Protein Science (1998), 7:2054–2064. Cambridge University Press. Printed in the USA.

Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity

PAUL C. SANSCHAGRIN AND LESLIE A. KUHN

Protein Structural Analysis and Design Laboratory, Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824-1319

(RECEIVED December 30, 1997; ACCEPTED May 13, 1998)

Abstract

Cluster analysis is presented as a technique for analyzing the conservation and chemistry of water sites from independent protein structures, and applied to thrombin, trypsin, and bovine pancreatic trypsin inhibitor (BPTI) to locate shared water sites, as well as those contributing to specificity. When several protein structures are superimposed, complete linkage cluster analysis provides an objective technique for resolving the continuum of overlaps between water sites into a set of maximally dense microclusters of overlapping water molecules, and also avoids reliance on any one structure as a reference. Water sites were clustered for ten superimposed thrombin structures, three trypsin structures, and four BPTI structures. For thrombin, 19% of the 708 microclusters, representing unique water sites, contained water molecules from at least half of the structures, and 4% contained waters from all 10. For trypsin, 77% of the 106 microclusters contained water sites from at least half of the structures, and 57% contained waters from all three. Water site conservation correlated with several environmental features: highly conserved microclusters generally had more protein atom neighbors, were in a more hydrophilic environment, made more hydrogen bonds to the protein, and were less mobile. There were significant overlaps between thrombin and trypsin conserved water sites, which did not localize to their similar active sites, but were concentrated in buried regions including the solvent channel surrounding the Na^+ site in thrombin, which is associated with ligand selectivity. Cluster analysis also identified water sites conserved in thrombin but not trypsin, and vice versa, providing a list of water sites that may contribute to ligand discrimination. Thus, in addition to facilitating the analysis of water sites from multiple structures, cluster analysis provides a useful tool for distinguishing between conserved features within a protein family and those conferring specificity.

Keywords: binding site characterization; conserved water sites; drug design; hydration; water-mediated ligand recognition

Water plays a vital role in protein structure, acting as the driving force behind protein folding through the hydrophobic effect (Kuntz & Kauzmann, 1974; Eisenberg & McLachlan, 1986), and also mediates contacts between proteins and their ligands (Raymer et al., 1997). Bound water can be used by the protein to confer specificity of binding, as seen for the tryptophan repressor (Otwinowski et al., 1988; Joachimiak et al., 1994), or to allow the protein to bind a broader range of ligands by filling gaps between the protein and ligand, as seen for the major histocompatibility complex (Wilson & Fremont, 1993). Another role of water, which is important for thrombin and trypsin, is the structural stabilization of cavities (Rashin et al., 1986; Meyer, 1992; Williams et al., 1994). Water can also play an important role in protein catalysis, as in hydrolysis by serine proteases (Blow et al., 1969; Perona et al., 1993; Singer et al., 1993); proteins stripped of their primary hydration layer are observed to lose catalytic function (Rupley & Careri, 1991).

The ability to quantitatively define conserved bound water sites from crystallographic protein structures has a number of practical uses. One of these is ligand design (Lam et al., 1994; Mikol et al., 1995; Ladbury, 1996; Wang & Ben-Naim, 1996). If the conserved water molecules are known for a protein ligand-binding site, then ligand design can be improved by including polar atoms at appropriate locations in the ligand to form hydrogen bonds with the water molecules, or to displace them from the binding site. If a bound water site is conserved in a number of independent structures of the protein, it is likely an essential part of the protein; its structural or functional role can be tested by conservative mutagenesis of side chains that tether the water molecule.

The solvent environment surrounding a protein is generally divided into two classes: (1) bulk water that is fluid and not bound to the protein, and (2) water that is either partially or strongly

Reprint requests to: Leslie A. Kuhn, Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824-1319; e-mail: kuhn@agua.bch.msu.edu; web site: http://www.bch.msu.edu/labs/kuhn.

bound (Edsall & McKenzie, 1983; Otting et al., 1991; Badger, 1993; Levitt & Park, 1993; Ringe, 1995). A typical method for analyzing the contributions of bound solvent is to use molecular graphics to visualize the water bound in a single protein structure, or a small number of closely related structures, and their proximity to catalytic or ligand-binding residues. A second method is to examine the water sites in a number of homologous proteins, measuring the degree to which each site is observed in the independent structures. This approach has been used to study the solvation of FKBP12 complexes with the immunosuppressant FK506 (Faerman & Karplus, 1995) and the solvation of T4 lysozyme (Zhang & Matthews, 1994), and also to predict water sites in unrelated protein structures (Pitt et al., 1993) as well as water molecule conservation or displacement upon ligand binding (Raymer et al., 1997). Analyses of water sites in crystallographic protein structures remain subject to limitations in fitting and refinement (Levitt & Park, 1993; Karplus & Faerman, 1994), though these techniques are improving (Badger, 1993; Jiang & Brünger, 1994; Burling et al., 1996), and the limitations of water assignment in any single structure can be minimized through the use of multiple, independently solved structures as a knowledge base for analysis and design. In this study, we employ a statistical method, complete linkage hierarchical clustering, to define consensus water sites in thrombin, trypsin, and bovine pancreatic trypsin inhibitor (BPTI), with the goal of determining the extent to which water sites are conserved for each protein and between the two serine proteases, and their relationship to ligand binding.

Thrombin was chosen for several reasons: there are a number of structures solved to good resolution with different ligands bound, thrombin is an important pharmaceutical target for regulating blood coagulation, and highly conserved water molecules are known to surround the binding site of its allosteric regulator, Na^+ (Di Cera et al., 1995; Zhang & Tulinsky, 1997). Thrombin is a serine protease involved at the junction between the coagulation and anticoagulation pathways and initiates both processes (reviewed in Furie & Furie, 1988 and Esmon, 1992). In addition to binding its receptor, proteolytic substrates, and several physiological inhibitors, thrombin also binds exogenous inhibitors such as hirudin (produced as an anticoagulant by leeches) and a substrate transition-state analog, D-Phe-Pro-Arg chloromethyl ketone (PPACK). Thrombin contains two major ligand binding sites: the active site and the fibrinogen binding site, or exosite, which provides an additional binding surface, enhancing the affinity for the fibrinogen substrate and hirudin and its analogs (Vijayalakshmi et al., 1994). The variety of crystallographic thrombin:ligand complexes provides a basis for studying water sites that are conserved in thrombin regardless of ligand, as well as those water sites that are ligand-specific.

A second goal is to determine which bound water sites are shared by thrombin and trypsin, a serine protease not involved in blood coagulation, in order to identify the essential water sites in serine proteases and point to water molecules specific to thrombin or trypsin ligand-binding sites. Trypsin is a serine protease which proteolytically activates other digestive proteases. The loop that binds Na^+ in thrombin cannot bind Na^+ in trypsin due to the change in conformation and chemistry associated with the Tyr 255 to Pro sequence difference (Dang & Di Cera, 1996); thus, trypsin lacks the allosteric regulation by Na^+ binding found in thrombin. Several high-resolution trypsin structures are available in the Protein Data Bank (PDB) (Bernstein et al., 1977; Abola et al., 1987), and its water structure has also been studied via techniques including room-temperature and low-temperature X-ray crystallography (Earnest et al., 1991), neutron diffraction (Finer-Moore et al., 1992), and D_2O-H_2O difference neutron diffraction (Kossiakoff et al., 1992). Bovine pancreatic trypsin inhibitor (BPTI) is a natural inhibitor of trypsin, and its water interactions have been studied using NMR and molecular dynamics (van Gunsteren et al., 1983; Brunne et al., 1993; Denisov et al., 1996) and simultaneous NMR and X-ray diffraction refinement (Schiffer et al., 1994). Several high-resolution structures of BPTI are available in the PDB, as well as X-ray diffraction structures of the trypsin:BPTI complex, allowing us to examine the fate of water molecules bound to the free trypsin and BPTI structures upon complex formation. Given the extensive structural data available for serine proteases, they provide an ideal system for testing the cluster analysis approach to unbiased identification of conserved water sites and analyzing their environmental features.

