Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Role of water molecules in the structure and function of aspartic proteinases

B. V. L. S. Prasad and K. Suguna*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Correspondence e-mail: suguna@mbu.iisc.ernet.in

Received 15 August 2001
Accepted 30 October 2001

The solvent structure in the crystals of ten aspartic proteinases has been analyzed to find the possible roles of conserved water molecules in their structure and activity. 17 waters have been identified which are common to at least eight of the ten examined enzyme structures. These include the catalytic water molecule, whose direct involvement in the mechanism of action of aspartic proteinases has been proposed previously. There appears to be at least one more functionally important water molecule strategically located to stabilize the flexible 'flap' region during substrate binding. Many other waters stabilize the structure, whilst a few have been found to maintain the active-site geometry required for the function of the enzyme. In particular, two waters related by the approximate molecular dyad are involved in the formation and preservation of a network of hydrogen bonds extending from the active site.

1. Introduction

Water molecules are now considered to be an integral part of proteins, aiding in stabilizing the protein fold and participating in their function. Protein hydration is known to play a crucial role in biological processes. With the availability of a large number of crystal structures of proteins at higher resolutions, it is now possible to analyze and compare the solvent structure within each class of proteins. Often, meaningful conclusions about the role of the solvent can be drawn by looking at a group of related proteins rather than individual proteins. There have been a limited number of studies on the comparison of the solvent structure in related proteins. Investigations on the solvent sites of T4 lysozyme (Zhang & Matthews, 1994), ribonuclease A (Kishan *et al.*, 1995) and hen egg-white lysozyme (Biswal *et al.*, 2000) resulted in the identification of a number of invariant water molecules for each of these enzymes. In these cases, identical proteins were studied under varying crystallization conditions, solvent content, crystal packing or with minor mutations. Analysis of conserved water molecules in homologous proteins has been carried out only for two families, legume lectins (Loris *et al.*, 1994) and serine proteases (Sreenivasan & Axelsen, 1992; Krem & Enrico, 1998), in which a small portion of water molecules was found to be conserved in each family, in spite of significant variations in the sequence. These observations suggest that the positions of certain water molecules must be conserved for structural and/or functional reasons. In the present study, we extend this type of analysis to the aspartic proteinase family.

Aspartic proteinases take part in activities as diverse as gastric digestion (pepsin and gastricsin), maintenance of blood pressure (renin), milk clotting (chymosin), protein turnover

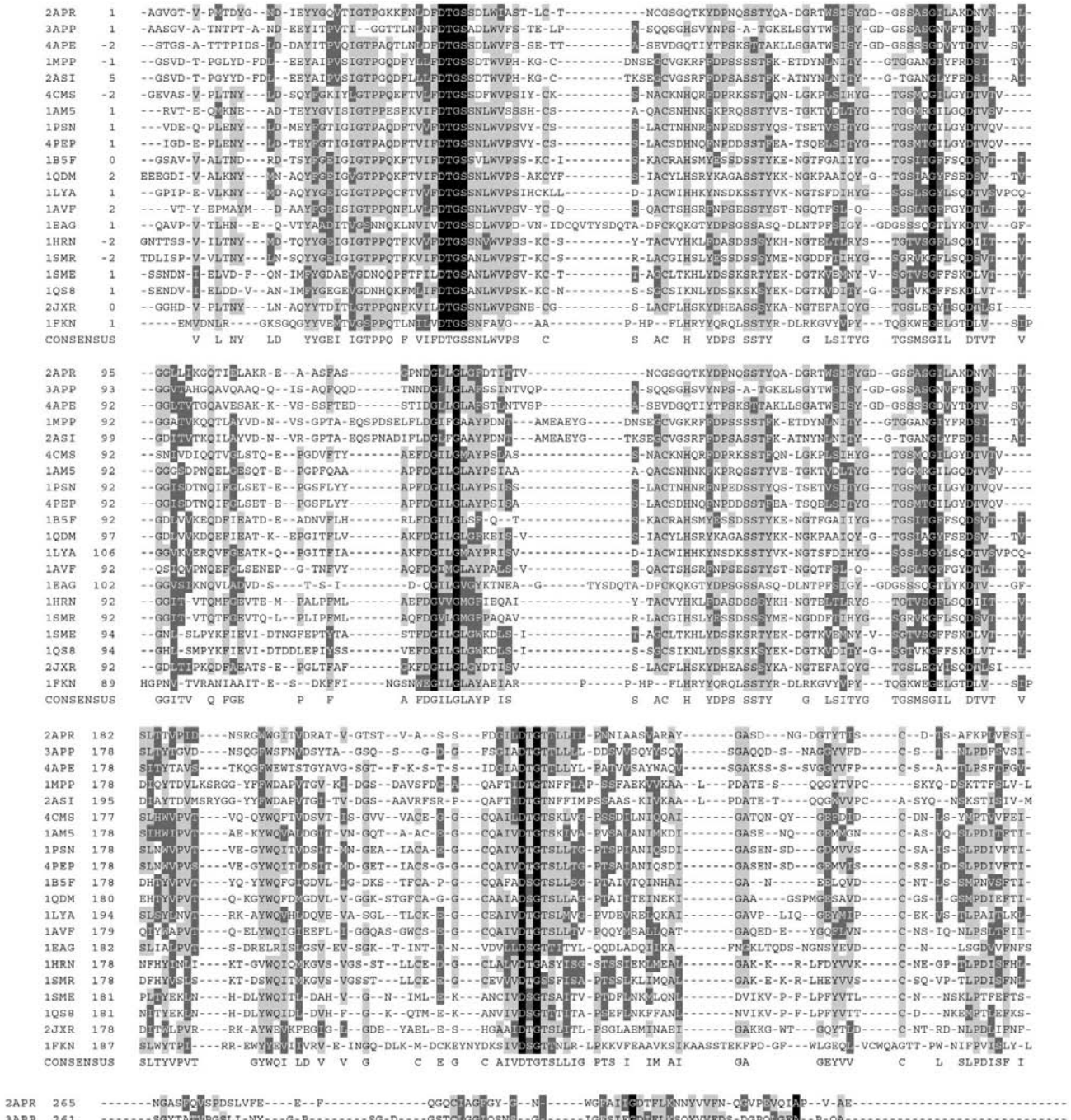

Figure 1
Structure-based sequence alignment of aspartic proteinases.

Table 1
Aspartic proteinases from different sources whose crystal structures are available in the PDB.

a, Ods-Pss-Lyw-Cha-Vas-Lyt; *b*, Ppp-Cpp-Cha-Ip3; *c*, Piv-His-Pro-Phe-His-Lpl-Tyr-Tyr-Ser; *d* and *e*, Iva-Val-Val-Sta-Ala-Sta; *f*, Mor-Phe-Nle-Chf-Nme; *g*, Glu-Val-Asn-Lol-Alq-Ala-Glu-Phe.

