Annu. Rev. Biophys. Biomol. Struct. 2004. 33:343–61
doi: 10.1146/annurev.biophys.33.110502.140414

First published online as a Review in Advance on February 3, 2004

THE ROLE OF WATER IN PROTEIN-DNA RECOGNITION

B. Jayaram and Tarun Jain
Department of Chemistry and Supercomputing Facility for Bioinformatics and Computational Biology, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India; email: bjayaram@chemistry.iitd.ac.in

Key Words hydrogen bonds, water-mediated recognition, electrostatic screening, packing densities, thermodynamic view, protein-DNA interactions

■ **Abstract** Is it by design or by default that water molecules are observed at the interfaces of some protein-DNA complexes? Both experimental and theoretical studies on the thermodynamics of protein-DNA binding overwhelmingly support the extended hydrophobic view that water release from interfaces favors binding. Structural and energy analyses indicate that the waters that remain at the interfaces of protein-DNA complexes ensure liquid-state packing densities, screen the electrostatic repulsions between like charges (which seems to be by design), and in a few cases act as linkers between complementary charges on the biomolecules (which may well be by default). This review presents a survey of the current literature on water in protein-DNA complexes and a critique of various interpretations of the data in the context of the role of water in protein-DNA binding and principles of protein-DNA recognition in general.

CONTENTS

SCOPE

The physiologically relevant B-form of DNA is unstable without water. Regulatory proteins and enzymes that bind to DNA specifically are also associated with solvent water. What becomes of water around DNA and the protein consequent to binding, does it assist the binding, and if so, how? These are the main issues addressed in this review. We start with a brief layout of the current models for protein-DNA recognition, advert to the inferences on water around DNA and proteins, and examine the evidence on the role of water in protein-DNA complexes from structural and thermodynamic perspectives.

CURRENT MODELS FOR PROTEIN-DNA RECOGNITION

How to recognize DNA of specified length and base sequence is the problem that has occupied some of the best minds in structural and molecular biology for over five decades now with no solution as yet (14, 71, 75, 95, 103, 123, 126, 131, 136). Proteins routinely recognize DNA specifically to regulate gene expression. A resolution of this problem has immediate implications in the design of DNA binding drugs and in demystifying noncovalent intermolecular recognition. Advances in X-ray and NMR studies have led to structural characterization of close to 600 protein-DNA complexes (5, 6, 15), which revealed several general features on the mode of protein-DNA interactions, and attempts have been made to classify these structures into different groups on the basis of DNA binding motifs such as helix-turn-helix, leucine zipper, and zinc finger (40, 74). Inspection of protein-DNA complexes at an atomic level reveals that contacts between DNA and protein could be explained in terms of direct hydrogen bonds, water-mediated hydrogen bonds, van der Waals, electrostatic, and hydrophobic contacts. Role of DNA structure and its sequence-dependent structural adaptation to facilitate protein binding, contribution of released DNA-bound counterions, and interfacial waters to the thermodynamics of binding are less well understood. Despite the numerous significant efforts to delineate forces responsible for specific binding, no simple rules for recognition transferable across systems have evolved. DNA recognition is still considered idiosyncratic and motif/case specific (58, 76, 81, 88, 94, 96, 119, 129, 135). We have developed an atlas of the binding free-energy components for over 120 protein-DNA complexes (19, 20, 50–52), investigated the role of waters in protein-DNA recognition (110), and more recently identified a set of supramolecular synthons transferable across all protein-DNA complexes. A summary of the above-cited studies with a special focus on the role of waters in protein-DNA recognition is presented below.

Hydrogen bond formation solved the mysteries associated with secondary structure formation [alpha helices and beta sheets in proteins (97, 98) and double-helix formation from single strands in DNA (138)]. The grooves of DNA are rich in hydrogen bond functional groups (118). The AT base pair provides N3 (H-bond

acceptor) and O2 (acceptor) atoms in the minor groove and N7 (acceptor), NH_2 (6-amino donor), and O4 (acceptor) atoms/groups in the major groove. The GC base pair provides NH_2 (2-amino donor), N3 (acceptor), and O2 (acceptor) in the minor groove and N7 (acceptor), O6 (acceptor), and NH_2 (4-amino donor group) in the major groove. If hydrogen bond formation were the sole mechanism of recognition, this information should suffice to distinguish GC, CG, and AT/TA base pairs from each other because of the order in which the acceptor and donor functional groups appear in the grooves. If this view is supplemented by hydrophobic contacts to thymine methyl groups in the major groove, the AT base pair can be distinguished from TA. The first crystal structure of a protein-DNA complex, e.g., the EcoRI endonuclease-DNA complex (82), showed 12 hydrogen bonds between guanines of DNA and arginines of protein and between adenines of DNA and glutamates and arginines of protein. Systematic mutational studies undertaken (EcoRI* activity investigations) further corroborated the view that hydrogen bonds could explain specific binding in this system. For a short time it seemed that the protein-DNA recognition problem was solved. Subsequent structural studies clearly indicated hydrogen bonds between amino acids and base pairs (such as the preference of glutamine for adenines in some repressor-operator complexes and that of arginine for guanines in zinc fingers) (21, 124), but no clear amino acid base correlation or rule emerged. Also, far greater numbers of intermolecular hydrogen bonds are observed in protein-DNA complexes than in permanent and nonobligate protein-protein complexes (1.4 per 100 $Å^2$ compared with 0.7 and 1.1, respectively) (58). Then came the exception, namely, the *trp* repressor-operator-specific complex, which showed not even one direct hydrogen bond to the bases (93, 121). Several water-mediated contacts between protein and DNA were observed, however. The series of studies following this indicated that hydrogen bond formation [called the direct, or digital code (131)] might be just one aspect of the many facets of protein-DNA recognition. More recently, C─H $\cdots$ O hydrogen bonds have also been advanced as contributing to specific binding (33, 78, 137).

Sequence-dependent DNA structure and its flexibility as an alternative mechanism for recognition by proteins started gaining ground at about this time. Noncontacted bases in 434 repressor-operator complexes contributed to binding, implying a role for base sequence–induced DNA structure (65). Crystal structures of specific complexes of DNA with catabolite gene activator protein (where DNA bends by $\sim 90^\circ$) (116) and integration host factor (where DNA takes a U-turn) (111), and TATA binding protein (29, 60, 61, 91) vividly illustrate the sequence-dependent DNA bendability/deformation or structural adaptation. As not all sequences could assume these shapes without significant energy expense, it became evident that intrinsic flexibility of DNA structure should play a role in recognition (16, 18, 37, 92, 148). This is considered an indirect/analog code for DNA recognition (131). Integration host factor (IHF) for instance is reported to recognize its cognate sites through indirect readout, with DNA twist playing a major role (77). It is apparent that proteins use electrostatic (including hydrogen bonding) and van der Waals

interactions to overcome energy penalty for deformation, with the right sequences presumably requiring less deformation energy.