Methods

Structure selection

Thrombin, trypsin, and BPTI structures were selected from the Protein Data Bank based upon the absence of unusual crystallization conditions (e.g., low pH), sequence insertions, deletions, or point mutations, and a resolution of $\leq$2.0 Å for trypsin and BPTI and $\leq$2.4 Å for thrombin. Ligand-free trypsin structures were selected; no ligand-free thrombin structures were available, but 6 of the 10 structures we analyzed have no ligand in either the active site or the exosite. The availability of 10 thrombin structures for analysis helped compensate for their somewhat lower resolution. Visual screening of the superimposed structures for each protein eliminated those with regions of large structural deviation likely to affect water site conservation, and only structures with refined water were included. The quality of water refinement was assessed using a mobility measure designed to normalize and combine the crystallographic temperature factor (*B*-value) and the occupancy (Craig et al., 1998). This facilitates comparison of atomic mobility between protein structures that were solved using different refinement protocols, in particular, those structures in which occupancy as well as *B*-value were allowed to vary during refinement.

$$\text{Mobility}_{\text{water molecule}} = \frac{B\text{-value}_{\text{water molecule}}/\text{Average } B\text{-value}_{\text{all waters in structure}}}{\text{Occupancy}_{\text{water molecule}}/\text{Average Occupancy}_{\text{all waters in structure}}}. \quad (1)$$

Using this normalization, a water molecule (or other atom) with a high degree of rigidity has a mobility value near 0, an atom with average mobility relative to other atoms in the structure has a mobility value of 1, and if an atom's mobility value is x, then it is x times as mobile as the average atom. In practice, the mobility of a water molecule is determined from its oxygen atom, since hydrogen atoms are not assigned in the majority of structures. Histograms of the water mobility values for each structure showed whether there were a number of water sites with unusually high mobility (>2); a preponderance of water sites with high mobility values was found, by analysis of inter-water distances, to indicate water molecules placed too close to each other (<2.6 Å). Such structures were excluded from our analysis. As an example, Figure 1 compares the mobility distributions of water sites in two BPTI structures. Structures selected using all the above criteria are presented in Table 1.

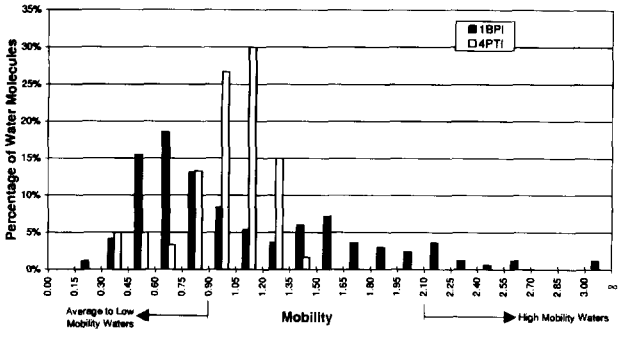

Fig. 1. Mobility distributions of two BPTI structures, used as a quantitative tool to screen for structures with uncertain water positions. The 4PTI distribution is narrow and shows that most water molecules have near-average mobility, and none are highly mobile. The 1BPI distribution is broad and has an extended right tail, showing the presence of a number of water molecules with high mobility values (≥ 2; at least twice as mobile as average). Further analysis showed that one-half of these high-mobility water sites, which were physically overlapping, could be explained by the occupancies of the overlapping sites summing to ≤ 1.0, suggesting that they represent alternate locations of a single water molecule; however, the use of multiple locations for a single water molecule would introduce a statistical bias into cluster analysis and has been avoided in this study.

Hierarchical clustering

The following steps were performed independently for the thrombin, trypsin, and BPTI structural sets (Table 1). The chosen structures were superimposed using main-chain least-squares superposition in InsightII (Molecular Simulations, Inc., San Diego, California) to transform the water coordinates into the same reference frame. There are several ways to cluster elements (e.g., water sites, which may themselves be single elements or clusters of elements) based on the matrix of inter-element distances: (1) the minimum distance between any pair of cluster elements (single linkage), (2) the distance between the cluster centroids, which are the mean x, y, z coordinates of the elements of each cluster (average linkage), or (3) the maximum distance between any pair of cluster elements (complete linkage). Complete linkage clustering was chosen because it produces compact, globular clusters and allows specification of a maximum diameter for any cluster by defining the maximum distance between cluster elements.

Clustering is an iterative process. Initially, a matrix of the distances between all pairs of elements is calculated. For the example shown in Figure 2, this matrix consists of all distances between the five water molecules. Complete linkage clustering begins by placing the two closest elements together into a cluster; 6PTI 108 and 4PTI 108 are less than the maximal distance of 2.4 Å apart and are grouped into a cluster (arbitrarily numbered 109). Next, the distance between this cluster and each of the remaining data elements is computed; for complete linkage clustering, the distance between an element and a cluster is defined as the maximum distance between that element and all the elements in the cluster. In Figure 2, the distance between elements 4PTI 139, 9PTI 103, and 6PTI 238 (which are not yet clustered) and cluster 109 is calculated as the distance to 4PTI 108, since it is the furthest element of cluster 109. This process is repeated until all elements are clustered into a single cluster or, as in this case, no further elements can be clustered without exceeding the selected maximum distance. Any elements not yet clustered due to this distance threshold are considered to define single-element clusters; for example, cluster 134 in Figure 2 consists only of water 6PTI 238.

A maximum cluster diameter of 2.4 Å was chosen, resulting in clusters with a maximum inter-water distance of 2.4 Å. This value was chosen because water molecules have an approximate effective radius of 1.6 Å, which includes the radius of the oxygen atom and a correction for the contribution of the hydrogen atoms, whose positions are typically unknown. Thus, if two water molecules are placed with their oxygen atoms at a center-to-center distance of 2.4 Å, their radii will overlap by 50%. This almost always prevents water sites from the same structure being included in the same cluster, since at <2.4 Å apart, they would be positioned too closely. Complete linkage clustering results in the set of maximally dense clusters (in terms of the average number of water molecules per cluster), which will be referred to as "microclusters" to emphasize that all water molecules within a cluster physically overlap. Clustering was performed using the SPlus statistical package (Mathsoft, Cambridge, Massachusetts). Our WatCH (Waters Clustered Hierarchically) software converts the SPlus clustering data into a PDB-formatted file containing the microcluster elements indexed by cluster number, a PDB-formatted file with the microcluster centroids, and a list of each microcluster and its constituent water molecules.

Crystal contact calculations

To observe the effect of crystal contacts on water sites, crystal contacts in the seven thrombin structures in space group C_2 were calculated using Chain (Sack, 1988), where interactions were included for symmetry mate atoms within 4.0 Å. Crystal contact residue and atom lists were generated for each of the crystallographic structures, with water sites represented by the microclusters observed in that protein. The number of times each microcluster appeared in a crystal contact was calculated for the seven thrombin structures, and software was developed to convert Chain's crystal contact lists and our microcluster lists into InsightII subsets to enable visualization of the spatial relationship between crystal contacts and microcluster conservation.