No	PDB code	Name (pH at which crystals were grown)	Proteinase type	Resolution (Å)	No. of residues	Sequence identity† (%)	R.m.s. deviation†	Pairs aligned†	No. of waters in PDB	No. of primary hydration waters	Invariant waters†	Stamp score
1	2apr	Rhizopuspepsin (6.0)	Native	1.8	325	100	0.000	1931	373	352	352	10.00
2	3app	Penicillopepsin (4.6)	Native	1.8	323	42.50	0.977	1297	318	340	134	8.010
3	4ape	Endothiapepsin (4.5)	Native	2.2	330	41.79	1.105	1344	343	216	81	8.016
4	1mpp	Pusillopepsin (4.5)	Native	2.0	357	31.82	1.483	1376	221	220	72	7.447
5	2asi	Mucorpepsin (4.0)	Native	2.15	356	31.75	1.483	1364	140	134	42	7.647
6	4cms	Chymosin (5.6)	Native	2.2	320	35.59	1.171	1352	131	126	62	8.006
7	1am5	Atlantic cod pepsin (5.4)	Native	2.16	324	35.84	0.922	1139	161	137	50	7.791
8	1psn	Human pepsin (5.0)	Native	2.2	326	40.63	1.168	1382	136	114	62	8.114
9	4pep	Porcine pepsin (2.0)	Native	1.8	326	39.08	1.242	1385	187	160	78	7.995
10	1b5f	Cardosin (5.5)	Native	1.72	320	34.04	1.201	1369	528	242	107	7.906
11	1qdm	Prophytepsin	Zymogen	2.3	332	34.44	1.349	1343	—	—	—	7.337
12	1lya	Cathepsin D	Native‡	2.5	338	33.80	1.171	1365	23	—	—	7.992
13	1avf	Gastricsin	Intermediate	2.36	320	32.71	1.289	1294	467	222	70	7.134
14	1eag	*Candida* Sap2	Complexed with *a*	2.1	342	32.72	1.253	1336	119	127	53	7.769
15	1hrn	Human renin	Complexed with *b*	1.8	327	28.01	1.465	1395	366	186	70	7.676
16	1smr	Rat renin	Complexed with *c*	2.0	335	30.69	1.637	1402	156	131	61	7.499
17	1sme	Plasmepsin II (*P. falciparum*)	Complexed with *d*	2.7	329	25.55	1.461	1331	115	56	19	7.577
18	1qs8	Plasmepsin II (*P. vivax*)	Complexed with *e*	2.5	330	28.90	1.596	1315	154	84	34	7.249
19	2jxr	Saccharopepsin	Complexed with *f*	2.4	330	36.77	1.274	1387	119	89	47	7.970
20	1fkn	Memapsin 2	Complexed with *g*	1.9	385	24.49	1.159	1211	529	378	81	7.260

† With respect to rhizopuspepsin. ‡ Very few waters in the PDB.

(cathepsin D), parasite viability (retropsins and plasmepsins) *etc*. Initially reported proteinase crystal structures (Davies, 1990) were of mammalian and fungal origin. The structures of retroviral enzymes were solved subsequently. The Protein Data Bank (Berman *et al.*, 2000) now contains aspartic proteinases from 20 different sources such as vertebrates, plants and yeast, plasmepsins from malarial parasites which degrade host haemoglobin and a more recently solved membrane-associated enzyme memapsin 2 that is implicated in Alzheimer's disease (Hong *et al.*, 2000). These are available either in the native form or as zymogens or complexes.

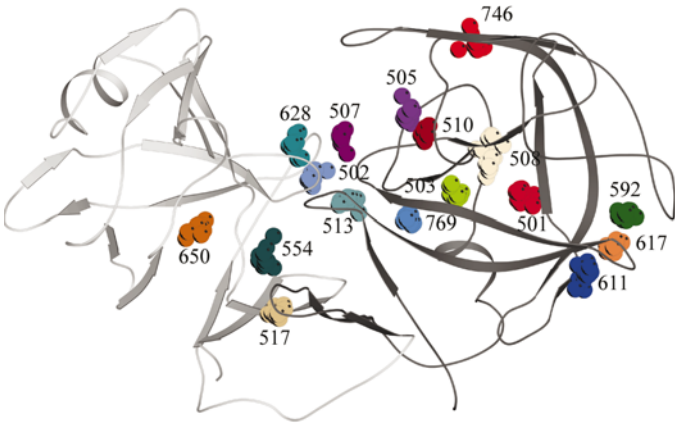

Figure 2
Locations of the 17 most commonly found water molecules in the crystal structures of ten native aspartic proteinases superimposed on rhizopuspepsin. All figures were generated using the program *MOLSCRIPT* (Kraulis, 1991).

According to the structural classification of proteins (SCOP; Murzin *et al.*, 1995), aspartic proteinases belong to the all-β protein class, acid proteinase fold and acid proteinase superfamily, which is divided into the pepsin-like and the retroviral families. The present analysis is confined to the monomeric pepsin-like aspartic proteinases found in vertebrates, plants, fungi and other microbial sources, but does not include the dimeric retroviral enzymes. Each protein consists of about 325 amino-acid residues (except the plant proteins, which have an additional saposin-like domain) which fold into two domains with the active site situated between them. They make use of two aspartates, one from each domain, for their function. The geometry of the active site is well conserved among various aspartic proteinases but interestingly, their specificities show wide variation.

The present analysis intends to find the involvement of water molecules, apart from the ubiquitous water that is implicated in the catalytic activity of the enzyme, in the structural and functional aspects of aspartic proteinases. The roles of 17 waters that are found to be most common to the enzymes have been thoroughly investigated. The protein environment and the conserved features pertaining to these waters have been examined. In certain cases, it has become possible to attribute specific roles to water molecules as catalytic, structural or functional, although these roles are generally strongly correlated.

2. Methods

Crystal structures of the aspartic proteinases from 20 different sources are available in the PDB (Table 1). Of these, ten

Table 2
Hydrogen bonds between conserved waters and the protein.

Numbering is as in rhizopuspepsin.

