doi:10.1006/jmbi.2001.5154 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **314**, 619–632

JMB

Do Water Molecules Mediate Protein-DNA Recognition?

Ch. Koti Reddy, Achintya Das and B. Jayaram*

Department of Chemistry
Indian Institute of Technology
Hauz Khas, New Delhi
110016, India

A comprehensive analysis of interfacial water molecules in the structures of 109 unique protein-DNA complexes is presented together with a new view on their role in protein-DNA recognition. Location of interfacial water molecules as reported in the crystal structures and as emerging from a series of molecular dynamics studies on protein-DNA complexes with explicit solvent and counterions, was analyzed based on their acceptor, donor hydrogen bond relationships with the atoms and residues of the macromolecules, electrostatic field calculations and packing density considerations. Water molecules for the purpose of this study have been categorized into four classes: viz. (I) those that contact both the protein and the DNA simultaneously and thus mediate recognition directly; (II) those that contact either the protein or the DNA exclusively *via* hydrogen bonds solvating each solute separately; (III) those that contact the hydrophobic groups in either the protein or the DNA; and, lastly (IV) those that contact another water molecule. Of the 17,963 crystallographic water molecules under examination, about 6% belong to class I and 76% belong to class II. About three-fourths of class I and class II water molecules are exclusively associated with hydrogen bond acceptor atoms of both protein and DNA. Noting that DNA is polyanionic, it is significant that a majority of the crystallographically observed water molecules as well as those from molecular dynamics simulations should be involved in facilitating binding by screening unfavorable electrostatics. Less than 2% of the reported water molecules occur between hydrogen bond donor atoms of protein and acceptor atoms of DNA. These represent cases where protein atoms cannot reach out to DNA to make favorable hydrogen bond interactions due to packing/structural restrictions and interfacial water molecules provide an extension to side-chains to accomplish hydrogen bonding.

Keywords: hydrogen bonds; extended side-chains; electrostatic screening; water-mediated recognition; packing density

*Corresponding author

Introduction

Macromolecular crystallography has implicated water molecules as important contributors to stability and specificity.[1–3] As their resolution improves beyond roughly 3 Å, crystal structures of protein-DNA complexes often reveal ordered water molecules at protein-DNA interfaces.[4] These water molecules in the complex may be remnants of the solvation shells of individual macromolecules before binding, they may occur to fill the gaps arising from imperfect matches of protein and DNA surfaces to sustain a certain threshold packing density, or they may play a more active functional role as mediators of protein-DNA recognition. In this study we explore the raison d'être of interfacial water molecules *via* a structural analysis of 109 protein-DNA complexes and arrive at some new perspectives on their function. We note, in particular, that a great majority of the observed water molecules do not merely continue to solvate the macromolecules in the complex but are located strategically at the interface primarily to buffer the electrostatic repulsions between phosphate groups of DNA and the electronegative atoms on the protein. A small percentage of water molecules are also noticed to act as linkers (extended side-chains), joining the hydrogen bond donor atoms of the protein to the acceptors atoms of DNA. We further analyze the ''ordered water molecules'' in terms of atom and residue-wise solvation trends, packing densities with and without

E-mail address of the corresponding author:
bjayaram@chemistry.iitd.ac.in

0022-2836/01/030619–14 $35.00/0

water molecules, alignment of water dipoles in the electrostatic field due to macromolecules, and comment on their implication to protein-DNA recognition.

Water in Protein and Nucleic Acid Structures

There is a large body of literature[1–3,5–10] that reviews the hydration patterns of both proteins and nucleic acids as observed using structural techniques. They have also been the topic of extensive theoretical studies.[11–13] Numerous examples have been reported of the structural and functional importance of water molecules around biomolecules. They are associated with the native structure of the proteins and in many cases they are implicated as having a direct bearing on molecular recognition and catalysis.[14,15] Structural studies point to a major role of water in protease-inhibitor binding and in antigen-antibody recognition.[16,17] One of the ordered water molecules seen in the complexes of HIV protease with peptide ligands has guided the design of a novel tightly bound inhibitor.[18,19] Water molecules have been shown to be crucial in defining the substrate specificity of bacterial arabinose-binding protein[20] and glutamate dehydrogenase.[21] Water is known to play a catalytic role in the hydrolysis of carboxypeptidase A.[22] In this case, the general base catalysis is triggered by the activation of a water molecule by a glutamate sidechain.

Some thorough reviews on DNA hydration are available.[1,5,6] First observed in the minor groove of the d(CGCGAATTCGCG) dodecamer,[23] where a pattern of water molecules link together the N-3 of an adenine base and O-2 of the thymine base of the adjacent T.A pair in the minor groove of *B*-form DNA, the spine of hydration seems to be a common feature to the A + T-rich regions[24] and is presumed to stabilize the DNA conformation. The discovery of structured water around a *B*-DNA fragment further catalyzed the quest for understanding DNA hydration structure. A structural analysis of crystalline CGATTAATCG[25] shows the spine of hydration in the narrow regions of the minor groove of the double helix, and ribbons of water in the wider sections. Similarly, in the structures of crystalline CCAACITTGG[26] the spine appears in the narrow center and is terminated by the presence of a hydrated ion. The hydration of DNA has been the focus of several computer simulation studies.[27–31]

In *Z*-DNA and *A*-DNA crystals, water molecules can bridge between adjacent phosphate groups. A spine of hydration is found in the major grooves of *Z*-DNA duplexes, with water molecules bridging the O-2 atoms of successive cytidine bases.[32,33] Analysis of water distributions around phosphate groups revealed that water molecules are concentrated in six hydration sites per phosphate group, and that the positions and occupancies of these sites are dependent on the conformation and type of nucleotide.[34] Right-handed DNA duplexes assume a *B*-form at high water activity and an *A* form at reduced levels. A free energy analysis[35] of molecular dynamics trajectories of *A* and *B*-forms of DNA in water and in mixed solvent systems revealed that the conformational preferences of DNA were due to a fine electrostatic balance between inter-phosphate repulsions, counterion-DNA attractions and solvation/desolvation energetics. The ordered water molecules thus may have to be viewed from both structural and energetic perspectives.

The crystal structure of the complex of $tRNA^{Gln}$ and its cognate tRNA synthetase shows that water molecules occur in the minor groove of the double-stranded RNA helix of the tRNA, near the amino acid acceptor stem, and may help to establish the tRNA's identity.[36]

Water in Protein-DNA Complexes

It has long been proposed that water molecules could participate in hydrogen bonding networks that link side-chain and main-chain atoms with the functional groups on the bases, and the anionic oxygen atoms of the phosphodiester backbone.[2,23,25,37–41] In the structure of trp repressor-DNA complex,[42,43] there are very few direct contacts between the protein and the base-pairs, and these contacts do not seem to be important for base-sequence recognition. There are, however, three ordered water molecules at the protein-DNA interface that hydrogen bond with both the base-pairs and the protein side-chains. The bases involved in these water-mediated interactions are among the most important in specifying the repressor's affinity for the operator sequence.

NMR studies of the Antennapedia homeodomain indicate that at least two amino acid side-chains at the protein-DNA interface are in close proximity to water molecules. The importance of these water molecules for binding and recognition was highlighted by the crystal structure of the paired homeodomain bound to DNA. Remarkably, in this structure, there are 18 ordered water molecules at the protein-DNA interface.[44] Stability and specificity are reported to be conferred by a network of water-mediated protein-DNA hydrogen bonds in the estrogen receptor-DNA complex.[45] Thus, water is considered to play a role in furnishing DNA binding specificity to nuclear hormone receptors.

Potentially specific interactions between protein and DNA were identified some years ago in terms of the pattern of hydrogen bonding of donor and acceptor sites in the major and minor grooves of DNA with that of side-chains of the amino acid residues on protein.[46] It is further proposed[47] that protein atoms involved in binding to DNA occupy positions normally occupied by water molecules in unbound DNA.

