Journal of Computer-Aided Molecular Design, 9 (1995) 521–531
ESCOM

J-CAMD 316

Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins

C.S. Poornima and P.M. Dean*

Drug Design Group, Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, U.K.

Received 6 August 1995
Accepted 2 September 1995

Keywords: Conserved water molecules; Drug design; Homologous proteins; Protein–ligand complex; Micro-grooves

Summary

Water molecules are known to play an important rôle in mediating protein–ligand interactions. If water molecules are conserved at the ligand-binding sites of homologous proteins, such a finding may suggest the structural importance of water molecules in ligand binding. Structurally conserved water molecules change the conventional definition of 'binding sites' by changing the shape and complementarity of these sites. Such conserved water molecules can be important for site-directed ligand/drug design. Therefore, five different sets of homologous protein/protein–ligand complexes have been examined to identify the conserved water molecules at the ligand-binding sites. Our analysis reveals that there are as many as 16 conserved water molecules at the FAD binding site of glutathione reductase between the crystal structures obtained from human and *E. coli*. In the remaining four sets of high-resolution crystal structures, 2–4 water molecules have been found to be conserved at the ligand-binding sites. The majority of these conserved water molecules are either bound in deep grooves at the protein–ligand interface or completely buried in cavities between the protein and the ligand. All these water molecules, conserved between the protein/protein–ligand complexes from different species, have identical or similar apolar and polar interactions in a given set. The site residues interacting with the conserved water molecules at the ligand-binding sites have been found to be highly conserved among proteins from different species; they are more conserved compared to the other site residues interacting with the ligand. These water molecules, in general, make multiple polar contacts with protein-site residues.

Introduction

The structural significance and functional rôle of water molecules in protein structures are not yet fully understood. Several structural rôles have been suggested for water molecules conserved in homologous families of proteins [1–5]. Water molecules have been observed to be involved in the formation of β-sheets [6] and they play an important rôle in mediating protein–ligand interactions [7–9]. If some of these water molecules are structurally and functionally important in ligand binding, they are likely to be conserved at the ligand-binding sites of homologous proteins. These structurally conserved water molecules change the contour of the binding-site surface. Therefore, the presence of conserved water molecules is crucial in site-directed ligand design. Our analyses of water binding at the ligand-binding sites in high-resolution crystal structures reveal that some of these water molecules are bound in deep micro-grooves on the binding-site surface and often make multiple hydrogen bonds [9,11]. This suggests that such water molecules are probably strongly bound to the binding-site surface and may not be displaced during ligand binding; they can be structurally significant and are often conserved. In this study, we have analysed the water molecules at the ligand-binding sites of structurally homologous proteins to see if some of these are conserved.

Several crystallographically ordered water molecules are found at the ligand-binding sites of many protein–ligand complexes. In different complexes of the same protein crystallized in the same crystal form [8,12], water molecules are often observed to occupy equivalent positions at the ligand-binding sites. When a given protein is crystallized in different crystal forms [13], the majority of

*To whom correspondence should be addressed.

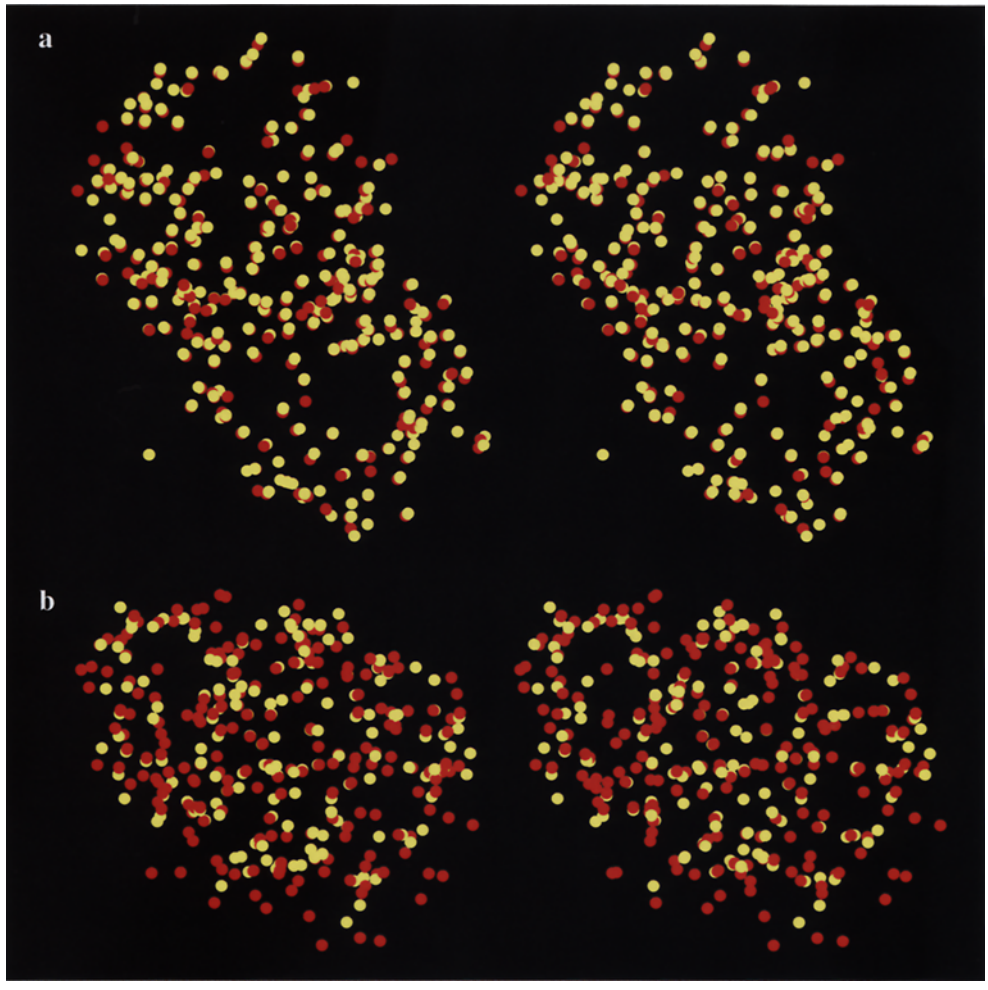

Fig. 1. (a) Overall solvent structure of the ABP–Ara complex superposed on the overall solvent structure of the ABP–Gal complex. Water molecules in ABP–Ara and in ABP–Gal are shown in red and yellow, respectively. Both complexes have been obtained from *E. coli* and crystallized in the same crystal form. The rms coordinate difference for main-chain atoms between the two structures is 0.19 Å. (b) Overall solvent structure of the GBP_{eco}–Gal complex obtained from *E. coli*, superposed on the overall solvent structure of the GBP_{styp}–Gal complex obtained from *S. typhimurium*. Water molecules in GBP_{eco}–Gal and in GBP_{styp}–Gal are shown in red and yellow, respectively. The rms coordinate difference for main-chain atoms between the two structures is 0.79 Å.

the hydration sites have been observed to be frequently occupied. Therefore, if a protein is crystallized with different ligands in the same crystal form, it is likely that the water molecules occupy equivalent positions at the ligand-binding sites. Furthermore, the number of conserved water molecules between any two homologous structures of a family of proteins have been observed to decrease with the decrease in sequence identity [6]. Therefore, water molecules conserved at the binding sites are probably structurally significant and likely to have some functional rôle in ligand recognition and binding. Such evolutionarily conserved water molecules at the ligand-binding sites further emphasize the importance of incorporation of water molecules in site-directed ligand design strategies.

