Buried Water in Homologous Serine Proteases†

Uma Sreenivasan and Paul H. Axelsen*

Department of Biochemistry and Molecular Biology, Mayo Clinic and Foundation, Rochester, Minnesota 55905

Received July 6, 1992; Revised Manuscript Received September 30, 1992

ABSTRACT: Buried water molecules in the structurally homologous family of eukaryotic serine proteases were examined to determine whether buried waters and their protein environments are conserved in these proteins. We found 16 equivalent water sites conserved in trypsin/ogen, chymotrypsin/ogen, elastase, kallikrein, thrombin, rat tonin and rat mast cell protease, and 5 additional water sites in enzymes which share the primary specificity of trypsin. Based on an alignment of 30 serine protease sequences, it appears that the protein environments of these 21 conserved buried waters are highly conserved. The protein environments of buried waters are comprised primarily of atoms from highly conserved residues or main chain atoms from nonconserved residues. In one instance, the protein environment of a water is conserved even in the presence of an unlikely Pro/Ala substitution. We also note 3 instances in which a histidine side chain substitutes for water, suggesting that the structural role of water at these sites is satisfied by the presence of an alternative hydrogen bonding partner. Buried waters appear to be integral structural components of these proteins and should be incorporated into protein structures predicted on the basis of sequence homology to this family, including the catalytic domains of coagulation proteases.

The structural significance of water molecules sequestered from bulk solvent within a protein matrix is not well understood. These "buried" waters have been found by X-ray crystallography in many proteins and may occur either individually or in small clusters. In structurally homologous proteins, they are typically found at equivalent sites, suggesting that they have a common role in the structure or function of such proteins (Edsall & McKenzie, 1983). In some cases they are associated with residues in the active site of an enzyme and possibly involved in mediating enzyme–substrate interactions (Finney, 1977). Small clusters of water have been noted to occur at the interface of protein folding domains, leading to speculation that buried water may facilitate interdomain motions (Rashin, 1986). Meyer et al. (1988) have proposed that internal water in serine proteases forms channels by which water displaced upon substrate binding can leave the protein. Buried waters are often associated with buried hydrophilic groups, and an important role in the protein folding process has been suggested (Finney, 1977; Edsall & McKenzie, 1983).

Serine proteases from eukaryotes have an especially large number of buried waters (Bode & Schwager, 1975; Kossiakoff, 1982; Sawyer et al., 1978; Birktoft & Blow, 1972). Buried waters at equivalent sites are highly conserved in different crystal structures of trypsin or trypsinogen (trypsin/ogen) (Finner-Moore et al., 1992; Kossiakoff et al., 1992). Many of these are also found at equivalent locations in crystal structures of the related enzymes chymotrypsin and elastase (Edsall & McKenzie, 1983). Water molecules are found in association with the buried β-turn in trypsin (Roseman, 1988), and two small clusters of water are found between the β-barrel substructures (Edsall & McKenzie, 1983).

Crystal structures are now available for seven members of the eukaryotic serine protease family, namely, trypsin/ogen, chymotrypsin/ogen, elastase, kallikrein, thrombin, rat tonin, and rat mast cell protease. The structural similarity of these proteins has encouraged several attempts to predict the structure of other serine proteases in this family using known structures as templates. This approach has been reasonably successful in the case of a prokaryotic trypsin (Greer, 1981a; Read, 1984) as well as in other protein families (Blundell et al., 1983; Weber, 1990) and has been applied to the catalytic domains of coagulation factors IXa Xa, and XIIa (Miyata et al., 1991; Geddes et al., 1989; Cool et al., 1985; Greer, 1981b; Furie, 1982). However, we do not know whether this approach can reliably predict whether buried waters present in a template structure should also be placed in a predicted structure.

The location of buried waters is vitally important for any study of serine protease structure and function which involves molecular mechanical simulations of protein dynamics. Buried waters in protein simulations must be initialized in rational positions; it is unreasonable to expect waters from the bulk phase to diffuse into buried sites in the course of such simulations, and we cannot assume that schemes intended for initializing water at the surface of a protein can reliably initialize water in buried sites. Without the correct placement of buried waters in a model structure, a simulation force field may induce unnatural hydrogen bond formation and locally inaccurate structure (Axelsen et al., 1987). Rashin (1986) has provided an algorithm for the detection of internal cavities in proteins and the prediction of possible water locations although their method does not address the probability of the occurrence of water in a given protein cavity. In two cases, Wade et al. (1991) have demonstrated that molecular mechanics calculations can correctly predict whether a protein cavity will be occupied by a water molecule. However, this type of analysis is not practical for proteins the size of serine proteases which contain many buried waters.

In lieu of such an energetic analysis, we have sought patterns which describe the occurrence of buried waters in serine proteases and patterns which describe the protein environment in which buried water is found. We aim to infer the location of buried waters in model protein structures on the basis of these patterns. In particular, we are interested in model

† U.S. is supported by a training grant from the National Heart Lung and Blood Institute. P.H.A. is supported by a grant from the National Institutes of Health (RO1 HL47469), by the L. P. Markey Charitable Trust, and by a grant-in-aid from the Minnesota Heart Association.

* To whom correspondence should be addressed. Tel (507) 284-6643; Fax (507) 284-9349.

