Journal of Computer-Aided Molecular Design, 9 (1995) 513–520
ESCOM

J-CAMD 315

Hydration in drug design. 2. Influence of local site surface shape on water binding

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: Drug design; Hydration; Protein–ligand complex; Deep grooves; Shallow grooves

Summary

If water molecules are strongly bound at a protein–ligand interface, they are unlikely to be displaced during ligand binding. Such water molecules can change the shape of the ligand binding site and thus affect strategies for drug design. To understand the nature of water binding, and factors influencing it, water molecules at the ligand binding sites of 26 high-resolution protein–ligand complexes have been examined here. Water molecules bound in deep grooves and cavities between the protein and the ligand are located in the indentations on the protein-site surface, but not in the indentations on the ligand surface. The majority of the water molecules bound in deep indentations on the protein-site surface make multiple polar contacts with the protein surface. This may indicate a strong binding of water molecules in deep indentations on protein-site surfaces. The local shape of the site surface may influence the binding of water molecules that mediate protein–ligand interactions.

Introduction

The explicit incorporation of water molecules into ligand design is a complex problem [1–4]. Our long-term objective is to be able to expel, retain or move water molecules dynamically within the site while a de novo design program is being executed. Before that scenario is possible, more basic information about crystallographically retained water molecules at the sites in protein–ligand complexes is desired. In the previous paper [5], it has been shown that there are many water molecules acting as bridges between the ligand and its site. Careful analysis revealed that the water molecules form multiple hydrogen bonds to link the ligand and the protein. Furthermore, these bridging water molecules are frequently found linked in chains to other bridged and nonbridged water molecules. Comparison between the hydration sites in empty binding sites and those found in protein–ligand complexes reveals that few, but not all water molecules are displaced by ligand binding. Often some of the water molecules at the binding sites are conserved in structurally similar proteins in different species [6].

There may be several physical and chemical factors that can influence the water binding in proteins. Kuhn et al. [7] have found that the protein surface shape influences water binding to a greater extent than does its chemical nature. This study revealed that groove surfaces are preferred for water binding compared to non-groove surfaces. Cavities in proteins have been studied in detail [8–10]; they have been characterised with respect to size, shape, area, water content and amino acid preference [8,9]. Larger cavities or grooves with a rich polar surface have been found to be preferably hydrated [9,10].

Although cavities and water binding to such cavities have been studied extensively, water-bound grooves and cavities at the protein–ligand interface have not been investigated in detail. If water molecules are bound to the protein surface at the interface between the protein and the ligand, they can effectively change the shape of the binding site and consequently modify the complementarity of this site. Therefore, it may be interesting to discern whether the water binding in the grooves and cavities between the protein and the ligand is influenced by any local surface shape of protein or ligand, or if such water binding is incidental. It is desirable to understand which water molecules are likely to be displaced or remain strongly bound to the protein surface on ligand binding. This information about water binding at the

*To whom correspondence should be addressed.

TABLE 1
PROTEIN–LIGAND COMPLEXES SELECTED FOR THIS STUDY

Protein–ligand complex	*R*-factor	Resolution (Å)	DGW[a]	SGW[b]	Ref.
Lactate dehydrogenase with NADH and oxamate (9ldt)	0.23	2.0	7	2	17
L-Arabinose binding protein with D-galactose (5abp)	0.13	1.8	2	0	18
Chloramphenicol acetyl transferase with chloramphenicol (3cla)	0.15	1.75	9	0	19
Thymidylate synthase with thymidine-5'-monophosphate and dihydrofolate (1tys)	0.18	1.8	12	4	20
Cholesterol oxidase with FAD (1coy)		1.7	12	0	21
Staphylococcal nuclease with 5'-deoxythymidine (1snc)	0.16	1.65	9	0	22
D-Glucose binding protein with D-galactose (2gbp)	0.14	1.9	4	0	23
Adipocyte lipid-binding protein with hexadecane sulfonic acid (1lic)	0.19	1.6	4	0	24
Glutathione reductase with flavin adenine dinucleotide (1ger)	0.16	1.85	14	2	25
H-*ras* p21 protein with guanosine-5'-triphosphate (5p21)	0.19	1.35	7	3	26
Alcohol dehydrogenase with NADH and DMSO	0.17	1.8	7	3	27
Actindin with the inhibitor ([*N*-(L-3-*trans*-carboxyoxirane-2-carbonyl)-L-leucyl]-amido-(4-guanido) butane (1aec)	0.14	1.86	4	2	28
Guanylate kinase with guanosine monophosphate (1gky)	2.0	2.0	6	3	29
Thermolysin with *N*-(1-(2-(*R*,*S*)-carboxy-4-phenylbutyl)- cyclopentyl-carbonyl)-(*S*)-tryptophan (1thl)	0.16	1.7	1	4	30
Triphosphate isomerase with phosphoglycolohydroxamate (7tim)	0.18	1.9	2	3	31
Nucleoside diphosphate kinase with 2'-deoxythymidine diphosphate (1ndc)	0.18	2.0	4	5	32
Phospholipase A2 with the transition-state analogue (1pob)	1.7	2.0	2	3	33
Human neutrophil elastase with methoxysuccinyl-Ala-Ala-Pro-Ala chloromethyl ketone (1hne)	0.16	1.84	2	4	34
Holo-D-glyceraldehyde-3-phosphate dehydrogenase (1gd1)	0.17	1.8	13	4	35
Proteinase A with a tetrapeptide (5sga)	0.11	1.8	2	4	36
Dihydrofolate reductase with NADPH (3dfr)	0.15	1.7	5	7	37
Ribonuclease T_1 with guanylic acid (1rnt)	0.19	1.9	1	3	38
FK506-binding protein (1fkf)	0.17	1.7	1	6	39
Citrate synthase coenzyme A (2cts)	0.16	2.0	2	1	40
Ribonuclease SA with 2'-guanosine-monophosphate (1gmr)	0.14	1.77	0	9	41
HIV-protease with acetyl pepstatin (5hvp)	0.17	2.0	3	6	42

