An Analysis of the Relationship between Hydration and Protein-DNA Interactions

Juliana Woda,* Bohdan Schneider,# Ketan Patel,* Kavin Mistry,* and Helen M. Berman*§

*Department of Chemistry and §Waksman Institute, Rutgers University, Piscataway, New Jersey 08854-8087 USA, and #J. Heyrovsky Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, CZ-18223 Prague, Czech Republic

ABSTRACT Eleven protein-DNA crystal structures were analyzed to test the hypothesis that hydration sites predicted in the first hydration shell of DNA mark the positions where protein residues hydrogen-bond to DNA. For nine of those structures, protein atoms, which form hydrogen bonds to DNA bases, were found within 1.5 Å of the predicted hydration positions in 86% of the interactions. The correspondence of the predicted hydration sites with the hydrogen-bonded protein side chains was significantly higher for bases inside the conserved DNA recognition sequences than outside those regions. In two CAP-DNA complexes, predicted base hydration sites correctly marked 71% (within 1.5 Å) of protein atoms, which form hydrogen bonds to DNA bases. Phosphate hydration was compared to actual protein binding sites in one CAP-DNA complex with 78% marked contacts within 2.0 Å. These data suggest that hydration sites mark the binding sites at protein-DNA interfaces.

INTRODUCTION

The influence of water on the conformation and interactions of nucleic acids has been the subject of many investigations (Berman, 1991, 1994; Westhof, 1993) ever since the structure of DNA was discovered to be dependent on the relative humidity (Franklin and Gosling, 1953). Recently, the availability of a large number of single crystal structures of oligonucleotides has made it possible to use knowledge-based approaches to predict hydration sites around the bases in DNA helices (Schneider et al., 1993; Schneider and Berman, 1995). These studies have established that the positions of the hydration sites are dependent on base type and DNA conformational class. Therefore, changes in base sequence and base morphology result in different hydration patterns. Further analysis of the hydration sites around DNA phosphate groups has also revealed that these hydration sites also depend on conformation and sequence (Schneider et al., 1998).

It has been proposed that positions of protein-nucleic acid hydrogen-bonding interactions are "marked" by DNA hydration. That is, protein atoms involved in binding to DNA occupy positions normally occupied by water molecules in unbound DNA (Seeman et al., 1976). To test this hypothesis, we have used the hydration site prediction method mentioned above to examine the interface of several protein-DNA complexes. The positions of the predicted hydration sites for DNA were compared with the crystallographically observed positions of protein side chains that bind to DNA. Special attention was paid to the conserved regions of binding to determine whether these sites are preferentially marked.

MATERIALS AND METHODS

Eleven structures were selected for study (Table 1). Several contain the helix-turn-helix motif: CAP-DNA$_{GCE}$ (Parkinson et al., 1996a); CAP-DNA$_{CON}$ (Parkinson et al., 1996b); 434 repressor-operator OR1 (Aggarwal et al., 1988); 434 repressor-operator OR2 (Shimon and Harrison, 1993); 434 Cro-operator OR1 (Harrison et al., 1988); engrailed homeodomain-DNA (Kissinger et al., 1990); lambda repressor (Beamer and Pabo, 1992); Mat α-2 homeodomain (Wolberger et al., 1991); and the trp repressor-operator (Otwinowski et al., 1988). The GAL-4 protein-DNA complex (Marmorstein et al., 1992) contains the leucine zipper motif and the ZIF268-DNA complex (Pavletich and Pabo, 1991) contains the zinc finger.

The intermolecular distances up to 3.5 Å between the DNA and the protein were calculated using the program BANG (Carrell, 1979) and DIST (Cohen et al., 1995) and potential hydrogen bonds were identified between hydrophilic atoms of the protein amino acid residues and the DNA bases and phosphates.

In all 11 protein-DNA complexes, the hydration sites of the DNA bases were predicted using the method developed by Schneider et al. (1993). The bases in the DNA molecules of crystalline protein-nucleic acid complexes were overlapped by the hydrated building blocks that have been derived for each base from higher resolution B-DNA oligomer crystal structures (Schneider and Berman, 1995). After superposition, the overall distribution of the water molecules around a DNA sequence was Fourier-averaged to obtain pseudoelectron densities. The positions of hydration sites were determined by manual-fitting the highest pseudoelectron densities using the program CHAIN (Sack, 1988). The hydration sites predict positions where water occurs with the highest probability.