Not all protein-DNA complexes are highly hydrated at the interface. The structure of the TATA box-binding protein (TBP) bound to DNA exhibits a very hydrophobic interface.[48–50] TBP interacts along the length of the minor groove of DNA, which is splayed open and curves away from the protein. Many hydrophobic amino acids on the surface of the protein are in van der Waals contact with the edges of the base-pairs and the ribose sugar moieties. As the minor groove is normally highly hydrated, many water molecules must be displaced and the driving force for complex formation would seem to be primarily entropic.[51]

Overall, the occurrence of several crystallographically ordered water molecules in protein-DNA complexes necessitates a molecular explanation for their presence and their implication to protein-DNA recognition. Here, we focus on water molecules observed at the interface of protein-DNA complexes and attempt to elucidate some general principles concerning their structural organization as pertinent to binding.

Structural analysis of 109 protein-DNA complexes: a new view on the role of water

A search through the RCSB Protein Data Bank†[52] revealed 587 protein-nucleic acid complexes, of which 568 were X-ray/NMR structures and the rest were theoretical models. Of these, 175 were unique complexes, the rest being repetitions and mutants. Some of these had gaps in their deposited coordinates and a few others did not contain any water molecules. Those structures were excluded, leaving 109 unique protein-DNA complexes for our study here. Table 1 shows their PDB codes and the number of water molecules reported in each case.

Hydrogen atoms were added to the complex and water oxygen atoms and minimized for 500 steps (50 steps of steepest descent and 450 steps of conjugate gradient) keeping the rest of the atoms fixed using the AMBER version 6 suite of programs.[53] To relax any further steric clashes yet stay close to the crystal structure, a restrained, all-atom minimization was carried out, wherein the complex was initially subjected to a restraint of 25 kcal/mol (1 cal = 4.184 J), which was relaxed gradually over 500 steps of minimization. Further analysis of water molecules is described briefly below.

Hydrogen bonding criterion

The criterion adopted to identify atoms that are hydrogen bonded to a reference water molecule is shown in Figure 1. We construct a cone about $H_{(1)}$ such that an acceptor atom $X_{(1)}$ on protein, DNA or any other water molecule that falls in that cone can make a hydrogen bond with $H_{(1)}$. The first condition required for $H_{(1)}$ to be hydrogen bonded to the acceptor atom $X_{(1)}$ is that OX (r in Figure 1) be less than 3.3 Å.[54] This distance ensures that each water molecule is assigned to at least one or the other atom of protein or DNA or another water molecule. The second condition is that $\theta_{(1)}$ be between 0 and 30°. For a donor atom $X_{(2)}$ to be hydrogen bonded to the oxygen atom of water, the angle $X_{(2)}OH_{(2)}$ i.e. $\theta_{(2)}$ or the angle $X_{(2)}OH_{(1)}$ should be between 80 and 140°. This is because of the orientation of the lone pairs in a tetrahedral geometry around the oxygen atom in water. The above conditions for hydrogen bond identification are liberal enough and err, if at all, on the right side. This procedure is repeated with $H_{(2)}$ and with the two lone pairs of oxygen atoms. The above assignment scheme partitions all water molecules in the system uniquely to a specific solute atom based on proximity criterion.

Classification of water molecules in protein-DNA complexes

Depending on the acceptor, donor hydrogen bond relationships, we classify the water molecules into four broad, mutually exclusive classes. Table 2 shows the classification scheme. Water molecules at the interface that make hydrogen bonds with protein as well as DNA are the most interesting for this study. We classify them as class I water molecules. Again, in class I, based on the type of the atom, i.e. acceptor or donor with which the water molecule is in contact through hydrogen bonds, we divide them further into four categories. In class II, we include water molecules that contact the atoms of either protein or DNA but not both. Here again, depending on the type of the atoms with which the water molecule is in contact, these water molecules are pooled under four categories. Water molecules proximal to hydrophobic atoms of either protein or DNA are dealt with separately and are included under class III to gauge their frequency of occurrence. Finally, we club water molecules hydrogen bonding with other water molecules as in the bulk solvent in class IV. The different classes defined above as determined for water molecules in the trp repressor-operator complex are illustrated in Figure 2. The results of the above classification in 109 protein-DNA complexes are shown in Table 3.

Do water molecules mediate protein-DNA interactions?

Class I water molecules are the mainstay behind the proposal of water-mediated protein-DNA interactions. Quite intriguingly, their number is very small (5.5 %). In class I, the role of category 1 and 4 water molecules is straightforward. We would not expect either a direct hydrogen bond or a favorable electrostatic interaction between two acceptor or two donor atoms. Here, the role of water appears

† http://www.rcsb.org/pdb

Table 1. A classification of water molecules reported in the crystal structures of protein-DNA complexes

Class I	*Hydrogen bonded to both protein and DNA simultaneously*	
Category 1:	Acceptor of protein-Water-Acceptor of DNA	(A_P-W-A_D)
Category 2:	Acceptor of protein-Water-Donor of DNA	(A_P-W-D_D)
Category 3:	Donor of protein-Water-Acceptor of DNA	(D_P-W-A_D)
Category 4:	Donor of protein-Water-Donor of DNA	(D_P-W-D_D)
Class II	*Hydrogen bonded to either protein or DNA*	
Category 1:	Acceptor of protein-Water	(A_P-W)
Category 2:	Donor of protein-Water	(D_P-W)
Category 3:	Acceptor of DNA-Water	(A_D-W)
Category 4:	Donor of DNA-Water	(D_D-W)
Class III	*Proximal to hydrophobic atoms of either protein or DNA*	
Category 1:	Hydrophobic atom of protein-Water	($H\phi_P$-W)
Category 2:	Hydrophobic atom of DNA-Water	($H\phi_D$-W)
Class IV	*Hydrogen bonded to other water molecules*	(W-W)

to be to reduce the electrostatic repulsions between the two acceptor or donor atoms (Figure 2(a)). Category 4 water molecules are fewer because of the smaller number of accessible hydrogen bond donor atoms on the DNA and so are category 2 water molecules. Category 3 water molecules can be considered to be mediating hydrogen bond interactions when packing requirements do not allow direct contacts between protein and DNA (Figure 2(b)). These water molecules appear to be acting as extended side-chains of protein to reach out to DNA or *vice versa*. However, the frequency of their occurrence in the crystallographic water molecules is very low (1.8 %).

The majority of the water molecules (76 %) fall under class II, suggesting that the main role of water is to solvate the protein and the DNA atoms at the interface than to mediate a hydrogen bond interaction. They may be playing a role in filling the void spaces at the protein-DNA interface. Also, the contribution of acceptor atoms (categories 1 and 3) is quite large, indicating clearly a dielectric screening role for these water molecules. Even if some of these water molecules turn out to be cations,[55–58] the inferences remain valid.

Our analysis, not surprisingly, shows that there are not many water molecules under class III. Crystal studies might not report them, as they might be more mobile and/or disordered and thus less easy to observe crystallographically, or they might have been lost to bulk upon complexation. Results from molecular dynamics simulations (presented below and in Table 5) support these statistics and inferences, and suggest that not many water molecules stay close to hydrophobic groups at protein-DNA interfaces. Class IV water molecules are not of immediate interest to protein-DNA recognition at the level of a primary contact analysis. They may be important, however, in analyzing higher levels of water organization such as clusters and ordered networks around biomolecules.