Embedding basic residues on the protein to recognize the polyanionic DNA appears to be one of the simplest strategies conceivable. Histone proteins in fact do exploit this mode of binding, albeit nonspecific, for some extremely efficient packing of DNA (70). The regulatory proteins and enzymes do not necessarily carry a large net positive charge to facilitate long-range Coulomb attraction, and some DNA binding proteins even carry a net negative charge [e.g., EcoRI endonuclease (−4), *met* repressor (−8)]. Detailed structural analysis of the DNA binding proteins does indicate an overall asymmetry in charge distribution irrespective of the net charge, with the DNA binding side containing an excess positive charge and the farther side an excess negative charge (52). This seems to suggest an anchoring (parking)/orientational role for charged residues, such that the negatively charged side of the protein turns away and the positively charged side faces DNA during binding. Also asymmetric neutralization of phosphates by basic residues is observed to facilitate DNA bending (35). Cation-π interactions add a new phenomenological view to the analysis (144), although theoretical analyses of the binding energetics based on molecular mechanics include the C─H $\cdots$O and cation-π interactions implicitly in the van der Waals and electrostatic terms.

Steric complementarity, static or dynamic, as a primary mode of recognition has a 100-year history in the context of enzyme-substrate interactions (lock-and-key fit and variations thereof) (67). In reference to protein-DNA complexes, the DNA binding proteins have concealed thus far signatures of any such negative imprints of the shapes of DNA grooves or backbone conveying complementarity messages. Detailed examination of the protein structures does reveal that a substructure or a secondary element [alpha helices in a majority of the cases and beta sheets in a few cases (101)] is typically involved in a snug fit in the major groove of DNA, wherein these make good van der Waals contacts. We have previously noted that van der Waals component of the protein-DNA interaction energy correlates well with loss in solvent-accessible surface area of DNA upon protein binding (52). The solvent-accessible area per base pair in canonical B-DNA is around 400 Å^2, and in protein-DNA complexes this varies from 90 to 200 Å^2 (A. Das & B. Jayaram, unpublished data). For an enveloping mode of binding with tight packing, the fraction of DNA surface covered by the protein is high. Endonucleases fall in this category, binding to short oligonucleotide (∼hexamer) sequences. Other DNA binding proteins such as repressors, zinc fingers that cover longer stretches of DNA (14–18 bp) only partly, have a smaller coefficient. The extent of coverage of DNA per base pair by the protein and the strength of van der Waals interactions are correlated to the motif employed.

A role for small ions as regulators of specificity in protein-DNA recognition was proposed a couple of decades ago (109). Salt effects in general oppose binding between two oppositely charged biomolecules owing to ion atmosphere and screening. Ion atmosphere around DNA has an additional dimension. It is now well established that DNA being a polyanion is surrounded by a sheath of counterions,

called condensed counterions, some of which are displaced upon protein binding, and this is thought to contribute favorably to binding through an increased entropy. This ion condensation around naked DNA and their release upon protein/drug binding is known collectively in the literature as polyelectrolyte effect (1, 49, 79, 108). Detailed theoretical analyses of free-energy components to binding indicate that enthalpy losses upon ion release override small entropy gain (52, 63, 85). A structural interpretation of counterion release and its implication to sequence-specific recognition of DNA are far from full comprehension (45, 46, 147).

Water molecules are observed often at the interface between the two biomolecules in many protein-DNA crystal structures, the classic case being the *trp* repressor-operator complex (47, 93, 117). Immobilized waters as extended side chains of proteins strategically located to recognize DNA became an attractive proposal and was invoked quite often in the literature. Waters are versatile in their solvation ability and assigning a specific functional role is difficult, nay, hazardous, especially if the binding process is occurring in aqueous medium. Thermodynamic studies indicate that water release upon binding from the surfaces of nonpolar atoms at the interface (hydrophobic effect) does contribute favorably to the energetics of binding (31, 36, 56, 73, 125). Sequence dependence in ordered water location at the interface is unclear at this stage. Whether waters could mediate recognition remains a question and this is taken up as the main theme of this review.

Thermodynamic studies, both experimental and theoretical, on protein-DNA complexes dramatically illustrate the necessity to consider the diverse competing effects such as van der Waals (steric complementarity), electrostatics including hydrogen bonding (electrostatic complementarity), adaptation expense, and entropy losses—all of which add up to –9 to –17 kcal (57)—in constructing a structure-based interpretation of binding free energies and hence recognition mechanism in protein-DNA complexes (9, 52, 55, 56, 68). This view is further strengthened by the observation that hydrogen bonding and hydrophobic interactions alone cannot explain the pattern of evolutionary conservation of base pairs in the binding site, suggesting a cumulative contribution of different types of interactions to specific recognition (84). The unbound protein and DNA are surrounded by solvent and small ions. Upon binding, the conformations of both protein and DNA may change to accommodate each other either to form hydrogen bonds or to pack well at the interface, and in the process some solvent molecules and small ions may be released. A schematic of the binding reaction is shown in Figure 1. In a holistic view of DNA-protein recognition, the mechanistic details between the initial and final states are numerous.

Is there a common theme for recognition that is transferable across all specific protein-DNA complexes? Whether there has to be a common mode or a balance between multiple modes for recognition is a moot issue (76). A new analysis (T. Jain, N. Latha & B. Jayaram, unpublished data) of the pattern of atoms occurring at the interface of 120 protein-DNA complexes resulted in the identification of 11 energetically favorable supramolecular synthons that are transferable across all specific protein-DNA complexes. [A synthon is the smallest substructural unit

that contains the maximum structural information and is stabilized by noncovalent interactions (89)]. These synthons are unique combinations of C, N, and O atoms and account for almost all the observed contacts in protein-DNA complexes. An interpretation of their occurrence in functional terms may call for a new view on intermolecular recognition. Further examination of these ideas is in progress.

WATER AROUND DNA

DNA hydration is the subject of several thorough reviews (2–4, 7, 62, 139). First observed in the minor groove of the d(CGCGAATTCGCG) dodecamer (22, 90), the spine of hydration seems to be a common feature of the AT-rich regions (62) 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 at the atomic level. A structural analysis of crystalline CGATTAATCG (106) 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 structure of crystalline CCAACITTGG (72) the spine appears in the narrow center.

A spine of hydration is also found in the major grooves of Z-DNA duplexes, with water molecules bridging the O_2 atoms of successive cytidines (12, 32). Analysis on water distributions around phosphate groups revealed that waters are concentrated in six hydration sites per phosphate and that the positions and occupancies of these sites are dependent on the conformation and type of nucleotide (3, 115). A suggestion has also been advanced that ordered water molecules may help establish the identity of tRNAs (112). Right-handed DNA duplexes assume a B form at high water activity and an A form at reduced levels. A free-energy analysis (54) of molecular dynamics trajectories of A and B forms of DNA in water and in mixed solvent systems modeled in atomic detail revealed that conformational preferences of DNA were due to a fine electrostatic balance between interphosphate repulsions, counterion-DNA attractions, and solvation/desolvation energetics. The "ordered waters" thus may have to be viewed in their energetic perspective.

WATER AROUND PROTEINS

Hydration of proteins observed using crystallographic, NMR, and molecular simulation techniques has been the subject of several excellent reviews (59, 69, 100, 102, 113, 130, 141). Numerous examples of the structural and functional importance of water around proteins have been reported. Waters 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 (107). Structural studies point to a major role for water in protease-inhibitor binding (44) and in antigen-antibody recognition (8). 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 (66). Water molecules are also crucial in defining the substrate specificity of bacterial arabinose binding protein (107) and glutamate dehydrogenase (127). Water also plays a catalytic role in the hydrolysis of carboxypeptidase A (43). Reviewing the status on waters in proteins, Levitt & Park (69) observe that water fills all space not occupied by protein atoms since nature abhors vacuum and that most of the ordered waters on the surface and in narrow crevices are in rapid motion with exchange times less than 100 ps (24, 99), but water molecules in interior cavities exchange more slowly (10 ns to 10 ms).