Water	Protein atoms			
	Main chain	Side chain	Waters	Remarks on the residues
Wat510	Asp40 O			Asp/Asn. Val in 1b5f
	Asp40 N			
		Ser38 OG		Conserved
		Tyr77 OH		Conserved. No density in 1avf
Wat505	Ser39 N			Ala in four proteins
	Ile130 O			Conserved in about half of the proteins
		Asp40 OD2		Asp/Asn. Val in 1b5f
Wat503	Gly124 N			Conserved
	Gly124 O			
			Wat589	
Wat502	Thr36 O			Conserved
	Trp195 N			Tyr in 1fkn
	Leu217 O			Hydrophobic in all
Wat611	Gly24 O			Conserved
			Wat552	
Wat592	Thr65 O			Ser/Thr
			Wat617	Water is conserved
Wat507		Asp35 OD1		Conserved
		Asp35 OD2		
		Asp218 OD1		Conserved
		Asp218 OD2		
			Wat533	
			Wat694	
			Wat858	
Wat554	Pro316 O			Pro only in 3app, 2apr, 4ape
			Wat559	
Wat508	Leu41 N			Conserved in many
	Thr103 O			Not conserved
		Glu105 OE1		Not conserved
			Wat589	
Wat517	Leu158 N			Mainly Leu or Met/Tyr
	Leu158 O			
	Leu304 O			Leu or hydrophobic
			Wat691	
Wat513	Tyr17 N			Phe in 1sme, 1qs8
	Thr219 O			Conserved, Ser in six proteins
		Asp33 OD2		Mainly hydrophobic (Asp only in 2apr, 4ape; Asn in 3app)
			Wat515	
Wat769	Gly126 O			Gly/Ala. Ser in 1b5f
		Thr138 OG1		Thr/Pro
		Gln313 OE1		Not conserved
			Wat519	
			Wat535	
			Wat572	
Wat617	Gly101 N			Not conserved
		Asn91 OD1		Not conserved
			Wat592	Water is conserved
.Wat501	Thr103 N			Not conserved
	Lys137 O			Not conserved
		Asn142 ND2		Asn in 12 proteins
			Wat516	
Wat628	Asp189 O			Not conserved
	Arg192 N			Not conserved
	Trp194 O			Tyr in 1fkn
			Wat646	
Wat650	Pro187 O			Pro in 11 proteins
	Thr198 N			Not conserved
			Wat608	
			Wat685	
			Wat526	
			Wat738	
Wat746	Ser74 N			Not conserved
	Ser74 O			
		Thr132 OG1		Not conserved

native forms (1–10 in Table 1) containing the coordinates of water molecules have been used in the present analysis for identifying the conserved water molecules. The resolution of these structures varies from 2.2 to 1.75 Å. In this resolution range, the positions of solvent molecules are determined with reasonable accuracy; in particular those that are close to the protein atoms constituting the first hydration shell. Structures of zymogens and complexes (11–20 in Table 1) were subsequently used for examining these waters and for structure-based multiple sequence alignment. In the absence of the native structures, the complexed forms at the highest available resolution have been considered in this set. 1eag, 1smr, 1sme, 1qs8, 2jxr and 1fkn fall into this category. In the case of the native structure of human renin (1bss), as the water positions were not determined, the complexed form, 1hrn, which has the highest resolution and a large number of waters, has been selected. A typical set of complexes has also been selected for comparing the invariant waters between the free and the bound forms of the enzymes. The PDB codes of these complexes are 1bxq (resolution 1.41 Å, complex of 3app), 2er7 (1.6 Å, 4ape), 3apr (1.8 Å, 2apr) and 1czi (2.3 Å, 4cms).

Waters which make hydrogen bonds (distance < 3.6 Å, angles N—H$\cdots$Wat $> 120^\circ$ and C—O$\cdots$Wat $> 90^\circ$) with the protein atoms, referred to as the primary hydration waters, were identified for all the proteinases. Rhizopuspepsin (2apr) has been taken as the reference molecule. All other proteinases with primary hydration waters were superimposed on 2apr with the program *ALIGN* (Cohen, 1997) using the main chain and the C^β atoms for alignment. The aligned proteinases along with their primary hydration waters were used for finding the conserved waters. Waters of the proteinases within 1.8 Å of those of 2apr having at least one common hydrogen bond to the protein were considered as conserved or equivalent waters. While the distance criterion was generally followed, in certain instances waters were considered equivalent if

similar hydrogen bonds were formed even if the distance criterion was not satisfied. Such cases arise owing to variations in local sequence and/or conformations. Water–water hydrogen bonds were assigned based on a distance cutoff of 3.6 Å only.

Multiple structure alignment of all 20 proteinases was performed using the program *STAMP* (Russel & Barton, 1992). The resultant structure-based sequence alignment (Fig. 1) has been extremely useful in comparing not only the individual residues but also the regions of proteins around the water molecules, as discussed later. The surface accessibility of each primary hydration water was computed in the presence of only the protein atoms, using the molecular-surface package of Connolly (1983). The probe radius was taken to be 1.2 Å. Waters with accessible surface area of less than 2.0 Å^2 were considered as internal waters.

2APR	3APP	4APE	2ASI	1MPP	4CMS	1AM5	1PSN	4PEP	1B5F
510	**723**	**534**	**453**	**632**	**610**	**614**	**415**	**334**	**417**
−0.4155	1.6201	−0.2225	−0.7448	−0.8575	−0.3210	−0.7541	−0.5072	−0.8375	−0.4565
0.000	0.864	0.055	0.581	0.000	6.106	0.839	0.311	0.073	0.743
505	517	535	406	631	506	636	430	333	644
−0.6432	−0.1798	−0.5028	−0.4242	−0.6821	−0.7797	−0.3167	−0.3080	−0.8163	0.9900
6.255	10.859	9.743	10.302	6.585	19.521	13.719	7.612	4.172	20.169
503	**501**	538	**405**	**634**	**507**	**729**	**421**	**330**	**446**
−0.7704	−1.1873	−0.7832	−0.1662	0.0195	−0.6574	0.2421	−0.3861	−0.9868	−0.5686
1.352	0.800	0.891	0.605	0.161	0.266	0.398	1.066	0.708	1.154
502	**504**	**533**	**404**	**637**	**612**	**602**	**414**	**328**	**424**
−0.8293	−0.7633	−0.7832	−1.1017	−0.7990	−0.7797	−0.9036	−0.8422	−1.1635	−0.7487
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
611	586	818	401	621	530		**406**	377	**444**
0.2195	1.9118	0.6986	−0.0898	0.0780	−0.5045	—	0.7525	−0.0443	−0.3949
7.603	12.483	14.729	10.934	8.646	4.604		0.605	6.177	1.903
592	692	656	408		520	635	451	376	461
0.9759	0.7449	−0.2625	−0.6771	*	−0.1987	1.6309	0.4431	0.4724	−0.3719
24.605	23.313	30.656	16.047		12.419	10.274	7.985	6.419	10.688
507	539	502	421	626	505	621	437	355	416
−0.5929	0.1735	−0.5028	−0.9268	−0.3313	0.0766	−0.4550	−0.0068	0.1253	−0.5139
20.600	22.571	16.859	24.952	22.902	27.736	27.278	22.375	24.630	19.929
554	785	582	438	724	552		450	445	704
1.9383	−0.1152	−0.1024	0.5514	0.4873	0.3824	—	1.0546	0.7920	−0.0541
30.610	33.935	19.638	29.163	30.568	35.715		35.303	33.040	29.815
									795
									−0.3030
									34.148
508	**507**	**536**	**424**	**633**	**613**	677	**422**	**332**	**445**
−0.5797	−0.5594	−0.5829	−0.9114	−0.5067	0.9023	−0.5270	0.0087	−0.8313	−0.3411
1.212	0.760	0.963	1.387	1.456	0.405	2.007	1.005	1.309	0.968
								337	
								−0.7229	
								2.717	
517			**418**	**700**	**581**	**626**	**402**	**364**	**403**
−0.1943	*	*	−1.0145	−0.6236	−0.4433	0.2571	−0.4109	−0.1543	−0.9444
0.402			0.327	0.790	0.000	0.000	0.000	0.000	0.000
513	**532**	526	420	**610**			497	351	**468**
−0.3243	0.3027	−0.3827	−1.0002	−0.2728	—	—	0.5204	0.0695	−0.5186
2.484	1.824	2.735	0.000	0.000			2.267	3.180	1.677
769	**529**	559			**527**	**605**	**416**	**336**	**433**
−0.5664	0.2240	−0.0623	*	*	−0.5657	−1.0513	−0.5122	−0.7629	−0.6422
1.830	1.073	2.693			1.340	0.496	0.607	0.740	1.297
617	701[b]	655	409	703	544	616	417	341	420
1.1392		0.0178	−1.0371	−0.2728	−0.5045	−1.4195	0.1819	−0.6239	0.2501
30.106	34.031	11.960	12.895	6.598	8.674	11.027	12.486	12.094	11.143
501	510	561	411	661		620	417	341	672
−0.9299	−0.5614	−0.6230	−0.9874	−0.6821	—	0.3935	−0.2933	−0.5972	−0.6500
7.483	6.977	4.041	10.136	4.361		10.869	4.968	7.404	5.161
628	508	**587**	429	**719**	**598**		424	344	**414**
−0.5180	−0.5432	−0.4628	−0.9494	−0.6236	−0.7797	—	−0.3344	−0.5799	−0.6704
4.650	2.517	0.000	9.762	0.000	1.627		6.606	7.614	0.000
			430	**720**					
			−0.6658	−0.3898					
			2.860	0.072					
650	509	578			534		440	360	423
−0.3177	−0.4685	−0.7031	*	*	−0.7797	*	−0.1097	0.2604	−0.5838
17.271	11.530	10.955			15.295		8.482	14.417	13.122
746	577		**427**	**666**	621		442		613
1.4742	1.6312	—	−0.2144	−0.3898	0.2906	—	0.2696	*	0.5251
17.978	17.454		0.047	0.023	33.746		27.389		30.362