Evaluation of bound water environments

The degree of conservation of the water microclusters, each representing a favored site for water binding, was calculated as the number of individual water molecules contained in the microcluster divided by the number of structures used for clustering. To assess the influence of the shape and chemistry of the water binding site on its conservation in different structures, the values of eight environmental features were calculated for each microcluster: atomic density (ADN), measured by the number of protein atoms within van der Waals packing distance, 3.6 Å, of the water molecule, which correlates with whether the site is in a groove (high density of protein neighbors) or a protrusion (low density of neighbors; Kuhn et al., 1992); local atomic hydrophilicity (AHP), measured by the sum of the atomic hydrophilicity values of all protein and water atoms within 3.6 Å of the water site (Kuhn et al., 1995); crystallographic temperature factor (B-value; BVAL), a measure of the atom's thermal mobility and spread in electron density, read from the protein's PDB file; the number of hydrogen bonds to neighboring protein atoms (PrHBD) and to neighboring water molecules (WatHBD), using a distance of ≤ 3.5 Å between donor and acceptor atoms; the water site mobility (MOB), a normalized mea-

Table 1. *Database of thrombin, trypsin, BPTI, and trypsin:BPTI structures for analysis of conserved water sites*

PDB code	Ligand binding site: Active site	Ligand binding site: Fibrinogen binding site (exosite)	Resolution (Å)	*R*-factor	Main-chain RMSD (Å)	Number of crystallographic bound waters	Crystallographic space group
Thrombin structures[a]							
1HAI	PPACK		2.4	0.14	0.000	194	C 2
1ABJ	PPACK		2.4	0.14	0.694	196	P 21 21 21
1PPB	PPACK		1.9	0.16	0.802	409	P 21 21 21
1TMB	Cyclotheonamide A	Hirugen	2.3	0.14	0.560	239	C 2
1HAH		Hirugen	2.3	0.11	0.345	205	C 2
1TMT	——CGP50,856——		2.2	0.18	0.458	111	P 21 21 2
1ABI	——Hirulog-3——		2.3	0.13	0.409	246	C 2
1THR		Hirullin	2.3	0.16	0.350	190	C 2
1THS		MDL-28050	2.2	0.16	0.439	140	C 2
1IHS	——Hirutonin-2——		2.0	0.17	0.481	146	C 2
Trypsin structures[b]							
1TPO			1.7	0.18	1.395	84	P 21 21 21
2PTN			1.6	0.19	0.103	82	P 21 21 21
3PTN			1.7	0.20	0.266	82	P 31 2 1
BPTI structures[c]							
4PTI			1.5	0.16	0.000	60	P 21 21 21
5PTI[d]			1.0/1.8	0.20/0.20	0.403	63	P 21 21 21
6PTI			1.7	0.16	0.436	73	P 21 21 2
9PTI			1.2	0.17	0.418	67	P 21 21 21
Trypsin/BPTI complex structures[e]							
2PTC			1.9	0.19	0.343/0.479	157	I 2 2 2
1TPA			1.9	0.18	0.336/0.638	159	I 2 2 2

[a]Superpositions and RMSD values are relative to 1HAI.
[b]Superpositions are relative to 1TPO, except for 1TPO which is relative to 1HAI.
[c]Superpositions are relative to residues 1–56 of 4PTI.
[d]Resolution and *R*-factor are for X-ray diffraction/neutron diffraction data.
[e]RMSDs are reported for the trypsin chain of the complex superimposed on 1TPO and for the BPTI chain of the complex superimposed on 4PTI.

sure of thermal mobility (see Structure selection, above); the summed *B*-values for all protein atoms within 3.6 Å of the water site (TPrBVAL); and the average *B*-value for these neighboring protein atoms (AvgPrBVAL). Several of these features are related, and our goal was to see which ones correlate best with the degree of water conservation.

For each microcluster, the value for each of the eight features was averaged over the individual environments of its water molecules. To assess the correlation between conservation of the microclusters and their environments, feature values were also averaged over all microclusters with a given degree of conservation (e.g., those containing waters from 6 of 10 structures). Averaging and subsequent plotting were done using Excel for Windows, version 5.0 (Microsoft Corp., Redmond, Washington). Visual analysis, color-coding, and graphics rendering of clustering results was performed using InsightII (Molecular Simulations, Inc., San Diego, California).

Calculation of overlapping microclusters between thrombin and trypsin

Distances were calculated between the centroids of microclusters in the superimposed structures of thrombin and trypsin, and overlapping microclusters were defined as those with a centroid to centroid distance of ≤ 1.8 Å. With a maximum diameter of 2.4 Å for each microcluster, the microclusters' radii overlap by 50% when their centroids are within 1.8 Å. To analyze the effect of using different overlap criteria and provide a list of microcluster overlaps between thrombin and trypsin using a less stringent criterion, overlapping microclusters were also tabulated for thresholds up to 2.4 Å (where two microclusters just touch). To determine the significance of the observed number of overlapping sites between thrombin and trypsin, in a separate experiment microcluster centroids were randomly placed in the thrombin structure at the density of microclusters experimentally observed for thrombin, and the same was done for trypsin; then, the number of overlaps between thrombin and trypsin microclusters was calculated from these random distributions. Because many of the water sites in thrombin and trypsin are buried in the proteins, the microcluster density was calculated for each protein based on the number of microclusters per Å^3 of protein volume, which was calculated separately for the thrombin and trypsin structures using the PQMS routine of the Molecular Surface Package, version 2.6 (Connolly, 1993; http://www.biohedron.com). Random placement of microcluster centroids and subsequent counting of overlaps were repeated 100 times to obtain statistical means and standard deviations for the number of overlaps as a function of overlap criterion (1.8