[a] Number of water molecules bound in deep grooves / deep micro-grooves at the protein–ligand interface.
[b] Number of water molecules bound in shallow grooves at the protein–ligand interface.

ligand binding site is crucial in ligand design and can in part be analysed from the hydration sites at the ligand binding sites of high-resolution protein–ligand complexes. A detailed study on the hydrogen-bonding interactions of the water molecules with the protein and the ligand was presented in the first paper of this series [5]. In the present study, water molecules at the ligand binding sites of 26 high-resolution crystal structures have been analysed to see if the surface shape properties of a local hydration subsite are responsible for water binding in grooves and cavities at the protein–ligand interface. Results of our previous study on hydrogen-bonding interactions of binding-site water molecules, together with the present work, may suggest a few common factors that are responsible for the binding of water molecules trapped at the protein–ligand interface.

Materials and Methods

Data set

X-ray atomic coordinates of the protein–ligand complexes chosen for this study were taken from the Brookhaven Protein Data Bank and are listed in Table 1. All protein–ligand complexes were refined to ≤2.0 Å.

All protein residues that lie within a distance of 7.0 Å from a ligand atom have been considered. Symmetry-related residues were generated wherever needed; for example, where the binding site is formed by residues from different subunits.

Surface calculation

Molecular surfaces and areas were calculated analytically by using Connolly's MS program [11]. This program calculates the molecular surface accessible to a probe sphere of a given radius. It calculates both contact surfaces: the convex regions formed by van der Waals surfaces of atoms, and the re-entrant surfaces, i.e., the concave regions traced by the probe sphere as it rolls along the grooves [12,13] between the van der Waals surfaces of the atoms.

All residues within a distance of 7.0 Å from each ligand atom were selected in generating the molecular surface. Contact and re-entrant surfaces were generated at once for all the selected residues of the protein, the ligand

atoms and other heteroatoms (as defined by the PDB), but water oxygen atoms were excluded. The atomic radii chosen were the same as those used by Kuhn et al. [7].

The *Shape* suite of programs [7] has been used to determine the nature of water binding at the ligand binding sites. This suite of programs consists of two programs, *Surfractal* and *Maxaccess*. *Surfractal* calculates the fractal atomic density around each surface point. *Maxaccess* analytically calculates the radius of the largest sphere, tangential to a given surface point that does not penetrate the protein surface.

Surfractal densities and *Maxaccess* radii were calculated at every surface point of the molecular surface generated. The average surfractal density of all surface points (including protein as well as ligand surface points) within a distance of 2.0 Å of a given water molecule was used to assess the location of the bound water molecule. The approximate surface area associated with each water molecule was calculated by adding together the surface areas associated with each surface point within a distance of 2.0 Å of a given water molecule. All water molecules within a distance of 3.5 Å from ligand atoms were considered.

Three groove types with bound water molecules are schematically illustrated in Fig. 1. Water molecules at the binding sites are considered to be bound in shallow grooves or deep grooves, or completely buried in cavities between the ligand and protein surfaces, based on the calculated average surfractal density corresponding to a given water molecule and the number of surface points contributing to the calculated average surfractal densities. The following scheme has been used in classifying the groove-bound water molecules. A water molecule was considered to be completely buried (Fig. 1a) if the average surfractal density was greater than 2.6. If the average surfractal density of a water molecule lies between 2.45 and 2.6, such a water molecule is generally found to be bound in a deep groove (Fig. 1b). The average surfractal densities of water molecules bound to shallow grooves (Fig. 1c) are always found to be ≤ 2.45. The assessment was also based on visual inspection of computer graphics of the proteins.

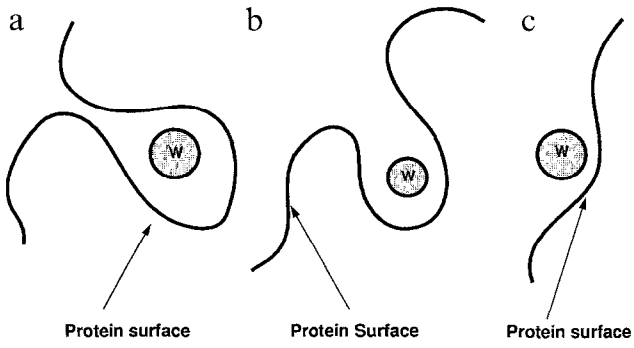

Fig. 1. Schematic illustration of water molecules bound in grooves or cavities. (a) Water molecule completely buried in a cavity; (b) water molecule bound in a deep groove; and (c) water molecule bound in a shallow groove.