0006-2960/92/0431-12785$03.00/0

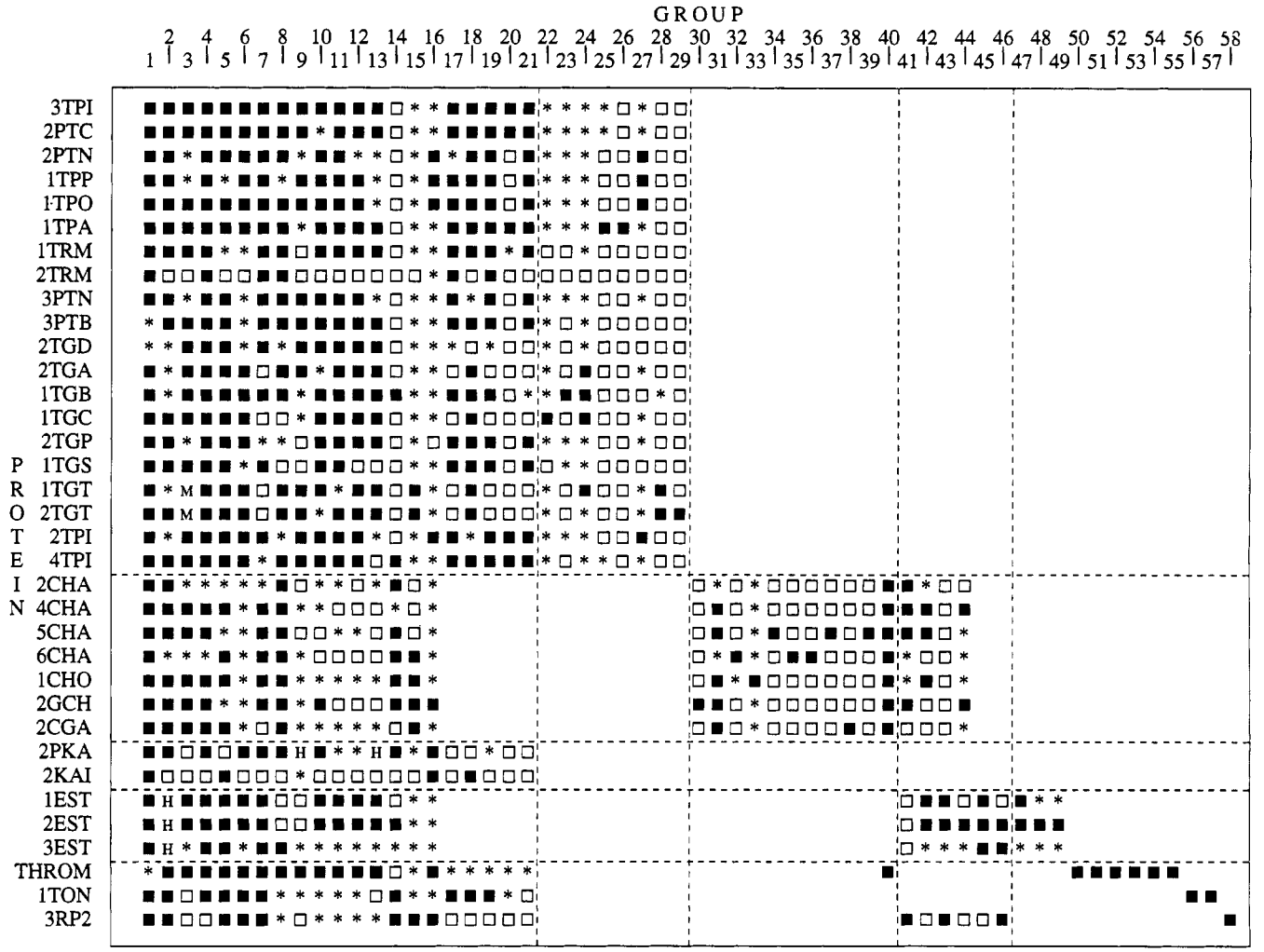

FIGURE 1: Buried waters in 35 serine protease crystal structures grouped according to equivalent structural sites. (■) Buried water present at a given site; (□) water missing from a given site, but cavity present; (*) water present but not buried; (H) water replaced by a histidine side chain; (M) water replaced by methanol.

structures of the coagulation proteases. Therefore, our examination included all available crystallographic structures for eukaryotic serine proteases as well as the amino acid sequences for 16 coagulation proteases in this family. We address four main questions.

To what extent are equivalent sites in known structures uniformly occupied by water? Is the protein environment of water preserved in the known structures? On the basis of sequence analysis, to what extent are the protein environments of buried waters conserved in serine proteases whose crystal structures are not yet known? Finally, are there equivalent sites in different proteins where amino acid side chains substitute for buried waters?

METHODS

Sequence and Structure Alignments. The sequences of 30 serine proteases (trypsin, chymotrypsin, elastase, kallikrein, thrombin, protein C, factors VII and IX–XII, tonin, and mast cell protease) from multiple species (Table I) were aligned manually. This alignment is in agreement with that reported by Greer (1990) except in isolated regions of nonhomologous insertions.