In this study, we have made an attempt to identify the structurally conserved water molecules at the ligand-binding sites of five sets of homologous proteins from different species. These structurally conserved water molecules have been examined to see if the nature of binding is similar to that of the water molecules considered to be 'strongly bound' in our previous studies [9,11].

Methods

Data set

Five different sets of homologous proteins from different species were chosen to identify the water molecules conserved at the ligand-binding sites during evolution. Atomic coordinates of all these protein–ligand complexes were obtained from the crystallographic data in the Brookhaven Protein Data Bank. All these crystal structures have been refined to a high resolution of 2.0 Å or better and are listed in Table 1, along with relevant refinement details.

If the ligand-binding site is located at the junction of two or more monomers, we have taken care to include all

the residues from different symmetry-related monomers that form the binding site.

Alignment, superpositions and identification of conserved water molecules

In this study, all protein residues interacting with the ligand, directly or indirectly through mediating water molecules, are referred to as 'site residues'. Amino acid sequences at the binding site region are aligned manually. For any two given structures, all the water molecules within a distance of 7.0 Å from the ligand atoms were extracted and superposed, using the transformation matrix obtained by least-squares superposition for main-chain atoms of the site residues. Water molecules are considered to be conserved if the distance between a given pair of water molecules from two different structures is less than 1 Å after least-squares superposition, and if they exhibit identical or equivalent polar and nonpolar interactions. All those protein atoms whose surface points are within a distance of 2.0 Å from a given water molecule, hence contributing to the water-bound solvent-accessible protein surface (generally groove surfaces), are compared between the two structures. This method of comparing the molecular surfaces around a conserved hydration site (in addition to the comparison of the hydrogen-bonding schemes) was useful in identifying those conserved water molecules that lacked polar interactions within the well-defined hydrogen-bonding distance of 3.5 Å, or did not have any polar interactions at all.

Molecular surfaces and solvent accessibility

Solvent-accessible surfaces for the site residues were generated using Connolly's MS program [23] employing a probe sphere with a radius of 2.0 Å. This choice of size of the probe sphere was found to be optimal and is in general agreement with the size of the probe used in other surface studies of proteins [4,24]. Water molecules bound in grooves and cavities at the binding-site surface have been assessed as such, by following the same method as described in the second paper of this series [10]. Fractal atomic densities (surfractal index) of surface points were calculated using the program *Surfractal* from the *Shape* suite of programs [26]. The classification scheme was based on the value of the average surfractal density of all the surface points within a distance of 2.0 Å from a given water molecule of interest, and on visual inspection of molecular graphics using the program Midas [10].

Results

Comparison of conserved hydration sites between structures from the same and from different species

Two water molecules at the sugar-binding site of arabinose binding protein (ABP) have been observed to

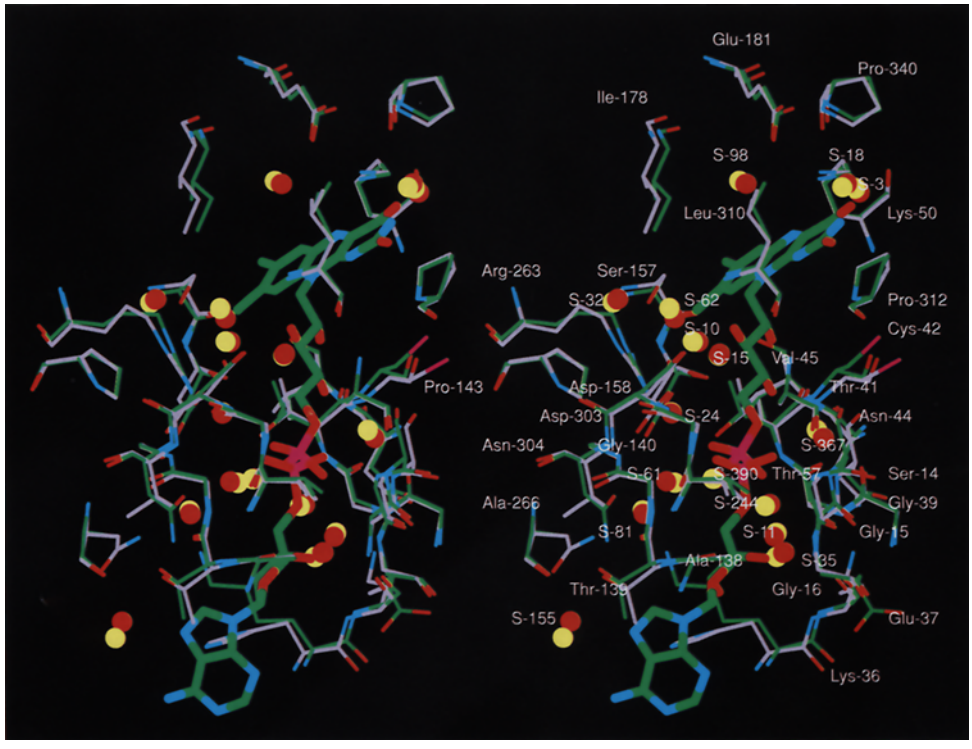

Fig. 2. Water molecules with the interacting site residues at the FAD binding site of glutathione reductase, conserved between the structures GR_{eco} and GR_{hum}. Water molecules in GR_{eco} and in GR_{hum} are shown as red and yellow spheres, respectively. The ligand is shown in bold. Carbon atoms of the site residues in GR_{eco} and those of the ligand are shown in green. Carbon atoms of the site residues in GR_{hum} are shown in grey. Colour code: red, oxygens; cyan, nitrogens.