The determination of grooves and cavities in macromolecules is probe-sphere dependent [11,14–16]. To resolve this problem, a set of completely buried water molecules were chosen and the probe radius was varied from 1.0 to 2.0 Å in intervals of 0.1 Å. The chosen set included all the completely buried water molecules. The surface areas of the protein and ligand surfaces associated with each water molecule showed little variation in the range 1.2–1.4 Å, and the number of buried water molecules remained constant. Protein and ligand surface areas and some of the associated water molecules decreased dramatically for probe radii greater than 1.6 Å, as the probe sphere cannot access grooves that are smaller than its own radius. The water molecules that are buried for the probes of radius greater than 1.2 Å become more solvent accessible. A probe radius of 1.2 Å appeared to be optimal, in accordance with previous studies on cavities and surfaces [8].

Results

Water molecules bound in deep grooves or buried between the protein and ligand surfaces are illustrated schematically in Figs. 2a and b. Similarly, water molecules bound in shallow grooves between the protein and ligand surfaces, or in a shallow groove on the ligand surface, are shown in Fig. 2c. The sizes and shapes of the grooves in a protein structure generally vary. This is also true for grooves and cavities between protein and ligand surfaces. Often, the indentations on either protein surface or ligand surface bind one or two water molecules. We shall refer to such indentations as 'micro-grooves'.

Ligand binding sites in proteins are located on the protein surface or in clefts on this surface. When a protein–ligand complex is formed and crystallized in solution, the polar and charged groups on the complex surface are generally solvated. However, some of the water molecules are trapped between the protein and the ligand. Here, we have made an attempt to compare and understand the differences between the water binding in deep grooves or cavities at the protein–ligand interface (as in Fig. 2a), and those bound in shallow grooves (as in Fig. 2c) at the protein–ligand interface (occurring towards the surface of the complex) or at the ligand surface.

Structural association of water molecules with protein

When either a single water molecule or a cluster of water molecules are buried between the protein and ligand molecular surfaces, the approximate area of the molecular surface exposed to the water molecules can always be calculated. The surface areas of the protein and ligand surfaces exposed to each water molecule are calculated by taking into account the area associated with all

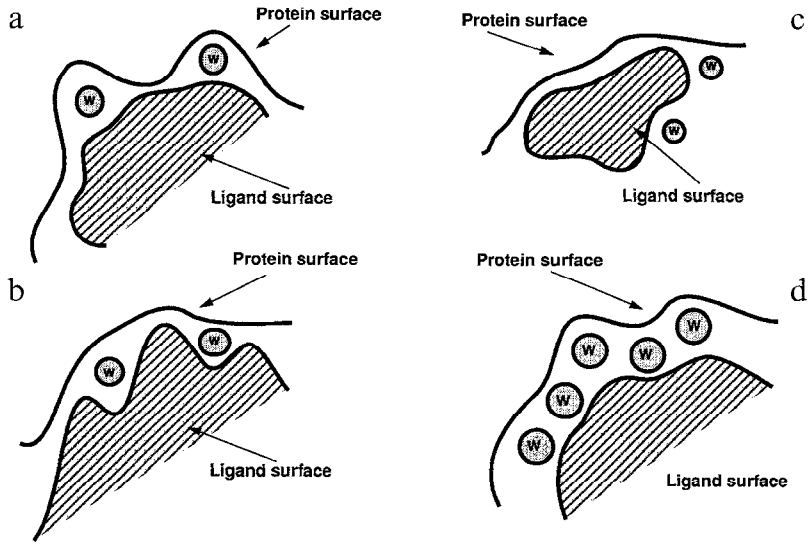

Fig. 2. Schematic illustration of water molecules bound in different types of grooves between protein and ligand. The hatched surfaces represent the ligand surface. (a) Water molecules bound in an indentation on the protein-site surface, where the protein surface area exposed to the water molecules is far larger than the ligand surface area; (b) water molecules bound in indentations on the ligand surface, where the ligand surface area exposed to the water molecule is larger than the protein-site surface area; (c) water molecules bound in shallow grooves at the protein–ligand interface and on the ligand surface; and (d) water molecules bound in clusters in elongated grooves with micro-grooves.

the surface points within a distance of 2.0 Å from the water molecule. These are calculated individually for all the water molecules bound in both deep grooves and shallow grooves.

In Fig. 3, the surface area of the protein exposed to a water molecule is plotted against the surface area of the ligand exposed to the same water molecule. Triangles correspond to water molecules bound in deep grooves or cavities (see Fig. 2a) at the protein–ligand interface, and square boxes correspond to water molecules bound in shallow grooves, either on the ligand surface or at the protein–ligand interface, as shown in Fig. 2b. The difference in the two distributions can be assessed using a two-dimensional Kolmogorov–Smirnov (K–S) test [43,44]. The K–S statistic D is a measure of maximum difference between any two given distributions [45]. The statistic D in Fig. 3 is 0.49 and the probability that the two distributions are identical is $P = 1.45 \times 10^{-8}$; the two distributions in Fig. 3 are thus different.