In summary, the following trend on the number of water molecules emerges from Table 3: class II $\gg$ IV $\gg$ I > III. Also, around 92 % of the total PDB water molecules fall in either class II or IV, which have no direct role in mediating protein-DNA contacts or in extended hydrogen bonds. In class I, the trend for the number of water molecules is observed to be: A_P-W-$A_D \gg D_P$-W-$A_D \gg A_P$-W-$D_D > D_P$-W-D_D. Category 3 water molecules in class I (i.e. D_P-W-A_D) can be thought of as hydrogen bond linkers. Though protein and DNA could form a hydrogen bond directly, distances permitting, these water molecules seem to be playing a mediatory role. Note that the number of these water molecules is small.

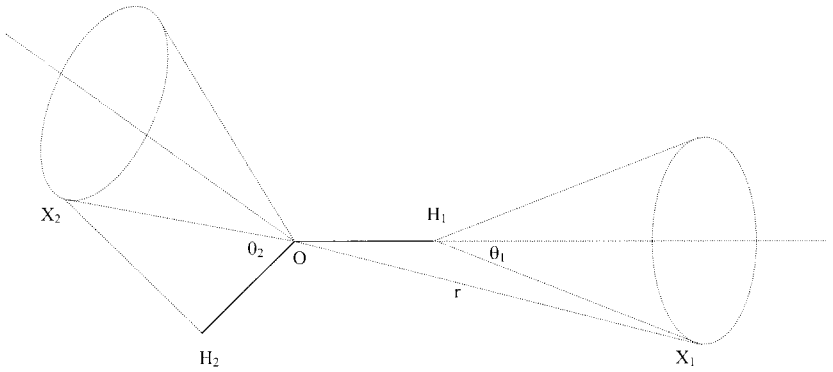

Figure 1. Structural criterion used for identifying hydrogen bonds. $H_{(1)}$, $H_{(2)}$ and O denote atoms of the water under examination. $X_{(1)}$, $X_{(2)}$ denote any atom on protein, DNA or another water molecule.

Table 2. Number of water molecules reported in the crystal structures of protein-DNA complexes examined

Study no.	PDB code	Number reported	Study no.	PDB code	Number reported
1	1A02	88	56	1HLO	18
2	1A0A	80	57	1HWT	58
3	1A6Y	234	58	1IF1	10
4	1A73	782	59	1IHF	384
5	1AAY	148	60	1JMC	90
6	1AIS	288	61	1LAU	161
7	1AKH	50	62	1LMB	140
8	1AM9	287	63	1MDY	25
9	1AOI	13	64	1MEY	132
10	1AWC	46	65	1MHD	24
11	1AZP	132	66	1MNM	53
12	1B01	26	67	1NFK	317
13	1B3T	182	68	1OTC	23
14	1B94	350	69	1PAR	45
15	1BC7	137	70	1PDN	13
16	1BER	466	71	1PER	40
17	1BF4	114	72	1PUE	143
18	1BHM	215	73	1QAJ	406
19	1BL0	144	74	1QBJ	244
20	1BNZ	99	75	1QP9	96
21	1BP7	7	76	1QRV	116
22	1BPY	321	77	1QUM	315
23	1BVO	36	78	1RCN	68
24	1BY4	230	79	1RPE	36
25	1CA5	102	80	1SKN	28
26	1CDW	280	81	1SVC	126
27	1CEZ	471	82	1TC3	49
28	1CF7	75	83	1TRO	316
29	1CIT	38	84	1TSR	384
30	1CKQ	151	85	1UBD	87
31	1CKT	74	86	1VAS	143
32	1CLQ	221	87	1WET	92
33	1CMA	36	88	1XBR	286
34	1CW0	206	89	1YRN	58
35	1D3U	180	90	1ZME	157
36	1D66	51	91	2BDP	468
37	1DEW	444	92	2BOP	242
38	1DH3	11	93	2BPA	178
39	1DMU	255	94	2DGC	92
40	1DP7	124	95	2DRP	57
41	1DSZ	338	96	2GLI	44
42	1DUX	225	97	2IRF	397
43	1ECR	49	98	2NLL	236
44	1EFA	62	99	2OR1	44
45	1EGW	656	100	2RAM	128
46	1EQZ	349	101	2UP1	144
47	1ERI	61	102	3CRO	25
48	1F3I	434	103	3HDD	53
49	1F66	325	104	3KTQ	149
50	1FJL	185	105	4CRX	473
51	1GDT	29	106	6CRO	7
52	1GLU	41	107	6MHT	148
53	1HAP	151	108	6PAX	84
54	1HCQ	158	109	9ANT	38
55	1HCR	16		**Total**	**17,963**

Molecular Dynamics Analysis of Interfacial Water Molecules

To address issues related to resolution in the reported crystal structures, the existence of multiple minima for water locations and orientations, and to ensure the general validity of the above results at ambient temperature, molecular dynamics (MD) simulations were performed on 35 of these complexes with solvent and counterions. Table 4 shows the PDB codes of these complexes and the number of water molecules considered in each case. The complexes were prepared for molecular dynamics simulations as follows: the hydrogen-minimized structures from above were taken as the starting point. Enough counterions were added to the protein-DNA complex to ensure electroneutrality of the system, followed by the addition of a 7.0 Å shell of water molecules. The system was then subjected to energy minimization. First, 500 steps of water minimization was carried out with a 25 kcal/mol restraint on the complex and ions. This was followed by 500 steps of minimization where the restraints were relaxed gradu-

Table 3. Number of water molecules observed under each class and category in the crystal structures of protein-DNA complexes

Class/Category			Number	Percentage
Class I		1. A_P-W-A_D	562	3.13
		2. A_P-W-D_D	92	0.51
		3. D_P-W-A_D	332	1.85
		4. D_P-W-D_D	4	0.02
		Total	990	5.51
Class II	Protein-water	1. A_P-W	7458	41.52
		2. D_P-W	2166	12.06
		Total	9624	53.58
	DNA-water	3. A_D-W	3230	17.98
		4. D_D-W	799	4.45
		Total	4029	22.43
		Total	13,653	76.01
Class III		1. Hϕ_P-W	318	1.77
		2. Hϕ_D-W	54	0.30
		Total	372	2.07
Class IV		W-W	2948	16.41
Total			17,963	100.00

ally with the restraints on the ions being released faster than on the complex. Then a 100-step all-atom minimization was carried out. Since we were interested only in the water molecules around the complex, the complex was subjected to a restraint of 1000 kcal/mol and SHAKE applied to all bonds for computational expediency. The system was then heated to 300 K over 35 ps with a restraint of 25 kcal/mol on the water molecules and ions that was relaxed gradually as above. A time step of 1 fs was used throughout. The system was then equilibrated for 15 ps followed by a 100 ps data collection phase. This procedure was repeated for each

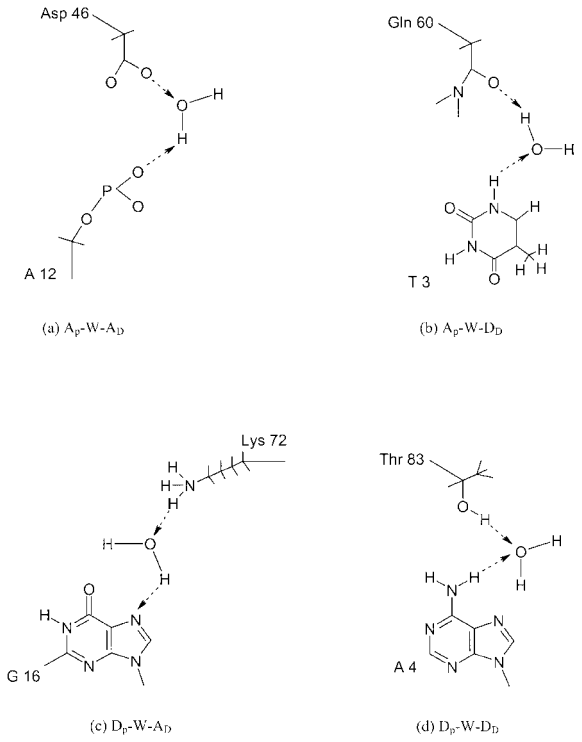

Figure 2. Illustrations of different categories of class I water molecules as emerging from molecular dynamics simulation studies on the trp repressor-operator complex.