X-ray crystallographic structures of trypsin, trypsinogen, chymotrypsin, chymotrypsinogen, elastase, kallikrein, rat tonin, and rat mast cell protease were obtained from the Brookhaven Protein Data Bank (PDB; Bernstein et al., 1977) and are referred to by their four-character symbols. The crystal structure of thrombin was generously provided by W. Bode (Bode et al., 1989). In all, there are 35 serine protease structures (34 from PDB and thrombin). Atom names are referenced according to IUPAC nomenclature (IUPAC–IUB, 1970). Graphical inspection of the structures was done with the aid of QUANTA and all coordinate manipulations were

Table I: Serine Protease Sequences Utilized in the Sequence Alignment

protein	ref	rotein	ref
human factor IX	*a*	human factor VII	*d*
sheep factor IX	*a*	bovine thrombin	*d*
pig factor IX	*a*	human thrombin	*d*
rabbit factor IX	*a*	porcine kallikrein	*d*
guinea pig factor IX	*a*	elastase	*d*
rat factor IX	*a*	chymotrypsinogen	*d*
mouse factor IX	*a*	dog chymotrypsinogen	*c*
cow factor IX	*a*	human trypsin	*d*
dog factor IX	*b*	human trypsin 2	*c*
human factor X	*c*	trypsinogen	*d*
bovine factor X	*c*	rat trypsin	*d*
human factor XI	*c*	rat trypsin 2	*c*
human factor XII	*c*	mouse trypsin	*c*
bovine protein C	*a*	rat tonin	*d*
human protein C	*a*	rat mast cell protease	*d*

[a] Sarkar et al., 1990. [b] Evans et al., 1989. [c] Devereaux et al., 1984. [d] Bernstein et al., 1977.

performed using CHARMM (Molecular Simulations Inc., Waltham, MA). Bound ligands were deleted from the structures, and hydrogens were added using the CHARMM H-build utility. The hydrogen positions so generated were then subjected to 10 steps of steepest descent and 100 steps of ABNR energy minimization.

Identification of Equivalent Buried Water Sites. The accessible surface area for each atom in each protein crystal structure was calculated by the method of Lee and Richards (1971) using a probe sphere of radius 1.4 Å (accuracy = 0.05). Water residues accessible to the probe (i.e., with any atom bearing a surface area greater than 0.001 Å^2) were deleted. This process of surface area calculation and deletion was repeated until all probe-accessible water molecules had been deleted. The remaining waters were classified as buried. The probe radius of 1.4 Å was chosen to represent the radius of a molecule of water. CHARMM coordinate manipulation utilities were used for the above calculations.

Main-chain atoms in the crystal structures were aligned by a least squares algorithm in QUANTA. Buried water molecules at equivalent positions in different protein structures were identified by visual inspection and assigned to groups. Each group consists of a set of water molecules, each from a different protein structure, which have equivalent locations within these structures.

Analysis of Protein Environments. We define the protein environment of a water molecule to be the set of non-hydrogen protein atoms within 5 Å of the water oxygen. Protein atoms beyond 5 Å were not considered because these distances could permit another atom to situate closer to the water.

Atoms which comprise the protein environment of a given water were divided according to whether they were part of a generic or a nongeneric residue. A generic residue is defined as one which is identical at homologous positions in the 30 serine protease sequences listed in Table I. All other residues are nongeneric. The individual atoms in these categories were subdivided into main-chain and side-chain atoms. We further divided nongeneric side chain atoms into those which involved atoms with chemically similar counterparts in all proteins (i.e., the same atom type at the same side-chain position and with comparable polarity) versus those which have a chemically dissimilar or no counterpart in some proteins. Dissimilar atoms may be expected to have different hydrogen bonding capacities. This scheme yields the category headings of Table III.

RESULTS

Equivalent Water Sites. Our analysis identified buried waters at 58 nonequivalent sites in the 35 serine protease crystal structures we examined (groups 1–58, Figure 1). A group is comprised of the set of buried waters at an equivalent site in each protein. To identify these groups, a residue identifier for one of the waters in each group is given in Table II.

Two difficulties were encountered in the course of this analysis. First, water that was buried in some structures was exposed in other structures by our criteria ("*", Figure 1). This arises from slight variations in structure and occurs even among different crystals of the same protein. Second, in many cases the space occupied by a water molecule in one protein is preserved in a second protein, but is not occupied by a water molecule; i.e., some protein structures contain cavities that are occupied by water in other structures. We presume that this is a result of partial occupancy or an inability to unambiguously resolve the water oxygen. These spaces have been marked with an open square in Figure 1. No attempt was made to measure cavity volume or artifically place a water in these empty cavities. Rather, we identified their capacity to accommodate a water molecule by graphically superimposing their structure onto cavities which did contain a water.

Table II: Group Identifiers

group	water[a]	group	water[a]	group	water[a]
1	3TPI-401	21	3TPI-562	40	5CHA-544
2	3TPI-405	22	1TGC-412	41	5CHA-498
3	3TPI-406	23	1TGB-414	42	2EST-27
4	3TPI-407	24	1TGT-416	43	2EST-28
5	3TPI-408	25	1TPA-481	44	2EST-29
6	3TPI-409	26	1TPA-608	45	2EST-26
7	3TPI-410	27	2TPI-755	46	2EST-22
8	3TPI-430	28	2TGT-876	47	2EST-35
9	3TPI-431	29	2TGT-893	48	2EST-36
10	3TPI-434	30	2GCH-87	49	2EST-46
11	3TPI-435	31	5CHA-531	50	THR-305
12	3TPI-516	32	6CHA-553	51	THR-324
13	3TPI-704	33	6CHA-593	52	THR-334
14	1TGB-401	34	5CHA-627	53	THR-411
15	2TGT-473	35	5CHA-637	54	THR-454
16	2TPI-477	36	6CHA-695	55	THR-618
17	3TPI-402	37	5CHA-702	56	1TON-3
18	3TPI-415	38	2CGA-712	57	1TON-95
19	3TPI-445	39	5CHA-730	58	3RP2-328
20	3TPI-541				

[a] Waters are identified by the four-character PDB symbol for the protein in which they occur and by the residue number.