Figure 3
Normalized *B* factor and accessible surface area (in Å^2) of the conserved water molecules. Normalized *B* factor $B_i' = (B_i - \langle B \rangle)/\sigma(B)$, where B_i is the *B* factor of each atom, $\langle B \rangle$ is the mean *B* factor of the protein molecule and $\sigma(B)$ is the standard deviation of the B factors. —, not found. * indicates that water cannot be found owing to the presence of a side chain in the vicinity. Bold numbers indicate waters with accessible surface areas of <2.0 Å^2.

3. Results and discussion

There are 42 water molecules that are conserved in at least five of the ten enzymes analyzed. We discuss in detail the 17 waters (Fig. 2) present in a minimum of eight proteinase structures, since it becomes difficult to draw any general conclusions about the rest of the waters. These 17 water molecules are listed in Fig. 3 along with their normalized *B* factors and accessibilities. Table 2 lists the hydrogen bonds in rhizopuspepsin involving the 17 conserved waters. Rhizopuspepsin has been taken as the representative of aspartic proteinases throughout the discussion and unless specifically mentioned the residue numbers and the water numbering correspond to those of rhizopuspepsin. While identification of a few well conserved waters was straightforward, difficulties have been encountered with regard to the others, as selecting waters from the electron-density maps is not a standard procedure but depends on the resolution, extent of refinement and individual judgement. Such difficulties are anticipated even while analyzing the same protein in different crystal structures. In the case of homologous structures, an additional factor of local sequence/structure variations come in the way of finding out the true and inherent characteristics of solvent molecules. Hence, critical examination of water–protein interactions and visual inspection using graphics were also carried out before attributing any role to the water molecules. The observations from the present analysis are as follows. An examination of Table 2 shows that the conserved water molecules make more hydrogen bonds to the main-chain atoms than to the side-chain atoms. It has also been observed that a significant number of water molecules are conserved in the vicinity of the active