The distribution for water molecules bound in deep grooves (Fig. 3) shows a shift towards the ordinate, indicating a larger portion of the protein surface being exposed to the bound water molecule as compared to the ligand surface. This feature can be quantified by computing the surface area of the protein exposed to a water molecule S_P, and the surface area of the ligand exposed to the same water molecule S_L; the difference in the surface area, $(S_P - S_L)$, is plotted against the frequency of occurrence in Fig. 4. The dark bars correspond to water molecules bound in deep grooves (Fig. 2a), while the lighter bars corresponds to water molecules bound in shallow grooves (Fig. 2b). These positive values of $(S_P - S_L)$ indicate that the protein surface area exposed to individual water molecules in deep grooves (or buried between the protein and the ligand) is larger than the corresponding ligand surface. This means that if a water molecule has an $(S_P - S_L)$ far larger than zero, it is likely to be bound in deep indentations on the protein surface, as shown in Fig. 2a. The difference $(S_P - S_L)$ is far above zero for many deep-groove water molecules in the data set. On the other hand, negative values of $(S_P - S_L)$ indicate that the surface

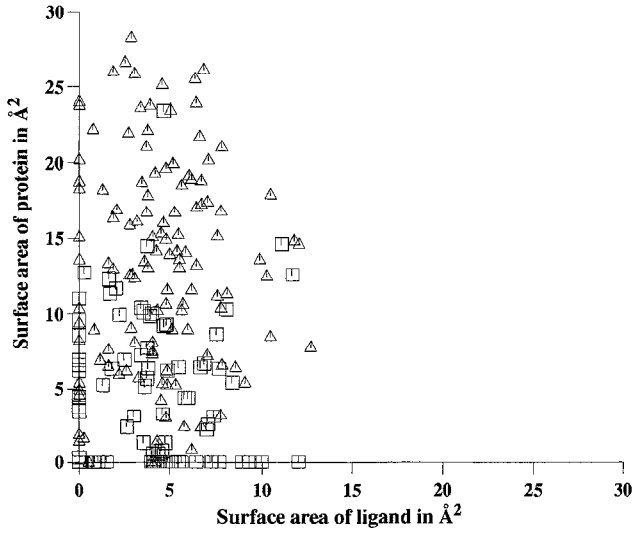

Fig. 3. Plot of the protein surface area exposed to a water molecule versus the ligand surface area exposed to the same water molecule. Water molecules in deep grooves (indentations) between the protein and the ligand are represented by triangles; water molecules bound in shallow grooves at the protein–ligand interface are represented by square boxes.

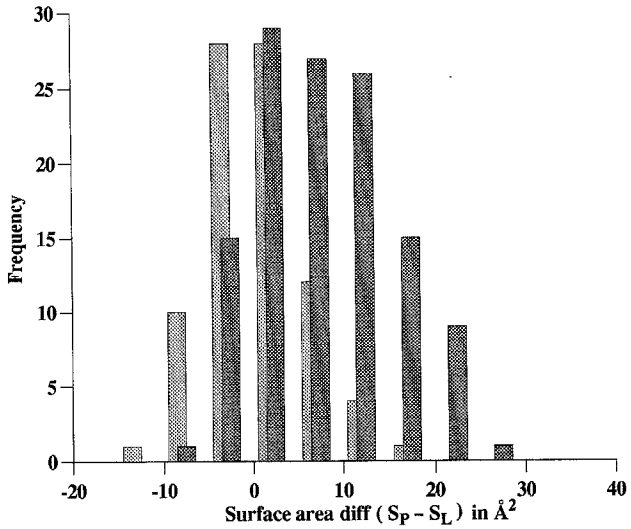

Fig. 4. Plot of difference in protein and ligand surface areas exposed to the same water molecule, against the frequency of occurrence. The dark bars correspond to water molecules bound in deep grooves or indentations between the protein and ligand surfaces. The lighter bars correspond to water molecules bound in shallow grooves at the protein–ligand interface.

area of the ligand exposed to a water molecule is larger than the surface area of the protein exposed to the same water molecule. A typical scenario is shown in Fig. 2b, where the water molecule is bound in the deep indentations on the ligand surface. For water molecules bound in deep grooves, there are very few cases for which $(S_P - S_L)$ is negative. This suggests that if a water molecule is bound in a deep groove or buried between the ligand and protein surfaces, in most of the cases in the data set, the water molecules are bound in deep indentations on the protein surface. This corresponds to the situation shown in Fig. 2a, but not to that shown in Fig. 2b. This is an important observation for ligand design. Design strategies therefore need to concentrate on placing water molecules on the protein surface; not on the ligand surface.

In Fig. 3, there are many data points corresponding to zero surface area of the ligand. This shows that there are water molecules bound in deep grooves at the protein–ligand interface, which do not make any contact with the ligand. This feature in Fig. 3, together with the population distribution for water molecules bound in deep grooves (see Fig. 4), suggests that the local surface shape of the hydration subsites (indentations on protein surface) on the protein-site surface may be responsible for water binding in deep micro-grooves.