We found 29 groups of equivalent internal water molecules in the 20 trypsin/ogen structures (groups 1–29, Figure 1). In 1TGT and 2TGT, a methanol molecule occupies the site of an otherwise conserved buried internal water ("M", group 3). Methanol in the solvent may have exchanged with a buried water at this site since Kossiakoff (1982) has noted that the amide protons of peptide residues at this site do not exchange whereas exchange is observed for water molecules at this site. For our purposes we regarded this site as one in which buried internal water is conserved. In eight cases (groups 14, 15, 22, 23, 25, 26, 28, and 29), we have identified a group despite finding only one or two crystal structures with a water molecule at this site. However, in each case there is a cavity at the corresponding site in the other trypsin/ogen structures and it is unoccupied, i.e., the waters are "missing" (see the Discussion).

It is apparent that 16 of the 29 waters (groups 1–16, Figure 1) found in trypsin/ogen are also present in chymotrypsin/ogen, elastase, kallikrein, thrombin, tonin, and rat mast cell protease. This includes three exceptions in which a histidine side chain replaces water ("H", groups 2, 9, and 13). These will be examined further below. Five additional waters (groups 17–21) are present in trypsin, thrombin, kallikrein, and tonin and are associated with residues surrounding the primary substrate binding site (Asp-189). The primary substrate binding sites of chymotrypsin, elastase, and rat mast cell protease also contain distinct sets of buried water molecules (groups 41–46). Finally, there are 31 buried internal water molecules unique to individual serine proteases (groups 22–40 and 47–58). These waters are associated with portions of these proteins that have no discernible sequence homology to each other. There were no such waters for kallikrein. In summary, we find 16 buried waters to be at equivalent locations in all eukaryotic serine proteases of known structure (groups 1–16) and an additional 5 waters that occupy equivalent locations among proteins sharing the primary specificity of trypsin (groups 17–21).

Protein Environment of Conserved Buried Water. We consider atoms in the protein environment of each water to

Table III: Atoms in the Protein Environment of Conserved Buried Water Molecules in Serine Protease Crystal Structures

group number	no. of residues[a] – generic[b]/total	no. of non-H atoms within 5 Å of the water oxygen: generic residues[b] main-chain atoms	generic residues[b] side-chain atoms	nongeneric residues[b] main-chain atoms	nongeneric residues[b] side-chain atoms: similar	nongeneric residues[b] side-chain atoms: dissimilar
1	7/12	16	4	6	0	3
2	1/8	3	2	13	5	6
3	4/10	12	0	10	6	3
4	1/9	3	1	13	2	10
5	2/9	2	5	8	4	5
6	1/10	3	0	23	3	2
7	4/8	13	3	6	2	4
8	2/9	3	2	18	6	3
9	2/8	4	2	16	2	3
10	0/8	0	0	21	6	2
11	2/9	4	1	16	4	0
12	1/8	4	1	19	1	2
13	1/7	0	3	16	3	2
14	4/7	14	2	5	1	2
15	2/6	4	4	14	4	2
16	1/9	1	0	15	8	8
17	1/11	2	0	14	6	7
18	0/9	0	0	25	5	4
19	1/8	4	2	16	2	4
20	0/9	0	0	31	3	4
21	0/10	0	0	22	3	1
total		92	32	327	76	77
%		15	5	54	13	13

[a] The number of residues with at least one atom in the protein environment of the water. [b] A generic residue is one which is identical at homologous positions in all 30 serine protease sequences (Sarkar et al., 1990) (Figure 1). All other residues are nongeneric.

be conserved if they are in generic residues, are in the main chain of nongeneric residues, or are chemically similar to all corresponding side-chain atoms in nongeneric residues. In other words, the number of atoms remaining (i.e., dissimilar side-chain atoms) is a measure of the extent to which a protein environment is not conserved. It is important to note that this measure is derived from a sequence alignment which included a number of coagulation proteases whose structure is not known. It is, therefore, of value when seeking to infer whether buried waters in a known structure are likely to be present at the equivalent site in one of these coagulation proteases.

We confined this analysis to 21 waters: the 16 buried waters common to all the serine proteases crystal structures analyzed (groups 1–16) and the 5 water molecules present in proteases with trypsin-like primary substrate specificity (groups 17–21), i.e., arginine or lysine at S_1 (Schecter & Berger, 1967). The latter were included because the aforementioned coagulation enzymes possess the same primary specificity as trypsin.

Our results for these 21 waters are summarized in Table III. The analysis yielded a total of 604 atoms in the environment of the 21 buried waters for an average of 29 per water (range 16–38). The majority of atoms in the protein environment of these buried waters (87%) are from generic residues and main-chain or chemically similar side chain-atoms of nongeneric residues. On average, dissimilar side-chain atoms comprise only 13% of the protein environment of conserved buried waters in serine proteases (range 0–35%). No pattern was found for the distribution of dissimilar side-chain atoms in the protein environment of these buried waters.