WATER IN PROTEIN-DNA BINDING

Systematic dehydration analysis of B-DNA suggests that waters from sugars can be removed as also from bases but not from phosphates (25–28, 41, 42, 114, 133, 134, 139). How proteins accomplish this to achieve specific recognition is one of many interesting questions in the molecular thermodynamics of protein-DNA binding. 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 oxygens of the phosphodiester backbone (118). Macromolecular crystallography provided the necessary supportive view that water molecules could act as major contributors to stability and specificity (47, 117, 145).

In the structure of *trp* repressor-DNA complex (93), direct contacts are observed only with the phosphate groups of DNA and these contacts do not seem important for base-sequence recognition. There are, however, three ordered water molecules at the protein-DNA interface that hydrogen bond with base pairs as well as with 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 (30). The importance of these water molecules for binding and recognition was highlighted by the crystal structure of the paired homeodomain bound to DNA (142). Remarkably, in this structure, there are 18 ordered water molecules at the protein-DNA interface. Stability and specificity are reported to be conferred by a network of water-mediated protein-DNA hydrogen bonds in estrogen receptor–DNA complex (64). A large number of water-mediated hydrogen bonds have been reported in the structure of the nucleosome core particle (17), which apparently provide further stability to direct interactions and enable formation of additional interactions between more distant elements. The binding of an 11-residue β hairpin of Smad3 MH1 in the major groove of DNA is buttressed by seven ordered water molecules (10). The DNA complex of *Hin* recombinase DNA binding domain contains two ordered water molecules. Systematic base mutational studies indicate that one of the waters is essential to stable complex formation, and the second plays an auxiliary role (13). On the basis of structural analysis, an interesting idea that has been advanced is that protein atoms involved in binding to

DNA occupy positions normally occupied by water molecules in unbound DNA (145).

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 hydrophobic interface (61, 91). TBP interacts along the length of the minor groove of DNA, which is splayed open and curves away from the protein. As the minor groove is normally highly hydrated, many water molecules must be displaced and the driving force for complex formation would seem primarily entropic.

Simulation studies (34, 38, 80, 120, 128, 132, 146) in general have supported the crystallographic/NMR observations on water-mediated hydrogen bonds.

RELATED INFORMATION FROM DRUG-DNA SYSTEMS

Although many protein-DNA contacts are mediated by water—interpreted as a promotional event to increase the effective surface (87)—fewer water-mediated contacts are observed in drug-DNA complexes. Waters are reported to mediate drug-DNA electrostatics in the major groove. High density of waters found in the minor groove in X-ray structures and molecular dynamics simulations is associated only with weakly bound solvent in solution (140). Other interpretations on role of waters as mediators of electrostatic interactions and screening repulsions between like charges have also been advanced in the context of drug-DNA binding (39). Thermodynamic studies on drug-DNA complexes indicate that water release is favorable to binding (11, 86; S.A. Shaikh & B. Jayaram, unpublished data).

Overall, occurrence of ordered waters in protein-DNA complexes necessitates a molecular explanation for their presence and their implication to protein-DNA recognition.

WATER AS HYDROGEN BOND DONOR AND ACCEPTOR AT INTERFACE

A comprehensive analysis of interfacial water molecules in 109 unique protein-DNA complexes that contact the protein and the DNA simultaneously and could mediate recognition presents a new view on their role in protein-DNA recognition (110). The interfacial water molecules form a small fraction (6%) of the crystallographically observed waters. Most of these waters occur between negative charges (partial or full) on protein and DNA. Noting that DNA is polyanionic, it is not surprising that a majority of the experimentally observed waters as well as those from molecular dynamics simulations should be located strategically so as to be involved in facilitating binding by screening unfavorable electrostatics. Just about one third of the interfacial waters occur between hydrogen bond donor atoms of protein and acceptor atoms of DNA. These represent cases in which

protein atoms cannot reach out to DNA to make favorable hydrogen bond interactions owing to packing/structural restrictions, and interfacial waters could act as linkers, providing an extension to side chains to accomplish hydrogen bonding.

WATER AS FILLER TO MAINTAIN PACKING DENSITIES AT INTERFACE

Density gradients are not sustainable for a system at equilibrium. Local density variations within the solvated macromolecular system could lead to transport of matter (passive transport), manifested via conformational transitions, interactions permitting, or diffusion of solvent. Calculated local densities in protein-DNA complexes in the absence of water are about 0.8 g/ml, but 1.0 g/ml with trapped waters, which is same as that of bulk solvent (110). During complexation in a solvent medium, large departures from bulk densities are not expected at the interface.

WATER AS BUFFER TO SCREEN UNFAVORABLE ELECTROSTATICS

An atom-wise contact analysis of interfacial waters indicates that waters predominantly interact with acceptor atoms of both protein and DNA (110). 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 and the oxygens of the side chains are the main contributors, and in the case of DNA, phosphate oxygens 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 few and those on the protein make favorable interactions with DNA any way and hence do not appear to prefer solvation at the interface to any significant extent. This again is an indication of the electrostatic role of the intervening water molecules.

A residue-wise contact analysis indicates that on the protein side Glu and Asp are the main residues interacting with water, followed by Ser, Thr, Asn, and Gln. For DNA, the backbone hydration dominates as expected (110). This provides an alternative view in favor of the electrostatic buffering action of water.

DNA is characterized by deep potentials in the grooves and high fields near the phosphates (53, 104, 105, 122). Electrostatic field calculations indicate that waters contacting protein and DNA simultaneously are generally found in regions of higher fields than the rest, and their average orientations are perpendicular to the field. Higher fields in this case imply a stronger repulsive force between the two acceptor atoms of protein and DNA, and perpendicular alignment is expected from both hydrogens of mediating water involved in hydrogen bonding with the acceptor atoms. The waters that do not contact protein and DNA simultaneously show no preferential direction, as the dipole would depend on the type of residues

in the vicinity. These observations once again bring to the fore the buffering action of such interfacial waters that occur between acceptor atoms.

In a nutshell, the abovecited data extracted from crystallographic waters lead us to the conclusion that interfacial waters mainly act to decrease the electrostatic repulsions between the electronegative atoms/like charges in protein-DNA complexes, thus favoring binding besides maintaining liquid densities at the interface. Analyses of water behavior in molecular dynamics simulations on 35 systems lead to identical conclusions.

WATER IN THE THERMODYNAMICS OF PROTEIN-DNA BINDING

On the energetic front, both experiment and theory clearly indicate that water release from nonpolar atoms makes a favorable contribution to the binding free energy of protein-DNA complexes (36, 47, 52, 56, 125). Complex formation between FpG and the THF-containing duplex at 15°C exhibits an unfavorable association enthalpy ($\Delta H = +7.5$ kcal/mol) that is more than offset by a favorable association entropy ($T\Delta S = +17.0$ kcal/mol). This, coupled with the large negative heat capacity (-0.67 kcal/deg/mol), is consistent with a significant contribution of water release from the buried nonpolar surface (83).

Protein-DNA binding in a majority of cases is characterized by favorable Coulomb interactions (direct electrostatics) and unfavorable desolvation electrostatics. The exception to the rule comes from *trp* repressor-operator, wherein direct electrostatics is unfavorable and desolvation electrostatics is favorable when interfacial waters are not considered (52). A detailed interaction energy analysis with interfacial waters indicates that direct as well as net electrostatics are favorable in this system. This is a clear case of water-mediated protein-DNA binding, where unfavorable electrostatics is mitigated by interfacial waters. Another example of this is the *PvuII*-DNA complex (52).