The distribution for water molecules bound in shallow grooves (Fig. 2c) is very different from that of water molecules bound in deep grooves (Fig. 4). The former distribution peaks around zero, indicating that the surface areas of the protein and the ligand exposed to the same water molecule are nearly equal for many water molecules in the data set. The surface areas of ligands exposed to water molecules bound in shallow grooves are often larger than those of proteins. Furthermore, a significant population on the abscissa (Fig. 3) suggests the presence of a number of water molecules bound in shallow grooves, but not making any contact with the protein surface. These features in Figs. 3 and 4 indicate that binding of water molecules in the shallow grooves at the ligand surface, or at the protein–ligand interface, is probably not influenced by any special physical feature of the protein or ligand surfaces.

In order to understand the nature of the deep-groove or micro-groove surface, the number of polar contacts made by each water molecule with the protein surface is plotted against the frequency of occurrence in Fig. 5. Dark and lighter bars correspond to water molecules bound in deep and shallow grooves, respectively. In calculating the number of polar contacts made by each water molecule with the protein surface, all polar atoms within a distance of 4 Å from the water molecule are considered. Most of the water molecules bound in deep grooves have two or more such polar contacts. On the other hand, the majority of the water molecules bound in shallow grooves have zero to two polar contacts with the protein surface. There are only few water molecules bound in shallow grooves that have multiple polar contacts with the protein surface.

Water molecules at the binding site of chloramphenicol acetyl transferase

The binding site of the binary complex of chloramphenicol acetyl transferase bound to chloramphenicol presents an interesting feature. There are several ordered water molecules buried between the protein surface and

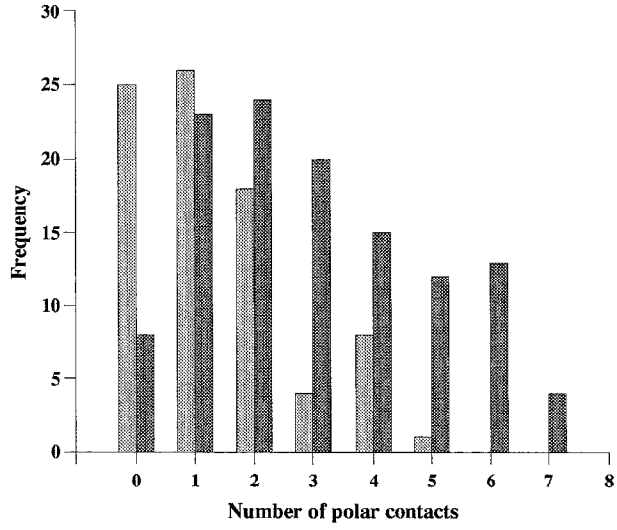

Fig. 5. Plot of the number of polar contacts that a water molecule makes with the protein-site surface against the frequency of occurrence. The dark bars correspond to water molecules bound in deep micro-grooves or indentations between protein and ligand. The lighter bars correspond to water molecules bound in shallow grooves at the protein–ligand interface or on the ligand surface.