Table 4. Number of water molecules considered in the structures of protein-DNA complexes studied *via* molecular dynamics simulations

Study no.	PDB code	Number
1	1A0A	5373
2	1AKH	4698
3	1B94	7786
4	1BER	8829
5	1BHM	6714
6	1BRN	4713
7	1CA5	3228
8	1CMA	4461
9	1A02	7836
10	1BC7	3519
11	1CDW	6099
12	1CKQ	8418
13	1DUX	6189
14	1ECR	5664
15	1GLU	5001
16	1LAU	4305
17	1MNM	7809
18	1PER	3993
19	1RPE	4173
20	1RVA	7458
21	1TSR	11,367
22	1UBD	4560
23	1VAS	4506
24	1ZQC	4549
25	2BOP	4830
26	2HAP	5910
27	2NLL	7110
28	2OR1	4086
29	2UP1	4878
30	3CRO	4389
31	3PVI	5802
32	1ERI	7455
33	1NFK	11,283
34	1TRO	6129
35	1LMB	4911
	Total	**210,031**

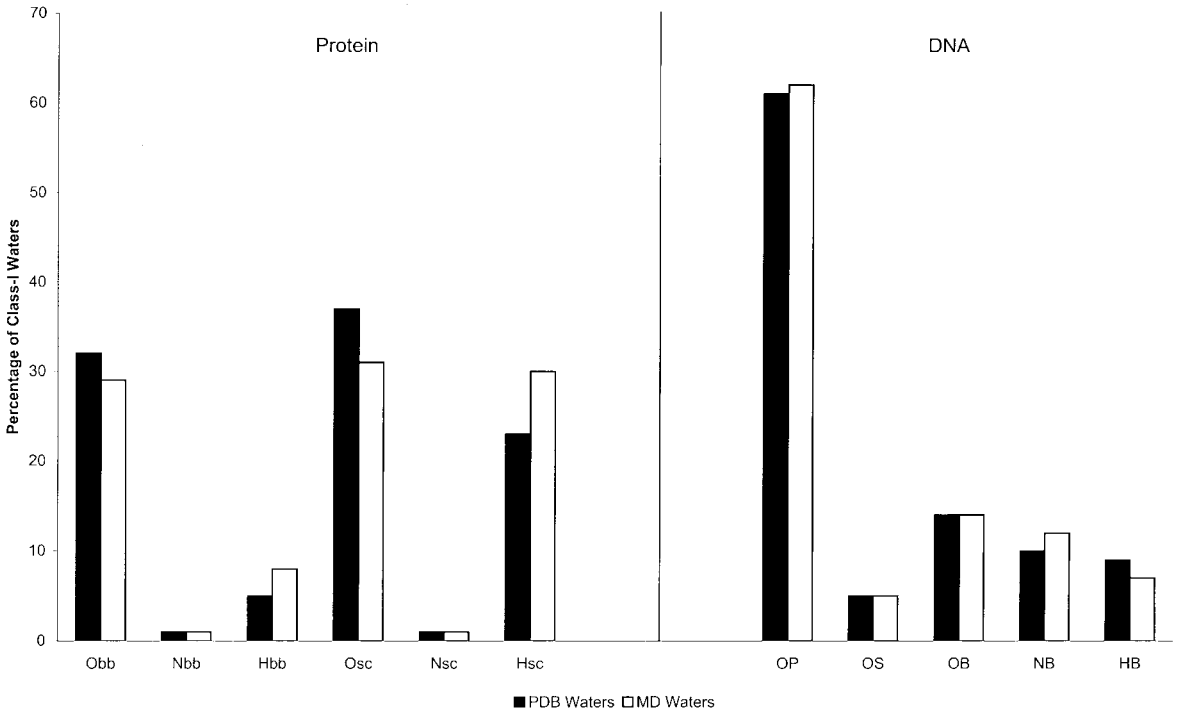

Figure 3. An atom-wise contact analysis of interfacial water molecules. The percentage of class I water molecules is shown against each acceptor and donor atom on protein and DNA. Obb, Nbb, Hbb and Osc, Nsc, Hsc denote oxygen, nitrogen and hydrogen atoms of the backbone and side-chains in proteins, respectively. OP and OS refer to oxygen atoms of phosphate and the sugar moiety of DNA. OB, NB and HB refer to oxygen, nitrogen and hydrogen atoms of DNA bases.

of the 35 complexes under investigation (Table 4). Typical run-time on a four-processor R10000 Origin 200 was about 46 hours for an average-sized system (1NFK, ~12,000 atoms). The simulation run lengths could be an issue in water analysis. The thermal motion of water may be regarded as of two types:[59] rapid oscillations about temporary equilibrium positions (characterized by a relaxation time τ_V); and slower displacements of equilibrium positions (characterized by a relaxation time τ_D). For liquid water, τ_V is about 0.2 ps and τ_D is close to 10 ps. Also, the dielectric relaxation time of water (τ_d) related to the rotational motion, is in the same range as τ_D. For ordered water, these

Table 5. Simulation averages for the number of water molecules observed under each class and category in the 35 protein-DNA complexes studied *via* molecular dynamics

Class/Category			Number	Percentage
Class I		1. A_P-W-A_D	1029	0.49
		2. A_P-W-D_D	126	0.06
		3. D_P-W-A_D	651	0.31
		4. D_P-W-D_D	21	0.01
		Total	1827	0.87
Class II	Protein-water	1. A_P-W	49,168	23.41
		2. D_P-W	13400	6.38
		Total	62,568	29.79
	DNA-water	3. A_D-W	21,381	10.18
		4. D_D-W	1764	0.84
		Total	23,145	11.02
		Total	85,714	40.81
Class III		1. Hϕ_P-W	588	0.28
		2. Hϕ_D-W	147	0.07
		Total	735	0.35
Class IV		W-W	121,755	57.97
Total			210,031	100.00