Amino Acid Side Chains That Substitute for Buried Water. We included groups 2, 9, and 13 among our list of conserved buried waters, but in each case there is an important exception. Histidine side chains apparently substitute for water residue 405 (3TPI, group 2) in elastase and water residues 431 and 704 (3TPI, groups 9 and 13, respectively) in kallikrein.

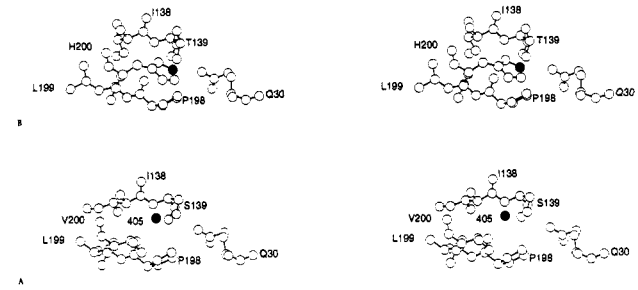

FIGURE 2: Comparison of the protein environments of water-405 in trypsin (A) and His-200 of elastase (B). The water:O and His:Nε2 atoms are shown as filled circles.

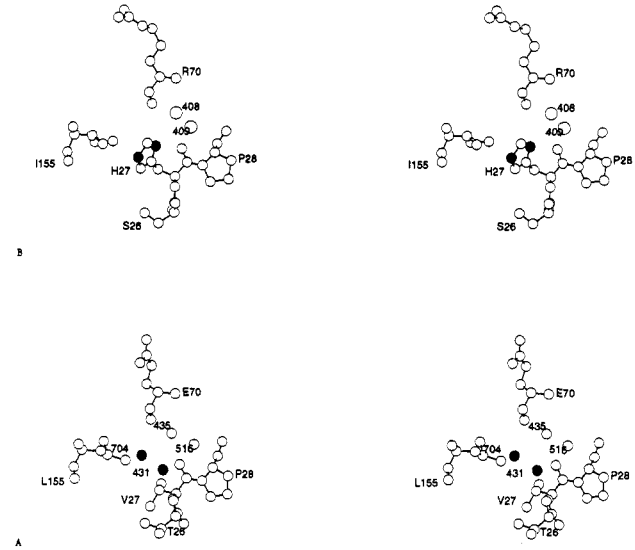

FIGURE 3: Comparison of the protein environments of waters-431 and -704 in trypsin (A) and His-27 of kallikrein (B). The water:O, His:Nδ1, amd His:Nε2 atoms are shown as filled circles.

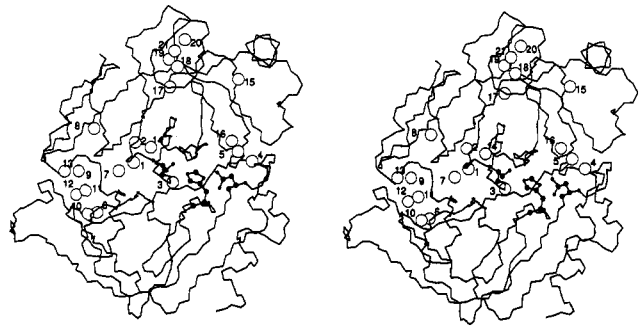

FIGURE 4: Stereoview of the set of 21 highly conserved buried waters in eukaryotic serine proteases. The trypsin backbone is represented as a stick drawing, with the catalytic triad at the center (filled circles). Water molecules are represented as open circles and are numbered with the group identifiers utilized in Figure 1.

His-200 in elastase occupies the site of the Val-200 side chain and water-405 (group 2) in trypsin, and atom His-200:Nε2 occupies a position equivalent to the oxygen of water-405 (Figure 2). A detailed comparison of the protein atoms surrounding water-405 and His-200:Nε2 indicates that their protein environments are similar (Table IV). The root mean square deviation for the least squares fit of main-chain atoms within 5 Å of water-405 and His-200:Nε2 is 0.48 Å. His-200 in elastase occupies a volume equivalent to that occupied by Val-200 and water-405 in trypsin. Both the oxygen of water-405 and His-200:Nε2 can hydrogen bond to 139:Oγ. Water-405 can form one additional hydrogen bond in trypsin to 139:N, which is not possible for His-200:Nε2. However, this

Table IV: Comparison of the Protein Environment of Water 3TPI-405 with His-200 (Elastase)

(A) Volumes (Å^3)[a]

3TPI		2EST	
residues within 5 Å of water-405 ($=v_a$)	889.7	same residues as for 3TPI ($=v_d$)	963.1
without water-405 ($=v_b$)	873.4		
occupied by water-405 ($=v_a - v_b$)	16.4		
without water-405 and Val-200 ($=v_c$)	787.6	without His-200 ($=v_e$)	856.0
occupied by water-405 and Val-200 ($=v_a - v_c$)	102.1	occupied by His-200 ($=v_d - v_e$)	107.1

(B) Hydrogen Bonds[b]

	3TPI			2EST		
	atoms (D–A)	distance (Å)	angle (D–H–A)	atoms (D–A)	distance (Å)	angle (D–H–A)
equivalent	139:Oγ–405:O	2.9	114.1	139:Oγ–200:Nε2	3.0	124.1
unconserved	139:N–405:O	4.0	93.5	139:N–200:Nε2	>5	

[a] Volumes were measured using 0.5-Å grid maps. [b] Distance cutoff D–A < 4 Å; angle cutoff D–H–A > 90°.