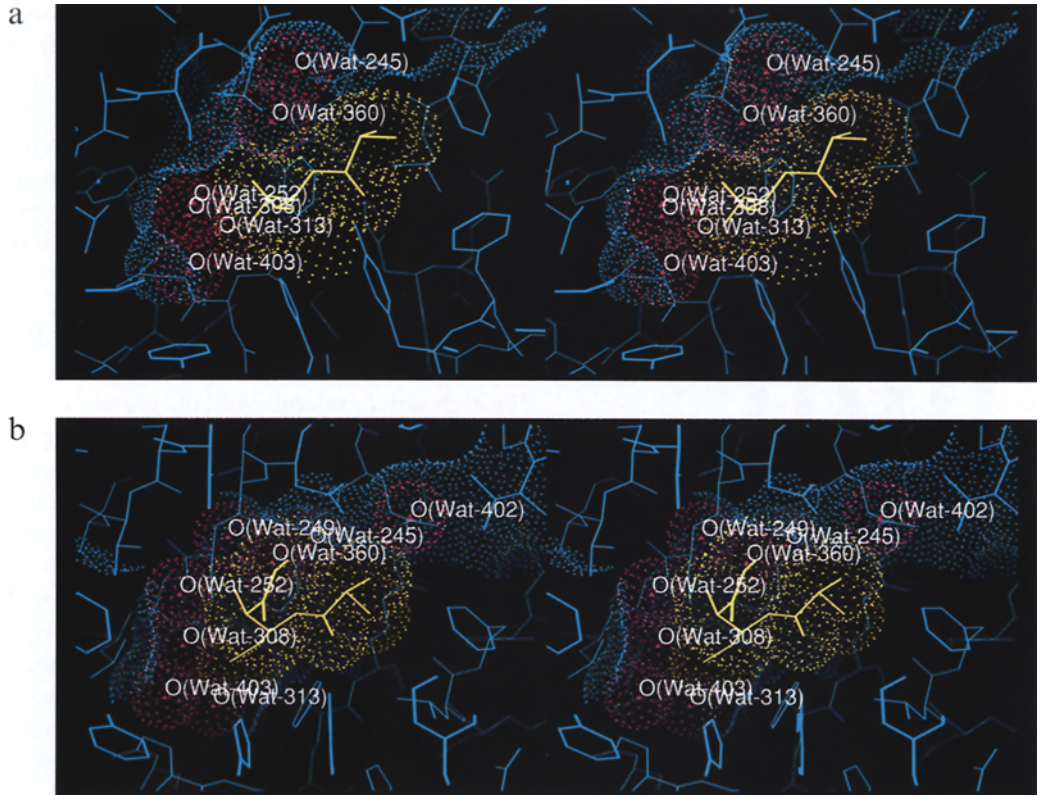

Fig. 6. Ligand binding site of chloramphenicol acetyl transferase with the ligand chloramphenicol; these compounds are depicted in dotted molecular surfaces with bound water molecules at the site. All the amino acid residues from different symmetry-related monomers forming the binding site have been considered in generating the molecular surface of the site. The dotted surface in cyan corresponds to the protein surface and the dotted surface in yellow corresponds to the ligand surface. Water molecules bound between the ligand and the protein are shown in magenta dotted spheres. (a) Water molecules Wat^{245} and Wat^{252} bound in deep micro-grooves and (b) layer of water molecules bound in the elongated groove between the enzyme and the ligand chloramphenicol.

the chloramphenicol surface. Two clusters of water molecules are present that form a water layer between the enzyme and the ligand (see Fig. 6b). The water molecule Wat^{308} forms hydrogen bonds with Wat^{313}, Wat^{252} and Wat^{403}. Similarly, Wat^{245}, Wat^{249} and Wat^{360} form a cluster. Figures 6a and b show the dotted molecular surface of the protein-site surface and the ligand surface with clusters of water molecules at the binding site, in two different orientations. All these water molecules are buried completely between the enzyme and the substrate in an elongated groove. The average surfractal densities of all the water molecules at the binding site are greater than 2.6. The residues presenting the protein-site surface are predominantly hydrophobic and water molecules, in particular Wat^{245}, Wat^{252} and Wat^{320}, are bound in these hydrophobic deep indentations (micro-grooves in Fig. 6a).

Visual inspection of binding sites revealed that the deep grooves on the protein surface at the protein–ligand interface are not always the ideal cases shown in Fig. 2a. As mentioned before, the shapes and sizes of these grooves and cavities vary [9,10] and are often elongated [8]. These elongated groove surfaces often contain micro-grooves or indentations (as shown in Fig. 2d), in which water molecules can be bound singly or in clusters. A typical example of clusters of water molecules is found in the elongated groove, containing micro-grooves, at the binding-site surface of chloramphenicol acetyl transferase. Wat^{245} and Wat^{252} are bound in deep micro-grooves (Fig. 6a). Other water molecules, bound in the elongated groove between the enzyme and the chloramphenicol, form clusters (see the discussion below). In such clusters, the water molecules bound in deep micro-grooves make multiple polar contacts with the protein surface, while others in the cluster, making secondary contact with the protein surface, have only one polar contact. This explains the higher frequency of occurrence (see Fig. 5) of water molecules bound in deep grooves (darker bars) that make one polar contact. It also explains the broader distribution of the data with values >0 in Fig. 4 for deep-groove water molecules. For those water molecules occurring in clusters in elongated grooves, the surface areas of the protein and ligand exposed to the individual water molecules can be approximately equal. However, a larger population occurring far above zero (see Fig. 4) indicates that a significant number of water molecules in the data set are bound in deep micro-grooves.

Discussion

Our analyses reveal that the water molecules bound in deep grooves and cavities at the protein–ligand interface are exposed more to the protein surface than to the ligand surface. More precisely, water molecules generally occur in deep indentations or micro-grooves on the protein-site surface, but not in deep indentations on the ligand surface. Furthermore, most of the water molecules that are bound in deep indentations or micro-grooves on the protein-site surface make multiple polar contacts with the groove surface of the protein. Williams et al. estimated an approximate energy of stabilization of 0.6 kcal/mol associated with each additional polar contact of a water molecule with the groove surface [10]. Thus, the multiple polar contacts of water molecules in deep micro-grooves suggest a stronger binding of these molecules. Our study suggests that local curvatures on the protein-site surface may be helpful in physically holding the water molecules more firmly. Thus, the local surface shape of the proteins may influence the binding of water molecules mediating protein–ligand interactions. The occurrence of these water molecules in cavities or deep micro-grooves at the protein–ligand interface need not necessarily be incidental.

Thermodynamic perturbation theory calculations by Wade et al. [46] on cavity water molecules in sulphate-binding proteins show that the free energy of binding of the water molecule in a polar cavity is more favourable by −10 kcal/mol. Their study also reveals that the free energy of water binding to a nonpolar cavity is slightly unfavourable (by 0.2 kcal/mol). Furthermore, recent surveys on cavities in proteins also suggest that the polar cavities are preferably hydrated [8,9]. Molecular dynamics simulations of bovine pancreatic trypsin inhibitor [47,48] predict short residence times for surface waters, while NMR experiments [49] suggest longer residence times for buried water molecules and shorter residence times for surface water molecules.