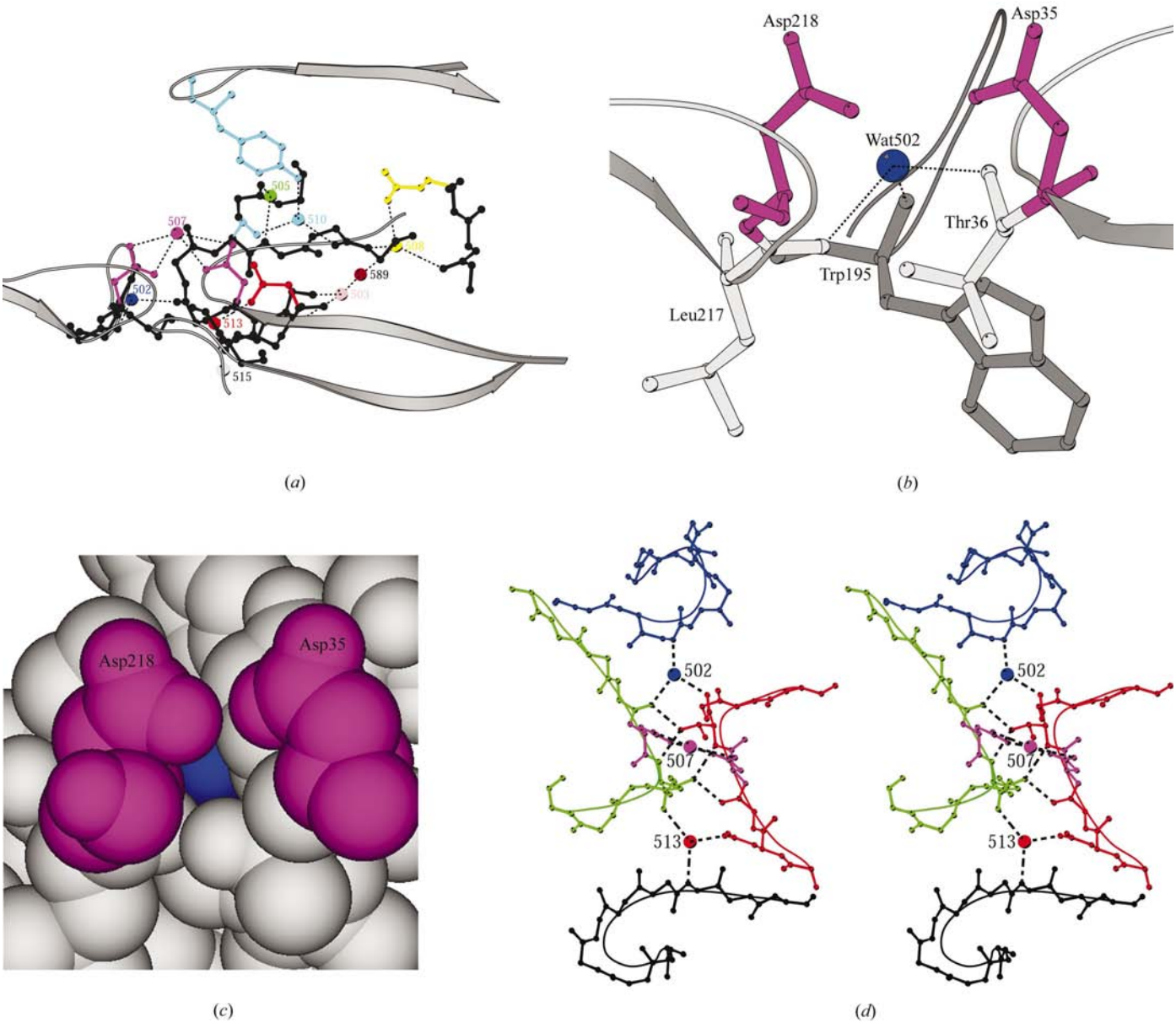

Figure 4
Conserved waters at the active site of rhizopuspepsin. The active aspartates are shown in magenta. (*a*) Network of hydrogen bonds in the vicinity of the active site in the binding cleft formed by conserved water molecules. Side chains and the water molecules they are interacting with are shown in the same colour. (*b*) The position of the buried conserved water molecule Wat502 with respect to the active aspartates. Wat502 bridges the two lobes of the enzyme through hydrogen bonds. (*c*) CPK representation of the atoms around the cavity in which Wat502 gets buried. (*d*) Stereoview of the 'fireman's grip' between the φ-loops (red and green) viewed down the approximate dyad that relates the N- and the C-terminal domains of rhizopuspepsin. Wat502 and Wat513, also related by the twofold axis, extend the 'fireman's grip' by making additional hydrogen bonds.

site. A majority of the conserved waters are located in the N-terminal domain. This perhaps reflects the differences in the conformational stability between the two domains as observed in porcine pepsin, whose N-terminal domain denatures while the C-terminal domain stays intact during alkaline inactivation of the enzyme (Lin *et al.*, 1993). None of the 17 waters is associated with the flexible subdomain of the C-terminal lobe (residues 200–215 and 223–300). No conserved water clusters or proton wires as identified in serine proteases (Sreenivasan & Axelsen, 1992; Krem & Enrico, 1998) have been found in aspartic proteinases.

3.1. The catalytic water

The role of Wat507, which makes hydrogen bonds with the active aspartates Asp35 and Asp217 (Fig. 4*a*) in the catalytic mechanism of aspartic proteinases, is well known. It is found in all pepsins including the retroviral proteinases and also in the activation intermediate 2 of human progastricsin, which represents a structural state at least one step prior to the formation of the mature enzyme in the activation pathway. This water could be located in the electron-density maps of high-resolution crystal structures of proteinases (Suguna, Padlan *et al.*, 1987). Its involvement in a nucleophilic attack on

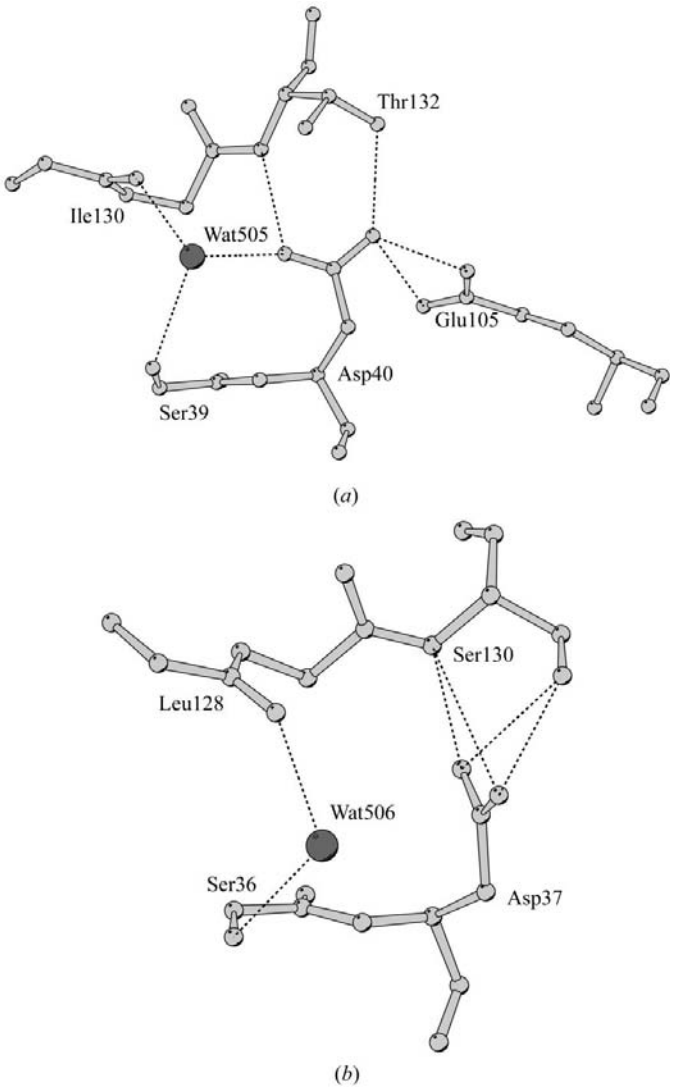

Figure 5
Change in the conformation of side chains owing to sequence changes in the vicinity as illustrated in (*a*) Asp40 of rhizopuspepsin and (*b*) the corresponding Asp37 in chymosin.

the carbonyl carbon of the scissile bond during catalysis was proposed based on the crystal structure of rhizopuspepsin with a reduced inhibitor (Suguna, Bott *et al.*, 1987). This water is displaced by small inhibitors designed to mimic the substrate when they are bound to the enzymes and is also found to be absent in the proenzymes. However, in the recently solved crystal structure of a new type of inhibitor complexed with saccharopepsin (Li *et al.*, 2000) it was observed that the water remains in place. The inhibitor binds as a long α-helix making no direct interactions with the active aspartates.

3.2. Waters at the active site related by the approximate twofold axis

Wat502, a totally buried water, is located close to the active site, making hydrogen bonds with the main-chain atoms of three conserved residues, Thr36, Leu217 and Trp195, acting as a bridge between the two domains (Figs. 4*b* and 4*c*). There are two tryptophans in the vicinity of this water which remain aromatic residues in other aspartic proteinases, except *Candida* Sap2. In human renin and rat renin, the corresponding water is slightly displaced as the loop containing Trp195 has a different structure. Nevertheless, Wat···Thr O and Wat···Trp N hydrogen bonds are conserved.