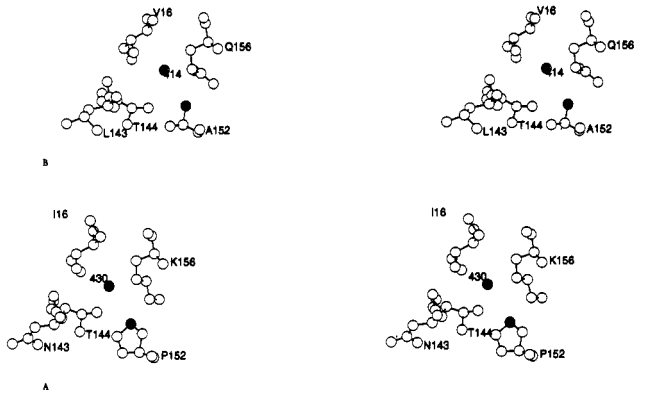

FIGURE 5: The protein environment of waters-430 in trypsin (A) -414 in elastase (B). Pro-152:C-γ (trypsin) and Ala-152:C-β (elastase) occupy homologous positions with respect to the water. Referenced atoms are shown as filled circles.

hydrogen bond is compensated for by a hydrogen bond between 139:N and 198:O in both enzymes. All other hydrogen bonds in the protein environment of 405:O and His-200:Nε2 are conserved (data not shown). Thus, Val-200 and water-405 in trypsin appear to be replaced by His-200 in elastase, with conservation of local structure and hydrogen bonds in the two enzymes.

The other instance in which histidine and water occupy equivalent positions occurs in kallikrein and trypsin (groups 9 and 13, Figure 2). Atoms Nδ1 and Nε2 in His-27 of kallikrein occupy positions equivalent to the oxygens of waters-431 (group 9) and -704 (group 13) in trypsin (Figure 3). A detailed comparison of the protein atoms surrounding the oxygens of waters-431 and -704 and atoms Nδ1 and Nε2 in His-27 indicates that their protein environments are similar (Table V). The root mean square deviation for the least squares fit of main-chain atoms located within 5 Å of waters-431 and -704, and Nδ1 and Nε2 in His-27, is 0.60 Å. His-27 of kallikrein occupies a volume equivalent to that occupied by Val-27 and waters-431 and -704 in trypsin. Waters-431 and -704, and Nδ1 and Nε2 in His-27, form equivalent hydrogen bonds to the other waters in the cluster as well as to neighboring protein atoms. Waters-431 and -704 in trypsin do, however, form additional hydrogen bonds which are missing in kallikrein. Other hydrogen bonds between backbone atoms comprising the protein environment of 431:O and 704:O, and Nδ1 and Nε2 in His-27, are conserved (data not shown). Thus, a single histidine residue in kallikrein replaces a valine and two water molecules in trypsin, with conservation of local structure and most hydrogen bonds in the two proteins.

DISCUSSION

It has been proposed that conservation of structure in homologous proteins is determined by a "core of conserved protein sequences" and that a quantitative relationship exists between the extent of homology in these core conserved sequences and the extent of structural homology (Chothia & Lesk, 1986).

We have identified a set of buried waters that are highly conserved in the family of eukaryotic serine proteases and which may be useful to consider as an additional component of the conserved structural core (Figure 4). Finner-Moore et al. (1992) recently reported that buried waters at equivalent sites in various crystal forms of trypsin/ogen are highly conserved. Earlier, in their review of water in proteins, Edsall and McKenzie (1983) noted the equivalence of several internal waters in trypsin, chymotrypsin, and elastase. On the basis of our analysis of buried waters at equivalent sites in 7 serine proteases of known structure, and an alignment of 30 serine protease sequences, we conclude that a set of 21 buried waters is highly likely to be found in homologous serine proteases of unknown structure.

In support of our conclusion, was have also noted that the protein environments of equivalent buried waters in serine proteases of known crystal structure are highly conserved. The majority of atoms in these buried water environments are part of the protein backbone. In these cases the substitution of one residue for another may have little consequence on the water environment, and the physical environment of buried waters may be conserved to a greater degree than the amino acid sequence. This tendency for protein environments to be conserved strongly suggests that buried waters preserved in serine proteases of known structure will also be present in the 16 coagulation enzymes of unknown structure whose sequences were considered in our analysis.

Some reservation about this conclusion may be raised in the case of groups 4, 16, and 17 where 24–35% of the protein environment is comprised of dissimilar side-chain atoms. However, these water sites are conserved despite the same degree of dissimilarity among just the known crystal structures. A particularly interesting example involves the water of group 8 where proline substitutes for alanine at position 152 of elastase and trypsin, respectively. Upon close inspection, the C-γ of Pro and the C-β of Ala occupy equivalent positions relative to the water (Figure 5). Both moieties are aliphatic, so that the physical environment of water in group 8 as well as the water itself is conserved despite a nonconservative residue substitution.