TABLE 1
LIST OF HOMOLOGOUS PROTEINS FROM DIFFERENT SPECIES COMPLEXED WITH THE SAME LIGANDS

Protein	Ligand	PDB code	Source	Resolution (Å)	Refinement program used	No. of water molecules[a]	Ref.
Glutathione reductase	FAD	1ger	*Escherichia coli*	1.86	X-PLOR	16	14
	FAD + NADH	1grb	*Homo sapiens*	1.85	TNT		15
D-Galactose/glucose-binding protein	D-Galactose	2gbp	*Escherichia coli*	1.9	PROLSQ	4	8
	D-Galactose	1gca	*Salmonella typhimurium*	1.7	X-PLOR		16
Citrate synthase	Coenzyme A	2csc	Chicken heart muscle (*Gallus gallus*)	1.7	TNT	4	17
	Coenzyme A	2cts	Pig heart (*Sus scrofa*)	2.0	DERIV/EREF		18
Lactate dehydrogenase	NADPH	9ldt	Porcine muscle (*Sus scrofa*)	2.0	PROLSQ	2	19
	No ligand	6ldh	Dogfish (*Squalis acanthias*)	2.0			20
Glutathione S-transferase	S-glutathionyl-10-hydroxy-9,10-dihydrophenanthrene	2gst	Rat liver (*Rattus rattus*)	1.8	X-PLOR	3	21
	No ligand	1hna	Human muscle (*Homo sapiens*)	1.85	X-PLOR		22

[a] Number of water molecules conserved at the binding site (results from the present study).

occupy the same position in all three complexes of ABP with arabinose, galactose and fucose [8]. All these complexes of ABP have been obtained from the same species (*E. coli*) and crystallized in the same crystal form. A recent study on complexes of adiposite lipid-binding protein also provided a similar observation. A network of 10 water molecules has been observed to be conserved at the fatty-acid binding site of these complexes [12]. All these complexes are crystallized in the same crystal form and the proteins have been obtained from the same species, i.e., mouse. These observations may raise the following questions. Do the water molecules occupy the same positions in different structures when a protein is obtained from the same species and crystallized in the same crystal form? To what extent are the hydration sites conserved between any two homologous protein structures from different species? In order to answer these questions we examined the hydration sites in the ABP complexes obtained from the same species (*E. coli*), and in the homologous structures of galactose complexed to glucose-binding protein (GBP) obtained from different species.

The two structures, arabinose-binding protein complexed with arabinose (ABP–Ara complex) and with galactose (ABP–Gal complex), have a 100% sequence identity. These structures show an rms coordinate difference of 0.19 Å for main-chain atoms after least-squares superposition. Figure 1a shows the superposition of the overall solvent structure of the ABP–Ara complex with the solvent structure of the ABP–Gal complex. Figure 1a is highly compact, and the majority of the solvent sites are well superposed when considering their coordinate errors. After the least-squares superposition, 196 of the 236 solvent sites determined in the ABP–Gal complex lie within a distance of 1.0 Å from the corresponding solvent sites in the ABP–Ara complex. This suggests that the majority of the hydration sites are conserved between the ABP–Ara and ABP–Gal complexes. Both complexes have been obtained from the same species and crystallized in the same crystal form. In Fig. 1b, the overall solvent structure of GBP complexed with glucose (GBP–Glu complex) obtained from *E. coli* [16] is superposed on the overall solvent structure of GBP complexed with galactose (GBP–Gal complex) obtained from *S. typhimurium* [16]. The two complexes of GBP have 94% sequence identity and show an rms coordinate difference of 0.79 Å after least-squares superposition. In this case, the superposed solvent structures are more scattered and only 79 of the 214 solvent sites determined in the GBP–Glu complex lie within a 1.0 Å sphere of equivalent solvent sites in the GBP–Gal complex. Although the occupancy of a hydration site depends on several factors [13], our analysis suggests a considerable decrease in the number of conserved hydration sites between the crystal structures of a protein obtained from different species. This observation is in qualitative agreement with results of Loris et al. [5] in their study of conserved water molecules in legume lectin crystal structures.

Our analysis of solvent structures of ABP and GBP complexes suggests the following. Firstly, the majority of the water molecules are likely to occupy the same positions in the two structures of a protein crystallized in the same crystal form. Therefore, it is reasonable to expect the water molecules in different complexes of a protein obtained from the same species to occupy the same positions at the ligand-binding sites. Secondly, the conserved hydration sites between the two structures of a protein obtained from different species can be significantly less. Furthermore, studies of Loris et al. [5] reveal that the number of conserved hydration sites decreases significantly with the decrease in sequence identity. These observations suggest that solvent sites that are conserved during evolution are probably structurally important. Therefore,

if the water molecules are conserved at the ligand-binding sites between the homologous protein structures from different species, such water molecules may be structurally more significant for ligand binding. This analysis forms the basis for the present study, aiming to identify the structurally conserved water molecules at the ligand-binding sites of homologous proteins during evolution.

Conserved water molecules at the ligand-binding sites

Five different sets of functionally similar homologous proteins have been chosen for this study. The number of conserved water molecules at the ligand-binding sites in each set of proteins are listed in Table 1.

Several studies [6,13,27,28] suggest that the majority of hydration sites are conserved between different structures of a protein from the same species. Our analysis, comparing the hydration sites of ABP and GBP, supports such observations. Especially the water molecules bound in grooves and those buried in cavities [27,28,30] are more likely to be conserved. High-resolution crystallographic data are available for many different protein–ligand complexes with different ligands or mutant proteins with the same ligand. We present the data for conserved water molecules at the ligand-binding sites for only one protein–ligand complex from two different species. Nevertheless, we have confirmed the presence of the conserved water molecules listed in Table 2 in all available high-resolution ($\leq$2.0 Å) crystal structures in all five sets of protein–ligand complexes.

The rms differences for the main-chain atoms of binding-site residues in each set varied between 0.72 and 0.17 Å. Table 3 lists details of rms differences observed for the main-chain atoms of site residues and also for the residues interacting with conserved water molecules in each set. The conserved water molecules and their interactions with polar atoms of the site residues in each set of protein–ligand complexes are listed in Table 2. All these conserved water molecules have an identical polar and nonpolar environment in any two structures compared. Protein residues at the binding sites are highly conserved among different species in each set. It is interesting to note that the protein residues interacting with the conserved water molecules at the binding site are 100% conserved in all the complexes chosen for this study, except in glutathione reductase. In this complex only ~90% of the residues interacting directly with the conserved water molecules are conserved. Site residues interacting with the conserved water molecules are more conserved compared with all the binding-site residues interacting with the ligand. The rms values given in Table 3 reflect the structural importance of the conserved water molecules at the binding sites.