To better appreciate motif dependence, if any, in protein-DNA binding energetics and the role of solvent in particular, we have computed separately the net electrostatic and nonelectrostatic contributions in ~100 protein-DNA complexes (A. Das & B. Jayaram, unpublished data). Net electrostatics refers to the sum of direct electrostatic (Coulomb) interactions computed with a dielectric of unity and desolvation expense [solvation energy of the complex − (solvation energy of DNA + solvation energy of the protein)] due to binding. Nonelectrostatic contributions refer to the van der Waals interactions between protein and DNA and a surface area–related cavitation contribution that includes loss in van der Waals interactions of solvent upon binding and a gain due to solvent release from the interface. Additional computational details are provided in References 51 and 52. The net electrostatic and the nonelectrostatic terms for each complex are normalized per monomer of the protein participating in binding and are depicted in Figure 2.

A clear DNA binding motif separation is discernible (Figure 2) when solvation effects are included in a steric versus electrostatic complementarity energy analysis of protein-DNA complexes. The clustering is reflective of a strategy of proteins to optimize both steric and electrostatic interactions to achieve specific binding despite diverse shapes. In addition to the net electrostatics, van der Waals, and cavitation components considered in Figure 2, other contributions such as small ion effects, deformation expense, and loss in rotational, translational, and vibrational entropies upon binding also add to the net binding free energies. These are computed [A. Das & B. Jayaram, unpublished data] as described in References 51 and 52. Figure 3 depicts a DNA binding protein motif-wise component analysis of the average binding energetics in over 100 protein-DNA complexes. In all cases, van der Waals (packing) and hydrophobic interactions are favorable to binding. The small ion effects are unfavorable as are the deformation and entropic contributions. The net electrostatic interactions, which include direct protein-DNA Coulomb interactions and desolvation contributions, are highly case specific. For leucine zippers and zinc fingers the net electrostatics turns favorable, whereas for enzymes the net electrostatics is unfavorable (Figure 2). Removal of charged residues from solvent exposure does cause some concern in this context. Burial of a large number of neutral polar groups/atoms carrying partial charges leads to unfavorable electrostatics, as is the case with enzymes. Energy analyses indicate that basic residues (carrying full charges) lose less owing to desolvation and gain more owing to favorable direct electrostatic interactions with DNA, leucine zippers being a case in point. Analysis of the interfacial residues in terms of those contacting bases versus phosphate backbone reveals that zinc fingers and homeodomains tend to use more basic residues to contact bases than do other motifs.

Water release from interface, in general, is favored by entropic terms and disfavored by enthalpic terms, with the former dominating. Water-mediated interactions would imply the opposite. Localization of water at the interface is unfavorable entropically (23). Thus the net enthalpy changes due to interfacial waters ($\Delta H^{net} = \Delta H_{P\text{-}w\text{-}D} - \Delta H_{P\text{-}w} - \Delta H_{D\text{-}w} - \Delta H_{P\text{-}D}$, in which $\Delta H_{P\text{-}w\text{-}D}$, $\Delta H_{P\text{-}w}$, $\Delta H_{D\text{-}w}$, and $\Delta H_{P\text{-}D}$ represent enthalpies associated with protein-DNA interactions in the presence of interfacial waters, protein-interfacial water interactions, DNA-interfacial water interactions, and protein-DNA interactions without interfacial waters, respectively) must be overridingly favorable, which seems unlikely, unless there is a large-scale reduction of repulsive interactions ($\Delta H_{P\text{-}D}$) between like charges on protein and DNA due to interfacial waters, as is the case with *trp* repressor-operator and *PvuII*-DNA complexes.

Water-mediated hydrogen bonds as a mode of recognition de-emphasize somewhat the role of steric complementarities in specific recognition in aqueous media. Even if the surfaces of the interacting molecules are not complementary, solvent water will occupy intervening space anyway. This as a principle for the design of small molecules to bind to proteins or DNA would require some clever strategies.

WATER IN THE KINETICS OF PROTEIN-DNA BINDING

An interesting suggestion advanced by Jen-Jacobson (56, 57) was that proteins could bind to DNA nonspecifically without water displacement, possibly in a three-dimensional diffusion mode (48, 143), followed by specific binding in a one-dimensional sliding motion in which waters are removed to provide room for direct contacts. Further studies of an atomic-level view of the kinetics of binding await.

CONCLUSIONS

Thermodynamic analyses of protein-DNA binding suggest that water release from protein-DNA interfaces is favorable to binding. Structural analyses of the waters remaining at the interface in protein-DNA complexes indicate that a majority of these waters facilitate binding by screening electrostatic repulsions between electronegative atoms/like charges of the protein and the DNA, in addition to maintaining liquid-state densities. A small fraction of the observed interfacial waters act as linkers to form extended hydrogen bonds between the protein and the DNA, compensating for the lack of a direct hydrogen bond.

Because the structures of protein-DNA complexes represent a convolution of folding and binding principles, a code for recognition remained refractory to elucidation. However, since the first crystal structure of a protein-DNA complex reported in 1986 (82), the picture of protein-DNA binding and protein folding and the focus on water *inter alia* are rapidly growing in detail and are likely to soon unveil the underlying molecular view of DNA recognition.

ACKNOWLEDGMENTS

Funding from the Department of Science & Technology, Indo-French Center for the Promotion of Advanced Research (IFCPAR), and Department of Biotechnology is gratefully acknowledged. BJ would like to thank Dr. D.L. Beveridge for many helpful discussions on the subject matter. Critical comments on the manuscript from Dr. D.L. Beveridge and Dr. M. Bansal, preparation of Figures 2 and 3 by Mr. A. Das, and help received from Ms. K. Bhushan, Ms. N. Latha, Mr. S. Bose, and Mr. P. Sharma in preparing the manuscript are gratefully acknowledged.

LITERATURE CITED

1. Anderson CF, Record MT. 1982. Polyelectrolyte theories and their application to DNA. *Annu. Rev. Phys. Chem.* 33:191–222
2. Berman HM. 1991. Hydration of DNA. *Curr. Opin. Struct. Biol.* 1:423–27
3. Berman HM. 1994. Hydration of DNA: take 2. *Curr. Opin. Struct. Biol.* 4:345–50