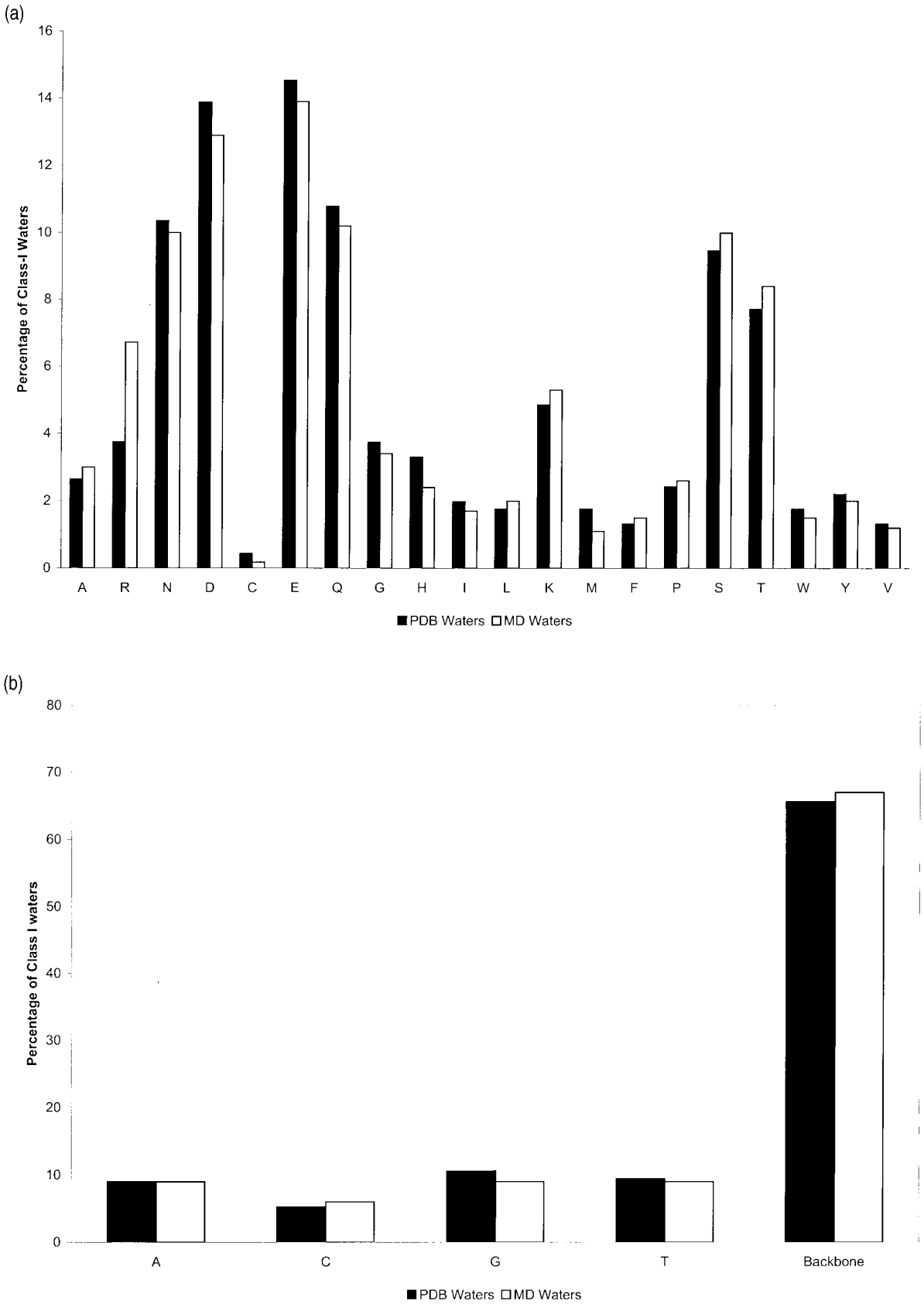

Figure 4. A residue-wise contact analysis of interfacial water molecules. A histogram view of the percentage of class I water molecules assigned to (a) each amino acid residue (denoted by the single-letter code) and (b) each constituent base and backbone of DNA.

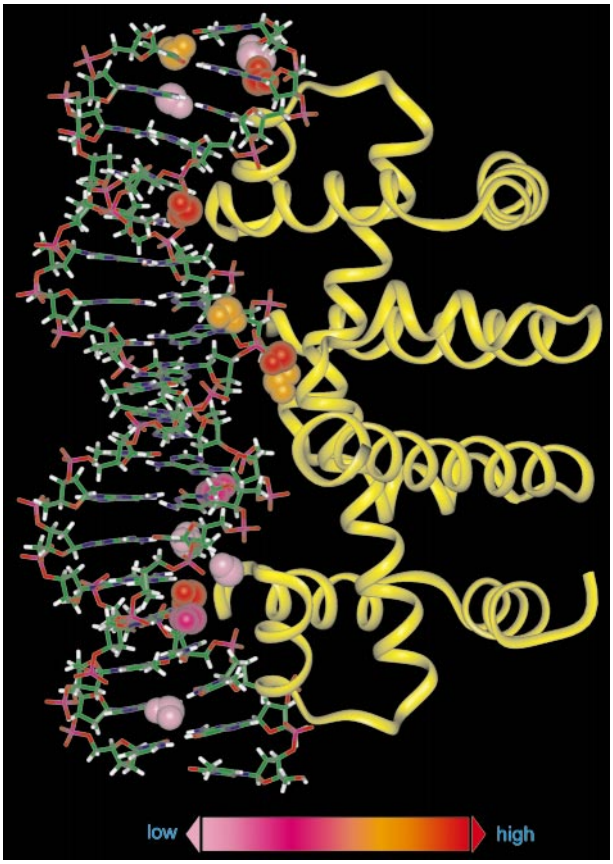

Figure 5. Class I water molecules (those contacting both protein and DNA) in the trp repressor-operator DNA complex colored by the magnitude of the electrostatic field.

relaxation times may be longer. The run lengths undertaken here are at least ten times longer than τ_D. However, as a check of convergence, five of these MDs were extended to 200 ps and one to 500 ps, and the results showed no significant variation. These results are presented in Table 5. Although the number of water molecules considered in the molecular dynamics simulations is an order of magnitude more than in crystal structure analyses (Table 3), the results are essentially identical and the inferences are unchanged.

Atom-wise contact analysis

An atom-wise contact analysis of interfacial water molecules is shown in Figure 3. That water molecules interact predominantly (to the extent of 90 % in DNA and 71 % in protein) with acceptor atoms of both protein and DNA is apparent. This provides further evidence that interfacial water appears primarily to reduce the electrostatic repulsions between acceptor atoms. Also, among the atoms of the protein, backbone oxygen atoms and the oxygen atoms of the side-chains are the main contributors. Whereas in the case of DNA, phosphate oxygen atoms are the principal contributors with the oxygen as well as the nitrogen of the bases participating equally. The accessible donor atoms on the DNA are very few, and those on the protein make favorable interactions with DNA anyway and hence do not appear to prefer solvation at the interface to any significant extent. This again is a pointer to the electrostatic role of the intervening water molecules.

Residue-wise contact analysis

A residue-wise contact analysis of interfacial water molecules is shown in Figure 4. It is interesting that adenine, guanine and thymine of DNA are relatively more hydrated than cytosine, while the backbone hydration dominates as expected. On the protein side, Glu and Asp are the main residues interacting with water. Of course, the contribution of Ser, Thr, Asn and Gln is considerable. This provides an alternative view in favour of the electrostatic buffering action of water. Data on the number of water molecules making hydrogen bonds to acceptor atoms of protein, donor atoms of protein, acceptor atoms of DNA, etc. for each of the 109 protein-DNA complexes is provided in the Supplementary Material. The data presented leads us to the conclusion that interfacial water molecules act mainly to decrease the electrostatic repulsions between the electronegative atoms in protein-DNA complexes.

Electrostatic Fields at Interfacial Waters

To further appreciate the location and orientation of interfacial water molecules, we computed the electrostatic fields at the water sites using finite-difference non-linear Poisson-Boltzmann (FDNLPB) methodology.[60–62] A grid size of 201^3 was used with a three-step focusing, where the extent of the box fill was varied from 40-80 %, leading to a final grid resolution of 2.5 grids/Å. A probe radius of 1.4 Å was used for the dielectric map and the interior and exterior dielectrics were set at 2 and 80. AMBER[63] charges and radii were used with a salt concentration of 0.145 M.