For convenience, the hydration sites conserved between the different structures are denoted by S-n; they are listed in Table 2. For example, hydration site S-3 in glutathione reductase corresponds to the hydration site Wat^{3} in GR_{eco} and to Wat^{501} in GR_{hum}.

Conserved water molecules at the FAD binding site of glutathione reductase

Glutathione reductase from *E. coli* (GR_{eco}) has a 52% sequence identity with glutathione reductase from human erythrocytes (GR_{hum}). However, the FAD binding site of the enzyme has been very well conserved between the structures. We have compared the solvent environment only at the FAD binding site of the enzyme, in order to identify the structurally conserved water molecules involved in FAD binding.

Between GR_{eco} and GR_{hum}, 16 water molecules are conserved at the FAD binding site of the enzyme. Of the 40 residues interacting with FAD molecules, 33 (~82%) have been conserved; the site residues interact with the ligand directly and indirectly through mediating water molecules. Main-chain atoms of the site residues show an rms deviation of 0.72 Å after the least-squares superposition. It is interesting to see that 29 (~90%) of the 32 residues involved in solvent-mediated interaction with the FAD molecule have been conserved between the two structures; the sequence identity is only 52%. All 16 pairs of conserved water molecules, together with residues involved in solvent binding, are shown in Fig. 2. The polar interactions of these water molecules in corresponding structures of GR_{eco} and GR_{hum} are listed in Table 2. Occasionally, substitutions have occurred among the residues interacting with these conserved water molecules. Of the 16 water molecules, nine make multiple polar contacts with surrounding residues within a distance of 3.5 Å. Note that the hydration sites, S-3 and S-244, are conserved without any polar contacts within 3.5 Å. However, Wat^{244} has two polar atoms within a distance of 4.0 Å. All these water molecules are bound in grooves or buried completely between the ligand and the protein (Table 2).

Conserved water molecules at the sugar-binding site of galactose-binding proteins

Galactose-binding protein from *E. coli* (GBP_{eco}) is evolutionarily not very distant from the homologous galactose/glucose-binding protein from *S. typhimurium* (GBP_{styp}). Both galactose/glucose-binding proteins have 309 residues, with 94% sequence identity. There are four water molecules at the sugar-binding site of the GBP_{eco}–Gal and GBP_{styp}–Gal complexes, and all these four water molecules are present in all high-resolution crystal forms of GBP so far reported. These four conserved water molecules, together with the binding-site residues, are shown in Fig. 3a. Site residues interacting with the sugar molecules are totally conserved in both structures, with an rms deviation of 0.15 Å. The water molecule at the conserved hydration site S-313 mediates the interaction of the hydroxyl O3 of galactose with the main-chain nitrogen of

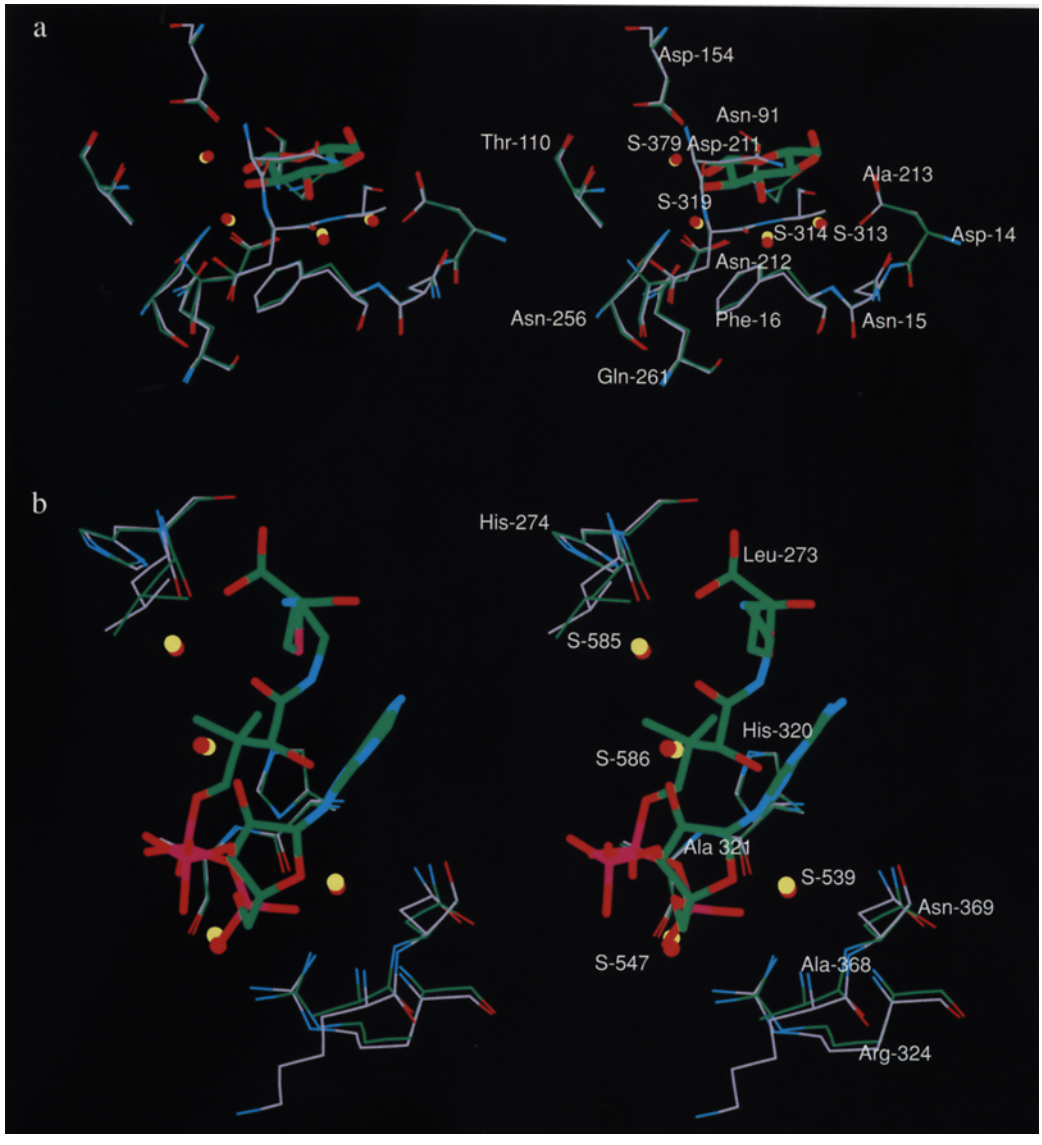

Fig. 3. Water molecules with the interacting site residues conserved between different structures from two different species. Conserved water molecules are shown as red and yellow spheres. (a) Conserved water molecules with site residues at the sugar-binding site of galactose/glucose-binding protein (GBP). (b) Conserved water molecules with site residues at the NADH binding site of citrate synthase.