4. Berman HM, Schneider B. 1998. Nucleic acid hydration. In *Handbook of Nucleic Acid Structure*, ed. S Neidle, pp. 295–312. Oxford, UK: Oxford Univ. Press
5. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–42
6. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, et al. 1977. The Protein Data Bank: a computer-based archive file for macromolecular structures. *J. Mol. Biol.* 112:535–42
7. Beveridge DL, Swaminathan S, Ravishankar G, Whitka JM, Srinivasan J, et al. 1993. Topics in molecular and structural biology. In *Water and Biological Macromolecules*, ed. E Westhof, 17:165–225. London: Macmillan
8. Bhat TN, Poljak RJ. 1994. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA* 91:1089–93
9. Bosch D, Campillo M, Pardo L. 2003. Binding of proteins to the minor groove of DNA: What are the structural and energetic determinants for kinking a base-pair step? *J. Comput. Chem.* 24:682–91
10. Chai J, Wu JW, Yan N, Massague J, Pavletich NP, Shi Y. 2003. Features of a Smad3 MH1-DNA complex. *J. Biol. Chem.* 278:20327–31
11. Chaires JB. 1997. Energetics of drug-DNA interactions. *Biopolymers* 44:201–15
12. Chevrier B, Dock AC, Hartmann B, Leng M, Moras D, et al. 1986. Solvation of the left handed hexamer d(5BrCG5BrCG5BrCG) in crystals grown at room temperature. *J. Mol. Biol.* 188:707–19
13. Chiu TK, Sohn C, Dickerson RE, Johnson RC. 2002. Testing water-mediated DNA recognition by the Hin recombinase. *EMBO J.* 21:801–14
14. Choo Y, Klug A. 1997. Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.* 7:117–25
15. Clore GM, Gronenborn AM. 1994. Structures of larger proteins, protein-ligand and protein-DNA complexes by multidimensional heteronuclear NMR. *Prog. Biophys. Mol. Biol.* 62:153–84
16. Crothers DM. 1998. DNA curvature and deformation in protein-DNA complexes: a step in the right direction. *Proc. Natl. Acad. Sci. USA* 95:15163–65
17. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* 319:1097–113
18. Dickerson RE, Chiu TK. 1997. Helix bending as a factor in protein-DNA recognition. *Biopolymers* 44:361–403
19. Dixit SB, Arora N, Jayaram B. 2000. How do hydrogen bonds contribute to protein-DNA recognition? *Proceedings of the Eleventh Conversation in Biomolecular Stereodynamic: Journal of Biomolecular Structure and Dynamics, Conversation 11*, ed. RH Sarma, MH Sarma, pp. 109–12. New York: Adenine
20. Dixit SB, Jayaram B. 1998. Role of hydrogen bonds in protein-DNA recognition: a comparison of generalized born and finite difference Poisson-Boltzmann solvation treatments. *J. Biomol. Struct. Dyn.* 16:237–42
21. Dreier B, Segal DJ, Barbas CF III. 2000. Insights into the molecular recognition of the 5-GNN-3 family of DNA sequences by zinc finger domains. *J. Mol. Biol.* 303:489–502
22. Drew HW, Dickerson RE. 1981. Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* 151:535–56
23. Dunitz JD. 1994. The entropic cost of bound water in crystals and biomolecules. *Science* 264:670
24. Eisenberg D, Kauzmann W. 1969. Properties of liquid water. In *The Structure and Properties of Water*, 4:150–55. Oxford: Oxford Univ. Press
25. Falk M, Hartman KA, Lord RC. 1962. Hydration of deoxyribonucleic acid. I.

A gravimetric study. *J. Am. Chem. Soc.* 84:3843–46
26. Falk M, Hartman KA, Lord RC. 1963. Hydration of deoxyribonucleic acid. II. An infrared study. *J. Am. Chem. Soc.* 85:387–91
27. Falk M, Hartman KA, Lord RC. 1963. Hydration of deoxyribonucleic acid. III. A spectroscopic study of the effect of hydration on the structure of deoxyribonucleic acid. *J. Am. Chem. Soc.* 85:391–94
28. Falk M, Poole AG, Goymom CG. 1970. Infrared study of the state of water in the hydration shell of DNA. *Can. J. Chem.* 48:1536–42
29. Flatters D, Lavery R. 1998. Sequence-dependent dynamics of TATA-Box binding sites. *Biophys. J.* 75:372–81
30. Fraenkel E, Pabo CO. 1998. Comparison of X-ray and NMR structures of the Antennapedia homeodomain-DNA complex. *Nat. Struct. Biol.* 5:692–97
31. Garner MM, Rau DC. 1995. Water release associated with specific binding of gal repressor. *EMBO. J.* 14:1257–63
32. Gesner RV, Quigley GJ, Egli M. 1994. Comparative studies of high resolution Z-DNA crystal structures. Part 1: common hydration patterns of alternating dC-dG. *J. Mol. Biol.* 236:1154–68
33. Ghosh A, Bansal M. 1999. C─H. . .O hydrogen bonds in minor groove of A-tracts in DNA double helices. *J. Mol. Biol.* 294:1149–58
34. Giudice E, Lavery R. 2002. Simulations of nucleic acids and their complexes. *Acc. Chem. Res.* 35:350–57
35. Gurlie R, Zakrzewska K. 1998. DNA curvature and phosphate neutralization: an important aspect of specific protein binding. *J. Biomol. Struct. Dyn.* 16:605–18
36. Ha JH, Spolar RS, Record MT Jr. 1989. Role of hydrophobic effect in the stability of site-specific protein-DNA complexes. *J. Mol. Biol.* 209:801–16
37. Harrington RE, Winicov I. 1994. New concepts in protein-DNA recognition: sequence-directed DNA bending and flexibility. *Prog. Nucleic Acid Res. Mol. Biol.* 47:195–270
38. Harris LF, Sullivan MR, Popken-Harris PD. 1999. Molecular dynamics simulations in solvent of the bacteriophage 434 cI repressor protein DNA binding domain amino acids (r1-69) in complex with its cognate operator (OR1) DNA sequence. *J. Biomol. Struct. Dyn.* 19:1–17
39. Harris SA, Gavathiotis E, Searle MS, Orozen M, Laughton CA. 2001. Cooperativity in drug-DNA recognition: molecular dynamics study. *J. Am. Chem. Soc.* 123:12658–63
40. Harrison SC. 1991. A structural taxonomy of DNA-binding domains. *Nature* 353:715–19
41. Hearst JE. 1965. Determination of the dominant factors which influence the net hydration of native sodium deoxyribonucleate. *Biopolymers* 3:57–68
42. Hearst JE, Vinograd J. 1961. The net hydration of deoxyribonucleic acid. *Proc. Natl. Acad. Sci. USA* 47:825–30
43. Hidong K, Lipscomb WN. 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* 29:5546–55
44. Huang K, Anderson LW, Laskowski M Jr, James MNG. 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.* 4:1985–87
45. Hud NV, Feigon J. 1997. Localization of divalent metal ions in the minor groove of DNA A-tracts. *J. Am. Chem. Soc.* 119:5756–57
46. Hud NV, 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.* 286:651–60
47. Janin J. 1999. Wet and dry interfaces:

the role of solvent in protein-protein and protein-DNA recognition. *Structure* 7:R277–79
48. Jayaram B. 1996. Some energetic and kinetic aspects of protein-DNA interactions: a theoretical study on the λ repressor-operator complex. In *Biological Structure and Dynamics: Proceedings of the Ninth Conversation in Biomolecular Stereodynamics*, ed. RH Sarma, MH Sarma, 1:109–20. New York: Adenine
49. Jayaram B, Beveridge DL. 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
50. 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*) 361:249–58
51. Jayaram B, McConnell KJ, Dixit SB, Beveridge DL. 1999. Free energy analysis of protein-DNA binding: the Eco-RI endonuclease-DNA complex. *J. Comput. Phys.* 151:333–57
52. Jayaram B, McConnell KJ, Dixit SB, Das A, Beveridge DL. 2002. Free energy analysis of 40 protein-DNA complexes: a consensus view on the nature of binding at the molecular level. *J. Comput. Chem.* 23:1–14
53. Jayaram B, Sharp K, Honig B. 1989. The electrostatic potential of B-DNA. *Biopolymers* 28:975–93
54. Jayaram B, Sprous D, Young MA, Beveridge DL. 1998. Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.* 120:10629–33
55. Jen-Jacobson L. 1995. Structural perturbation approaches to thermodynamics of site specific protein-DNA interactions. *Methods Enzymol.* 259:305–44
56. Jen-Jacobson L. 1997. Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers* 44:153–80
57. Jen-Jacobson L, Engler LE, Jacobson LA. 2000. Structural and thermodynamic strategies for site-specific DNA binding proteins. *Struct. Fold Des.* 8(10):1015–23. Erratum. 2000. *Struct. Fold Des.* 8(12): 251
58. Jones S, van Heyningen P, Berman HM, Thornton JM. 1999. Protein-DNA interactions: a structural analysis. *J. Mol. Biol.* 287:877–96
59. Karplus PA, Faerman C. 1994. Ordered water molecules in macromolecular structure. *Curr. Opin. Struct. Biol.* 4:770–76
60. Kim JL, Nikolov DB, Burley SK. 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365:520–27
61. Kim Y, Geiger JH, Hahn S, Sigler PB. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365: 512–20
62. Kochoyan M, Leroy JL. 1995. Hydration and solution structure of nucleic acids. *Curr. Opin. Struct. Biol.* 5:329–33
63. Kombo DC, Jayaram B, McConnell KJ, Beveridge DL. 2002. Calculation of the affinity of the λ repressor-operator complex based on free energy component analysis. *Mol. Simul.* 28:187–211
64. Kosztin D, Bishop TC, Shulten K. 1997. Binding of the estrogen receptor to DNA. The role of waters. *Biophys. J.* 73:557–70
65. Koudelka GB, Harrison SC, Ptashne M. 1987. Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro. *Nature* 326:886–88
66. Lam PYS, Jadhav PK, Eyermann CJ, Hodje CN, Ru Y, et al. 1994. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* 263:380–84
67. Lehninger AL, Nelson DL, Cox MM. 1993. In *Principles of Biochemistry*, 8:198–204. New Delhi: CBS Publ.
68. Lesser DR, Kupriewski MR, Jen-Jacobson L. 1990. The energetic basis of specificity in the Eco RI endonuclease–DNA interaction. *Science* 250:776–86

69. Levitt M, Park BH. 1993. Water: Now you see it, now you don't. *Structure* 1:223–26
70. Lewin B. 2000. Nucleosomes. In *Genes VII*, 19:567–600. Oxford: Oxford Univ. Press
71. Luisi B. 1995. DNA-protein interactions at high resolution. In *DNA-Protein: Structural Interactions*, ed. DMJ Lilley, 1:1–48. Oxford: Oxford Univ. Press
72. Lipanov A, Kopka ML, Kaczor G, Rzeskowiak M, Quintana J, Dickerson RE. 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* 32:1373–80
73. Lundback T, Hard T. 1996. Sequence-specific DNA-binding dominated by dehydration. *Proc. Natl. Acad. Sci. USA* 93:4754–59
74. Luscombe NM, Austin SE, Berman HM, Thornton JM. 2000. An overview of the structures of protein-DNA complexes. *Genome Biol.* 1:1–37
75. Luscombe NM, Laskowski RA, Thornton JM. 2001. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* 29:2860–74
76. Luscombe NM, Thornton JM. 2002. Protein–DNA interactions: amino acid conservation and the effects of mutations on binding specificity. *J. Mol. Biol.* 320:991–1009
77. Lynch TW, Read EK, Mattis AN, Gardner JF, Rice PA. 2003. Integration host factor: putting a twist on protein-DNA recognition. *J. Mol. Biol.* 330:493–502
78. Mandel-Gutfreund Y, Margalit H, Jernigan R, Zhurkin V. 1998. A role for C─H. . .O interaction in protein-DNA complexes. *J. Mol. Biol.* 277:1129–40
79. Manning GS. 1978. The molecular theory of polyelectrolyte solutions with application to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* 11:179–246
80. Marco E, Garcia-Nieto R, Gago F. 2003. Assessment of molecular dynamics simulations of the structural determinants of DNA-binding specificity for transcription factor Sp1. *J. Mol. Biol.* 328:9–32
81. Matthews BW. 1988. Protein-DNA interaction. No code for recognition. *Nature* 335:294–95
82. McClarin JA, Frederick CA, Wang BC, Greene P, Boyer HW, et al. 1986. Structure of the DNA-Eco RI endonuclease recognition complex at 3 Å resolution. *Science* 234:1526–41
83. Minetti CASA, Remeta DP, Zharkov DO, Plum GE, Jhonson F, et al. 2003. Energetics of lesion recognition by a DNA repair protein: thermodynamic characterization of formamidopyrimidine-glycosylase (FpG) interactions with damaged DNA duplexes. *J. Mol. Biol.* 328: 1047–60
84. Mirny LA, Gelfand MS. 2002. Structural analysis of conserved base pairs in protein–DNA complexes. *Nucleic Acids Res.* 30:1704–11
85. Misra VK, Hecht JL, Sharp KA, Friedman RA, Honig B. 1994. Salt effects on protein-DNA interactions: the lambda cI repressor and *Eco* RI endonuclease. *J. Mol. Biol.* 238:264–80
86. Misra VK, Honig B. 1995. On the magnitude of the electrostatic contribution to ligand-DNA interactions. *Proc. Natl. Acad. Sci. USA* 92:4691–95
87. Moravek Z, Neidle S, Schneider B. 2002. Protein and drug interactions in the minor groove of DNA. *Nucleic Acids Res.* 30:1182–91
88. Nadassy K, Wodak SJ, Janin J. 1999. Structural features of protein–nucleic acid recognition sites. *Biochemistry* 38:1999–2017
89. Nangia A, Desiraju GR. 1998. Supramolecular synthons and pattern recognition. *Top. Curr. Chem.* 198:57–95
90. Narayana N, Ginell SL, Russu I, Berman HM. 1991. Crystal and molecular structure of a DNA fragment:

d(CGTGAATTCACG). *Biochemistry* 30: 4449–55
91. Nickolov DB, Chen H, Halay ED, Hoffman A, Roeder RG, Burley SK. 1996. Crystal structure of a human TATA box–binding protein/TATA element complex. *Proc. Natl. Acad. Sci. USA* 93:4862–67
92. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB. 1998. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA* 95:11163–68
93. Otwinowski Z, Schevitz RW, Zhang RG, Lawson CL, Jochiamiak A, et al. 1988. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* 335:321–29
94. Pabo CO, Nekludova L. 2000. Geometric analysis and comparison of protein-DNA interfaces: Why is there no simple code for recognition? *J. Mol. Biol.* 301:597–624
95. Pabo CO, Sauer RT. 1992. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* 61:1053–95
96. Patikoglou G, Burley SK. 1997. Eukaryotic transcription factor-DNA complexes. *Annu. Rev. Biophys. Biomol. Struct.* 26:289–325
97. Pauling L, Corey RB. 1951. Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc. Natl. Acad. Sci. USA* 37:235–40
98. Pauling L, Corey RB, Branson HR. 1951. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chains. *Proc. Natl. Acad. Sci. USA* 37:205–11
99. Peon J, Pal SK, Zewail AH. 2002. Hydration at the surface of the protein monellin: dynamics with femtosecond resolution. *Proc. Natl. Acad. Sci. USA* 99:10964–69
100. Pettitt BM, Makarov VA, Andrews BK. 1998. Protein hydration density: theory, simulations and crystallography. *Curr. Opin. Struct. Biol.* 8:218–21
101. Phillips SEV. 1991. Specific β-sheet interactions. *Curr. Opin. Struct. Biol.* 1:89–98
102. Prasad BV, Suguna K. 2002. Role of water molecules in the structure and function of aspartic proteinases. *Acta Crystallogr. D* 58:250–59
103. Ptashne M, ed. 1992. *A Genetic Switch Phage λ and Higher Organisms*. Cambridge, MA: Blackwell Sci.
104. Pullman A, Pullman B. 1981. Molecular electrostatic potential of the nucleic acids. *Q. Rev. Biophys.* 14:289–380
105. Pullman B. 1983. Electrostatics of polymorphic DNA. *J. Biomol. Struct. Dyn.* 1:773–94
106. Quintana JR, Grzeskowiak K, Yanagi K. Dickerson RE. 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.* 225:379–95
107. Quiocho FA, Wilson DK, Vyas NK. 1989. Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* 340:404–7
108. Record MT Jr, Anderson CF, Lohman T. 1978. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.* 11:103–78
109. Record MT Jr, Anderson CF, Mills P, Mossing M, Roe JH. 1985. Ions as regulators of protein-nucleic acid interactions in vitro and in vivo. *Adv. Biophys.* 20:109–35
110. Reddy CK, Das A, Jayaram B. 2001. Do water molecules mediate protein-DNA recognition? *J. Mol. Biol.* 314:619–32
111. Rice PA, Yang SW, Mizuuchi K, Nash HA. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87:1295–306
112. Rould MA, Perona JJ, Soll D, Steitz TA. 1989. Structure of *E. coli* glutaminyl tRNA synthetase complexed with tRNAGln and ATP at 2.8 Å resolution: implications for tRNA discrimination. *Science* 246:1135–42
113. Rupley JA, Careri G. 1991. Protein

hydration and function. *Adv. Protein Chem.* 41:937–72
114. Saenger W. 1983. Water and nucleic acids. In *Principles of Nucleic Acid Structure*, 17:368–84. New York: Springer-Verlag
115. Schneider B, Patel K, Berman HM. 1998. Hydration of the phosphate group in double helical DNA. *Biophys. J.* 75:2422–34
116. Schultz SC, Shields GC, Steitz TA. 1991. Crystal structure of a CAP-DNA complex: The DNA is bent by 90 degrees. *Science* 253:1001–07
117. Schwabe JW. 1997. The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.* 7:126–34
118. Seeman NC, Rosenberg JM, Rich A. 1976. Sequence specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* 73:804–8
119. Selvaraj S, Kono H, Sarai A. 2002. Specificity of protein–DNA recognition revealed by structure-based potentials: symmetric/asymmetric and cognate/non-cognate binding. *J. Mol. Biol.* 322:907–15
120. Serf S, Nilsson I. 1999. Structure, interaction, dynamics and solvent effects on the DNA-EcoRI complex in aqueous solution from molecular dynamics simulation. *Biophys. J.* 77:1782–800
121. Shakked Z, Guerstein G, Frollow F, Rabinivich D, Joachimiak A, Sigler PB. 1994. Determinants of repressor-operator recognition from the structure of the trp operator binding site. *Nature* 368:469–73
122. Sharp K, Honig B. 1990. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biomol. Struct.* 19:301–32
123. Sigler PB. 1991. Protein-nucleic acid interactions. *Curr. Opin. Struct. Biol.* 1:61–62
124. Sinden RR. 1994. DNA-protein interactions. In *DNA Structure and Function*, 8:287–325. San Diego: Academic
125. Spolar RS, Record MT Jr. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263:777–84
126. Steitz TA. 1990. Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophys.* 23:205–80
127. Stillman TJ, Baker PJ, Britton KL, Rice DW. 1993. Conformational flexibility in glutamate dehydrogenase: role of water in substrate recognition and catalysis. *J. Mol. Biol.* 234:1131–39
128. Suenaga A, Yatsu C, Komeiji Y, Uebeyasi M, Meguro T, Yamato I. 2000. Molecular dynamics simulation of trp repressor-operator complex. Analysis of hydrogen bond patterns of protein-DNA interactions. *J. Mol. Struct.* 526:209–18
129. Suzuki M, Brenner SE, Gerstein M, Yagi N. 1995. DNA recognition code of transcription factors. *Protein Eng.* 8:319–28
130. Teeter MM. 1991. Water-protein interactions: theory and experiment. *Annu. Rev. Biophys. Biomol. Struct.* 20:577–600
131. Travers A, ed. 1993. *DNA-Protein Interactions*. London: Chapman & Hall. 1st ed.
132. Tsui V, Radhakrishnan I, Wright PE, Case DA. 2000. NMR and molecular dynamics studies of the hydration of a zinc finger DNA complex. *J. Mol. Biol.* 302:1101–17
133. Tunis M-JB, Hearst JE. 1968. On the hydration of DNA. I. Preferential hydration and stability of DNA in concentrated trifluoracetate solution. *Biopolymers* 6:1325–44
134. Tunis M-JB, Hearst JE. 1968. On the hydration of DNA. II. Base composition dependence of the net hydration of DNA. *Biopolymers* 6:1345–53
135. von Hippel PH. 1994. Protein-DNA recognition: new perspectives and underlying themes. *Science* 263:769–70
136. von Hippel PH, Berg OG. 1986. On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. USA* 83:1608–12
137. Wahl MC, Sundaralingam M. 1997. C─H. . .O hydrogen bonding in biology. *Trends Biochem. Sci.* 22:97–102
138. Watson JD, Crick FHC. 1953. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171:737–38

139. Westhof E, Beveridge DL. 1990. Hydration of nucleic acids. In *Water Science Reviews*, ed. F Franks, pp. 24–136. Cambridge, UK: Cambridge Univ. Press
140. Williams HE, Searle MS. 1999. Structure, dynamics and hydration of the nogalamycin-d(ATGCAT)2 complex determined by NMR and molecular dynamics simulations in solution. *J. Mol. Biol.* 290:699–716
141. Williams MA, Goodfellow JM, Thornton JM. 1994. Buried waters and internal cavities in monomeric proteins. *Protein Sci.* 3:1224–35
142. Wilson DS, Geunther B, Desplan C, Kurian J. 1995. High resolution crystal structure of a paired (pax) class homeodomain dimer on DNA. *Cell* 82:709–19
143. Winter RB, Berg OG, von Hippel PH. 1981. Diffusion driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli lac* repressor-operator interactions: kinetic measurements and conclusions. *Biochemistry* 20: 6961–77
144. Wintjens R, Lievin J, Rooman M, Buisine E. 2000. Contribution of cation-pi interactions in the stability of protein-DNA complexes. *J. Mol. Biol.* 302:395–410
145. Woda J, Schneider B, Patel K, Mistry K, Berman HM. 1998. An analysis of the relationship between hydration and protein-DNA interactions. *Biophys. J.* 75:2170–77
146. Yang L, Beard WA, Wilson SH, Roux B, Broyde S, Schlick T. 2002. Local deformations revealed by dynamics simulations of DNA polymerase β with DNA mismatches at the primer terminus. *J. Mol. Biol.* 321:459–78
147. Young MA, Jayaram B, Beveridge DL. 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
148. Young MA, Ravishanker G, Beveridge DL, Berman HM. 1995. Analysis of local helix bending in crystal structures of DNA oligonucleotide and DNA-protein complexes. *Biophys. J.* 68:2454–68