Table V: Comparison of the Protein Environment of Waters 3TPI-431 and 3TPI-704 with His-27 (Kallikrein)

(A) Volumes (Å^3)[a]

3TPI		2PKA	
residues within 5 Å of waters-405 and -704 ($=v_a$)	1240.5	same residues as for 3TPI ($=v_e$)	1432.6
without water-431 ($=v_b$)	1225.0		
occupied by water-431 ($=v_a - v_b$)	15.5		
without waters-405 and -704 ($=v_c$)	1209.4		
occupied by waters-405 and -704 ($=v_a - v_c$)	31.1		
without Val-27, waters-405 and -704 ($=v_d$)	1132.9	without His-27 ($=v_f$)	1331.6
occupied by Val-27, waters-405 and -704 ($=v_a - v_d$)	107.6	occupied by His-27 ($=v_e - v_f$)	101.0

(B) Hydrogen Bonds[b]

	3TPI			2PKA		
	atoms (D–A)	distance (Å)	angle (D–H–A)	atoms (D–A)	distance (Å)	angle (D–H–A)
equivalent	431:O–435:O	3.1		27:Nδ1–408:O	3.3	161.5
	431:O–27:O	3.0		27:Nδ1–27:O	2.9	116.3
	431:O–516:O	3.4		27:Nδ1–409:O	3.5	114.0
	704:O–71:Oδ1	2.8		71:Nδ1–27:Nϵ2	2.8	157.0
	704:O–70:O	3.0		27:Nδ1–70:O	3.5	104.6
unconserved	27:N–431:O	2.9	154.6			
	25:N–431:O	3.0	172.8			
	25:N–704:O	3.0	159.4			
	435:O–704:O	3.3			>5	

[a] Volumes were measured using 0.5-Å grid maps. [b] Distance cutoff D–A < 4 Å; angle cutoff D–H–A > 90°. D–H–A angles which involve a water proton are not reported.

The degree to which buried waters and their protein environments are conserved suggests that buried waters are important components of protein tertiary structure. However, buried waters are not consistently present in all crystal structures, and differences exist even among different crystal structures of the same protein. The absence of a water in a protein crystal cavity could reflect the inability to crystallographically resolve a water which is, in fact, present. We find no obvious correlation between crystallographic resolution and the number of missing waters in a given structure. Local disorder does not explain the absence of water from a given cavity since atoms comprising the protein environments of these cavities have, in some cases, lower temperature factors when the cavity is empty than when occupied by a water molecule (data not shown). Whether a water molecule is found buried in a protein should depend on the relative free energy of solvation of water in bulk solution or in the protein cavity. Wade et al. (1991) have compared the energetics of water situated in a protein cavity known to be occupied by water to its energetics in a cavity which is sufficiently large, but probably unoccupied. They have shown that there are reasonable thermodynamic grounds for supposing that water is indeed absent from sufficiently large protein cavities. An energetic analysis of this type is beyond the scope of this work, but the highly conserved nature of water and cavity protein environments, coupled with at least one crystal structure with water present in a given cavity, suggests that the cavity has reasonably favorable solvation energetics for water.

Since the protein environments of buried waters in eukaryotic serine proteases are highly conserved, we might expect the occupancy of protein cavities to vary according to the activity coefficient of water in the bulk solvent. Proteins crystallize under varied solvent conditions. Thus, the absence of water in some cavities may be explained by supposing that the energetics of some crystallizing liquors favor occupancy whereas others do not. Situations giving rise to partial occupancy may also arise, or changes in the activity coefficient of water during the crystallization process may result in waters which, though not energetically favored, are kinetically trapped in or excluded from the folded and crystalline protein. Compared to crystallizing liquors, we can safely presume that bulk water activity will be greater under noncrystallizing circumstances. Under ordinary solvent conditions it is quite possible that water will occupy many of the cavities which are crystallographically empty. Alternatively, water may play an important role in the initial process of folding, but may be expendable in the final folded protein.

The preponderance of main-chain atoms in the protein environment of buried waters is consistent with a role for water to solvate polar main-chain atoms that are not involved in secondary structure (Edsall & McKenzie, 1983). In contrast to peptide groups in protein secondary structures which are hydrogen bonded and therefore relatively nonpolar (Roseman, 1988), water may be necessary to reduce the energetic cost of burying peptide groups in the core of a protein which do not have hydrogen bond partners. This would explain the prevalence of buried water in between the two β-barrels found in serine proteases (Finner-Moore et al., 1992; Bartunik et al., 1989; Meyer et al., 1988), inasmuch as these structures have fragmented and irregular secondary structure. Meyer et al. (1988, 1990) noted that internal waters in serine proteases form channels which link the active site to the surface and proposed that these channels provide a mechanism for release of water molecule upon substrate binding. Another possibility is that buried waters may facilitate interdomain motions (Rashin, 1986). Although we do not have evidence that this particular type of motion actually occurs in serine proteases, it is plausible, and the evidence with regard to the dynamics of whole domains is incomplete (Brunger et al., 1987; Huber & Bennett, 1983).

The pattern suggesting that a core of buried waters are conserved across the entire serine protease family is very strong, but there are noteworthy exceptions. In three instances histidine side chains substitute for buried water and a valine side chain, and the histidine side chain forms at least one hydrogen bond with a likely hydrogen bond partner for the water. This exception was observed in two different proteins at three different water sites. Finner-Moore noted a similar replacement of histidine for two waters and an asparagine side chain, although by the criteria used in our work, the two

waters were not buried. It may be significant that the histidine side chain is found to substitute for water in each case, suggesting that the presence of a hydrogen bonding group is critically important at these sites. It would appear that the presence of water at certain sites in the protein fold is so important that changes in sequence which preclude the situation of a water at that site must be compensated for by substitutions that preserve the functional role of the water. These examples of histidine side chains acting as replacements for buried water suggest a stabilizing role for water as a hydrogen bonding partner in the interior of proteins. It is important to note that this is the only exception in a set which also considers the amino acid sequences of many serine proteases for which crystal structures are not yet known.