The magnitude of the field at the water molecules was computed by averaging the field at the oxygen atom and the two hydrogen atoms. Figure 5 shows the fields at class I water molecules in the trp repressor-operator DNA complex. Figure 6(a) shows a water molecule from the same system in a high field making contacts with a phosphate oxygen atom and the side-chain of an aspartate residue. Figure 6(b) shows a water molecule in low field, in the groove between N-7 of adenine and an arginine. For free DNA, the maximum fields are at the phosphate groups while the potential maxima are concentrated in the grooves.[61,64] This indicates that water molecules should prefer the backbone, which shows up in our analysis. The field around DNA in protein-DNA complexes is affected by the local environ-

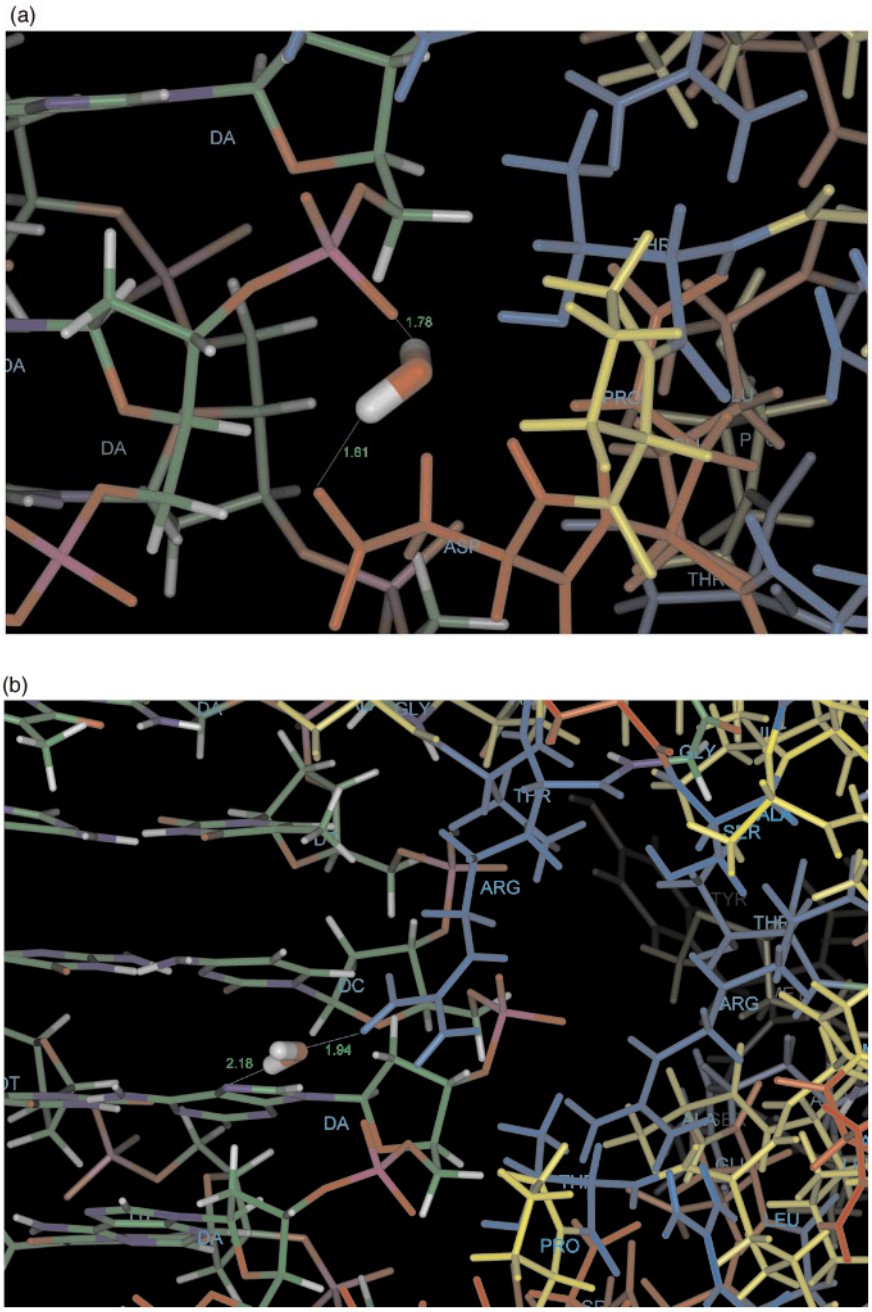

Figure 6. A closer view of a class I water molecule of the trp-repressor-operator DNA complex in (a) high field (H-bonded to a phosphate oxygen atom and Asp) and (b) in low field (between Arg and N-7 of A). The DNA is colored according to atom and the acidic, basic and neutral residues of the protein are shaded red, blue and yellow, respectively.

ment. In general, fields at water sites near the phosphate groups in protein-DNA complexes are higher than in the grooves. However, fields in the grooves are increased drastically due to the occurrence of clusters of charged amino acid residues in some cases. Higher fields imply stronger force, attractive between opposite charges and repulsive between two acceptor atoms. This observation once again brings out the buffering action of such interfacial water molecules that occurs between acceptor atoms.

Densities in Protein-DNA Complexes

Pursuing the idea that density gradients are not sustainable for a system at thermal equilibrium, and that local density variations within the solvated macromolecular system could lead to transport of matter, manifested *via* conformational transitions/structural rearrangements (interactions permitting), or diffusion of solvent, we computed the densities in the structures of protein-DNA complexes emerging from the MD trajectories. The

occupied volume of each protein-DNA complex was determined by a Monte Carlo procedure within a box bounding the extents of the system after adding 1.8 Å to the radius of each atom. The results are reported in Table 6. Local densities in protein-DNA complexes in the absence of water are about 0.8 g/ml but 1.0 g/ml with trapped water molecules, the same as that of bulk solvent. The hypothesis is that during complexation in a solvent medium, large departures from bulk densities are not to be expected at the interface. The molecular dynamics simulation data on all the 35 complexes supports this view. Table 6 reports the densities with and without interfacial water molecules for the complexes as well as protein and DNA individually.

Two concerns arise in any structural analysis. One is the dependence of the inference on the data size. Second is the thermodynamic relevance. About the data on the number of water molecules reported in crystal studies (Table 1), certainly there is no uniformity in either the resolution achieved or the criterion adopted. Regardless of this, there is no reason to believe that crystal studies have chosen to report only the water molecules that are sandwiched between two acceptor atoms, which is the main observation here. These water molecules constitute an overridingly large proportion of the reported water molecules with an unambiguously identifiable role in electrostatic screening. Moreover, the molecular dynamics studies, which consider about 0.2 million water molecules (Tables 4 and 5), corroborate our findings. On the energetic front, both experiment and theory clearly indicate that water release from non-polar atoms makes a favorable contribution to the binding free energy of protein-DNA complexes.[65–74] Work on the contribution of interfacial water molecules to binding free energies is in progress. We find the preferential location of water molecules between protein and DNA atoms, which ought to repel each other with the attendant solvation energetics that has but to be favorable, a new view on the role of interfacial waters.

Conclusions

A structural analysis based on the chemical identity of macromolecular atoms proximal to the interfacial water molecules observed in the crystal studies of 109 protein-DNA complexes leads unequivocally to the inference that a great majority

Table 6. Densities (in g/ml) of protein-DNA complexes with and without interfacial water molecules

		Complex		Protein		DNA	
Study no.	PDB code	Without water	With water	Without water	With water	Without water	With water
1	1A02	0.83	1.02	0.78	0.96	0.86	0.96
2	1A0A	0.80	1.01	0.75	0.98	0.85	1.09
3	1AKH	0.84	1.05	0.76	0.95	0.85	1.00
4	1B94	0.85	1.05	0.84	0.99	0.86	0.93
5	1BC7	0.82	1.06	0.79	0.97	0.81	0.99
6	1BER	0.81	1.04	0.81	0.98	0.84	1.07
7	1BHM	0.81	1.03	0.81	0.97	0.84	1.08
8	1BRN	0.83	1.05	0.82	1.02	0.83	0.91
9	1CA5	0.78	1.06	0.76	0.98	0.79	0.98
10	1CDW	0.74	1.07	0.77	1.00	0.76	1.09
11	1CKQ	0.81	1.03	0.84	0.98	0.82	0.93
12	1CMA	0.79	1.01	0.79	0.96	0.81	0.97
13	1DUX	0.80	1.08	0.82	0.99	0.81	1.04
14	1ECR	0.80	0.99	0.79	0.95	0.83	0.93
15	1ERI	0.80	1.02	0.85	0.97	0.81	0.92
16	1GLU	0.82	1.05	0.77	0.95	0.83	1.00
17	1LAU	0.83	1.00	0.83	0.99	0.71	0.92
18	1LMB	0.78	1.06	0.79	0.96	0.82	1.01
19	1MNM	0.83	1.01	0.77	0.95	0.84	0.97
20	1NFK	0.82	1.02	0.81	0.99	0.83	0.94
21	1PER	0.78	1.07	0.82	0.95	0.80	0.98
22	1RPE	0.82	1.06	0.81	0.96	0.85	0.98
23	1RVA	0.80	1.02	0.84	0.97	0.82	0.95
24	1TRO	0.75	1.06	0.77	0.98	0.78	1.06
25	1TSR	0.81	1.05	0.83	1.03	0.83	1.01
26	1UBD	0.82	1.07	0.76	0.95	0.85	1.04
27	1VAS	0.76	1.07	0.79	0.97	0.74	1.05
28	1ZQC	0.80	1.02	0.80	0.99	0.81	0.97
29	2BOP	0.81	1.07	0.82	0.98	0.83	1.02
30	2HAP	0.82	1.04	0.76	0.95	0.83	1.03
31	2NLL	0.77	1.08	0.76	1.04	0.82	1.01
32	2ORI	0.81	1.06	0.80	0.96	0.84	1.00
33	2UP1	0.74	1.05	0.79	1.00	0.63	0.95
34	3CRO	0.81	1.05	0.79	0.95	0.83	0.99
35	3PVI	0.79	1.02	0.81	0.96	0.81	0.94
	Average	0.80	1.04	0.80	0.98	0.81	0.99