Phe[16]. This has been observed before in crystal structure studies of these complexes. The present analysis shows that, in addition to S-313, there are three more water molecules conserved at the sites S-314, S-319 and S-379. They make weak polar contacts (3.5 Å < polar contact distance < 4.0 Å) with O1 and O3 of glucose/galactose and multiple and strong polar contacts with the site residues. The details of these interactions are listed in Table 2. All these four water molecules at the sugar-binding site of GBP are bound in deep grooves between the protein and the ligand.

Conserved water molecules at the coenzyme A binding site of citrate synthase

Four hydration sites are conserved between the structures of citrate synthase from chicken (CS_{chi}) and from pig (CS_{pig}) at the coenzyme A binding site. The site residues are highly conserved. The rms difference for least-squares superposition of the main-chain atoms of the site residues is 0.26 Å. Water molecules at the sites S-539 and S-547 are within well-defined hydrogen-bonding distance (< 3.0 Å) from the ligand atoms AO2 and AO4, respectively. S-547 makes another strong polar contact with the oxygen AO1 of coenzyme A. The water molecule at S-539 is bound in a deep groove and the water molecules at S-547 and S-585 are bound in shallow grooves. The water molecule at S-586 is also bound in a shallow groove and does not have any strong polar contact with the ligand atoms. However, all four hydration sites are conserved in all available high-resolution crystal structures of citrate synthase complexes and make multiple polar contacts with site residues.

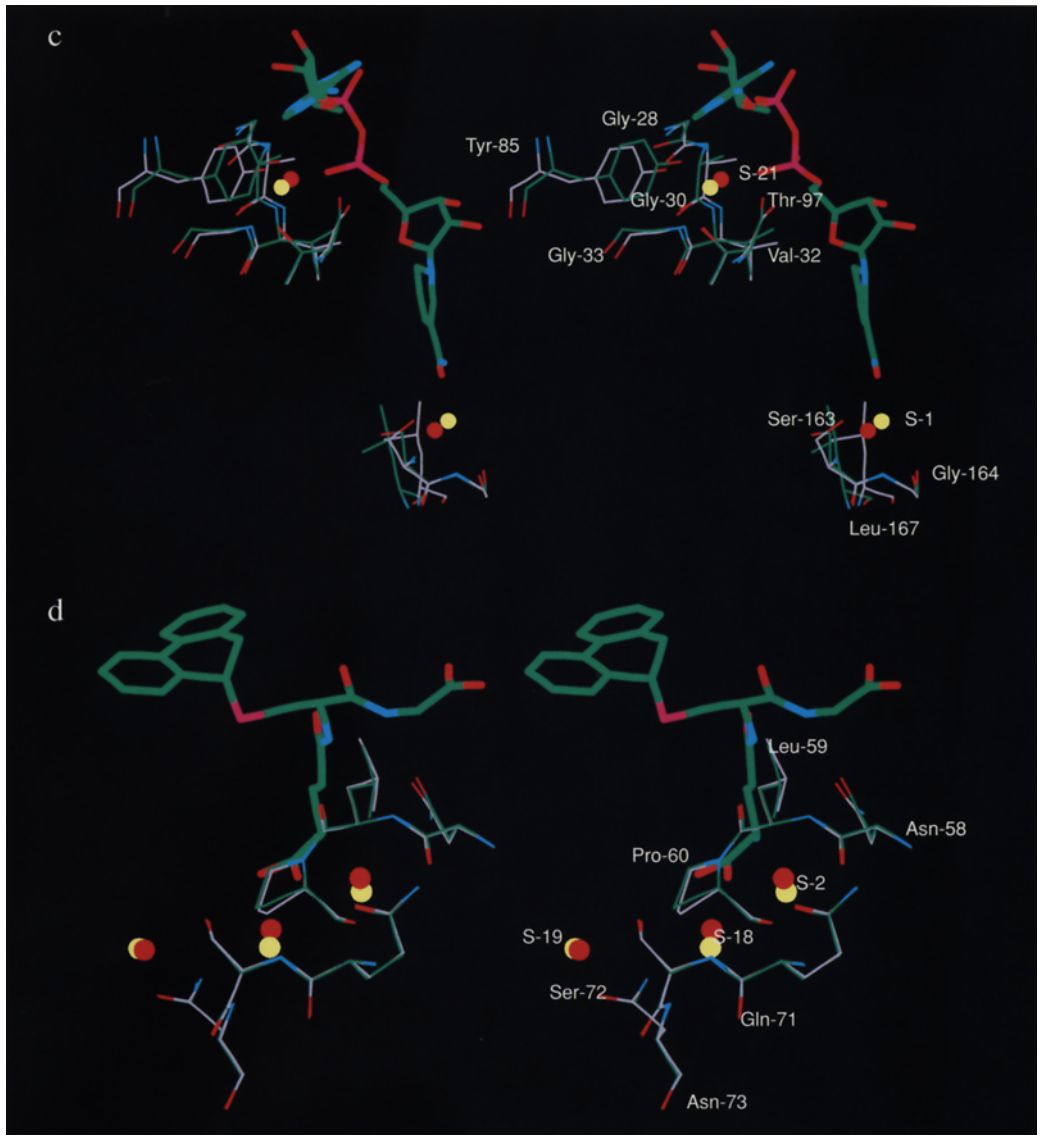

Fig. 3 (continued). (c) Conserved water molecules with site residues at the NADH binding site of lactate dehydrogenase. (d) Conserved water molecules with site residues at the binding site of glutathione S-transferase (GST). Carbon atoms of the site residues are shown in green and grey. The colour code and other features are as described in the legend of Fig. 2.

Conserved water molecules at the NADH binding site of lactate dehydrogenase

Lactate dehydrogenase from pig (LD_{pig}) is homologous with lactate dehydrogenase from dogfish (LD_{df}), with a sequence identity of 76%. The NADH binding site of lactate dehydrogenase is well conserved between these high-resolution structures, with an rms difference of 0.7 Å for the binding-site residues. Two water molecules have been conserved at the NADH binding site (Fig. 3b). These water molecules, at the hydration sites S-1 and S-21, make strong polar contacts with the oxygen atoms NO7 of the nicotinamide moiety and NO2 of the pyrophosphate (distance < 3.0 Å). The water molecules at sites S-1 and S-21 are buried between the protein residues and the NADH molecule in LD_{pig}; water molecules Wat^{474} and Wat^{333} at the equivalent sites in LD_{df} are bound in deep grooves and they have identical multiple polar contacts with the binding-site residues in the absence of the NADH ligand.