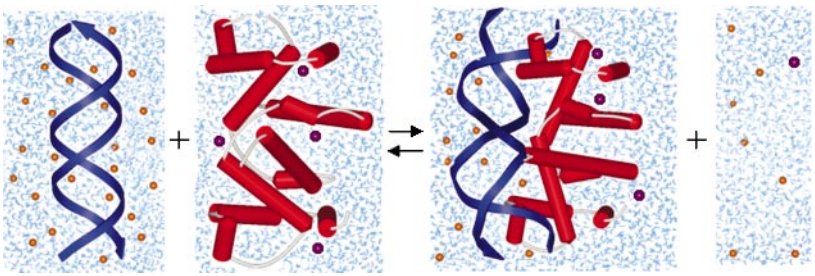

Figure 1 A schematic representation of protein-DNA binding. The process may be fully described by the following phenomenological equation:

$$[\text{DNA} + \text{x condensed counterions}]_{aq+salt} + [\text{Protein}]_{aq+salt} = [\text{DNA*. Protein*. y condensed counterions}]_{aq+salt} + (\text{x-y}) \text{ counterions} + \text{solvent molecules}$$

The asterisk (*) refers to the structural variations between the native protein/DNA and that in the complex upon binding. The binding process as illustrated is accompanied by the release of water molecules and counterions as well as changes in DNA and protein conformations.

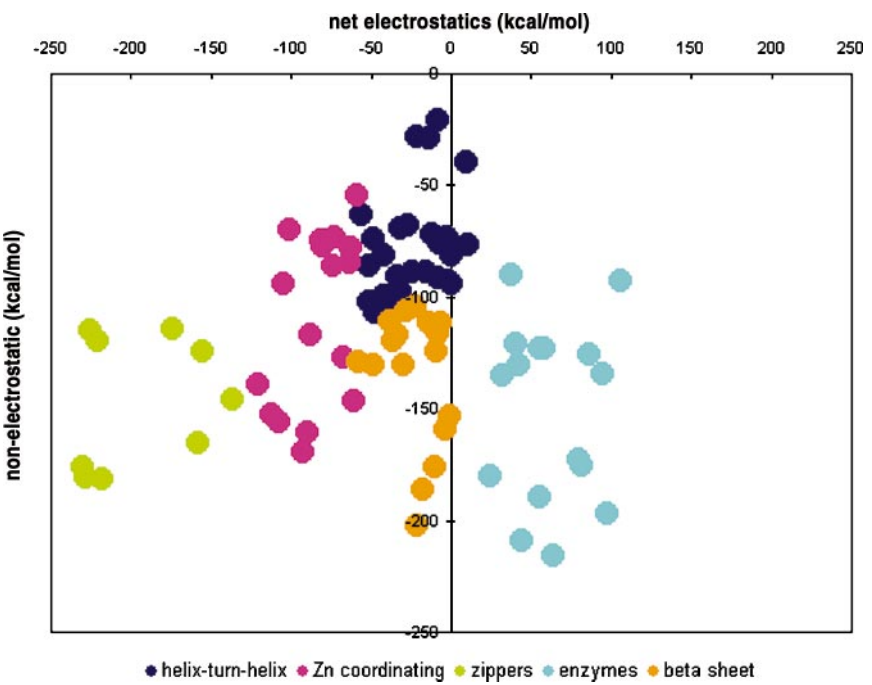

Figure 2 Net electrostatic contribution to protein-DNA binding is shown against non-electrostatic (van der Waals and hydrophobic) contributions for ~100 complexes color-coded according to DNA binding motif. Both electrostatics and van der Waals components include desolvation contributions.

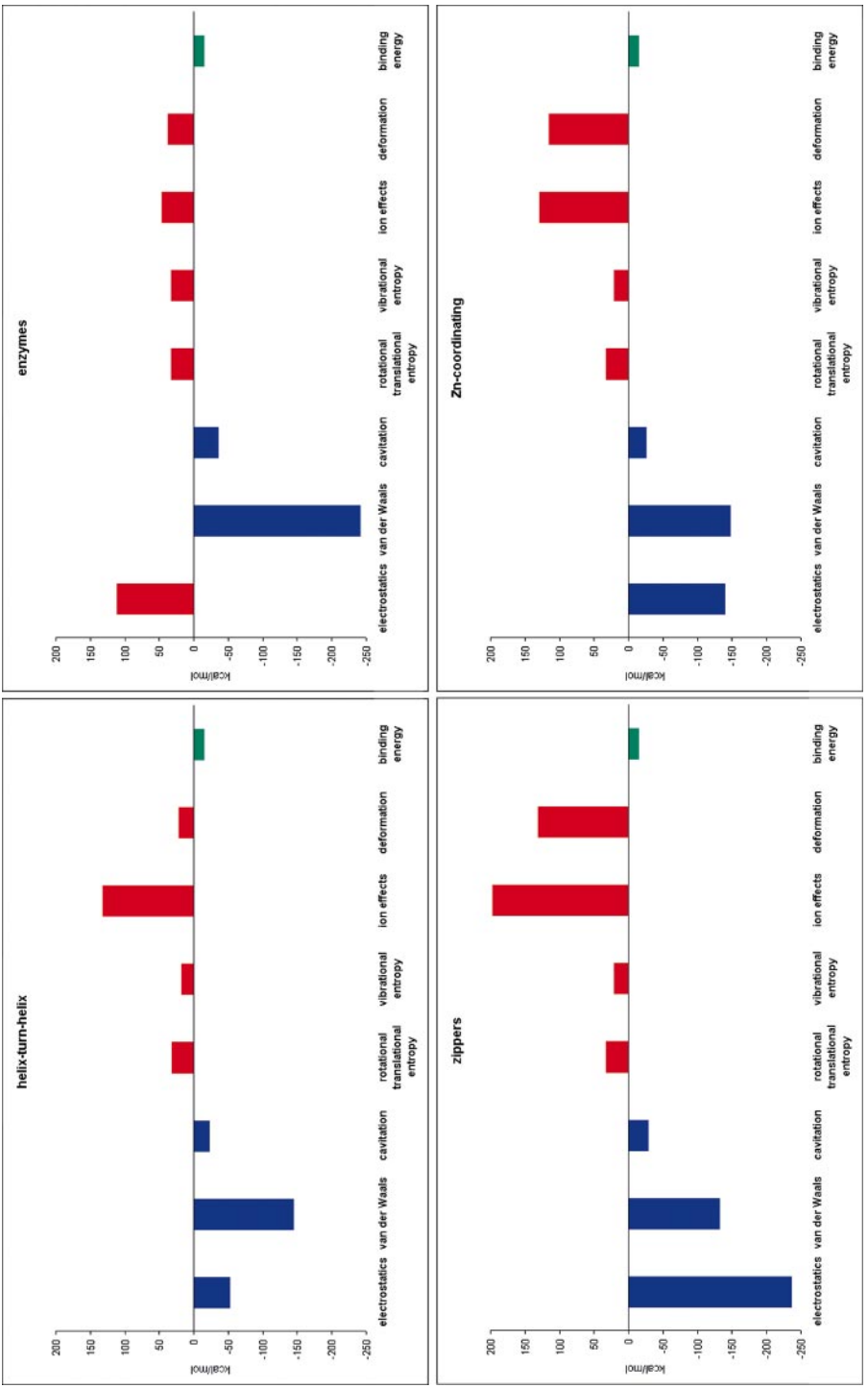

Figure 3 A free-energy component analysis of protein-DNA binding averaged motif-wise in 100 complexes.