The phosphate hydration sites were also predicted for the CAP-DNA$_{GCE}$ complex (Parkinson et al., 1996a) using the hydration model by Schneider et al. (1998). First the phosphate groups in the backbone were classified according to whether they were in the BI or BII conformations. BI conformation was defined as having the backbone torsion angles ζ (C3′-O3′-P-O5′) $> 240°$ and ϵ (C4′-C3′-O3′-P) $< 210°$; the BII conformation was defined with $\zeta < 210°$ and $\epsilon > 210°$. Hydrated phosphate building blocks were then used to derive the hydration sites around the phosphates using a procedure analogous to the one used for predicting base hydration.

Distances between the predicted hydration sites and the protein atoms in contact with DNA were measured in all complexes. If the distance between the amino acid atom and the hydration site was within 1.5 Å for bases and 2.0 Å for phosphates, the amino acid position was considered as "marked" by the hydration site. Water molecules that bridge protein and nucleic acid atoms in protein-nucleic acid complexes were evaluated using the same criteria. The root-mean-square deviations (rmsd) between the hydration

Received for publication 29 May 1998 and in final form 10 August 1998.

Address reprint requests to Dr. Helen M. Berman, Department of Chemistry, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854-8087. Tel.: 732-445-4667; Fax: 732-445-4320; E-mail: berman@adenine.rutgers.edu.

0006-3495/98/11/2170/08 $2.00

TABLE 1 Structures of protein-DNA complexes used in this analysis

Structure	Resolution (Å)	R-Factor (%)	DNA-Binding Motif*	Reference
CAP-DNA (GCE)	2.5	19.7	HTH	Parkinson et al., 1996a
CAP-DNA (CON)	2.7	19.8	HTH	Parkinson et al., 1996b
434 Repressor-Operator (OR1)	2.5	17.9	HTH	Aggarwal et al., 1988
434 Repressor-Operator (OR2)	2.5	20.9	HTH	Shimon and Harrison, 1993
434 Cro-Operator (OR1)	2.5	22.0	HTH	Harrison et al., 1988
Lambda Repressor-Operator	1.8	18.9	HTH	Beamer and Pabo, 1992
Mat-Alpha2 Homeodomain-DNA	2.7	22.6	HTH	Wolberger et al., 1991
Gal4-DNA	2.7	23.0	L-zip	Marmorstein et al., 1992
Engrailed Homeodomain-DNA	2.8	22.5	HTH	Kissinger et al., 1990
Zif268-DNA	2.1	18.2	ZnF	Pavletich and Pabo, 1991
Trp Repressor-Operator	1.9	16.7	HTH	Otwinowski et al., 1988

*Protein motif bound to DNA. HTH, helix-turn-helix; L-zip, leucine zipper; ZnF, zinc-finger motif.

sites and either the experimentally observed water positions or the hydrogen-bonded protein atoms were calculated. A statistical zI test (Langley, 1971) was used to compare the number of marked protein binding positions to the number of unmarked sites to determine whether the number of predicted contacts is statistically significant.

RESULTS

The interface in CAP-DNA complexes

The predicted base hydration sites agree well with the protein contacts to the DNA bases in both CAP-DNA complexes analyzed (Fig. 1). The rmsd values are close to 1 Å for the unbent DNA sequences that bind to CAP, and the average rmsd is 1.4 Å for the two bent DNA sequences 5′-TGTGA-3′ and 5′-TCACA-3′. Fig. 1 illustrates the base and protein contact sites as well as how well the predicted hydration sites agree with actual positions of the contacting protein atoms.

The two CAP-DNA complexes have different DNA sequences and locally different basepair geometry, although their overall structures are virtually the same. A large bend of the DNA in both CAP-DNA complexes results in the fusion of hydration sites from various bases within the bend (Fig. 2). Although the deviation from a typical B type conformation is very large, 71% of the contacts between protein and base atoms are marked within 1.5 Å.

There are 20 contacts between protein atoms and phosphate charged oxygens that are shorter than 3.50 Å; there are none to either O5′ or O3′. These contacts are shown in Fig. 1 *a*, with the distances between protein atoms and their closest predicted hydration sites marked. Fifteen (75%) of the contacting protein atoms had a predicted hydration site within 2.0 Å, with eight atoms marked within 1.5 Å. The average distance from the hydration sites to protein atoms was 1.6 Å (estimated standard deviation, esd, 0.6 Å).