Conserved water molecules at the binding site of glutathione S-transferase

Three water molecules are conserved at the binding site of glutathione S-transferase between rat (GST_{rat}) and human (GST_{hum}). Of the 23 binding-site residues, 19 have been conserved. Site residues superpose with an rms difference of 0.96 Å for main-chain atoms. Water molecules at S-2 and S-18 mediate the interaction of site residues with the ligand oxygen atoms O1 and OT1, respectively. These water molecules make strong polar contacts, both with ligand atoms and site residues, where the hydrogen-bonding distance is often less than 3.0 Å. The water mol-

TABLE 2
EQUIVALENT CONSERVED WATER MOLECULES WITH THEIR POLAR INTERACTIONS AT THE LIGAND-BINDING SITES OF VARIOUS PROTEINS FROM DIFFERENT SPECIES

Water site	Conserved water molecule	Protein atoms/ residues	Conserved water molecule	Protein atoms/ residues	Distance[a] (Å)	Nature of water binding	No. of polar contacts[b]
Glutathione reductase 1ger			**1grb**				
S-3	Wat^{3}		Wat^{501}		0.28	Shallow groove water	0 (0)
S-10	Wat^{10}	O Val^{45} O Ser^{157} OG Ser^{157} N Asp^{158} OD1 Asp^{158}	Wat^{495}	O Val^{61} O Ser^{177} OG Ser^{177} N Asp^{178} OD1 Asp^{178}	0.10	Deep groove water	4 (5)
S-11	Wat^{11}	O Gly^{11} N Gly^{13} O Gly^{13} N Ser^{14} N Gly^{15} N Gly^{16} O Ala^{138}	Wat^{490}	O Gly^{27} N Gly^{29} O Gly^{29} N Ser^{30} N Gly^{31} N Gly^{32} O Ala^{155}	0.22	Buried water	4 (7)
S-15	Wat^{15}	N Asp^{303} OD1 Asp^{303} O Val^{308} N Leu^{310} O Leu^{310} N Ala^{314}	Wat^{491}	N Asp^{331} OD1 Asp^{331} O Ala^{336} N Leu^{338} O Leu^{338} N Ala^{342}	0.17	Buried water	3 (6)
S-18	Wat^{18}	O Pro^{340}	Wat^{489}	O Pro^{368}	0.25	Deep groove water	1 (1)
S-24	Wat^{24}	OG1 Thr^{41} N Gly^{141} O Gly^{141} NH2 Arg^{263} OD2 Asp^{303}	Wat^{496}	OG1 Thr^{57} N Gly^{158} O Gly^{158} NH2 Arg^{291} OD2 Asp^{331}	0.10	Buried water	3 (5)
S-32	Wat^{32}	OG Ser^{157} NH1 Arg^{263}	Wat^{494}	OG Ser^{177} NH1 Arg^{291}	0.24	Deep groove water	2 (2)
S-35	Wat^{35}	OE2 Glu^{34} N Glu^{37} O Glu^{37}	Wat^{538}	OE2 Glu^{50} N Lys^{53} O Lys^{53}	0.39	Deep groove water	2 (3)
S-61	Wat^{61}	N Thr^{139} N Gly^{140} O Val^{301} N Asp^{303} N Asn^{304}	Wat^{488}	N Thr^{156} N Gly^{157} O Val^{329} N Asp^{331} N Val^{332}	0.44	Buried water	4 (5)
S-62	Wat^{62}	NE Arg^{263} OD2 Asp^{303}	Wat^{500}	NE Arg^{291} OD2 Asp^{331}	0.44	Buried water	1 (2)
S-81	Wat^{81}	O Thr^{139} N Gly^{141} O Gly^{141}	Wat^{635}	O Thr^{156} N Gly^{158} O Gly^{158}	0.15	Shallow groove water	3 (3)
S-98	Wat^{98}	OE1 Glu^{181}	Wat^{569}	OE1 Glu^{201}	0.21	Buried water	1 (1)
S-155	Wat^{155}	O Ala^{266}	Wat^{796}	O Asn^{294}	0.83	Deep groove water (ligand groove)	1 (1)
S-244	Wat^{244}	O Gly^{40} ND2 Asn^{44}	Wat^{588}	O Gly^{56} ND2 Asn^{60}	0.13	Deep groove water	0 (2)
S-367	Wat^{367}	N Ser^{14} OG Ser^{14} O Gly^{39} N Thr^{41} N Cys^{42}	Wat^{982}	N Ser^{30} OG Ser^{30} O Gly^{55} N Thr^{57} N Cys^{58}	0.39	Buried water	4 (5)
S-390	Wat^{390}	N Thr^{41} OG1 Thr^{41}	Wat^{688}	N Thr^{57} OG1 Thr^{57}	0.30		1 (2)

[a] Distance between the conserved water molecules when superposed.
[b] Number of polar contacts conserved within 3.5 Å from the water molecule in both structures. The number inside parentheses indicates the number of polar contacts conserved within 4.0 Å from the water molecule.

TABLE 2 (continued)