The two domains of the aspartic proteinases are related by an approximate twofold axis passing through Wat507. The two active aspartates are related by this dyad. Interestingly, a conserved water molecule, Wat513, has been found to be related to Wat502 by the same dyad (Fig. 4*d*). Superposition of the N- and the C-terminal domains of rhizopuspesin (Suguna *et al.*, 1987) showed that Tyr17 and Thr219, which are hydrogen bonded to Wat513, are twofold related to Trp195 and Thr36, respectively, which make hydrogen bonds with Wat502. Wat513 makes a hydrogen bond with Asp33 OD2 in rhizopuspepsin, penicillopepsin and endothiapepsin. This residue is Asp/Asn only in these three fungal proteinases (Fig.1), which are specific for substrates with a lysine in the P1 position. It has been shown by mutational analysis and kinetic studies that Asp33 is required to provide the optimum electrostatic environment for substrate binding (Lowther *et al.*, 1995).

Wat502 and Wat513 along with the other hydrogen bonds bridge the two φ loops on which the active aspartates are located. It appears that the 'fireman's grip' (Pearl & Blundell, 1984) formed by four hydrogen bonds at the active site obtains additional stability from these two conserved water molecules (Fig. 4*d*). The rigidity of the active-site geometry and of the 'fireman's grip' are so essential for the activity of the enzyme that the collapse of the 'fireman's grip' in proplasmepsin II makes it inactive, as the active aspartates are pulled away from each other owing to a reorientation of the domains (Bernstein *et al.*, 1999; Khan *et al.*, 1999).

3.3. Other waters in the vicinity of the active site

Apart from the waters discussed above, three other water molecules (Wat503, Wat505, Wat508) located near the substrate-binding region form a network of hydrogen bonds extending into the N-terminal domain (Fig. 4*a*). Wat503 is hydrogen bonded to both the N and O atoms of Gly124, which is a totally conserved residue. Wat505, close to the active site, hydrogen bonds to Ser39 N and also to Asp40 OD2 (Fig. 5*a*). In chymosin, this Asp has a different conformation (Fig. 5*b*), enabling hydrogen bonding to a nearby Ser which replaces the Thr in rhizopuspepsin. Wat508 makes a conserved hydrogen bond with Leu41 N.

3.4. A water strategically placed to stabilize the active conformation of the 'flap'

A comparison of the native and the inhibitor-bound forms of aspartic proteinases has revealed that a hydrogen bond is formed between Wat510 and Tyr77 OH of the 'flap' in the complexes, even if not formed in the native structures. On the other hand, if the hydrogen bond already exists in the native structures, it becomes stronger in the complexes, as in rhizo-

puspepsin (Fig. 6*a*). In penicillopepsin, the hydrogen bond is absent in the native structure (Wat···Tyr77 OH is 4.83 Å), but strong hydrogen bonds are formed in the complexes (Fig. 6*b*). In chymosin, the tyrosine residue changes its conformation to enable the formation of the hydrogen bond (Fig. 6*c*) in the complex. Thus, Wat510 plays a role in bringing the 'flap' towards the substrate to enhance the interaction between the enzyme and the substrate. This 'cantilever' type (Harte *et al.*, 1990) movement of the 'flap' occurs in many complexes of aspartic proteinases and this water appears to be aiding in this mechanism.

The equivalent of Wat510 also exists in the complex of porcine pepsin with a large protein inhibitor from *Ascaris suum*, whose mode of binding is totally different from that of any of the known inhibitors. One of the β-sheets of the inhibitor joins the β-strands of the 'flap' to form a continuous β-sheet in the complex (Ng *et al.*, 2000). In the complex of saccharopepsin with its own inhibitor, which binds as a long

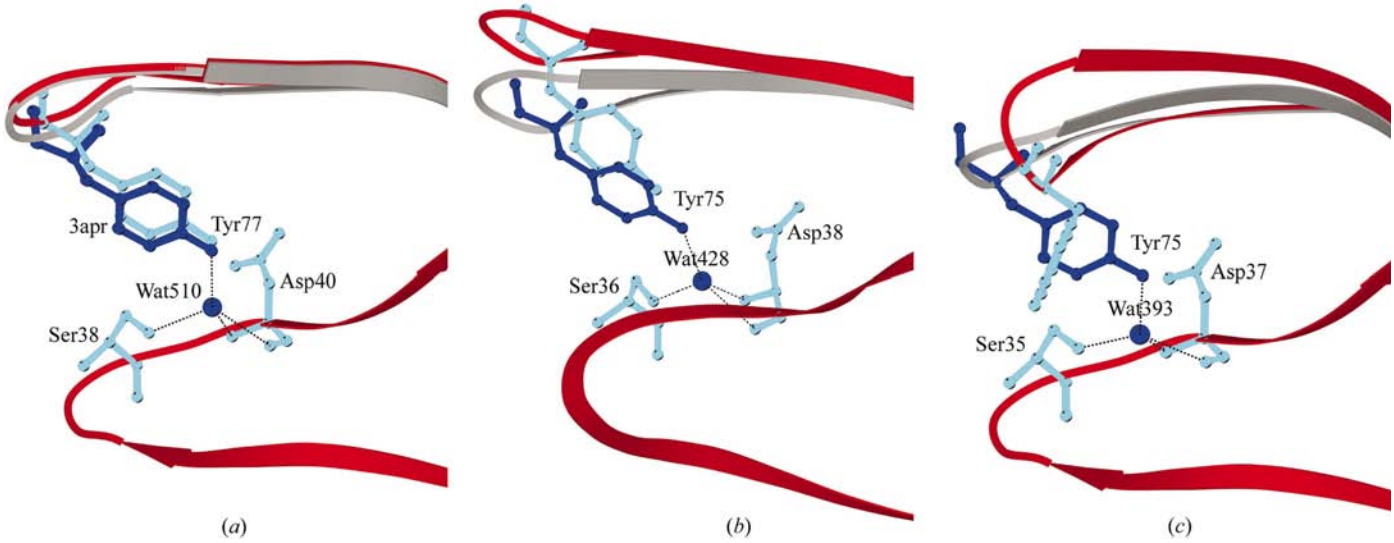

Figure 6
The residues in the native structures are shown in cyan and the tyrosine in the complexes in blue. (*a*) In rhizopuspepsin, the hydrogen bond between Tyr77 OH and Wat510 becomes stronger in the complex (3apr), (*b*) in penicillopepsin a new hydrogen bond is formed in the complex (1bxq) and (*c*) in chymosin the tyrosine changes conformation to enable the formation of the hydrogen bond in the complex (1czi).

