA), and native (NAT) bovine trypsin.[41] The only water of the 62 waters located in BA–trypsin that was not present in MIP-trypsin is in the MIP binding site. Since the BA–trypsin we used in our comparison was determined independently in a different laboratory than MIP–trypsin, the congruence of their water structures indicates the well ordered water molecules of a protein can be reliably determined in a high resolution X-ray study.

Internal water molecules are well-ordered and highly conserved

Positions of internal water molecules are much more highly conserved than the external water molecules. Because the internal water molecules are often integral to the structure, they generally have full occupancy and lower than average B-factors (Fig. 6). The internal water molecules have an average temperature factor comparable to the average temperature factor of the protein: $B = 16$ Å^2 compared to $B = 14$ Å^2 for the protein and $B = 31$ Å^2 for external water molecules. The higher conservation may be due in some cases to the independence of these water molecules from crystal packing effects (discussed below), or the independence from species difference, as internal amino acid residues tend to be better conserved between species than are the external residues.[43] Sixteen of the nineteen water molecules that were completely conserved among the compared structures are internal water molecules in MIP–trypsin.

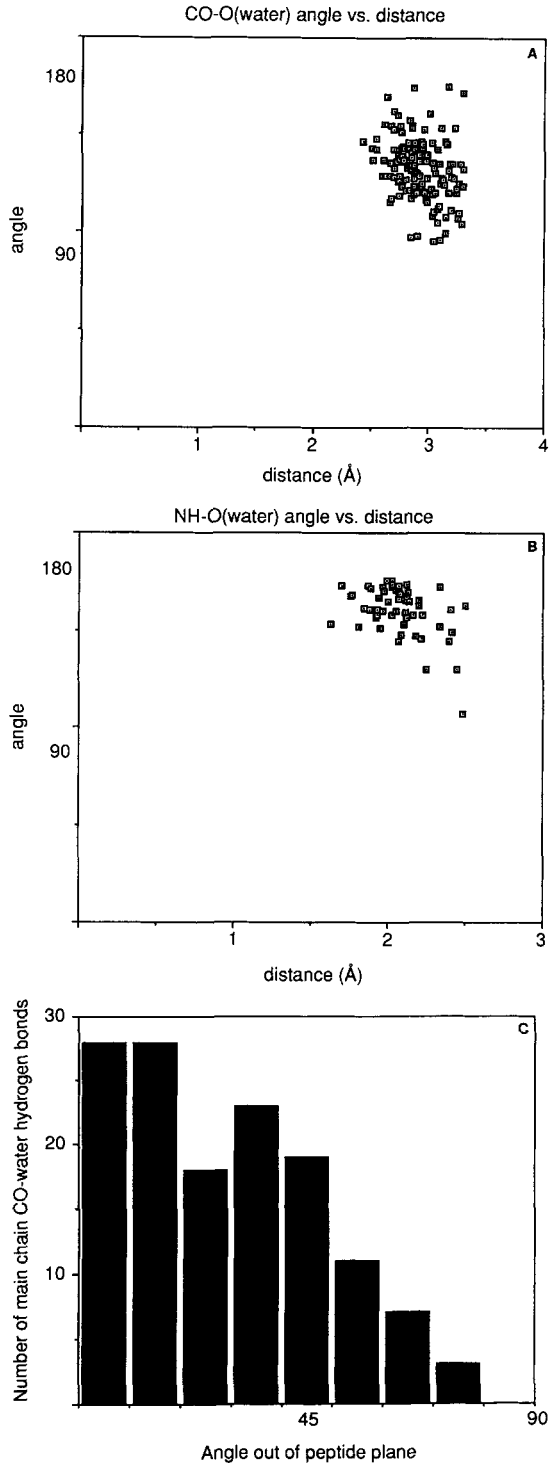

Fig. 5. Distribution of hydrogen bond distances and angles for water molecules forming hydrogen bonds with the main chain of MIP–trypsin. Hydrogen positions are based on the known stereochemistry of the peptide group. **A.** Angle vs. distance for CO–O hydrogen bonds. **B.** Angle vs. distance for NH–O hydrogen bonds. **C.** Angle of hydrogen bond out of the peptide N–C–O plane for water molecules hydrogen bonded to main-chain carbonyl groups.

Internal water molecules are predominantly hydrogen bonded to backbone atoms; 18 internal water molecules (50%) are hydrogen bonded only to backbone amides or carbonyls, while only four are hydrogen bonded exclusively to side chains. Twenty water molecules are hydrogen bonded to two or more backbone atoms.

Of the 30 internal water molecules in a mutant of anionic rat trypsin,[18] S195N, which has 65% sequence homology with bovine trypsin, 29 are present in the MIP–trypsin consensus structure. Side chains hydrogen bonded to the internal water molecules in the two structures are usually conserved or have substitutions that can accommodate hydrogen bonding interactions to the internal water molecules (Fig. 7). Because the conserved waters are adapting to different side chain types in rat compared to bovine trypsin, they overlap less well than the conserved waters of the bovine trypsin structures and consequently have a lower overlap coefficient (0.7 compared to 0.8). The average difference in the positions of conserved waters in the overlapped structures of MIP–trypsin and rat trypsin is 0.35 Å, compared to 0.15 Å for the conserved internal waters of MIP–trypsin and BA-trypsin.

The only incidence in rat trypsin of a side chain type that does not accommodate the bovine internal water structure is the substitution of His for Asp at residue 71. Histidine displaces a water in bovine trypsin that hydrogen bonds to NH-23 and O-52. In rat trypsin the peptide plane of residue 22 is rotated so that O-22 points in the opposite direction and hydrogen bonds to the histidine side chain while NH-23 hydrogen bonds to an external water. Absence of the internal water is therefore accompanied by a conformational change in the vicinity of the water (Fig. 8).