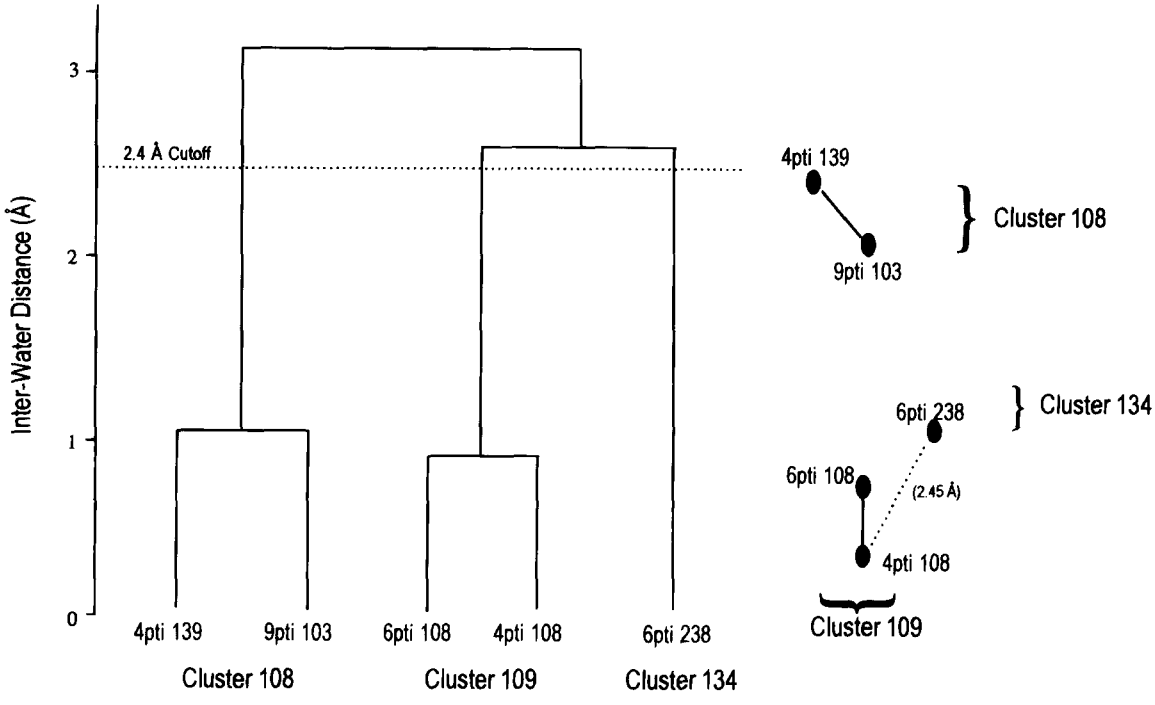

Fig. 2. Example of complete linkage clustering applied to water sites from several BPTI structures. A portion of the BPTI clustering tree is shown at left, based on the inter-water distances from the superimposed BPTI structures, shown at right. Note that 6PTI water 108 and 6PTI water 238 are not clustered together even though they are closer than the cutoff distance of 2.4 Å, since 6PTI 108 belongs to a cluster in which one water (4PTI 108) is too far from 6PTI 238 to meet the 2.4 Å threshold. This feature of complete-linkage clustering guarantees that no cluster contains water sites separated by more than 2.4 Å. At this distance, all water molecules in a microcluster are overlapping, and it is unlikely that more than one water molecule will be included from any given protein structure (they would be too close). Cluster numbers are arbitrary sequential indices, whereas individual water molecules are labeled by the residue number from the corresponding PDB file.

to 2.4 Å). For analyzing conserved water site proximity to functionally important sites (e.g., residues in the catalytic triad), a distance threshold of 3.6 Å from the microcluster centroid to the functionally important atom(s) was used. Interaction with an active-site or exosite ligand was determined by measuring the distance to all ligands bound in the superimposed structures.

Results

Clustering statistics

To identify shared versus unique conserved water sites for thrombin and trypsin, we performed complete linkage clustering on their water sites (Table 2). Clustering of 2,075 water sites from the ten thrombin structures yielded 708 microclusters with an average of 2.93 waters each, indicating that the average water site was observed in 29.3% of the structures. Of the 708 microclusters, 18.5% were found in at least half of the 10 structures. Clustering of 248 waters from the three trypsin structures yielded 106 microclusters, conserved on average in 78.0% of the structures. Of these microclusters, 56.6% were observed in all three structures. This high degree of conservation was surprising, but two of the three structures (PDB codes 1TPO and 2PTN) were solved by the same crystallographers and have very similar water sites; however, mobility plots (data not shown) indicated that the water assignments in both structures are reasonable. (We considered analyzing additional structures, but there were only three ligand-free, wild-type, bovine trypsin structures solved under typical crystallization conditions.) Given the similarity in water assignments for 1TPO and 2PTN, trypsin water sites were considered to be highly conserved if they appeared in all three structures. A similar analysis of BPTI clustered 263 water sites from four structures into 134 microclusters, with an average conservation of 49.0%. Of these microclusters, 54.5% were found in at least half of the BPTI structures.

Environmental analysis

Analysis of water site environments provided insights into the determinants of conserved water binding. All protein-bound microclusters, containing at least one water making direct contact ($\leq$3.6 Å) with the protein molecule, were analyzed. There were 521 protein-bound microclusters for thrombin, 98 for trypsin, and 117 for BPTI. Highly conserved water molecules occupied somewhat different environments than less conserved waters (Fig. 3). Conserved microclusters had more neighboring protein atoms (atomic density; ADN), more hydrogen bonds to the protein (PrHBD), and were in a more polar environment, indicated by more hydrophilic neighboring atoms (atomic hydrophilicity; AHP). The total *B*-value of neighboring protein atoms (TPrBVAL) also correlated positively with the degree of microcluster conservation, likely due to correlation with ADN, the number of neighboring protein atoms. As expected, other measures of mobility—the water *B*-value (BVAL), its mobility (MOB), and the average *B*-value of neighboring protein atoms (AvgPrBVAL)—were anti-correlated with

Table 2. *Clustering statistics*

Thrombin (10 superimposed structures)	
Number of water molecules	2,075
Number of water clusters	708
Average conservation (waters/cluster)	2.93 (29.3%)
Number of clusters with ≥50% conservation	131 (18.5%)
Number of clusters with 100% conservation	28 (4.0%)
Mean protein volume (Å^3)	38,272
Cluster density (clusters/Å^3)	0.01850
Conserved cluster[a] density (clusters/Å^3)	0.00342
Trypsin (3 superimposed structures)	
Number of water molecules	248
Number of water clusters	106
Average conservation (waters/cluster)	2.34 (78.0%)
Number of clusters with ≥50% conservation	82 (77.3%)
Number of clusters with 100% conservation	60 (56.6%)
Mean protein volume (Å^3)	27,211
Cluster density (clusters/Å^3)	0.00390
Conserved cluster[a] density (clusters/Å^3)	0.00301
BPTI (4 superimposed structures)	
Number of water molecules	263
Number of water clusters	134
Average conservation (waters/cluster)	1.96 (49.0%)
Number of clusters with ≥50% conservation	73 (54.5%)
Number of clusters with 100% conservation	18 (13.4%)
Mean protein volume (Å^3)	7,313
Cluster density (clusters/Å^3)	0.01832
Conserved cluster[a] density (clusters/Å^3)	0.00998

[a]Conserved clusters are those with ≥50% conservation.

water conservation. The number of hydrogen bonds to other water molecules (WatHBD) did not correlate strongly with conservation, suggesting that consensus water sites are not strongly stabilized by hydrogen-bonded water networks.

Effects of crystal contacts on bound water conservation

The effects of crystal contacts upon water binding were examined by spatially correlating water site conservation with contacts in the protein lattice. To address whether water sites were preferentially excluded from or trapped in these contacts, we visualized crystal contact residues for seven thrombin structures in the C_2 space group along with the locations of conserved water sites. Crystal contacts had fewer conserved water sites than surrounding areas, consistent with the observed expulsion of interfacial bound water upon dimerization of chymotrypsin (Blevins & Tulinsky, 1985).

Spatial analysis of the conserved microclusters

To explore how microclusters of different conservation levels are distributed spatially around the protein, molecular graphics visualization was used. For thrombin, a concentration of highly conserved microclusters (in ≥50% of the structures; yellow spheres in Fig. 4A) was found near the sodium site but not observed in the active site, perhaps due to water displacement by the presence of active-site ligands in 7 of the 10 structures. Other conserved microclusters were observed in deep grooves or cavities within the protein, as expected from the known correlation between water site conservation and groove topography (Kuhn et al., 1992) and previous studies on the conservation of buried waters in serine proteases (Rashin et al., 1986; Finer-Moore et al., 1992; Meyer, 1992; Sreenivasan & Axelson, 1992). When the exosite ligands were superimposed, a structurally conserved region, comprising the six N-terminal ligand residues (green tubes at the bottom right of Fig. 4A), and a structurally variable region, extending from the seventh residue to the C-terminus of the ligand (orange tubes at rightmost edge of Fig. 4A), were found; water sites associated with the structurally conserved region in the exosite ligands were also generally conserved. Similar patterns of buried water site conservation were observed for trypsin.