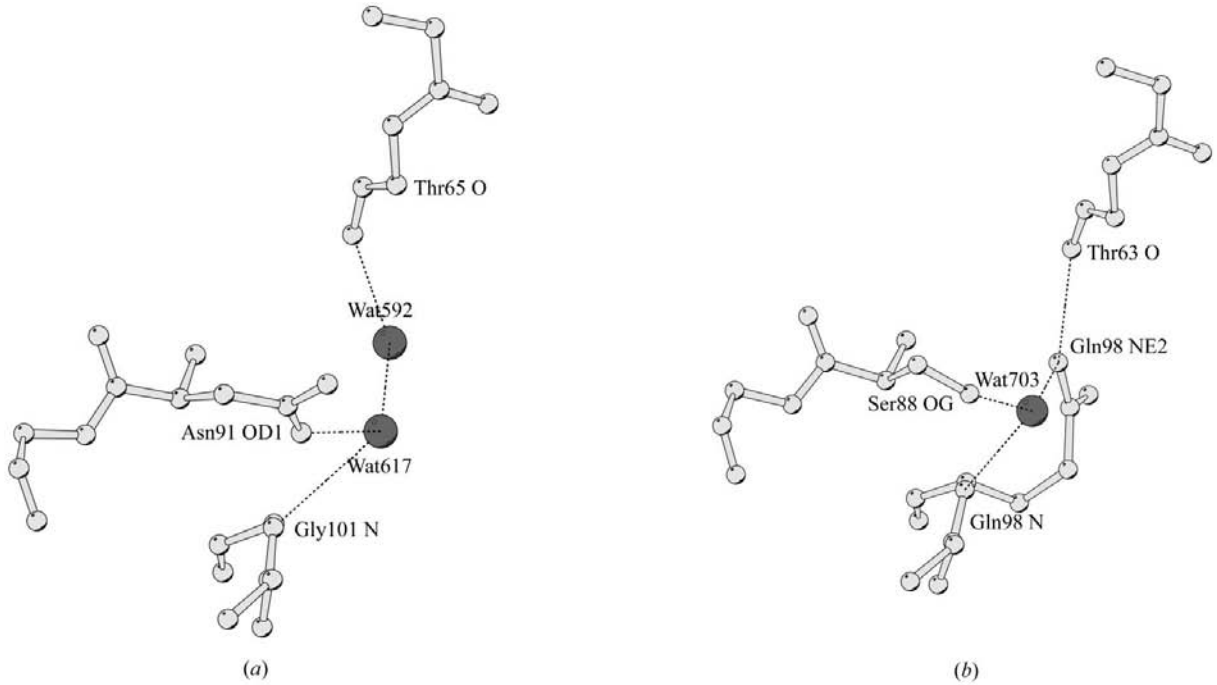

Figure 7
Wat592 and Wat617 in (*a*) rhizopuspepsin and (*b*) pusillopepsin, where a side chain occupies the position of one of the water molecules.

α-helix (Li *et al.*, 2000), the equivalent water has no direct hydrogen bond with the tyrosine, whereas the other two hydrogen bonds (Table 2) are present. Instead, the OH group of the tyrosine residue makes a hydrogen bond with Asp OD1 of the inhibitor and is connected to the equivalent of Wat510 by another water molecule.

The environment of Wat510 is mainly hydrophobic, consisting of residues Gly124, Leu41, Trp42, Ile75 and Leu122, of which Gly124 is totally conserved. Trp42 is conserved in all except in memapsin 2 (Fig. 1). However, in this case another tryptophan, Trp76, occupies this position retaining the local hydrophobic environment precisely and perfectly.

3.5. Waters near the 'flap'

The 'flap' is formed of residues between Thr65 and Asn91 which are connected by two water molecules Wat617 and Wat592 (Fig. 7*a*). This is the only water bridge found to be conserved in many aspartic proteinases. In penicillopepsin, Gln98 NE2 replaces Wat592 but the hydrogen-bonding pattern is conserved (Fig. 7*b*). Another conserved water, Wat746, is involved in hydrogen bonding with Ser74 located slightly away from the base of the 'flap' (Fig. 2 and Table 2).

3.6. Stabilization of a loop

In rhizopuspepsin, penicillopepsin and endothiapepsin, Wat628 stabilizes a loop by making three hydrogen bonds with its main-chain atoms (Fig. 8*a*). This water remains invariant in spite of the differences in loop lengths among various enzymes. In two cases, *Candida* Sap2 and cod pepsin, this water is displaced by the side chains of Ser190 and Gln191, respectively. In mucorpepsin and pusillopepsin, which have three extra residues in the loop, two water molecules are present instead of one, with a different hydrogen-bonding pattern (Fig. 8*b*). This is a good example to show how optimization of the protein structure is achieved by variation in the solvent structure. In the rest of the proteinases, there is a deletion compared with rhizopuspepsin in this loop (Fig. 1). Nevertheless, the water exists with a different hydrogen-bonding scheme (Fig. 8*c*). The base of this loop is stabilized by another conserved water, Wat650.

3.7. Other waters

In rhizopuspepsin, Wat611, a surface water makes two hydrogen bonds, one with Wat552, which is not conserved, and the other with the O atom of Gly24, which is totally conserved. The environment of this water has charged residues (Glu21, Asp24 and His26) in plasmepsin II of *P. falciparum*, whereas no charged residues are found close to this water in other proteinases. It is clear from Fig. 1 that the nature of these residues is unique to plasmepsins (Glu21, Asp24 and Gln26 in the *P. vivax* enzyme; a corresponding water is not present in the coordinate set).

Wat501 hydrogen bonds to the α-helix of the N-terminal domain. Wat554 hydrogen bonds to the carbonyl O atom of Pro316 on the C-terminus. Wat769 is located in a surface cavity. The side chain of Gln313 blocks the entry to the cavity such that the water gets buried. Since Gln is not a conserved residue, other side chains or in some instances water molecules are involved in blocking the cavity in other proteinases.