The poorest correlations among internal water molecules are those between the trypsinogens[15,17,40] and the other structures. Trypsinogen contains a very mobile region, the activation domain, which was not modeled by Fehlhammer et al.[40] in their unliganded trypsinogen and which was built in a different conformation from the mature protein in unliganded trypsinogen determined by Kossiakoff et al.[15] and in the DIP–trypsinogen structure.[17] The seven internal water molecules in DIP–trypsinogen that are not found in the trypsin structures lie in this mobile region of the apoprotein.

External water molecules and crystal-packing effects

Water molecules in the region between symmetry-related protein molecules comprise about one-half of the external water molecules in most of the crystal structures examined. Although these water molecules have approximately the same B-factors as external water molecules not at a crystal contact, for nonisomorphous structures they are much more

TABLE IV. Trypsin Structures Included in Comparisons

Name	Resolution	R-factor	No. of water molecules	References
MIP–trypsin	1.35	.15	211	11,12,13
GBZ–trypsin	2.00	.18	100	16
BA–trypsin	1.70	.18	62	38
Trypsin/BPTI	1.9	.19	157	39
DIP–trypsinogen	2.1	.18	73	15,17
Trypsinogen	1.8	.20	93	40
Native trypsin	1.7	.18	84	41
Rat trypsin	2.3	.13	149	18

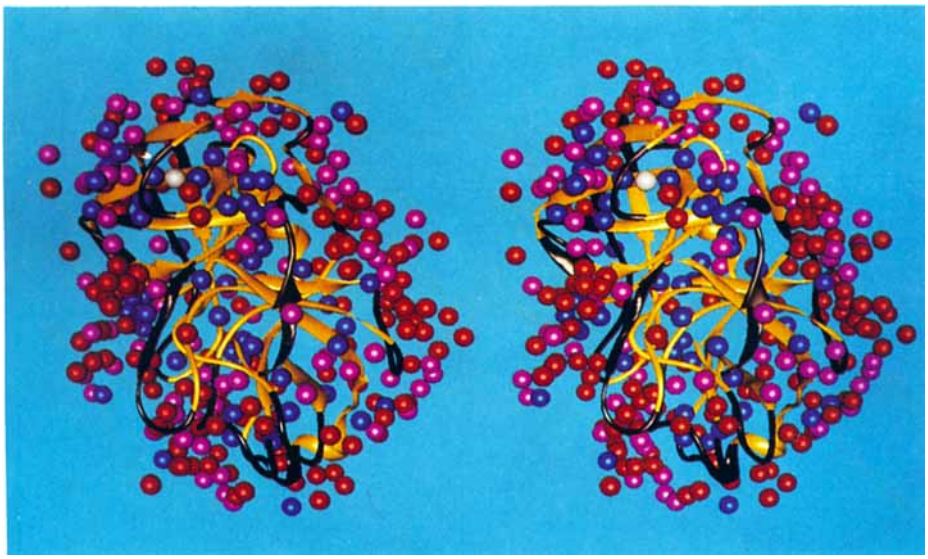

Fig. 6. Crossed-eye stereo ribbon[42] drawing of MIP–trypsin with water molecules shown as spheres of 1.8 Å radii color-coded by temperature factor. The blue spheres represent water molecules with the lowest B-factors (B<20 Å^2), magenta spheres represent water with B-factors between 20 and 35 Å^2, and red spheres represent water molecules with B-factors >35 Å^2. The crystal contact regions are colored black in the ribbon drawing.

poorly conserved. The variation in the correlation of the crystal contact water molecules is much more dramatic than the variation in correlation of external noninterface water molecules. In general, differences in crystal packing, more than any other factor, including species variation, accounts for variability in the solvent structure.

Networks of water molecules link hydrogen-bonding sites on the neighboring protein molecules. Hydrogen-bonding sites on the protein surface at the crystal interface regions are highly variable among the different crystal forms of trypsin. Distributions of hydrogen bond distances and angles for the interface waters alone were similar to the overall distributions (Fig. 5); therefore, there is no evidence that the hydrogen bonding of crystal contact water molecules is strained.

An exception to the nonconservation of crystal contact water molecules between nonisomorphous structures appears between trigonal and orthorhombic crystal forms. The correlations of the interface water molecules of the two trigonal trypsinogen structures[15,17,40] with water molecules of the three isomorphous orthorhombic structures (MIP-[11,12,13], BA-[38], and native bovine trypsin[41]) are nearly as high as, and in some cases higher than, the correlations of the noninterface external water molecules. This may be explained by the fact that while the trypsinogen crystal structures lie in a different space group than the MIP-, BA-, and native trypsin structures, one of the protein–protein interfaces in the two sets of structures is the same. In DIP–trypsinogen, 18 of 22 water molecules at the shared interface are conserved compared to 8 of 14 water molecules at nonshared interfaces. In trypsinogen, 25 of 41 water molecules at the shared interface are conserved compared to 17 of 34 water molecules at other interfaces.