Given the functional importance of Na^+ binding for switching between the coagulant (Na^+ bound) and anticoagulant (water bound) forms of thrombin (Di Cera et al., 1995), this region of the structure was analyzed in detail. The Na^+ sites in structures 1HAI and 1HAH assigned by Zhang and Tulinsky (1997), which were originally labeled as water molecules in the PDB structures and later confirmed by rubidium replacement to represent a Na^+ site (Di Cera et al., 1995), occur in two overlapping microclusters (centroids 1.2 Å apart) containing Na^+/water molecules from all 10 structures. The 38 water microclusters in the channel coupling the Na^+ site with the active site are >50% conserved on average, consistent with the recent discovery of this conserved solvent channel (Zhang & Tulinsky, 1997).

Overlapping water sites between thrombin and trypsin

To define water sites shared between these serine proteases involved in distinct biochemical pathways, overlaps between ≥50% conserved sites in thrombin and trypsin were evaluated (Table 3; Fig. 4B). The number of overlapping water microclusters in thrombin and trypsin, 37, is statistically significant, since 7.9 overlaps would be expected if the conserved water sites in thrombin and trypsin were distributed randomly (see Methods). Seven of the conserved microclusters were in the active-site region, four being near at least one of the catalytic triad residues. Three overlapping microclusters between thrombin and trypsin were near the Na^+ binding site of thrombin, with two more in the surrounding solvent channel. Conservation of solvent in this region (Fig. 4B, lower left), which regulates the coagulant/anticoagulant function of thrombin via Na^+ binding/displacement, suggests it is also important in trypsin.

To assess whether water site conservation between thrombin and trypsin is associated with conservation of nearby side chains and their conformations, we evaluated the neighborhoods of the 37 shared water sites in the context of PDB structures 1HAI (thrombin) and 1TPO (trypsin). Ninety-two percent of the shared water sites had chemically and conformationally similar environments, based on no more than one side-chain substitution and no more than one residue with a significant (1.5–2 Å) shift. Larger shifts were considered structurally dissimilar, yet even substituted side chains tended to be similar through the γ-carbon. Of the 37 shared sites, 38% were structurally very similar, with no side-chain substitutions and no positional shifts exceeding 1.5 Å. Thus, conserved protein structure between thrombin and trypsin largely accounted for their water site conservation, which can be considered a shared feature of their structure and function as serine proteases.

Several water sites were highly conserved in functionally important regions of thrombin or trypsin but not shared between the

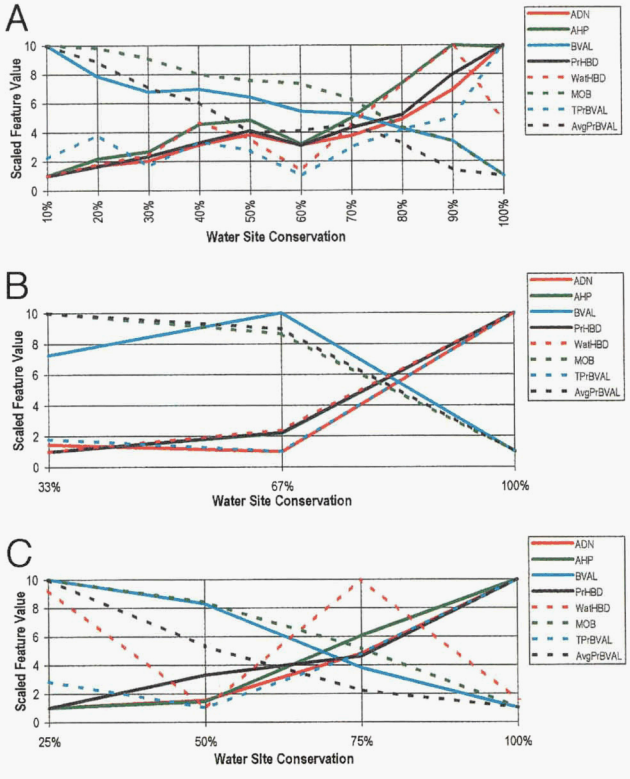

Fig. 3. Correlation between water site conservation and environmental features for (**A**) thrombin, (**B**) trypsin, and (**C**) BPTI. Shown are the average values of eight features for the water microclusters as a function of their degree of conservation. Features are abbreviated as follows: density of neighboring protein atoms (ADN), hydrophilicity of neighboring protein atoms (AHP), crystallographic temperature factor of the water molecule (BVAL), number of hydrogen bonds from the water molecule to the protein (PrHBD), number of hydrogen bonds from the water molecule to other waters (WatHBD), mobility of the water molecule (MOB), summed temperature factors of neighboring protein atoms (total protein *B*-value; TPrBVAL), and average temperature factor of neighboring protein atoms (AvgPrBVAL). The feature values have been averaged within microclusters as described in Methods, and normalized to range between 1 and 10 to allow visualization on the same plot. The curve of AHP for trypsin superimposes with the curve for PrHBD, and is therefore not apparent on the plot. Approximately linear correlation with conservation is seen for many of the features, as discussed in Results.

two enzymes, and therefore may contribute to their specificity differences (Table 4). Four more microclusters were specifically associated with active-site ligands in trypsin than were seen for thrombin, in part reflecting the larger inhibitor in trypsin; 13 residues of BPTI interact with trypsin, whereas the thrombin active-site ligands are three to seven residues long. Five Na^+ binding site and channel waters were shared between thrombin and trypsin (Table 3); however, 15 conserved sites in this region were found only in thrombin (Table 4). Combined with the five conserved exosite water positions found uniquely in thrombin and eight active-site water positions found uniquely in trypsin (Table 4), it is apparent that bound water can make a significant contribution to ligand specificity.

Contribution of conserved water molecules to the trypsin:BPTI complex

Trypsin provides an ideal system to test the applicability of a lock-and-key mechanism for the contributions of protein-bound