Water site	Conserved water molecule	Protein atoms/ residues	Conserved water molecule	Protein atoms/ residues	Distance[a] (Å)	Nature of water binding	No. of polar contacts[b]
Galactose/glucose-binding protein 2gbp			**1gca**				
S-313	Wat^{313}	OD1 Asp^{14} OD2 Asp^{14} N Asn^{15} N Phe^{16} ND2 Asn^{211}	Wat^{5}	OD1 Asp^{14} OD2 Asp^{14} N Asn^{15} N Phe^{16} ND2 Asn^{211}	0.13	Deep groove water	3 (5)
S-314	Wat^{314}	ND2 Asn^{211} N Asp^{212} N Ala^{213} OD1 Asp^{236}	Wat^{17}	ND2 Asn^{211} N Asp^{212} N Ala^{213} OD2 Asp^{236}	0.24	Deep groove water	3 (4)
S-319	Wat^{319}	ND2 Asn^{91} OD1 Asn^{256} ND2 Asn^{256} OE1 Gln^{261} NE2 Gln^{261}	Wat^{9}	ND2 Asn^{91} OD1 Asn^{256} ND2 Asn^{256} OE1 Gln^{261} NE2 Gln^{261}	0.12	Deep groove water	2 (5)
S-379	Wat^{379}	O Asn^{91} ND2 Asn^{91} OG1 Thr^{110} OD1 Asp^{154} OD2 Asp^{154}	Wat^{11}	O Asn^{91} ND2 Asn^{91} OG1 Thr^{110} OD1 Asp^{154} OD2 Asp^{154}	0.15	Deep groove water	3 (5)
Citrate synthase 2csc			**2cts**				
S-539	Wat^{539}	N His^{320} O His^{320} N Ala^{368} N Asn^{369} O Asn^{369}	Wat^{41}	O His^{320} O His^{320} N Lys^{368} N Asn^{369} O Asn^{369}	0.21	Deep groove water	2 (5)
S-547	Wat^{547}	O His^{320} O Ala^{321} NH1 Arg^{324} NH2 Arg^{324}	Wat^{49}	O His^{320} O Ala^{321} NH1 Arg^{324} NH2 Arg^{324}	0.34	Shallow groove water	2 (4)
S-585	Wat^{585}	O Leu^{273} ND1 His^{274}	Wat^{87}	O Leu^{273} ND1 His^{274}	0.24	Shallow groove water	1 (2)
S-586	Wat^{586}	NE2 His^{320} N Ala^{321}	Wat^{88}	NE2 His^{320} N Ala^{321}	0.43	Shallow groove water	1 (2)
Lactate dehydrogenase 9ldt			**6ldh**				
S-1	Wat^{1}	O Ser^{163} OG Ser^{163} O Gly^{164} N Leu^{167}	Wat^{474}	O Ser^{161} OG Ser^{161} O Gly^{162} N Leu^{165}	0.74	Buried water	2 (3)
S-21	Wat^{21}	O Gly^{28} N Gly^{30} N Val^{32} N Gly^{33} O Thr^{97} OG1 Thr^{97}	Wat^{333}	O Gly^{27} N Gly^{29} N Val^{31} N Gly^{32} O Thr^{95} OG1 Thr^{95}	0.72	Buried water	4 (5)
Glutathione S-transferase 2gst			**1hna**				
S-2	Wat^{2}	ND2 Asn^{58} N Leu^{59} O Leu^{59} N Pro^{60} O Pro^{60} N Gln^{71}	Wat^{291}	ND2 Asn^{58} N Leu^{59} O Leu^{59} N Pro^{60} O Pro^{60} N Gln^{71}	0.53	Deep groove water	2 (6)
S-18	Wat^{18}	OE1 Gln^{71} N Ser^{72} OG Ser^{72} N Asn^{73} ND2 Asn^{73}	Wat^{260}	OE1 Gln^{71} N Ser^{72} OG Ser^{72} N Asn^{73} ND2 Asn^{73}	0.61	Shallow groove water	3 (5)
S-19	Wat^{19}	OG Ser^{72} N Asn^{73} OD1 Asn^{73}	Wat^{287}	OG Ser^{72} N Asn^{73} OD1 Asn^{73}	0.28	Deep groove water	2 (3)

TABLE 3
BACKBONE RMS DEVIATIONS AFTER LEAST-SQUARES SUPERPOSITION OF CORRESPONDING PROTEIN RESIDUES AT THE BINDING SITES OF DIFFERENT PROTEINS IN EACH SET

Protein–ligand complex	PDB code	No. of equivalent residues[a]	No. of residues conserved[b]	Rms deviation of main-chain atoms[c] (Å)	No. of equivalent residues[d]	No. of residues conserved[e]	Rms deviation of main-chain atoms[f] (Å)
Glutathione reductase	1ger						
	1grb	32	29	0.31	40	33	0.72
D-Galactose/glucose-binding protein	2gbp						
	1gca	13	13	0.13	18	18	0.15
Citrate synthase	2csc						
	2cts	9	9	0.23	22	21	0.26
Lactate dehydrogenase	9ldt						
	6ldh	12	12	0.47	27	23	0.70
Glutathione S-transferase	2gst						
	1hna	6	6	0.07	23	19	0.96

[a] Number of residues interacting with water molecules at the binding site in both structures.
[b] Number of conserved residues interacting with water molecules at the binding site.
[c] Rms deviation of main-chain atoms of those residues interacting with water molecules.
[d] Number of residues interacting with the ligand directly and through mediating water molecules.
[e] Number of conserved residues interacting with the ligand directly and through mediating water molecules.
[f] Rms deviation of main-chain atoms of all residues interacting with the ligand directly and through mediating water molecules.

ecule at site S-2 is bound in a deep groove, whereas that at S-18 is bound in a shallow groove. The water molecule at S-19 is directed slightly away from the ligand, but conserved in all high-resolution crystal forms of glutathione transferase. All these water molecules are conserved in (GST_{hum}) in the absence of the ligand. The details of the interactions are listed in Table 3.

Discussion and Conclusions

Ligand-binding sites in five sets of protein–ligand complexes have been examined to identify evolutionarily conserved water molecules between different species. This study reveals the presence of conserved water molecules at the ligand-binding sites of homologous proteins and shows that the majority of these conserved water molecules share many common features. For all the structures studied here, these water molecules generally make multiple polar contacts with the site residues. Their polar interactions with site residues and ligand atoms are conserved among all the protein–ligand complexes in a given set. In addition, they are either buried in grooves at the protein-site surface or buried in cavities between the protein and the ligand. This shows that our predefined criteria for stronger water binding at the ligand-binding sites, based on the analyses of high-resolution crystal structures [9,10], are satisfied for most of the water molecules conserved at the ligand-binding sites. Most of the buried water molecules in internal cavities and grooves have been observed to make multiple polar contacts with the cavity surfaces [5,24,29], especially the conserved water molecules in homologous structures [5].

The water molecules at the ligand-binding sites in LD_{df} and GST_{hum} are conserved even in the absence of the ligand. Moreover, the site residues interacting with the conserved water molecules are 100% conserved among all the homologous structures in each set of proteins studied here. The site residues interacting with ligand are less conserved compared to those interacting with conserved water molecules at the site. The groove surfaces of proteins present the same (or similar) polar and nonpolar environments for a given pair of conserved water molecules in the respective structures. These observations may suggest the structural significance of these conserved water molecules. In summary, our present analyses of water binding at the ligand-binding sites in high-resolution crystal structures of homologous proteins suggest that water molecules bound in deep grooves or microgrooves, making multiple polar contacts, are likely to be conserved. If the water molecules are strongly bound to the site and unlikely to be displaced during ligand binding, they modify the shape and complementarity of the binding sites. Thus, these conserved water molecules at the binding sites may change the conventional definition of ligand-binding sites.