There are a few fully occupied water molecules with B-factors <25 Å^2 in the crystal interface that are well conserved between structures. The only three completely conserved external water mole-

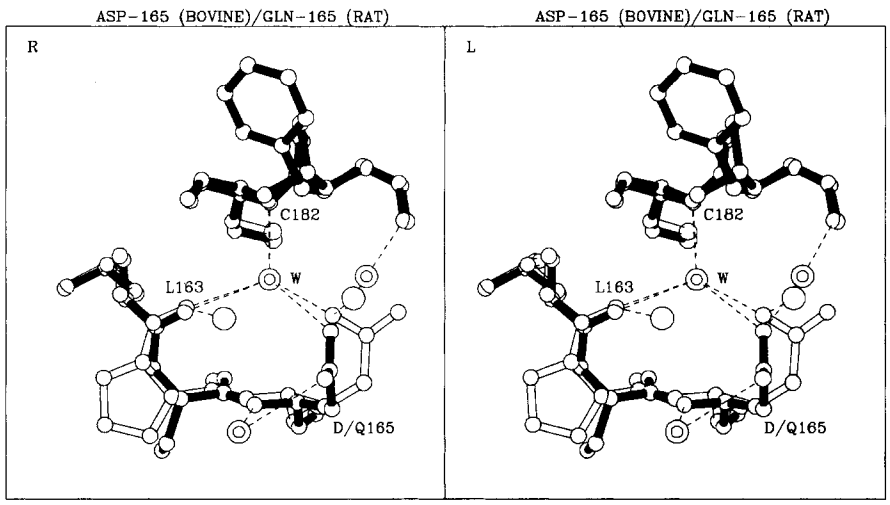

Fig. 7. PLUTO drawing of MIP–trypsin (filled bonds) around Asp-165 overlaid on the D-102/N rat trypsin structure (open bonds), where residue 165 is a glutamine. The side chain substitution at 165 accommodates internal water W, which hydrogen bonds to main-chain atoms O-163 and NH-182 in both structures.

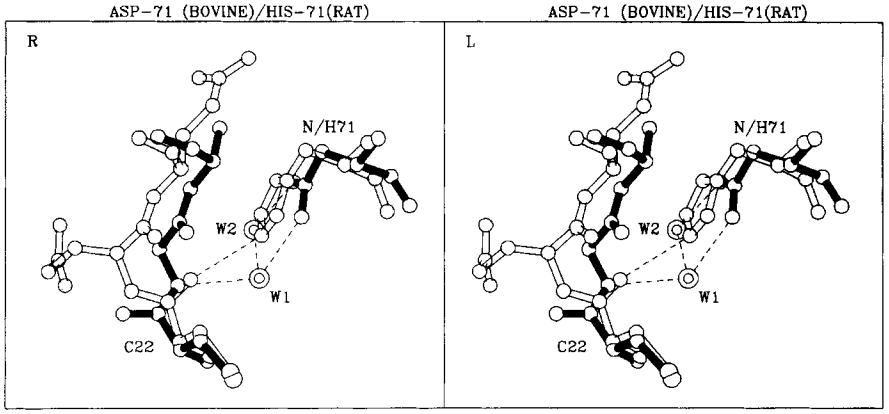

Fig. 8. PLUTO drawing of MIP–trypsin (filled bonds) around Asp-71 overlaid on the D102/N rat trypsin structure (open bonds), where residue 71 is a histidine. This is the only example of a side-chain substitution in rat trypsin that could not accommodate internal water molecules located in MIP–trypsin. Internal water W1 hydrogen bonds to the amide of Gly-23. In rat trypsin, the 22C–23N peptide plane is rotated 180°, so that O-22 hydrogen bonds to the His-Nε2, resulting in a localized conformational difference from the bovine structure.

cules are all at a crystal contact in the MIP–trypsin crystal structure. Two of these are part of a network of highly conserved water molecules at the calcium binding site (Fig. 9). Completely conserved water 1 of Figure 9, which is hydrogen bonded to the amide of residue 76, is an interface water in all the structures, suggesting that it may be a determinant of packing arrangements in trypsin crystals. The crystal contact, and the hydrogen bonds surrounding this water varied among the space groups, although the hydrogen bond to NH-76 was conserved. Two other completely conserved water molecules are interface water molecules in MIP-trypsin,[11–13] BA-trypsin,[38] native trypsin,[41] and rat trypsin.[18]

Water-mediated hydrogen bonds at the end of β-strands

Thirteen of the 19 totally conserved water molecules are hydrogen bonded to backbone atoms within or at the ends of the β-strands (Fig. 10). Eleven of these lie in clusters around the calcium-binding site. These water clusters have been described in detail for trypsin[10,44] and elastase.[24] One cluster of six water molecules mediates hydrogen bonds between segments of chain leading into two β-strands of one of the β-barrels in the structure (Fig. 11). These hydrogen-bonded structures may be analogous to folding intermediates for β-sheets or ribbons, reminiscent of the folding intermediates suggested by