Our work identifies a set of buried waters that are preserved in all serine proteases of known structure. These are likely to be preserved in serine proteases of unknown tertiary structure, especially in the 16 coagulation enzymes of unknown structure whose sequences were considered in our analysis. Although this approach may not predict all buried waters in a model protein structure, the high degree to which this set of buried waters is conserved in serine proteases strongly suggests that molecular mechanics simulations incorporate this set of buried waters into their initial structures.

ACKNOWLEDGMENT

The computer facilities of Dr. F. G. Prendergast and a critical reading of the manuscript by Dr. W. G. Owen are gratefully acknowledged. Computer equipment was provided in part by the Ahmanson Foundation.

SUPPLEMENTARY MATERIAL AVAILABLE

Tables providing the residue identifiers corresponding to individual waters in each group, tables of the atoms comprising the protein environments of each of the 21 groups of conserved buried water molecules, and a figure giving the alignment of the 30 serine protease sequences (18 pages). Ordering information is given on any current masthead page.

REFERENCES

Axelsen, P. H., Haydock, C., & Prendergast, F. G. (1987) *Biophys. J. 51*, 282a.
Bartunik, H. D., Summers, L. J., & Bartch, H. H. (1989) *J. Mol. Biol. 203*, 1045–1070.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.
Birktoft, J. J., & Blow, D. K. (1972) *J. Mol. Biol. 68*, 187–240.
Blundell, T., Sibanda, B. L., & Pearl, L., (1983) *Nature 304*, 273–275.
Bode, W., & Schwager, P. (1975) *J. Mol. Biol. 98*, 693–717.
Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt, G., & Bartunik, H. (1983) *J. Mol. Biol 164*, 237–282.
Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R., & Hofsteenge, J. (1989) *EMBO J. 8*, 3467–3475.
Brunger, A. T., Huber, R., & Karplus, M. (1987) *Biochemistry 26*, 5153–5162.
Chothia, C., & Lesk, A. M. (1986) *EMBO J. 5*, 823–826.
Cool, D. E., Edgell, C. S., Louie, G. V., Zoller, M. J., Brayer, G. D., & MacGillvray, R. T. A. (1985) *J. Biol. Chem. 260*, 13666–13676.
Devereaux, J., Haeberlie, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.
Edsall, J. T., & McKenzie, H. A. (1983) *Adv. Biophys. 16*, 53–183.
Evans, J. P., Watzke, H. H., Ware, J. L., Stafford, D. W., & High, K. A. (1989) *Blood 74*, 207–212.
Finner-Moore, J. S., Kossiakoff, A. A., Hurley, J. H., Earnest, T., & Stroud, R. M. (1992) *Proteins 12*, 203–222.
Finney, J. L. (1977) *Philos. Trans. R. Soc. London, B 278*, 3–32.
Furie, B., Bing, D. H., Feldman, R. J., Robison, D. J., Burnier, J. P., & Furie, B. C. (1982) *J. Biol. Chem. 257*, 3875–3882.
Geddes, V. A., Le Bonniec, B. F., Louie, G. V., Brayer, G. D., Thompson, A. R., & MacGillvray, R. T. A. (1989) *J. Biol. Chem. 264*, 4689–4697.
Greer. J. (1981a) *J. Mol. Biol. 153*, 1027–1042.
Greer, J. (1981b) *J. Mol. Biol. 153*, 1043–1053.
Greer, J. (1990) *Proteins 7*, 317–334.
Huber, R., & Bennett, W. S. (1983) *Biopolymers 22*, 261–279.
IUPAC–IUB (1970) *Biochemistry 9*, 3471–3479.
Kossiakoff, A. A. (1982) *Nature 296*, 713–721.
Kossiakoff, A. A., Sintchak, M. D., Shpungin, J., & Presta, L. G. (1992) *Proteins 12*, 223–236.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379-400.
Malcolm, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F., & Wilson, A. C. (1990) *Nature 345*, 86–89.
Meyer, E., & Geller, M. (1990) *Biol. Chem. Hoppe-Seyler 371*, Suppl., 59–64.
Meyer, E., Cole, G., Radhakrishman, R., & Epp, O. (1988) *Acta Crystallogr., Sect. B. 44*, 26.
Miyata, T., Sakai, T., Sugimoto, M., Naka, H., Yamamoto, K., Yoshioki, A., Fukui, H., Mitsui, K., Kamiya, K., Umeyama, H., & Iwanaga, S. (1991) *Biochemistry 30*, 11286.
Rashin, A. A. (1986) *Biochemistry 25*, 3619–3625.
Read, R. J., & James, M. N. G. (1988) *J. Mol. Biol. 200*, 523–551.
Read, R. J., Brayer, G. D., Juresek, L., & James, M. N. G. (1984) *Biochemistry 23*, 6570–6575.
Roseman, M. A. (1988) *J. Mol. Biol. 201*, 621–623.
Sarkar, G., Koeberl, D. D., & Sommer, S. S. (1990) *Genomics 6*, 133–143.
Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157.
Wade, R. C., Mazor, M. H., McCammon, A., & Quiocho, F. A. (1991) *Biopolymers 31*, 919–931.
Warshel, A. (1987) *Nature 330*, 15–16.
Weber, I. T (1990) *Proteins, 7*, 172–184.