of water molecules serve to buffer electrostatic repulsions between electronegative atoms of the protein and the DNA. About 2% of the observed water molecules act as linkers to form extended hydrogen bonds between the protein and the DNA, compensating for the lack of a direct hydrogen bond.

Acknowledgements

Funding from the Indo-French Center for the Promotion of Advanced Research (IFCPAR) is gratefully acknowledged. The authors thank Ms Preeti Misra for help in preparing the complexes, and Ms Parul Kalra for helpful discussions on the project.

References

1. Westhof, E. & Beveridge, D. L. (1990). Hydration of nucleic acids. In *Water Science Reviews 5* (Franks, F., ed.), pp. 24-136, Cambridge University Press, Cambridge.
2. Westhof, E. (1993). Structural water bridges in nucleic acids. In *Water and Biological Macromolecules* (Westhof, E., ed.), pp. 226-243, CRC Press, Boca Raton, FL.
3. Berman, H. M. & Schneider, B. (1999). Nucleic acid hydration. In *Handbook of Nucleic Acid Structure* (Neidle, S., ed.), pp. 295-312, Oxford University Press, Oxford.
4. Luisi, B. (1995). DNA-protein interactions at high resolution. In *DNA-Protein Structural Interactions* (Lilley, D. M. J., ed.), pp. 1-48, IRL Press, Oxford.
5. Berman, H. M. (1991). Hydration of DNA. *Curr. Opin. Struct. Biol.* **1**, 423-427.
6. Berman, H. M. (1994). Hydration of DNA: take 2. *Curr. Opin. Struct. Biol.* **4**, 345-350.
7. Kochoyan, M. & Leroy, J. L. (1995). Hydration and solution structure of nucleic acids. *Curr. Opin. Struct. Biol.* **5**, 329-333.
8. Pettitt, B. M., Makarov, V. A. & Andrews, B. K. (1998). Protein hydration density: theory, simulations and crystallography. *Curr. Opin. Struct. Biol.* **8**, 218-221.
9. Thanki, N., Thornton, J. M. & Goodfellow, J. M. (1988). Distributions of water round amino acid residues in proteins. *J. Mol. Biol.* **202**, 637-657.
10. Williams, M. A., Goodfellow, J. M. & Thornton, J. M. (1994). Buried waters and internal cavities in monomeric proteins. *Protein Sci.* **3**, 1224-1235.
11. Karplus, P. A. & Faerman, C. (1994). Ordered water molecules in macromolecular structure. *Curr. Opin. Struct. Biol.* **4**, 70-776.
12. Lounnas, V., Pettitt, B. M. & Phillips, G. N., Jr (1994). A global model of the protein-solvent interface. *Biophys. J.* **66**, 601-614.
13. Feig, M. & Montgomery, P. (1998). Crystallographic water sites from a theoretical perspective. *Structure,* **6**, 1351-1354.
14. Rupley, J. A. & Careri, G. (1991). Protein hydration and function. *Advan. Protein Chem.* **41**, 37-172.
15. Kuntz, I. D., Meng, E. C. & Shoichet, B. K. (1994). Structure-based molecular design. *Acct. Chem. Res.* **27**, 117-123.
16. Huang, K., Anderson, L. W., Laskowski, M., Jr & James, M. N. G. (1995). Water molecules participate in proteinase-inhibitor interactions: crystal structure of Leu 18, Ala 18 and Gly 18 variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. *Protein Sci.* **4**, 1985-1987.
17. Bhat, T. N. & Poljak, R. J. (1994). Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl Acad. Sci. USA,* **91**, 1089-1093.
18. Teeter, M. M. (1991). Water-protein interactions: theory and experiment. *Annu. Rev. Biophys. Biophys. Chem.* **202**, 577-600.
19. Lam, P. Y. S., Jadhav, P. K., Eyermann, C. J., Hodje, C. N., Ru, Y., Bacheler, L. T. *et al.* (1994). Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science,* **263**, 380-384.
20. Quiocho, F. A., Wilson, D. K. & Vyas, N. K. (1989). Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature,* **340**, 404-407.
21. Stillman, T. J., Baker, P. J., Britton, K. L. & Rice, D. W. (1993). Conformational flexibility in glutamate dehydrogenase: role of water in substrate recognition and catalysis. *J. Mol. Biol.* **234**, 1131-1139.
22. Hidong, K. & Lipscomb, W. N. (1990). Crystal structure of the complex of carboxypeptidase A with a strongly bound phosphonate in a new crystalline form: comparison with structures of other complexes. *Biochemistry,* **29**, 5546-5555.
23. Drew, H. W. & Dickerson, R. E. (1981). Structure of a *B*-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* **151**, 535-556.
24. Narayana, N., Ginell, S. L., Russu, I. & Berman, H. M. (1991). Crystal and molecular structure of a DNA fragment: d(CGTGAATTCACG). *Biochemistry,* **30**, 4449-4455.
25. Quintana, J. R., Grzeskowiak, K., Yanagi, K. & Dickerson, R. E. (1992). Structure of a *B*-DNA decamer with a central T-A step C-G-A-T-T-A-A-T-C-G. *J. Mol. Biol.* **225**, 379-395.
26. Lipanov, A., Kopka, M. L., Kaczor, G., Rzeskowiak, M., Quintana, J. & Dickerson, R. E. (1993). Structure of the *B*-DNA decamer C-C-A-A-C-I-T-T-G-G in two different space groups: conformational flexibility of *B*-DNA. *Biochemistry,* **32**, 1373-1380.
27. Mezei, M. & Beveridge, D. L. (1986). Structural chemistry of biomolecular hydration *via* computer simulation: the proximity criterion. *Methods Enzymol.* **127**, 21-47.
28. Beveridge, D. L., Swaminathan, S., Ravishankar, G., Whitka, J. M., Srinivasan, J., Prevost, C. *et al.* (1993). Molecular dynamics simulations on the hydration, structure and motions of DNA oligomers. In *Water and Biological Macromolecules* (Westhof, E., ed.), pp. 165-225, CRC Press, Boca Raton, FL.
29. Jayaram, B. & Beveridge, D. L. (1996). Modeling DNA in aqueous solutions: theoretical and computer simulation studies on the ion atmosphere of DNA. *Annu. Rev. Biophys. Biomol. Struct.* **25**, 367-394.
30. Rudnicki, W. R. & Pettitt, B. M. (1996). Modeling the DNA-solvent interface. *Biopolymers,* **4**, 107-119.
31. Makarov, V. A., Feig, M., Andrews, B. K. & Pettit, B. M. (1998). Diffusion of solvent around biomolecular solutes: a molecular dynamics simulation study. *Biophys. J.* **75**, 150-158.