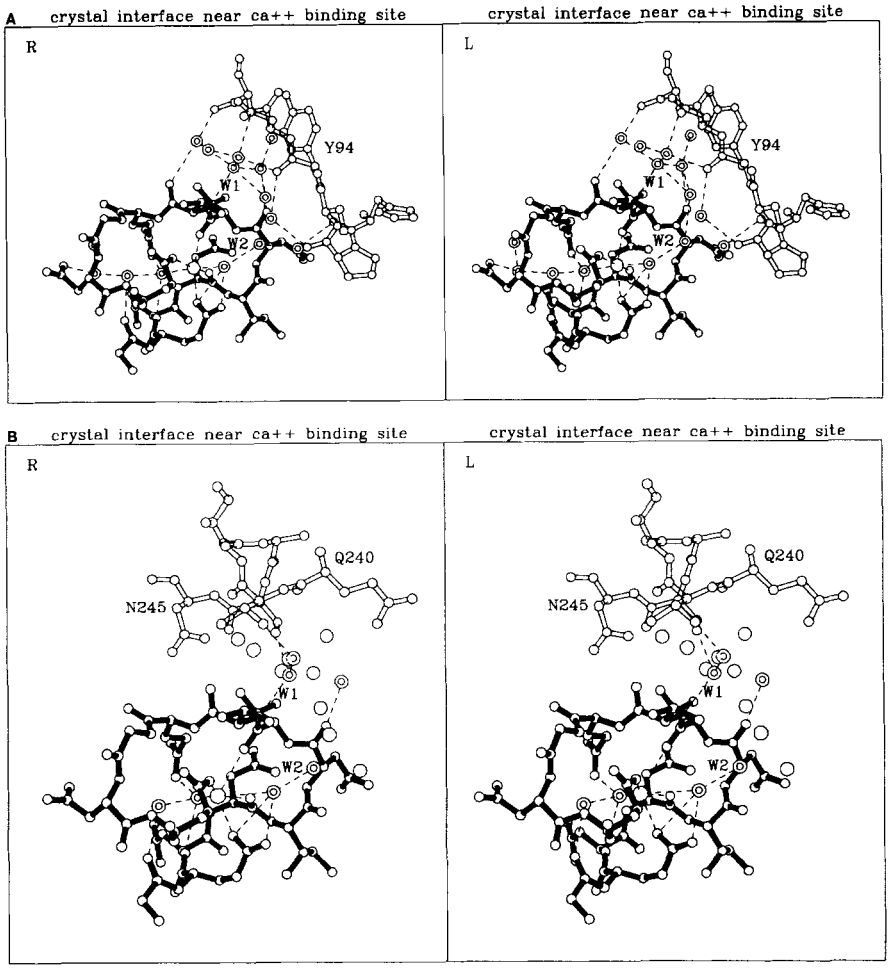

Fig. 9. **A.** Interface between symmetry related molecules near the calcium binding site in MIP–trypsin. A conserved network of internal water molecules around the calcium ion (large open circle) is connected through hydrogen bonds to a cluster of water molecules in the interface, including two water molecules, W1 and W2, present in all eight structures in the study. The symmetry related protein molecule is drawn with open bonds. **B.** Identical view of the calcium binding site in GBZ–trypsin. The calcium binding site is again adjacent to a symmetry related molecule, shown with open bonds, but the interface is completely different than that shown in A and crystal contact water W1 makes different hydrogen bonds than in MIP–trypsin. Water molecules from the MIP–trypsin structure are shown superimposed on the GBZ–trypsin interface as open circles.

water-inserted α-helical segments.[45] The cluster of water molecules is not necessarily involved in the folding of trypsin, but its hydrogen-bonding pattern suggests mechanisms by which water could assist in β-sheet formation. O-65a and NH-32 make the final hydrogen bond between the two β-strands composed of residues 31–33 and residues 65, 65a, and 66, respectively (Fig. 11). Residues Gln-30, Val-31, and Leu-66 have ϕ,ψ angles of β structure (Table V), but O-30 and NH-69 are 4.6 Å apart, not close enough to hydrogen bond. One water forms a hydrogen bond with both atoms, and a second water links the final β-sheet carbonyl, O-65a, to O-30. Thus, through hydrogen bonding these water molecules maintain the proper conformation and orientation of the two strands for pairing in a β-sheet.

Water molecules play a role in substrate binding

Meyer et al.[24] suggested that water channels in the trypsin family of serine proteases provide routes by which water displaced by substrate binding can exit the protein. Three clusters of ordered water molecules were identified in elastase that connect the active site or the binding pocket to the bulk solvent region. Analogous clusters are found in the trypsins we examined (see Fig. 10). Two completely conserved water molecules connect Ser-214 to the exterior of the protein. A second cluster connecting the specificity site to the outside of the protein consists of two water molecules, one completely conserved and the other present in seven of the eight structures in our study. The third channel lies in the

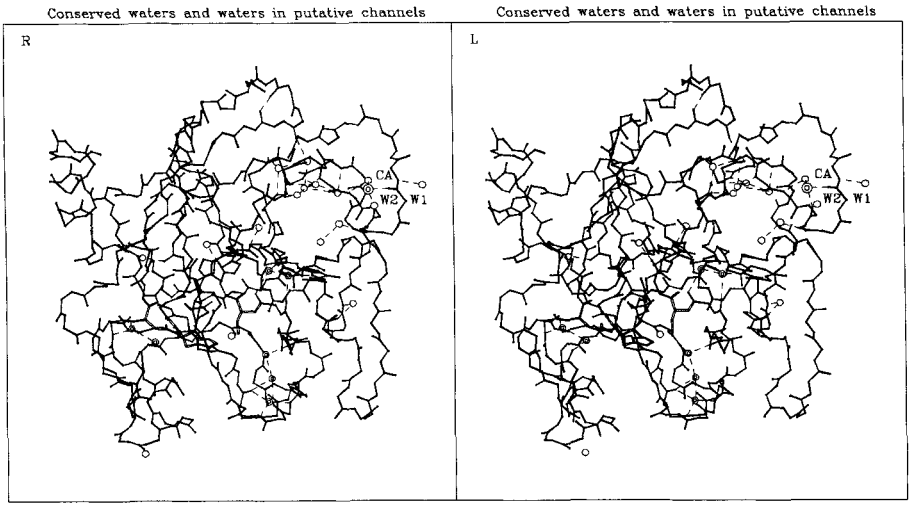

Fig. 10. Crossed-eye stereo view of the backbone atoms of MIP–trypsin showing completely conserved water molecules, the calcium ion, and putative channel water molecules (three of which are themselves completely conserved). The calcium ion is represented by the large concentric circles, the putative channel water molecules are represented by the small concentric circles, and the other conserved water molecules are shown as open circles. The side chains of the active site triad, His-57, Asp-102, and Ser-195, are shown with open bonds.