Our study on conserved water molecules at the ligand-binding sites of homologous proteins reveals that most of these water molecules (at the ligand-binding sites) are bound in deep grooves and/or buried between protein and ligand [6], and that they make multiple polar contacts. Buried water molecules have been observed to be conserved in the homologous family of serine proteases [50]. Conserved water molecules in a homologous family of legume lectins have zero or very little solvent accessibility [51]. Furthermore, the B-factors of these water molecules are comparable to the average B-factors of the protein atoms. X-ray and neutron diffraction experiments on crystals of trypsin also suggest that the buried water molecules are more conserved than the surface water molecules [52,53]. Buried water molecules observed in NMR experiments in FK506-binding protein [54], pancreatic trypsin inhibitor [55], thioredoxin [56] and interleukin-1β [57] occur at the same positions as in crystal structures. These water molecules have been observed to have long residence times, ranging from 10^{-2} to 10^{-8} s. These observations imply a functional and structural importance for water molecules bound in deep grooves and cavities in protein structures. These buried water molecules have often been observed to bridge secondary structural elements, forming multiple hydrogen bonds [58]. In the present case, the water molecules bound in deep micro-grooves may contribute to the stability of the protein–ligand complex by making multiple hydrogen bonds between them.

On the basis of our present analyses, we propose that deep grooves and micro-grooves on the protein-site surface, with a rich polar surface, are probably the preferred sites of strong water binding. Such water molecules are probably unlikely to be displaced in the process of ligand binding. Water molecules in shallow grooves at the protein–ligand interface are less tightly bound, and are probably displaced during ligand binding. Water molecules bound in deep indentations on the protein-site surface effectively change the shape of the ligand-binding site surface, thus presenting a relatively smoother new surface that is complementary to the ligand. Therefore, these water molecules can be structurally and functionally important for ligand design.

Current methods for de novo ligand design ignore the presence of water molecules in the site. In the light of this survey and from the observations described in the previous paper, it follows that water molecules form an integral part of the protein–ligand complexes. Two features stand out: firstly, the water molecules in the complex are often multiply hydrogen-bonded in bridging the ligand and the protein, sometimes through networks of water molecules; secondly, these water molecules are often located in micro-grooves within the site. Therefore, any realistic de novo ligand design program should attempt to incorporate these design features into its algorithm. As candidate structures are being grown, the newly created surface between the ligand candidate and the site needs to be assessed with regard to shape characteristics for water binding as well as hydrogen-bonding possibilities. The algorithm also needs to keep track of possible partial networks of water molecules being formed at the protein–ligand interface. Strategies of this type implicitly assume that the generated structures will dock to the site in the orientation and conformation in which they are grown by de novo generation. If an algorithm can be developed to incorporate these design features, its scoring function will need to be fine-tuned against known hydrated complexes. One way forward might be to use a neural network to create a scoring function from the hydrated protein–ligand complexes studied here.

In the third paper of this series on hydration, a detailed study on the water molecules conserved at the ligand-binding sites of structurally homologous proteins

is presented. The presence of such conserved water molecules at the ligand-binding sites confirms the structural importance of these water molecules, which play a significant role in mediating the protein–ligand interactions.