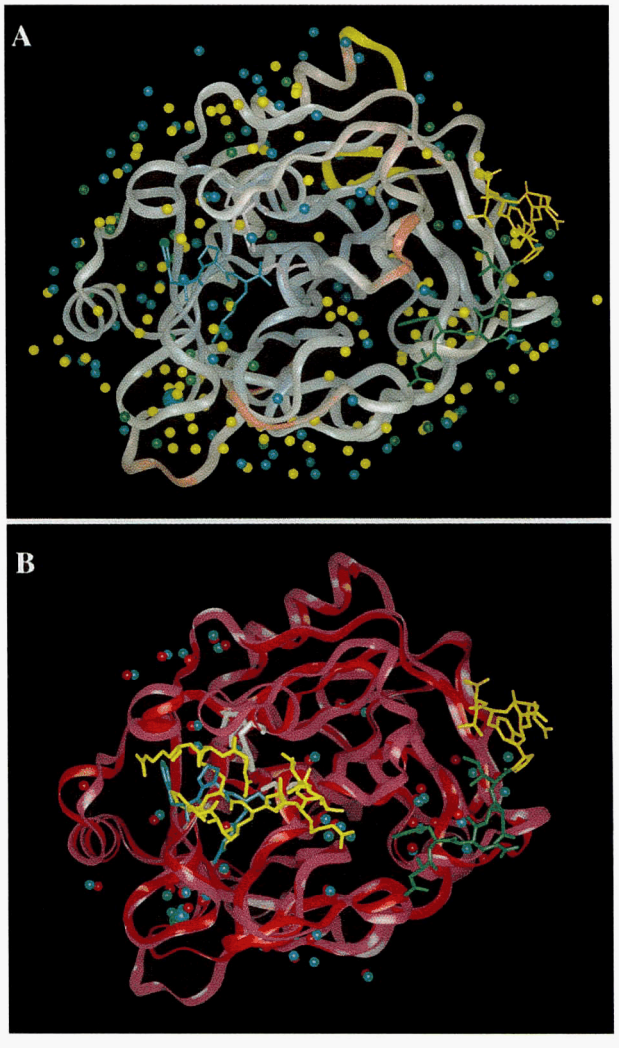

Fig. 4. Conserved water sites in thrombin and trypsin. **A:** Thrombin microclusters containing water sites from three of the ten structures are colored blue, sites found in four are green, and sites found in at least five are yellow. The backbone ribbon of 1HAI is shown colored by *B*-value (dark blue equals a *B*-value of 0, white ~30, red ~50, and yellow >50 $Å^2$). The catalytic triad Asp, Ser, and His side chains are rendered as pink tubes at center. PPACK, an active-site ligand from 1HAI, is shown in blue tubes, and hirugen, an exosite ligand from 1HAH, is shown in green (structurally conserved region) and orange tubes (structurally divergent region) at right. The sodium ion (labeled as water 410 in 1HAI) is rendered as a large blue sphere at lower left. Highly conserved water sites are concentrated near the sodium site and its channel, at lower left, and many others are buried. **B:** Overlapping conserved water sites between thrombin and trypsin. Water sites conserved in at least half of the structures of thrombin (water sites shown as blue spheres) or trypsin (red spheres) are shown. The backbone of thrombin (represented by 1HAI) is shown as a magenta ribbon, and trypsin (represented by 1TPO) is shown as a red ribbon. Catalytic triad side chains are shown by white tubes (center of figure), and PPACK (a thrombin active-site inhibitor) is shown in blue and hirugen (a thrombin exosite inhibitor) is shown in green and orange (structurally conserved and divergent regions, respectively). The binding epitope of the trypsin inhibitor, BPTI, is shown in yellow and superimposed from 2PTC; note the conformational similarity between PPACK and BPTI, extending downward from the proline residue in PPACK. The sodium ion from 1HAI is rendered as a large blue sphere at lower left. Overlaps between conserved thrombin and trypsin water sites are concentrated near the sodium site, despite trypsin having no known functional similarity here; these conserved water molecules form a bridge to the active site.

Table 3. *Overlapping conserved water sites between thrombin and trypsin*

Thrombin			Trypsin			Distance between thrombin and trypsin cluster centroids[d] (Å)
Cluster number	Percent conservation[a]	Representative water residue number[b]	Cluster number	Percent conservation[a]	Representative water residue number[c]	
1021	70	570	25	100	470	0.16
899 AL,[e] Na,[e] C[e]	100	417	45	100	415	0.30
996	100	423	22	100	717	0.41
954	100	461	5	100	430	0.43
1197	70	515	54	67	806	0.48
1017 AL,C	100	407	33 AL[e]	100	416	0.52
935 AS[e]	100	430	6 AS[e]	100	703	0.54
857 AS	100	445	20	100	701	0.56
951 AS	100	468	19	100	408	0.56
874	100	401	28	100	708	0.56
970	50	551 (1HAH)	72	100	752	0.57
885	100	404	18	100	721	0.62
888 AL,C	100	403	31	100	704	0.65
926	100	414	30	100	429	0.67
1214	80	480	21	67	751	0.72
1075	80	489	59	100	736	0.75
948	50	554 (1ABJ)	62	67	754	0.75
852	90	441	13	100	473	0.76
1016 AS	100	436	10 AL	100	410	0.77
963	100	439	17	100	722	0.78
1051	70	455	66	100	728	0.80
878	100	405	9	100	406	0.82
972 Na,C	90	448	35	100	705	0.86
1032	80	469	55	67	803	0.94
981	100	467	38	100	709	0.96
955	100	412	29	100	716	0.99
1139	70	537	42	100	530	1.01
1150 E[e]	90	507	8	67	738	1.06
832	50	458	65	67	801	1.10
1111	90	546	60	100	744	1.11
1086 Na,C	70	409 (1HAH)	44	100	562	1.14
890	90	406	24	100	516	1.29
870	90	443	4	100	726	1.32
916	80	452	34	100	604	1.40
921	100	413	1	100	746	1.58
1038	90	451	2	100	741	1.61
1108	80	539	16	100	725	1.66
1150	90	507	71	100	733	1.86
1259	70	426	52	67	750	1.93
1053	80	457	56	67	735	2.10
981	100	467	24	100	516	2.13
1170	60	494	66	100	728	2.15
964 Na,C	90	450	35	100	705	2.15
948	50	554 (1ABJ)	65	67	801	2.19
1032	80	469	37	100	720	2.25
827	100	446	18	100	721	2.29
995	90	431	27	67	743	2.30
1119	50	505	2	100	741	2.34
857	100	445	10	100	410	2.36

[a]Only waters with at least 50% conservation are tabulated.

[b]Representative thrombin waters are from 1HAI unless there is no member water from 1HAI, in which case the structure containing the representative water residue is noted.

[c]Representative trypsin waters are from 1TPO.

[d]A line divides the table into highly overlapping water microclusters with centroids ≤1.8Å apart (cluster radii overlapping by ≥50%), shown in the top section of the table, from somewhat overlapping microclusters with centroids 1.8–2.4 Å apart.

[e]Labels indicate overlapping clusters that interact with (are ≤3.6 Å from) exosite ligands (E), active site ligands (AL), active-site catalytic triad residues (AS), Na^+ site (Na), or Na^+ channel waters (C).

Table 4. *Functionally relevent conserved water sites unique to thrombin or trypsin*

	Cluster number	Percent conservation[a]	Representative water residue number[b]
Thrombin			
Active site catalytic triad residues			
	No nonoverlapping conserved water sites		
Active site ligands	1153	100	408
	1196	90	428
Exosite ligands	821	60	560
	949	50	576
	1179	70	415
	1241	50	490 (1HAH)
	1278	70	496 (1HAH)
Na^+ binding site	1195	80	418
	976	100	424
	1121	90	482
	838	100	514
Na^+ channel waters	1001	100	409
	1195	80	418
	976	100	424
	1196	90	428
	788	70	463
	914	100	464
	944	50	474
	1121	90	482
	915	90	497
	838	100	514
	1229	60	457 (1HAH)
Trypsin			
Active site catalytic triad residues			
	48	100	747
	80	100	702
Active site ligands	80	100	702
	61	100	710
	48	100	747
	64	67	807
	23	67	808
	77	100	805

[a]Only waters with at least 50% conservation and near a functional site in thrombin or trypsin are tabulated.

[b]Representative thrombin waters are from 1HAI unless there is no member water from 1HAI, in which case the structure of the representative water is noted. Representative trypsin waters are from 1TPO.

and ligand-bound water molecules to serine protease complex formation, because several high-resolution structures are available for ligand-free trypsin and BPTI structures and their complex. Using water microclusters identified for trypsin and BPTI, the conserved water sites from each protein were compared with water sites conserved in structures of the trypsin:BPTI complex (2PTC and 1TPA). The free trypsin structures were superimposed onto the trypsin chain in the 2PTC complex, and the free BPTI structures were superimposed onto the BPTI chain in 2PTC. Three conserved microclusters from the free structures overlapped with conserved water sites in the complex (large spheres in Fig. 5), two being contributed by trypsin and one by BPTI. Thus, three of the seven trypsin:BPTI interfacial water molecules were donated by the free proteins, whereas four were newly recruited or shuffled upon complex formation. This contrasts with the contributions of water molecules bound to the free structures of lysozyme and the D1.3 antibody, which contribute 20 of the 25 water molecules observed in the antibody:lysozyme interface (Braden et al., 1995). Thus, the hydration structure of the free protein and ligand and the creation of new environments favorable for water binding upon docking of the protein and ligand should both be considered in inhibitor design. A pattern recognition technique that predicts which water sites are favored upon protein-ligand binding is one approach for tackling this problem (Raymer et al., 1997).