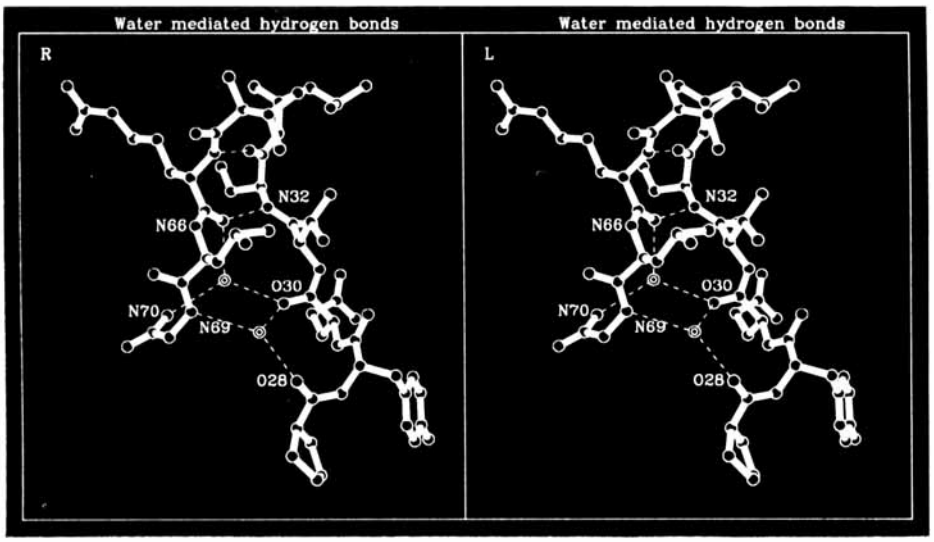

Fig. 11. Crossed-eye stereo plot of two water molecules hydrogen bonded to the ends of two β-strands, composed of residues 31–33, and 65–66, respectively. The hydrogen bonds between water molecules and main-chain atoms (dashed lines) stabilize residues 30 and 66 in an extended conformation.

binding pocket and is also well conserved. The three water molecules comprising this channel are absent only in the two trypsinogen structures, where the binding pocket is highly mobile.[15,17,40]

All water molecules in the three channels are internal. They are likely to be essential to the integrity of the structure and it can be argued that they are conserved for that reason alone. However, subtilisin, a serine protease with a different protein fold than that of trypsin, has eight well-ordered internal water molecules connecting its active site to the outside of the protein.[46] The existence of water channels in two serine proteases, which are not structurally homologous, provides a stronger argument that such channels provide conserved mechanisms for water displacement by substrate. Site-directed mutagenesis experiments, in which side chains displace water molecules in the channels, are one means of testing this hypothesis.

CONCLUSION

Most detailed knowledge about water structure at the protein surface derives from X-ray diffraction. Details of protein-bound water have been described

TABLE V. ϕ,ψ Angles for Ends of Two β-Strands

Residue	ϕ	ψ
Q30	−77	130
V31	−116	156
S32	−113	130
R65a	−112	112
L66	−102	159
G69	67	24

in several well-refined proteins of molecular weight less than 12,000 D.[1,3,33–36,47,48] It is thus encouraging that excellent agreement can be obtained between water structures determined independently by X-ray and neutron diffraction. X-ray determinations of water structure are likely, however, to be incomplete, because many bound water molecules scatter at a level similar to the noise level. Because of widely varying numbers of water molecules in X-ray water structure determinations for closely related proteins from different laboratories,[49] it is relevant to known quantitatively how reliably peaks can be interpreted as due to water molecules. We have estimated this from a comparison of independent X-ray and neutron diffraction results for MIP-trypsin. Most X-ray waters with good hydrogen bonding geometry, occupancies of at least 4 electrons and B-factors <40 Å^2 overlapped D_2O–H_2O solvent density. Even in our high resolution X-ray structure, with neutron diffraction solvent density available for comparison, only 211 waters, approximately one water per residue, were identified. These 211 waters included only 16 waters that were not hydrogen bonded to polar groups on the protein.

For the well-ordered water molecules most likely to be functionally important, there is excellent agreement among the high resolution X-ray structures of trypsins and other enzymes and zymogens in the trypsin family determined independently by different workers. Most exceptions are attributable to functional differences, such as different substrate specificity or the presence of different inhibitors. The comparisons of trypsin water structure described here support the view that bound water molecules in trypsin have important roles in substrate binding, structural stabilization, and protein–protein association.

Because of the greater degrees of freedom in a large protein such as trypsin, water structure is less determined by steric constraints of the molecular surface, and so may reflect the accommodation between polypeptide geometry and water geometry at the surface. Diffraction studies of small molecules such as cyclodextrins and vitamin B_{12} coenzyme[50,51] provide precise information on solvent interactions. However, hydrogen bonding geometries are often more distorted than seen in trypsin. The water molecules are an integral part of the lattice and are oriented to optimize hydrogen bonding subject to an often more constrained structure for the smaller molecules.