32. Chevrier, B., Dock, A. C., Hartmann, B., Leng, M., Moras, D., Thoung, M. & Westhof, E. (1986). Solvation of the left-handed hexamer d(5BrCG5BrCG5BrCG) in crystals grown at room temperature. *J. Mol. Biol.* **188**, 707-719.
33. Gesner, R. V., Quigley, G. J. & Egli, M. (1994). Comparative studies of high resolution *Z*-DNA crystal structures. Part 1: Common hydration patterns of alternating dC-dG. *J. Mol. Biol.* **236**, 1154-1168.
34. Schneider, B., Patel, K. & Berman, H. M. (1998). Hydration of the phosphate group in double helical DNA. *Biophys. J.* **75**, 2422-2434.
35. Jayaram, B., Sprous, D., Young, M. A. & Beveridge, D. L. (1998). Free energy analysis of the conformational preferences of *A* and *B* forms of DNA in solution. *J. Am. Chem. Soc.* **120**, 10629-10633.
36. Rould, M. A., Perona, J. J., Soll, D. & Steitz, T. A. (1989). Structure of *E. coli* glutaminyl tRNA synthetase complexed with tRNAGln and ATP at 2.8 Å resolution: implications for tRNA discrimination. *Science,* **246**, 1135.
37. Pullman, A., Pullman, B. & Berthod, H. (1978). An SCF ab initio investigation of the 'through-water' interaction of the phosphate anion with the Na+ cation. *Theor. Chim. Acta,* **47**, 175.
38. Kopka, M. L., Fratini, A. V., Drew, H. R. & Dickerson, R. E. (1983). Ordered water structure around a *B*-DNA dodecamer. a quantitative study. *J. Mol. Biol.* **163**, 129-146.
39. Saenger, W. (1987). Structure and dynamics of water surrounding macromolecules. *Annu. Rev. Biophys. Biomol. Struct.* **16**, 93-114.
40. Schwabe, J. W. R. (1997). The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.* **7**, 126-134.
41. Janin, J. (1999). Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure,* **7**, R277-R279.
42. Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q. *et al*. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. *Nature,* **335**, 321-329.
43. Shakked, Z., Guzikevich-Guerstein, G., Frollow, F., Rabinovich, D., Joachimiak, A. & Sigler, P. B. (1994). Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature,* **368**, 469-473.
44. Wilson, D. S., Geunther, B., Desplan, C. & Kurian, J. (1995). High resolution crystal structure of a paired (pax) class homeodomain dimer on DNA. *Cell,* **82**, 709-719.
45. Kosztin, D., Bishop, T. C. & Schulten, K. (1997). Binding of the estrogen receptor to DNA. The role of waters. *Biophys. J.* **73**, 557-570.
46. Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). Sequence specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA,* **73**, 804-808.
47. Woda, J., Schneider, B., Patel, K., Mistry, K. & Berman, H. M. (1998). An analysis of the relationship between hydration and protein-DNA interactions. *Biophys. J.* **75**, 2170-2177.
48. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature,* **365**, 512-520.
49. Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature,* **365**, 520-527.
50. Nickolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G. & Burley, S. K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl Acad. Sci. USA,* **93**, 4862-4867.
51. Dunitz, J. D. (1994). The entropic cost of bound water in crystals and biomolecules. *Science,* **264**, 670.
52. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al*. (2000). The protein data bank. *Nucl. Acids Res.* **28**, 235-242.
53. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. L. *et al*. (1999). *AMBER 6*, University of California, San Francisco.
54. Schneider, B., Cohen, D. & Berman, H. M. (1992). Hydration of DNA bases: analysis of crystallographic data. *Biopolymers,* **32**, 725-750.
55. Hud, N. V. & Feigon, J. (1997). Localization of divalent metal ions in the minor groove of DNA *A*-tracts. *J. Am. Chem. Soc.* **119**, 5756-5757.
56. Young, M. A., Jayaram, B. & Beveridge, D. L. (1997). Intrusion of counterions into the spine of hydration in the minor groove of *B*-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.* **119**, 59-69.
57. Hud, N. V., Sklenar, V. & Feigon, J. (1999). Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA *A*-tract bending. *J. Mol. Biol.* **286**, 651-660.
58. Denisov, V. P. & Halle, B. (2000). Sequence-specific binding of counterions to *B*-DNA. *Proc. Natl Acad. Sci. USA,* **97**, 629-633.
59. Eisenberg, D. & Kauzmann, W. (1969). *The Structure and Properties of Water*, chapt. 4, Oxford University Press, Oxford.
60. Gilson, M., Sharp, K. A. & Honig, B. (1988). Calculating the electrostatic potential of molecules in solultion: method and error assessment. *J. Comput. Chem.* **9**, 327-335.
61. Jayaram, B., Sharp, K. A. & Honig, B. (1989). The electrostatic potential of *B*-DNA. *Biopolymers,* **28**, 975-993.
62. Honig, B. H. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science,* **268**, 1144-1149.
63. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr, Ferguson, D. M. *et al*. (1995). A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **117**, 5179-5197.
64. Pullman, B. (1983). Electrostatics of polymorphic DNA. *J. Biomol. Struct. Dynam.* **1**, 773-794.
65. Ha, J.-H., Spolar, R. S. & Record, M. T. (1989). Role of hydrophobic effect in the stability of site specific protein-DNA complexes. *J. Mol. Biol.* **209**, 801-816.
66. Jayaram, B., DiCapua, F. M. & Beveridge, D. L. (1991). A theoretical study of polyelectrolyte effects in protein-DNA interactions: Monte Carlo free energy simulations on the ion atmosphere contribution to the thermodynamics of λ repressor operator complex formation. *J. Am. Chem. Soc.* **113**, 5211-5215.
67. Spolar, R. S. & Record, M. T., Jr (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science,* **263**, 777-784.
68. Jen-Jacobsons, L. (1995). Structural perturbation approaches to thermodynamics of site specific protein-DNA interactions. *Method Enzymol.* **259**, 305-344.

69. Jayaram, B. (1996). Some energetic and kinetic aspects of protein-DNA interactions: a theoretical study on the λ repressor-operator complex. *J. Biol. Struct. Dynam. Proceedings of the Ninth Conversation in Biomolecular Stereodynamics* (Sarma, R. H. & Sarma, M. H., eds), vol. 1, pp. 109-120, Adenine Press, New York.
70. Jayaram, B., Das, A. & Aneja, N. (1996). Energetic and kinetic aspects of macromolecular association: a computational study of λ repressor-operator complexation. *J. Mol. Struct. (TheoChem),* **361**, 249-258.
71. Nadassy, K., Wodak, S. J. & Janin, J. (1999). Structural features of protein-nucleic acid recognition sites. *Biochemistry*, **38**, 1999-2017.
72. Jayaram, B., McConnell, K. J., Dixit, S. B. & Beveridge, D. L. (1999). Free energy analysis of protein-DNA binding: the *Eco*RI endonuclease-DNA complex. *J. Comp. Phys.* **151**, 333-357.
73. Sheinerman, F. B., Norel, R. & Honig, B. (2000). Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* **10**, 153-159.
74. Jayaram, B., McConnell, K. J., Dixit, S. B., Das, A. & Beveridge, D. L. (2001). Free energy analysis of 40 protein-DNA complexes: a consensus view on the nature of binding at the molecular level. *J. Comput. Chem.* **in the press**.

Edited by B. Honig

(Received 20 November 2000; received in revised form 29 August 2001; accepted 8 October 2001)

http://www.academicpress.com/jmb

Supplementary Material is obtainable on IDEAL