Several structural rôles have been suggested for conserved water molecules in proteins. Structurally conserved water molecules are found at buried sites in homologous proteins and they have been suggested to play an important rôle in protein folding [2,3]. Buried water molecules in a structurally homologous family of serine proteases were found to be conserved [4]. Substitution of a histidine side chain for a conserved water molecule in serine proteases is suggested to be indicative of the structural importance of that conserved water molecule. A significant number of hydration sites are conserved in the family of legume lectins [5] and most of these conserved water molecules have low or zero solvent accessibility. One of

the conserved water molecules in legume lectins is thought to stabilize a β-hairpin. Six long-lived water molecules have been observed [30] in a recent study on the binary and ternary complexes of human dihydrofolate reductase with methotrexate (MTX) and NADPH. Two of these water molecules bridge the secondary structural elements in the crystal structures. Another two mediate the enzyme and MTX interactions, while the remaining two bridge the enzyme with NADPH. It is interesting that all these water molecules in general make multiple hydrogen bonds and five of them are buried. Some of these are conserved between vertebrate and bacterial species [30].

The present study, together with the previous studies on hydrogen-bonding interactions [9] and site-surface shape [10], suggests that some of the conserved water molecules may be structurally as important as any other protein-site atom and may be treated as such. Therefore it may be unwise to ignore the rôle of water molecules in designing a novel drug molecule. Inclusion of water molecules may affect the design process in different ways. Firstly, they can change the shape of the site and consequently affect the generation of possible molecular graphs to fit the site. Secondly, their presence may change the basic properties of the site used in drug design, such as electrostatics and hydrophobicity. Thus, water may affect the complementarity of feasible ligand structures. In effect, the inclusion of water molecules in ligand design may significantly modify the essential design parameters for an evolving novel ligand. At this stage, we can only speculate on how water molecules might be incorporated in a design program. Firstly, those water molecules making two or more hydrogen bonds with the site residues of an empty site can be considered to be fixed. Such water molecules can be treated as any other protein-site atom in structure generation. Secondly, water might be added after a structure generation run to examine where it can bind; then a 'water score' can be incorporated into the algorithm. Neither of these approaches is in any sense dynamic, since water is considered only at the start or end of structure generation. A more interesting approach would be to consider water bridging as an integral part of structure generation. Every time a terminus node is generated on the ligand, it should be marked as a potential hydrogen-bond acceptor or donor to be considered for water bridging. The local region of the site can then be searched for appropriate hydrogen bonds to a water molecule located in that vicinity. The water molecule then behaves rather like a special fragment, which can be added or discarded while generation proceeds.

Acknowledgements

C.S.P. thanks Dr. Cyrus Chothia for interesting discussions and Rhône-Poulenc Rorer for a postdoctoral research fellowship. P.M.D. thanks the Wellcome Trust for the award of a Principal Research Fellowship. Part of this work was carried out in the Cambridge Centre for Molecular Recognition.

References

1 Edsall, J. and McKenzie, H.A., Adv. Biophys., 16 (1983) 53.
2 Blake, C.C.F., Pulford, W.C.A. and Artymiuk, P.J., J. Mol. Biol., 167 (1983) 693.
3 Sreenivasan, U. and Axelsen, P.H., Biochemistry, 31 (1992) 12785.
4 Rashin, A.A., Iofin, M. and Honig, B., Biochemistry, 25 (1986) 3619.
5 Loris, R., Stas, P.P.G. and Wyns, L., J. Biol. Chem., 269 (1994) 26722.
6 Thanki, N., Umrania, Y., Thornton, J.M. and Goodfellow, J.M., J. Mol. Biol., 224 (1991) 669.
7 Finney, J.L., Phil. Trans. R. Soc. London Ser. B, 278 (1977) 3.
8 Vyas, N.K., Vyas, M.N. and Quiocho, F.A., Science, 242 (1988) 1290.
9 Poornima, C.S. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 500.
10 Poornima, C.S. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 513.
11 Dean, P.M., Barakat, M.T. and Todorov, N.P., In Dean, P.M., Jolles, G. and Newton, C.G. (Eds.) New Perspectives in Drug Design, Academic Press, London, 1995, pp. 155–183.
12 LaLonde, J.M., Bernlohr, D.A. and Banaszak, L.J., Biochemistry, 33 (1994) 4885.
13 Zhang, X.-J. and Matthews, B.W., Protein Sci., 3 (1994) 1031.
14 Mittl, P.R.E. and Schulz, G.E., Protein Sci., 3 (1994) 799.
15 Karplus, P.A. and Schulz, G.E., J. Mol. Biol., 210 (1989) 163.
16 Zou, J.-Y., Flocco, M.M. and Mowbray, S.L., J. Mol. Biol., 233 (1993) 739.
17 Karpusas, M., Holland, D. and Remington, S.J., Biochemistry, 30 (1991) 6024.
18 Remington, S.J., Wiegand, G. and Huber, R., J. Mol. Biol., 158 (1982) 111.
19 Dunn, C.R. and Holbrook, J.J., Phil. Trans. R. Soc. London Ser. B, 332 (1991) 177.
20 Abad-Zapatero, C., Griffith, J.P., Sussman, J.L. and Rossmann, M.G., J. Mol. Biol., 198 (1987) 445.
21 Ji, X., Armstrong, R.N. and Gilliland, G.L., Biochemistry, 32 (1993) 12949.
22 Raghunathan, S., Chandross, R.J., Kretsinger, H.R., Allison, T.J., Penington, C.J. and Rule, G.S., J. Mol. Biol., 338 (1994) 815.
23 Connolly, M.L., Science, 221 (1983) 709.
24 Hubbard, S.J., Gross, K.-H. and Argos, P., Protein Eng., 7 (1994) 613.
25 Hubbard, S.J. and Argos, P., Protein Sci., 3 (1995) 2194.
26 Kuhn, L.A., Siani, M.A., Pique, M.E., Fisher, C.L., Getzoff, E.D. and Trainer, J.A., J. Mol. Biol., 228 (1992) 13.
27 Finer-Moore, J.S., Kossiakoff, A.A., Hurley, J.H., Earnest, T. and Stoud, R.M., Proteins, 12 (1992) 203.
28 Kossiakoff, A.A., Sintchak, M.D., Shpungin, J. and Presta, L.G., Proteins, 12 (1992) 223.
29 Williams, M.A., Goodfellow, J.M. and Thornton, J., Protein Sci., 3 (1994) 1224.
30 Meiering, E.M. and Wagner, G., J. Mol. Biol., 247 (1995) 294.