Details of protein hydration we describe agree with most previous results.[1,3,36] Polar groups are directly hydrated and hydrophobic groups are not, except at crevices and crystal contacts where participation in large water clusters extending over polar regions provide some hydrogen-bonding interactions with other water molecules. Deviations from ideal values of protein–water hydrogen bond geometries involving well-ordered protein hydrogen-bond donors and acceptors can largely be accounted for by experimental uncertainty. Most water molecules occur in clusters, but the extent of clustering is not dramatic enough to implicate water–water interactions as a major factor in protein hydration. The determinants of hydration are thus primarily water–protein interactions, with perhaps a secondary role for interactions between bound water molecules. It may thus prove possible to ignore cooperativity of water binding in making theoretical placements of bound water, at least to a first approximation.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant GM24485 to R.M.S. and by postdoctoral fellowship GM11453 to T.E. Figure 6 was made by Julie Newdoll using the Midas Plus program written by Conrad Huang, Eric Pettersen and Greg Couch at the UCSF Computer Laboratory.

REFERENCES

1. Edsall, J.T., McKenzie, H.A. Water and proteins. II. The location and dynamics of water in protein systems and its relation to their stability and properties. Adv. Biophys. 16:53–183, 1983.
2. Bernal, J.D., Fowler, R.H. A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. J. Chem. Phys. 1:515–548, 1933.
3. Finney, J.L. The organization and function of water in protein crystals. In: "Water, A Comprehensive Treatise." Franks, F. (ed.). New York: Plenum Press, 1979:47–122.
4. Cooke, R., Kuntz, I.D. The properties of water in biological systems. Annu. Rev. Biophys. Bioeng. 3:95–126, 1974.
5. Careri, G., Gratton, E., Yang, P.-H., Rupley, J.A. Correlation of IR spectroscopic, heat capacity, diamagnetic susceptibility and enzymatic measurements on lysozyme powder. Nature 284:572–573, 1980.
6. Levitt, M., Sharon, R. Accurate simulation of protein dynamics in solution. Proc. Natl. Acad. Sci. U.S.A. 85:7557–7561, 1988.
7. Yang, P.-H., Rupley, J.A. Protein–water interactions. Heat capacity of the lysozyme–water system. Biochemistry 18:2654–2661, 1979.
8. Matthews, B.W. X-ray structure of proteins. In: "The Proteins." Neurath, H., Hill, R.L. (eds.). 3rd Ed. New York: Academic Press, 1977:403–590.
9. Shpungin, J., Kossiakoff, A.A. A method of solvent structure analysis for proteins using D_2O–H_2O neutron difference maps. Methods Enzymol. 127:329–342, 1986.
10. Kossiakoff, A.A., Sintchak, M.D., Shpungin, J., Presta, L.G. Analysis of solvent structure in proteins using neutron D_20–H_20 solvent maps: Pattern of primary and secondary hydration of trypsin. Proteins 12:223–236, 1992.
11. Chambers, J.L., Stroud, R.M. The accuracy of refined protein structures: Comparison of two independently refined