4. Internal waters

Internal waters have been identified in all ten native aspartic proteinase structures considered in the analysis. While the number of totally buried waters (accessible surface area of 0.0 Å^2) in each protein is about two to five, the number of waters with accessible surface area less than 2 Å^2 varies between 10 and 20. Six of the conserved waters, Wat510,

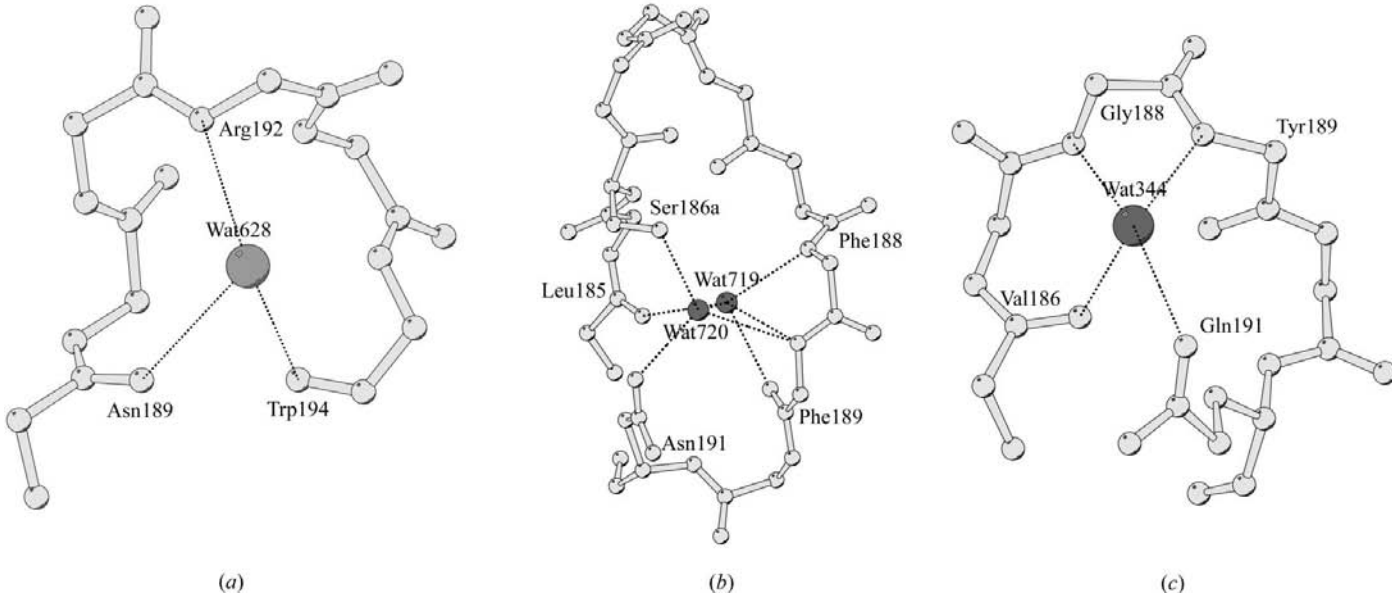

Figure 8
Stabilization of a loop by water molecules (*a*) in rhizopuspepsin, (*b*) in pusillopepsin with a longer loop and (*c*) in porcine pepsin with a shorter loop.

Wat502, Wat503, Wat508, Wat517 and Wat769, happen to be internal. Of these, Wat502 is totally buried, with an accessible surface area of 0.0 Å^2. Wat510 is buried in all aspartic proteinases. One exception to this is chymosin, as discussed earlier, where Tyr77 has a different conformation in the native structure, making this water more exposed than in the complex (Fig. 8*c*).

5. Conclusions

The importance of water molecules in aspartic proteinases has already been established by identifying the catalytic water (Suguna *et al.*, 1987) and by exploiting the position of a conserved water molecule (Wlodawer & Vondrasek, 1998) in the binding cleft of the HIV proteinase–inhibitor complexes in order to design new improved drugs for AIDS. The present extensive analysis of water molecules in the available crystal structures of aspartic proteinases has provided insights into the roles of a few other conserved waters in the structure and activity of the enzymes. Though the positions of water molecules were determined for individual proteins and characterized thoroughly in terms of hydrogen bonding, accessibility and environment, only by analyzing them together does it become possible to understand their importance. This analysis demonstrates how in aspartic proteinases a small number of conserved water molecules help in maintaining the active-site geometry required for activity, in fixing the active conformation of the flexible 'flap' region and in stabilizing the structure. This type of study is not useful only to appreciate the significance of water in protein structure, folding, stability and function, but can also aid in protein dynamics, homology modelling and drug design.

Facilities at the Distributed Information Centre and the Interactive Graphics Based Facility (both supported by the Department of Biotechnology) and the Supercomputer Education and Research Centre are acknowledged. We thank the Council of Scientific and Industrial Research for financial support.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bernstein, N. F, Cherney, M. M., Loetscher, H., Ridley, R. G. & James, M. N. G. (1999). *Nature Struct. Biol.* **6**, 32–37.
Biswal, B. K., Sukumar, N. & Vjayan, M. (2000). *Acta Cryst.* D**56**, 1110–1119.
Cohen, G. H. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Connolly, M. L. (1983). *J. Appl. Cryst.* **16**, 548–558.
Davies, D. R. (1990). *Annu. Rev. Biophys. Biophys. Chem.* **9**, 189–215.
Harte, W. E., Swaminathan, S., Mansuri, M. M., Martin, J. C., Rosenberg, I. E. & Beveridge, D. L. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 8864–8868.
Hong, L., Koetsch, G., Wu, S., Terzyan, S., Ghosh, A. K., Zhang, X. C. & Tang, J. (2000). *Science*, **290**, 150–153.
Khan, A. R., Khazanovich-Bernstein, N., Bergmann, E. M. & James, M. N. G. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 10968–10975.
Kishan, K. V. R., Chandra, N. R., Sudarsanakumar, C., Suguna, K. & Vijayan, M. (1995). *Acta Cryst.* D**51**, 703–710.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Krem, M. M. & Enrico, D. C. (1998). *Proteins Struct. Funct. Genet.* **30**, 34–42.
Li, M., Phylip, L. H., Lees, W. E., Winther, J. R., Dunn, B. M., Wlodawer, A., Kay, J. & Gustchina, A. (2000). *Nature Struct. Biol.* **7**, 113–117.
Lin, X., Loy, J. A., Sussman, F. & Tang, J. (1993). *Protein Sci.* **2**, 1383–1390.
Loris, R., Stas, P. P. G. & Wyns, L. (1994). *J. Biol. Chem.* **269**, 26722–26733.
Lowther, T., Majer, P. & Dunn, B. M. (1995). *Protein Sci.* **4**, 689–702.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Ng, K. K., Petersen, J. F., Cherney, M. M., Garen, C., Zalatoris, J. J., Rao-Naik, C., Dunn, B. M., Martnez, M. R., Peansky, R. J. & James, M. N. G. (2000). *Nature Struct. Biol.* **7**, 653–657.
Pearl, L. & Blundell, T. L. (1984). *FEBS Lett.* **174**, 96–101.
Russel, R. B. & Barton, G. J. (1992). *Proteins Struct. Funct. Genet.* **14**, 309–323.
Sreenivasan, U. & Axelsen, P. H. (1992). *Biochemistry*, **31**, 12785–12791.
Suguna, K., Bott, R. E., Padlan, E. A., Subramanian, E., Sheriff, S., Cohen, G. H. & Davies, D. R. (1987). *J. Mol. Biol.* **196**, 877–900.
Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D. & Davies, D. R. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 7009–7013.
Wlodawer, A. & Vondrasek, J. (1998). *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–84.
Zhang, N. J. & Matthews, B. W. (1994). *Protein Sci.* **3**, 1031–1039.