The consensus region (Parkinson et al., 1996a), indicated in boldface in Fig. 1 *a*, has four protein nitrogen atoms that contact phosphates of residues 4 and 6 (Fig. 3). The biochemically important contacts between R169 and T4 and its symmetrically related partner are both marked.

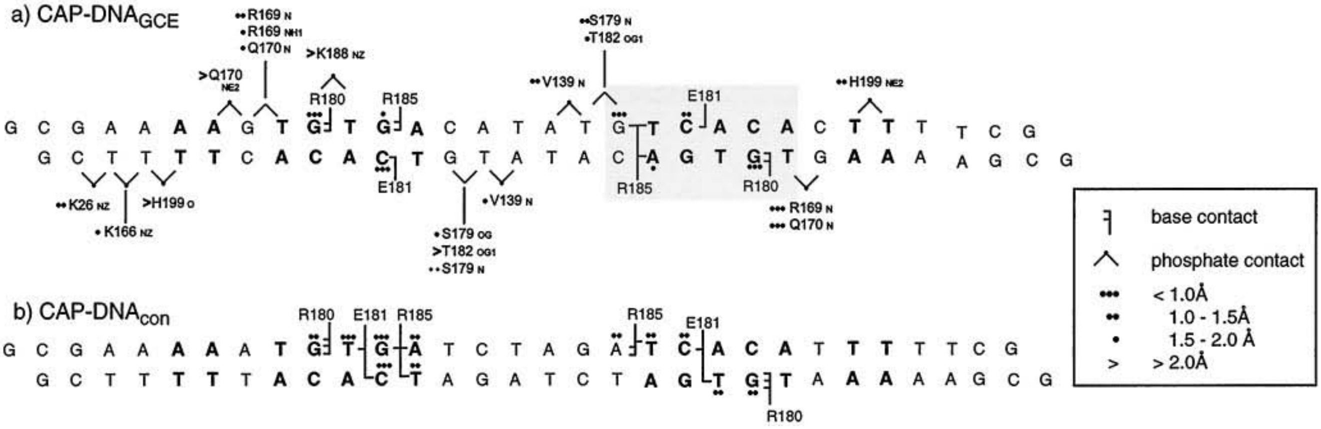


FIGURE 1 The DNA sequences in the two CAP-DNA complexes for which protein interaction sites were predicted. Sequence (*a*) (Parkinson et al., 1996a) shows both base and phosphate hydration while sequence (*b*) (Parkinson et al., 1996b) shows only base hydration. The consensus sequences are bold. The interacting protein residues are indicated by one-letter amino acid codes. Triple dots indicate that the predicted sites are within 1.0 Å of the observed sites; double dots indicate the agreement is between 1.0 and 1.5 Å; single dots 1.5–2.0 Å. (>), the hydration site is >2.0 Å from the observed protein atoms.

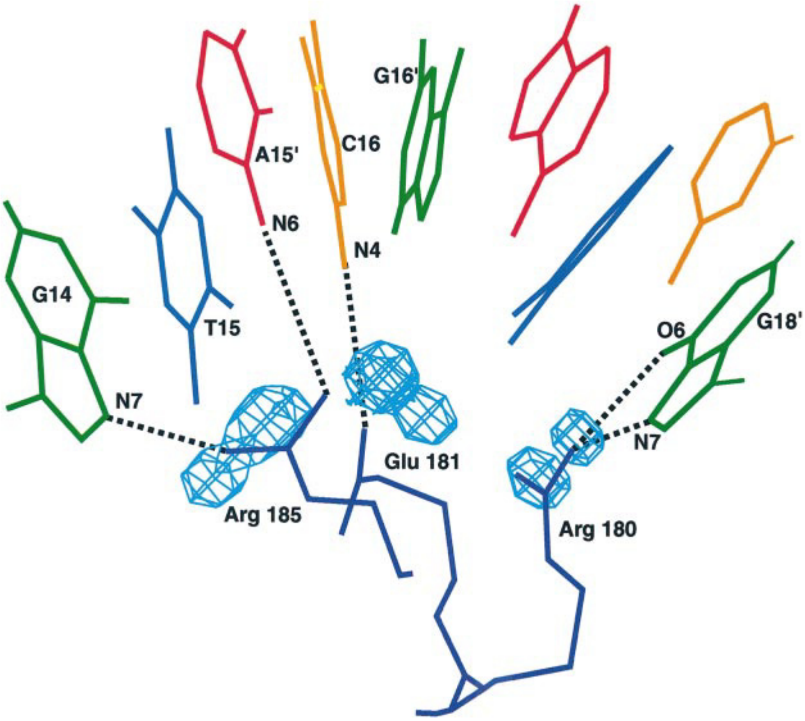


FIGURE 2 The interface between CAP and DNA_{GCE} in the high resolution complex (Parkinson et al., 1996a) showing base hydration in the bent part of the sequence. The predicted base hydration is drawn as pseudoelectron density in cyan and the interacting protein residues are shown in dark blue. The region shown in this figure is shaded in Fig. 1.