models of bovine trypsin. Acta. Crystallogr. B35: 1861–1874, 1979.
12. Chambers, J.L., Stroud, R.M. Difference Fourier refinement of the structure of DIP–trypsin at 1.5 Å with a minicomputer technique. Acta Crystallogr. Sect. B B33:1824–1837, 1977.
13. Stroud, R.M., Kay, L.M., Dickerson, R.E. The structure of bovine trypsin: Electron density maps of the inhibited enzyme at 5 Å and 2.7 Å resolution. J. Mol. Biol. 83:185–208, 1974.
14. Kossiakoff, A.A., Spencer, S.A. Direct determination of protonation states of aspartic acid-102 and histidine-57 in the tetrahedral intermediate of the serine proteases: Neutron structure of trypsin, Biochemistry 20:6462–6474, 1981.
15. Kossiakoff, A.A., Chambers, J.L., Kay, L.M., Stroud, R.M. Structure of bovine trypsinogen at 1.9 Å resolution. Biochemistry 16:654–664, 1977.
16. Mangel, W.F., Singer, P.T., Cyr, D.M., Umland, T.C., Toledo, D.L., Stroud, R.M., Pflugrath, J.W., Sweet, R.M. Guanidino-benzoyl trypsin: The structure of a catalytic intermediate. Biochemistry 29:8351–8357, 1991.
17. Jones, M.O., Stroud, R.M. Lack of the transition state stabilization site is a factor in the inactivity of trypsinogen, a serine protease zymogen. Structure of DFP inhibited bovine trypsinogen at 2.1 Å resolution. In preparation.
18. Earnest, T., Fauman, E., and Stroud, R. 1.59 Å structure of trypsin at 120 K: Comparison of low temperature and room temperature structures. Proteins, 10:171–187, 1991.
19. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.E., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
20. Quiocho, F.A., Wilson, D.K., Vyas, N.K. Substrate specificity and affinity of a protein modulated by bound water molecules. Nature 340:404–407, 1989.
21. Tsukada, H., Blow, D.M. Structure of α-chymotrypsin refined at 1.68 Å resolution, J. Mol. Biol. 184:703–711, 1985.
22. Blevins, R.A., Tulinsky, A. The refinement and the structure of the dimer of α-chymotrypsin at 1.67 Å resolution. J. Biol. Chem. 260:4264–4275, 1985.
23. Sawyer, L., Shotton, D.M., Campbell, J.W., Wendell, P.L., Muirhead, H., Watson, H.C., Diamond, R., Ladner, R.C. The atomic structure of crystalline porcine pancreatic elastase at 2.5 Å resolution. Comparisons with the structure of α-chymotrypsin. J. Mol. Biol. 118:137–208, 1978.
24. Meyer, E., Cole, G., Radhakrishnan, R. Structure of native porcine pancreatic elastase at 1.65 Å resolution. Acta. Crystallogr. B44:26–38, 1988.
25. Navia, M.A., McKeever, B.M., Springer, J.P., Lin, T.-Y., Williams, H.R., Fluder, E.M., Dorn, C.P., Hoogsteen, K. Structure of human netrophil elastase in complex with a peptide chloromethyl ketone inhibitor at 1.84 Å resolution. Proc. Natl. Acad. Sci. U.S.A. 86:7–11, 1989.
26. Hendrickson, W.A. Stereochemically restrained refinement of macromolecular structures. Methods Enzymol. 115:252–270, 1985.
27. Jones, T.A. Interactive computer graphics: FRODO. Methods Enzymol. 115:157–170, 1985.
28. Perry, K.M., Fauman, E.B., Finer-Moore, J.S., Montfort, W.R., Maley, G.F., Maley, F., Stroud, R.M. Plastic adaptation toward mutations in proteins: Structural comparison of thymidylate synthases, Proteins 8:315–333, 1990.
29. Lee, B.K., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55:379–400, 1971.
30. Connelly, M.L. Solvent accessible surfaces of proteins and nucleic acids. Science 221:709–713, 1983.
31. Henderson, R., Moffat, J.K. The difference Fourier technique in protein crystallography: Errors and their treatment. Acta Crystallogr. Sect. B B27:1414–1420, 1971.
32. Savage, H. Water structure in vitamin B_{12} coenzyme crystals. I. Analysis of the neutron and X-ray solvent densities, Biophys. J. 50:947–965, 1986.
33. Watenpaugh, K.D., Margulis, T.N., Sieker, L.C., Jensen, L.H. Water structure in a protein crystal: Rubredoxin at 1.2 Å resolution, J. Mol. Biol. 122:175–190, 1978.
34. Blake, C.C.F., Pulford, W.C.A., Artymiuk, P.J. X-ray studies of water in crystals of lysozyme. J. Mol. Biol. 167:693–723, 1983.
35. Teeter, M.M. Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. Proc. Natl. Acad. Sci. U.S.A. 81:6014–6018, 1984.
36. Baker, E.N., Hubbard, R.E. Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol. 44:97–179, 1984.
37. Kossiakoff, A.A., Shpungin, J., Sintchak, M.D. Hydroxyl hydrogen conformations in trypsin determined by the neutron diffraction solvent difference map method: Relative importance of steric and electrostatic factors in defining hydrogen bond geometries. Proc. Natl. Acad. Sci. U.S.A. 87:4468–4472, 1990.
38. Bode, W., Schwager, P. The refined crystal structure of bovine β-trypsin at 1.8 Å resolution. II. Crystallographic refinement, calcium binding site, benzamidine binding site, and active site at pH 7.0. J. Mol. Biol. 98:693–717, 1975.
39. Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J., Steigemann, W. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 Å resolution. J. Mol. Biol. 89:73–101, 1974.
40. Fehlhammer, H., Bode, W., Huber, R. Crystal structure of bovine trypsinogen at 1.8 Å resolution. II. Crystallographic refinement, refined crystal structure, and comparison with bovine trypsin. J. Mol. Biol. 111:415–438, 1977.
41. Marquart, M., Walter, J., Deisenhofer, J., Bode, W., Huber, R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Crystallogr. Sect. B., B39:480–490, 1983.
42. Ferrin, T. E., Huang, C. C., Jarvis, L. E., Langridge, R. The MIDAS database system, J. Mol. Graph. 6:2–12, 1988.
43. Hubbard, T.S.P., Blundell, T.L. Comparison of solvent-inaccessible cores of homologous proteins: Definitions for protein modelling. Protein Eng. 1:159–171, 1987.
44. Bartunik, H.D., Summers, L.J., Bartsch, H.H. Crystal structure of bovine β-trypsin at 1.5 Å resolution in a crystal form with low molecular packing density. J. Mol. Biol. 203:1045–1070, 1989.
45. Sundaralingam, M., Sekharudu, Y.C. Water-inserted α-helical segments implicate reverse turns as folding intermediates. Science 244:1333–1337, 1989.
46. McPhalen, C.A., James, M.N.G. Structural comparison of two serine proteinase–protein inhibitor complexes: Eglin-C-subtilisin Carlsberg and CI-2-Subtilisin Novo. Biochemistry 27:6582–6598, 1988.
47. Smith, J.L., Corfield, P.W.R., Hendrickson, W.A., Low, B.W. Refinement at 1.4 Å resolution of a model of erabutoxin b: Treatment of ordered solvent and discrete disorder. Acta Crystallogr. Sect. A 44:357–368, 1988.
48. Thanki, N., Thornton, J.M., Goodfellow, J.M. Distribution of water around amino acid residues in proteins. J. Mol. Biol. 202:637–657, 1988.
49. Scanlon, W.J., Eisenberg, D. Solvation of crystalline proteins. Solvent bound in sperm whale metmyoglobin type A crystals at 6.1 and 23.5°C. J. Phys. Chem. 1981: 3251–3256, 1981.
50. Savage, H. Water structure in vitamin B_{12} coenzyme crystals. II. Structural characteristics of the solvent networks, Biophys. J. 50:967–980, 1986.
51. Savage, H.F.J., Finney, J.L. Repulsive regularities of water structure in ices and crystalline hydrates. Nature 322: 717–720, 1986.