Discussion

Conservation of water sites in thrombin and trypsin

A number of water sites were conserved in at least half of the thrombin and trypsin structures, and several sites were found in all of the structures examined (Table 2). An earlier detailed study of the solvent structure of trypsin (Finer-Moore et al., 1992) defined 211 consensus water sites via high-resolution X-ray diffraction data for the waters' oxygen atoms, verified by D_2O-H_2O difference neutron scattering density for the waters' hydrogen atoms. We found significantly fewer consensus water sites, 60, perhaps due to comparing three structures. A key goal was to distinguish conserved water sites characteristic of serine proteases in general from those contributing to ligand specificity. Thirty-seven overlapping conserved water sites were found between thrombin and trypsin, four and a half times the number expected for a random distribution of water sites. Finer-Moore et al. (1992) evaluated similarity in solvent structure between pairs of eight trypsin and trypsinogen structures and also found significant similarity between them. Ten of the 37 shared sites we observed were in contact with ligands or associated with the solvent channel proximal to the Na^+ site (Table 3). This is consistent with the observation of Krem and Di Cera (1998) that one-third of the conserved internal water sites in serine proteases (Sreenivasan & Axelson, 1992) are located near the Na^+ site; they proposed that the water structure stabilizes this pocket associated with substrate specificity (Krem & Di Cera, 1998). We also found two water sites conserved between thrombin and trypsin in a channel leading from Ser 214, which interacts with Asp of the catalytic triad; this solvent channel has been proposed as an exit path for protons produced during catalysis (Meyer, 1992).

Conserved water sites and ligand specificity

To identify water sites that can contribute to substrate specificity, we analyzed water sites conserved in the functionally important regions of thrombin and trypsin but not conserved between the two enzymes. The 22 water sites conserved in the active site, Na^+ binding region, and exosite of thrombin but not in trypsin, and the eight active-site water molecules conserved in trypsin but not in thrombin (Table 4) are likely to contribute to their different substrate specificities. Design of thrombin inhibitors may be optimized by mimicking these water interactions, as has been achieved for HIV protease (Lam et al., 1994) and cyclophilin-A (Mikol et al., 1995). Our study of 20 nonhomologous proteins bound to diverse ligands showed that water molecules in ligand-binding sites can be displaced by similarly polar ligand atoms (Raymer et al., 1997), but also that water-mediated bridges between protein

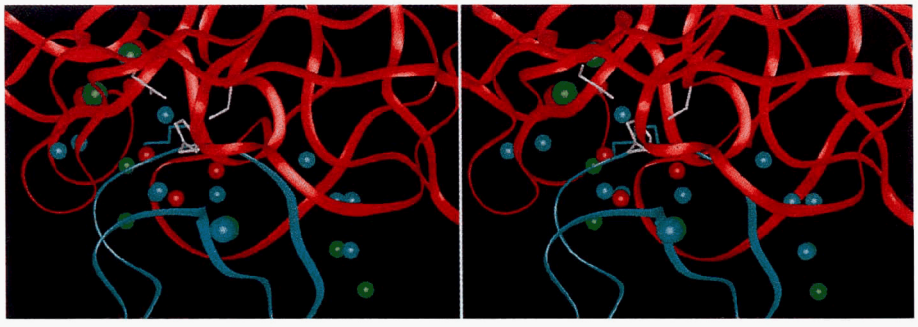

Fig. 5. Conservation of water sites in the trypsin:BPTI interface. This stereo close-up of the interface between trypsin (red ribbon, from PDB 1TPO) and BPTI (blue ribbon, from PDB 4PTI), superimposed onto the analogous chains of the trypsin:BPTI complex (PDB 2PTC), shows the conservation of water sites between the free structures and their complex. The orientation is approximately a 90° rotation about a horizontal axis relative to Figure 4. Conserved (≥50%) water sites from the free trypsin structure are shown as red spheres, those from BPTI are shown in blue, and interfacial waters found in both structures of the trypsin:BPTI complex (2PTC and 1TPA) are shown in green. Water sites overlapping between the complexes and free structures are rendered as large spheres, while non-overlapping sites are rendered as small spheres. The catalytic triad in trypsin is shown in white, and the side chain of the inhibitory Lys 15 from BPTI is shown in cyan. Of the seven trypsin:BPTI interfacial water sites, three are contributed by either trypsin or BPTI.

and ligand are ubiquitous, with 19 water-mediated hydrogen-bond interactions between proteins and small ligands, on average (A. Cayemberg & L.A. Kuhn, unpubl. results). Thus, the positions of conserved interfacial water molecules can be used to specify a template of favorable hydrogen bonds for ligands to satisfy, providing another strategy for optimizing ligand design.

Future directions

Water site conservation is being analyzed for other proteins, including the kringle domains of apolipoprotein(a) and plasminogen, as well as for structures solved in complex with an array of ligands, such as thymidylate synthase. Cluster analysis will also be applied to analyze the conservation of ligand atom positions and chemistry for these structures, which will then be used to develop a binding template, in combination with conserved water sites, for ligand screening and design. We will also test the ability of single-linkage clustering, which can define hydrogen-bonded chains or networks of water molecules in a single structure (rather than complete linkage, which is appropriate for defining water sites overlapping between structures), to locate functionally important water channels in proteins such as cytochrome *c* oxidase.

Conclusions

This study demonstrates hierarchical clustering as a useful tool for unbiased definition and analysis of consensus water sites when several independent structures of a protein are available; this approach is particularly useful for resolving the continuum of water site overlaps that occurs when a number of structures are superimposed. Analysis of colocalization between thrombin and trypsin water sites showed a small but significant number of overlaps predominantly surrounding the sodium ion site in thrombin and the corresponding region in trypsin. Cluster analysis of water sites and their environments also allowed us to identify the features associated with highly conserved water sites: (1) a high density of protein atom neighbors, indicating that the water is in a protein groove or cavity, (2) several hydrogen bonds to protein, (3) a hydrophilic environment, and (4) low thermal mobility of the site. Since cluster analysis is a general statistical method, it is also expected to be useful for analyzing the conservation of side-chain and ligand atom positions and their chemistries. The WatCH software developed to analyze conserved atom positions based on SPlus clustering trees, and the PDB-formatted files containing thrombin and trypsin conserved water microclusters will be made available upon request to the authors (sanschag@sol.bch.msu.edu and kuhn@agua.bch.msu.edu).

Acknowledgments

We would like to thank Cindy Fisher (Structural Bioinformatics, Inc.) and Alexander Tulinsky (Michigan State University) for their valuable feedback on the manuscript, and also acknowledge the American Heart Association, Michigan Affiliate (15GB978) and the National Science Foundation (BIR-9600831) for generously supporting this research.

References

Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng J. 1987. Protein Data Bank. In: Allen FH, Bergerhoff G, Sievers R, eds. *Crystallographic databases-information content, software systems, scientific applications*. Chester, UK: Data Commission of the International Union of Crystallography. pp 107–132.

Badger J. 1993. Multiple hydration layers in cubic insulin crystals. *Biophys J 65*:1656–1659.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Blevins RA, Tulinsky A. 1985. Comparison of the independent solvent structures of dimeric α-chymotrypsin with themselves and with γ-chymotrypsin. *J Biol Chem 260*:8865–8872.

Blow DM, Birktoft JJ, Hartley BS. 1969. Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature 221*:337–340.

Braden BC, Fields BA, Poljak RJ. 1995. Conservation of water molecules in an antibody-antigen interaction. *J Mol Recognit 8*:317–325.