The positions of the 57 crystallographically observed water molecules hydrogen-bonded to phosphate oxygens of the CAP-DNA_{GCE} complex are predicted well. The average distance between experimentally observed water sites to predicted hydration site is 1.6 Å (esd 0.6 Å). Twelve water molecules, which represent 20% of all phosphate-bound waters, are associated with ester oxygens; these contacts constitute <15% of contacts in higher resolution B-DNA

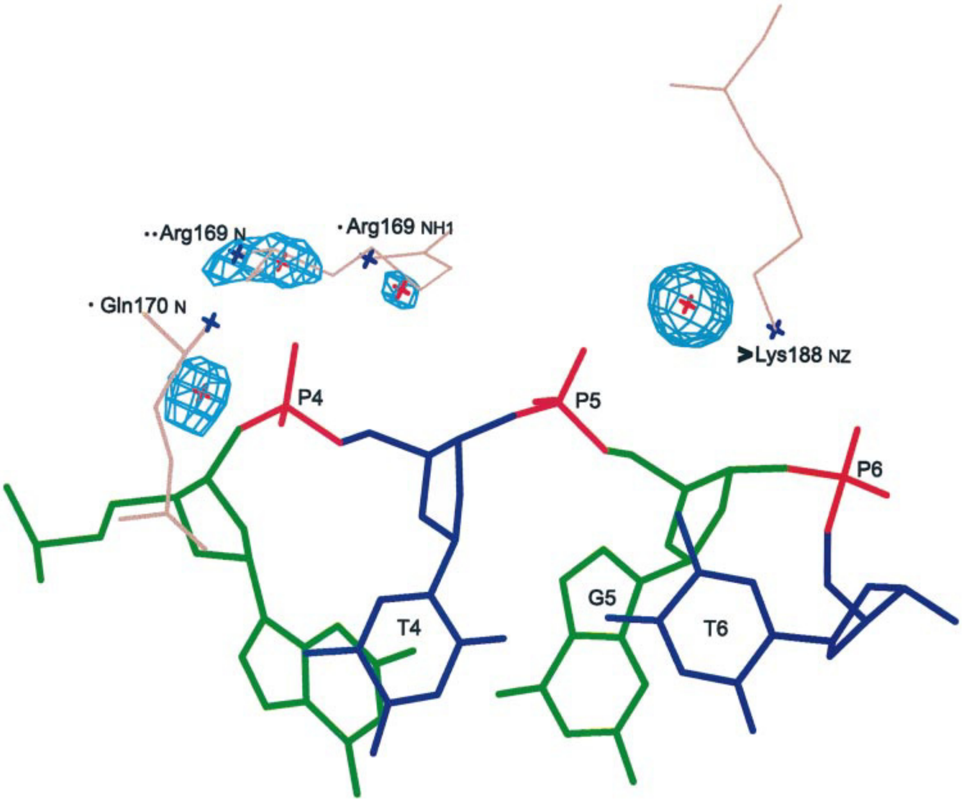


FIGURE 3 A view of the three residues in the consensus region for the high resolution CAP-DNA_{GCE} complex (Parkinson et al., 1996a). The predicted phosphate hydration is drawn as pseudoelectron density in cyan, the interacting protein residues are shown in dark brown, and the phosphate groups are red. The protein atoms that contact the DNA are shown as blue crosses. The predicted sites are the red crosses.

structures (Schneider et al., 1998). Overall, using 2.0 Å as a criterion, 71% of the water molecules bound to phosphate oxygens are predicted.

Base-protein interaction sites

Fig. 4 presents a summary of the analyses of base-protein interactions for the remaining nine protein-DNA complexes. For each DNA sequence, the crystallographically observed protein contacts, the indicators of how well the positions of protein atoms are predicted by the hydration sites and the rmsd values between the protein atoms and hydration sites are shown. Examples of the fit between the experimentally observed protein contact points and the predicted sites are shown in Fig. 5. The overall agreement is good.