APPENDIX

TABLE A. Overlap Coefficients and Percentage Matching Waters for Eight Sets of Water Molecules*

	MIP†	BA	NAT	BPTI	TG	DTG	GBZ	Rat	No.
MIP	—	.9,98	.9,97	.3,55	.5,71	.5,67	.5,62	.3,52	294
BA	—	—	.8,91	.4,56	.5,66	.4,52	.5,57	.5,60	78
NAT	—	—	—	.4,49	.5,60	.4,52	.4,53	.4,56	107
BPTI	—	—	—	—	.2,36	.2,40	.3,45	.3,42	152
TG	—	—	—	—	—	.4,49	.3,39	.2,38	131
DTG	—	—	—	—	—	—	.3,44	.3,43	91
GBZ	—	—	—	—	—	—	—	.3,48	122
Rat	—	—	—	—	—	—	—	—	187

*Symmetry-related water molecules within 4 Å of the protein are counted as independent observations, since the structures compared do not all belong to the same crystal form. The total number of water molecules, including symmetry-related water molecules within 4 Å of the protein, are also listed. Water molecules in the BPTI–trypsin structure that are in contact exclusively with inhibitor are excluded.

†Abbreviation used: MIP, MIP–trypsin; BA, BA–trypsin; NAT, native trypsin; BPTI, BPTI-trypsin; TG, trypsinogen; DTG, DIP–trypsinogen; GBZ, GBZ–trypsin; rat, D102N rat trypsin.

TABLE B. Overlap Coefficients and Percentage Matching Waters for the Eight Sets of Internal Water Molecules Compared to the Complete Set of Water Molecules for All Other Structures*

	MIP	BA	NAT	BPTI	TG	DTG	GBZ	Rat	No.	Avg B
MIP	—	.8,81	.9,89	.7,83	.5,70	.5,67	.7,81	.7,81	37	16
BA	1,99	—	1.,99	.8,95	.6,72	.7,90	.9,99	.8,95	22	9
NAT	1.,96	.9,92	—	.8,96	.6,76	.7,80	.9,96	.8,92	22	16
BPTI	.8,96	.8,92	.8,92	—	.5,72	.6,80	.8,92	.8,92	25	23
TG	.8,95	.8,86	.8,95	.7,91	—	.7,82	.8,86	.7,86	22	15
DTG	.6,74	.6,74	.6,73	.5,74	.5,60	—	.6,74	.6,74	23	20
GBZ	.9,1.	.9,96	.9,1.	.7,92	.6,78	.6,78	—	.8,96	27	11
Rat	.8,97	.8,90	.8,93	.8,90	.5,67	.5,73	.7,87	—	30	14

*The total number of internal water molecules and their average B-factor are given for each structure. Water molecules in the BPTI–trypsin structure that are in contact exclusively with inhibitor are excluded. Abbreviations are those used in Table V. All water molecules in a given structure are listed along the horizontal axis and the subset of internal water molecules are listed along the vertical axis.

TABLE C. Overlap Coefficients and Percentage Matching Waters for the Eight Sets of Interface Water Molecules Compared to the Complete Set of Water Molecules for All Other Structures*

	MIP	BA	NAT	BPTI	TG	DTG	GBZ	Rat	No.	Avg B
MIP	—	.3,37	.4,44	.1,14	.2,28	.2,25	.1,14	.1,15	163	31
BA	.8,91	—	.8,84	.1,28	.5,62	.4,41	.1,25	.2,34	32	21
NAT	.8,93	.5,61	—	.1,16	.4,50	.2,29	.1,23	.1,25	44	24
BPTI	.1,29	.1,7	.0,7	—	.1,14	.0,0	.0,14	.1,21	14	35
TG	.3,56	.2,22	.2,32	.1,13	—	.2,23	.1,14	.1,17	75	25
DTG	.5,72	.3,39	.3,41	.1,25	.4,50	—	.1,25	.1,22	36	20
GBZ	.1,27	.1,10	.1,16	.1,20	.1,12	.1,12	—	.1,20	49	28
Rat	.1,25	.1,11	.1,12	.1,25	.1,8	.0,5	.1,7	—	73	32

*Symmetry related water molecules are treated independently in the comparisons, since the eight structures compared do not all belong to the same space group. The total number of interface water molecules, with symmetry-related water molecules counted independently, and the average B-factor for the interface water molecules are also given. For the BPTI–trypsin structure, water molecules in contact exclusively with inhibitor are omitted. Abbreviations are the same as in Table V. All water molecules in a given structure are listed along the horizontal axis and the subsets of crystal interface water molecules are listed along the vertical axis.

TABLE D. Overlap Coefficients and Percentage Matching Waters for Eight Sets of Noncrystal Contact External Water Molecules Compared to the Complete Set of Water Molecules for All Other Structures*

	MIP	BA	NAT	BPTI	TG	DTG	GBZ	Rat	No.	Avg B
MIP	—	.2,19	.2,24	.2,32	.2,34	.1,14	.2,30	.3,45	94	31
BA	.8,96	—	.9,92	.4,58	.5,67	.2,33	.5,62	.5,63	24	20
NAT	.9,99	.6,59	—	.5,65	.5,65	.3,41	.5,65	.6,78	41	24
BPTI	.3,49	.2,24	.2,26	—	.1,25	.1,15	.1,26	.2,34	113	36
TG	.6,88	.4,47	.4,56	.3,53	—	.2,26	.2,41	.3,50	34	25
DTG	.4,62	.3,38	.4,53	.3,38	.3,43	—	.2,43	.3,47	32	22
GBZ	.6,78	.2,30	.3,48	.3,43	.3,43	.1,26	—	.3,48	46	29
Rat	.4,60	.1,15	.2,27	.1,24	.2,27	.1,14	.2,26	—	84	35

*The total number of non-interface external water molecules, and their average B-factor are given for each structure. For the BPTI–trypsin structure, water molecules in contact exclusively with inhibitor are omitted. Abbreviations are the same as in Table V. All water molecules in a given structure are listed along the horizontal axis and the subsets of noninterface external water molecules are listed along the vertical axis.