Acknowledgements

C.S.P. thanks 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 Dean, P.M., In Dean, P.M. (Ed.) Molecular Similarity in Drug Design, Blackie Academic & Professional, London, 1995, pp. 1–23.
2 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.
3 Ringe, D. and Petsko, G.A., In Dean, P.M., Jolles, G. and Newton, C.G. (Eds.) New Perspectives in Drug Design, Academic Press, London, 1995, 89–118.
4 Kuntz, I.D., Meng, E.C. and Shoichet, B.K., In Dean, P.M., Jolles, G. and Newton, C.G. (Eds.) New Perspectives in Drug Design, Academic Press, London, 1995, pp. 137–154.
5 Poornima, C.S. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 500.
6 Poornima, C.S. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 521.
7 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.
8 Rashin, A.A., Iofin, M. and Honig, B., Biochemistry, 25 (1986) 3619.
9 Hubbard, S.J., Gross, K.-H. and Argos, P., Protein Eng., 7 (1994) 613.
10 Williams, M.A., Goodfellow, J.M. and Thornton, J.M., Protein Sci., 3 (1994) 1224.
11 Connolly, M.L., Science, 221 (1983) 709.
12 Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
13 Richards, F.M., Annu. Rev. Biophys. Bioeng., 6 (1977) 151.
14 Connolly, M.L., J. Am. Chem. Soc., 107 (1985) 1118.
15 Eriksson, A.E., Baase, W.A., Wozniak, J.A. and Matthews, B.W., Nature, 355 (1992) 371.
16 Eriksson, A.E., Baase, W.A., Zhang, X.-J., Heinz, D.W., Blaber, M., Balwin, E.P. and Matthews, B.W., Science, 255 (1992) 178.
17 Dunn, C.R. and Holbrook, J.J., Phil. Trans. R. Soc. London Ser. B, 332 (1991) 177.
18 Quiocho, F.A., Wilson, D.K. and Vyas, N.K., Nature, 340 (1989) 404.
19 Leslie, A.G.W., J. Mol. Biol., 213 (1990) 167.
20 Fauman, E.B., Rutenber, E.E., Maley, G.F., Maley, F. and Stroud, R.M., Biochemistry, 33 (1994) 1502.
21 Li, J.Y., Vrielink, A., Brick, P. and Blow, D.M., Biochemistry, 32 (1993) 11507.
22 Loll, P.J. and Lattman, E.E., Proteins, 5 (1989) 183.
23 Vyas, N.K., Vyas, M.N. and Quiocho, F.A., Science, 242 (1988) 1290.
24 LaLonde, J.M., Bernlohr, D.A. and Banaszak, L.J., Biochemistry, 33 (1994) 4885.
25 Mittl, P.R.E. and Schulz, G.E., Protein Sci., 3 (1994) 799.
26 Pai, E.F., Krengel, U., Petsko, G.A., Goody, R.S., Kabsch, W. and Wittinghofer, A., EMBO J., 9 (1990) 2351.
27 Al-Karadaghi, S., Cedergren-Zeppezauer, E.S., Petratos, K., Hovmoeller, S., Terry, H., Dauter, Z. and Wilson, K.S., Acta Crystallogr., B50 (1994) 793.
28 Varughese, K.I., Su, Y., Cromwell, D., Hasnain, S. and Xuong, N.-H., Biochemistry, 31 (1992) 5172.
29 Stehle, T. and Schulz, G.E., J. Mol. Biol., 224 (1992) 1127.
30 Holland, D.R., Barclay, P.L., Danilewicz, J.C., Matthews, B.W. and James, K., Biochemistry, 33 (1994) 51.
31 Davenport, R.C., Bash, P.A., Seaton, B.A., Karplus, M., Petsko, G.A. and Ringe, D., Biochemistry, 30 (1991) 5821.
32 Cherfils, J., Morea, S., Lascu, I., Veron, M. and Janin, J., Biochemistry, 33 (1994) 9062.
33 White, S.P., Scott, D.L., Otwinowski, Z., Gelb, M.H. and Sigler, P.B., Science, 250 (1990) 1560.
34 Navia, M.A., McKeever, B.M., Springer, J.P., Lin, T.Y., Williums, H.R., Fluder, E.M., Dorn, C.P. and Hoogsteen, K., Proc. Natl. Acad. Sci. USA, 86 (1989) 7.
35 Skarzynsky, T., Moody, P.C.E. and Wonacott, A.J., J. Mol. Biol., 193 (1987) 171.
36 James, M.N.G., Sielecki, A.R., Brayer, G.D., Delbaere, L.T.J. and Bauer, C.A., J. Mol. Biol., 144 (1980) 43.
37 Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. and Kraut, J., J. Biol. Chem., 257 (1982) 13650.
38 Arni, R., Heinemann, U., Maslowska, M., Tokuoka, R. and Saenger, W., Acta Crystallogr., B43 (1987) 549.
39 Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J., Science, 252 (1991) 839.
40 Remington, S., Wiegand, G. and Huber, R., J. Mol. Biol., 158 (1982) 111.
41 Sevcik, J., Hill, C.P., Dauter, Z. and Wilson, K.S., Acta Crystallogr., B49 (1993) 257.
42 Fitzgerald, P.M.D., McKeever, B.M., Van Middlesworth, J.F., Springer, J.P., Heimbach, J.C., Leu, C.-T., Herber, W.K., Dixon, R.A.F. and Darke, P.L., J. Biol. Chem., 265 (1990) 14209.
43 Fasano, G. and Franceschini, A., Mon. Not. R. Astron. Soc., 225 (1987) 155.
44 Peacock, J.A., Mon. Not. R. Astron. Soc., 202 (1983) 615.
45 Ledermann, W., In Lloyd, E. (Ed.) Handbook of Applicable Mathematics, Vol. 6, Part A, Wiley, New York, NY, 1984, pp. 257–258.
46 Wade, R.C., Mazor, M.H., McCammon, J.A. and Quiocho, F.A., Biopolymers, 31 (1991) 919.
47 Levitt, M. and Sharon, R., Proc. Natl. Acad. Sci. USA, 85 (1988) 7557.
48 Brunne, R.M., Liepinsh, E., Otting, G., Wüthrich, K. and van Gunsteren, W.F., J. Mol. Biol., 231 (1993) 1040.
49 Otting, G., Liepinsh, E. and Wüthrich, K., Science, 254 (1991) 974.
50 Sreenivasan, U. and Axelsen, P.H., Biochemistry, 31 (1992) 12785.
51 Loris, R., Stas, P.P.G. and Wyns, L., J. Biol. Chem., 269 (1994) 26722.
52 Finer-Moore, J.S., Kossiakoff, A.A., Hurley, J.H., Earnest, T. and Stoud, R.M., Proteins, 12 (1992) 203.
53 Kossiakoff, A.A., Sintchak, M.D., Shpungin, J. and Presta, L.G., Proteins, 12 (1992) 223.
54 Xu, R.X., Meadows, R.P. and Fesik, S.W., Biochemistry, 32 (1993) 2473.
55 Otting, G. and Wüthrich, K., J. Am. Chem. Soc., 111 (1989) 1871.
56 Forman-Kay, J.D., Gronenborn, A.M., Wingfield, P.T. and Clore, G.M., J. Mol. Biol., 220 (1991) 209.
57 Clore, G.M., Bax, A., Wingfield, P.T. and Gronenborn, A.M., Biochemistry, 29 (1990) 5671.
58 Meiering, E.M. and Wagner, G., J. Mol. Biol., 247 (1995) 294.