Brunne RM, Liepinsh E, Otting G, Wüthrich K, van Gunsteren WF. 1993. A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. *J Mol Biol 231*:1040–1048.

Burling FT, Weis WI, Flaherty KM, Brünger AT. 1996. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science 271*:72–77.

Connolly ML. 1993. The molecular surface package. *J Mol Graph 11*:139–141.
Craig L, Sanschagrin PC, Rozek A, Lackie S, Kuhn LA, Scott JK. 1998. The role of structure in antibody cross-reactivity between peptides and folded proteins. *J Mol Biol 281*:183–201.
Dang QD, Di Cera E. 1996. Residue 225 determines the Na^+-induced allosteric regulation of catalytic activity in serine proteases. *Proc Nat Acad Sci USA 93*:10653–10656.
Denisov VP, Peters J, Hörlein HD, Halle B. 1996. Using buried water molecules to explore the energy landscape of proteins. *Nat Struct Biol 3*:505–509.
Di Cera E, Guinto ER, Vindigni A, Dang QD, Ayala YM, Wuyi M, Tulinsky A. 1995. The Na^+ binding site of thrombin. *J Biol Chem 270*:22089–22092.
Earnest T, Fauman E, Craik CS, Stroud RM. 1991. 1.59 Å structure of trypsin at 120 K: Comparison of low temperature and room temperature structures. *Proteins Struct Funct Genet 10*:171–187.
Edsall JT, McKenzie HA. 1983. 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.
Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature 319*:199–203.
Esmon CT. 1992. The protein C anticoagulant pathway. *Arterioscl Thromb 12*:135–145.
Faerman CH, Karplus PA. 1995. Consensus preferred hydration sites in six FKBP12-drug complexes. *Proteins Struct Funct Genet 23*:1–11.
Finer-Moore JS, Kossiakoff AA, Hurley JH, Earnest T, Stroud RM. 1992. Solvent structure in crystals of trypsin determined by X-ray and neutron diffraction. *Proteins Struct Funct Genet 12*:203–222.
Furie B, Furie BC. 1988. The molecular basis of blood coagulation. *Cell 53*:505–518.
Jiang J-S, Brünger AT. 1994. Protein hydration of penicillopepsin and neuraminidase crystal structures. *J Mol Biol 243*:100–115.
Joachimiak A, Haran TE, Sigler PB. 1994. Mutagenesis supports water mediated recognition in the trp repressor/operator system. *EMBO J 13*:367–372.
Karplus PA, Faerman CH. 1994. Ordered water in macromolecular structure. *Curr Opin Struct Biol 4*:770–776.
Kossiakoff AA, Sintchak MD, Shpungin J, Presta LG. 1992. Analysis of solvent structure in proteins using neutron D_2O-H_2O solvent maps: Pattern of primary and secondary hydration in trypsin. *Proteins Struct Funct Genet 12*:223–236.
Krem MM, Di Cera E. 1998. Conserved water molecules in the specificity pocket of serine proteases and the molecular mechanism of Na^+ binding. *Proteins Struct Funct Genet 30*:34–42.
Kuhn LA, Siani MA, Pique ME, Fisher CL, Getzoff ED, Tainer JA. 1992. The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. *J Mol Biol 228*:13–22.
Kuhn LA, Swanson CA, Pique ME, Tainer JA, Getzoff ED. 1995. Atomic and residue hydrophilicity in the context of folded protein structures. *Proteins Struct Funct Genet 23*:536–547.
Kuntz ID, Kauzmann W. 1974. Hydration of proteins and polypeptides. *Adv Protein Chem 28*:239–345.
Ladbury JE. 1996. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem Biol 3*:973–980.
Lam PY, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, Meek JL, Otto MJ, Rayner MM, Wong YN, Chang C-H, Weber PC, Jackson DA, Sharpe TR, Erickson-Viitanen S. 1994. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science 263*:380–384.
Levitt M, Park BH. 1993. Water: Now you see it, now you don't. *Structure 1*:223–226.
Meyer E. 1992. Internal water molecules and H-bonding in biological macromolecules: A review of structural features with functional implications. *Protein Sci 1*:1543–1562.
Mikol V, Papageorgiou C, Borer X. 1995. The role of water molecules in the structure based design of (5-hydroxynorvaline)-2-cyclosporin: Synthesis, biological activity, and crystallographic analysis with cyclophilin A. *J Med Chem 38*:3361–3367.
Otting G, Liepinsh E, Wüthrich K. 1991. Protein hydration in aqueous solution. *Science 254*:974–980.
Otwinowski Z, Schevitz RW, Zhang R-G, Lawson CL, Joachimiak A. 1988. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature 335*:321–329.
Perona JJ, Craik CS, Fletterick RJ. 1993. Locating the catalytic water molecule in serine proteases. *Science 261*:620–621.
Pitt WR, Murray-Rust J, Goodfellow JM. 1993. AQUARIUS2: Knowledge-based modeling of solvent sites around proteins. *J Comp Chem 14*:1007–1018.
Rashin AA, Iofin M, Honig B. 1986. Internal cavities and buried waters in globular proteins. *Biochemistry 25*:3619–3625.
Raymer ML, Sanschagrin PC, Punch WF, Venkataraman S, Goodman ED, Kuhn LA. 1997. Predicting conserved water-mediated and polar ligand interactions in proteins using a k-nearest-neighbors genetic algorithm. *J Mol Biol 265*:445–464.
Ringe D. 1995. What makes a binding site a binding site? *Curr Opin Struct Biol 5*:825–829.
Rupley JA, Careri G. 1991. Protein hydration and function. *Adv Protein Chem 41*:38–173.
Sack JS. 1988. CHAIN: A crystallographic modeling program. *J Mol Graph 6*:224–225.
Schiffer CA, Huber R, Wüthrich K, van Gunsteren WF. 1994. Simultaneous refinement of the structure of BPTI against NMR data measured in solution and X-ray diffraction data measured in single crystals. *J Mol Biol 241*:588–599.
Singer PT, Smalås A, Carty RP, Mangel WF, Sweet RM. 1993. The hydrolytic water molecule in trypsin, revealed by time-resolved Laue crystallography. *Science 259*:669–673.
Sreenivasan U, Axelson PH. 1992. Buried water in homologous serine proteases. *Biochemistry 31*:12785–12791.
van Gunsteren WF, Berendsen HJC, Hermans J, Hol WGJ, Postma JPN. 1983. Computer simulation of the dynamics of the hydrated protein crystals and its comparison with X-ray data. *Proc Nat Acad Sci USA 80*:4315–4319.
Vijayalakshmi J, Padmanabhan KP, Mann KG, Tulinsky A. 1994. The isomorphous structures of prethrombin2, hirugen-, and PPACK-thrombin: Changes accompanying activation and exosite binding to thrombin. *Protein Sci 2*:2254–2271.
Wang H, Ben-Naim A. 1996. A possible involvement of solvent-induced interactions in drug design. *J Med Chem 39*:1531–1539.
Williams MA, Goodfellow JM, Thornton JM. 1994. Buried waters and internal cavities in monomeric proteins. *Protein Sci 3*:1224–1235.
Wilson IA, Fremont DH. 1993. Structural analysis of MHC class I molecules with bound peptide antigens. *Semin Immunol 5*:75–80.
Zhang E, Tulinsky A. 1997. The molecular environment of the Na^+ binding site of thrombin. *Biophys Chem 63*:185–200.
Zhang X-J, Matthews BW. 1994. Conservation of solvent-binding sites in 10 crystal forms of T4 lysozyme. *Protein Sci 3*:1031–1039.