The DNA at the interface of OR1 is straight and the contacts are well predicted (Fig. 4 *a*). Fig. 5 *a* shows a typical example of how well the protein side chain fits into the predicted hydration site. The contact sites for OR2 are not predicted as well (Fig. 4 *b*) but are still within an acceptable range (Fig. 5 *b*). The Cro-DNA interface is very well marked, as shown in Figs. 4 *c* and 5 *c*. The lambda-DNA interface is characterized by both water-mediated and direct contacts (Fig. 4 *d*). Although the contact shown in Fig. 5 *d* is good, there are other contacts that are not as

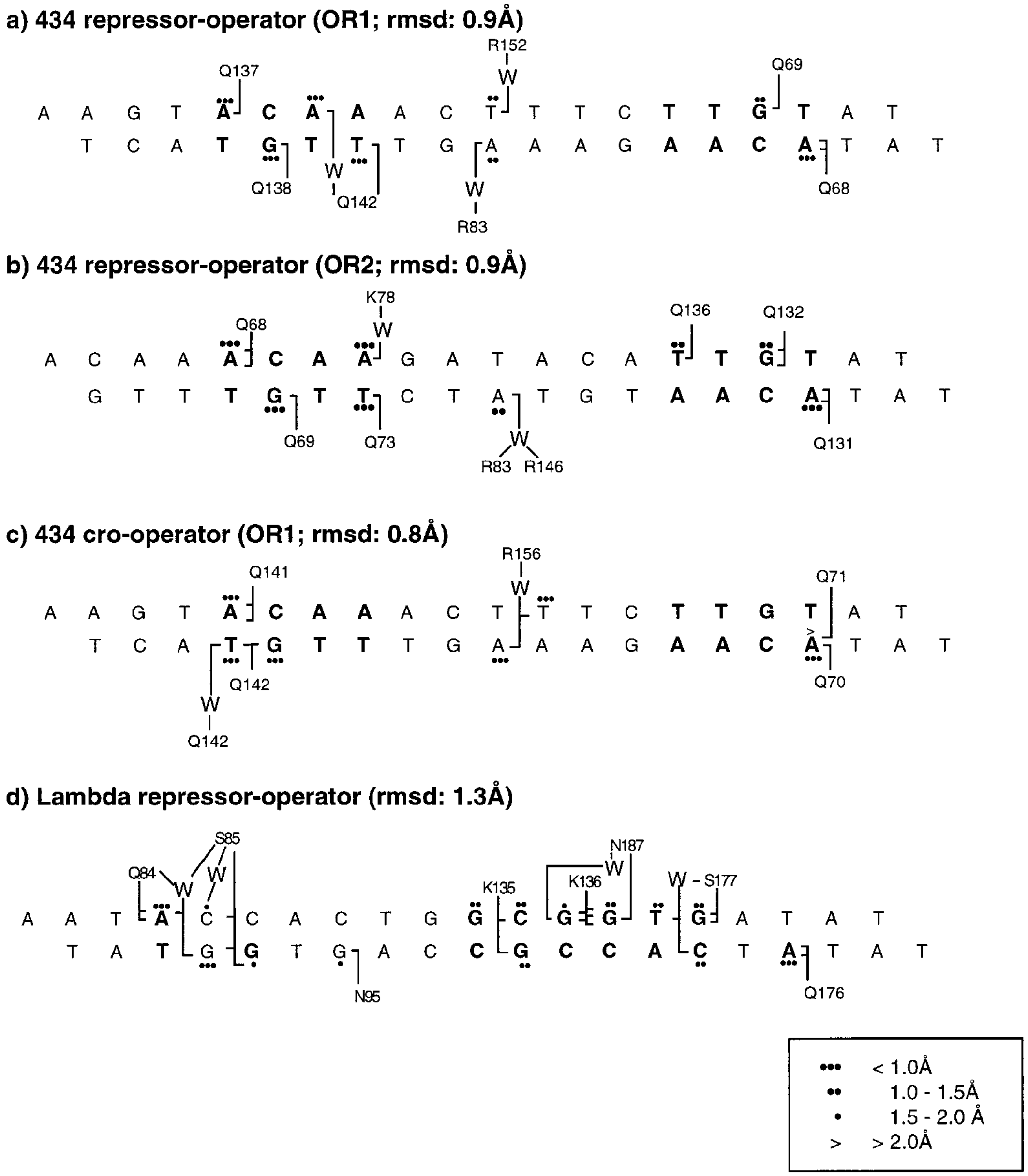
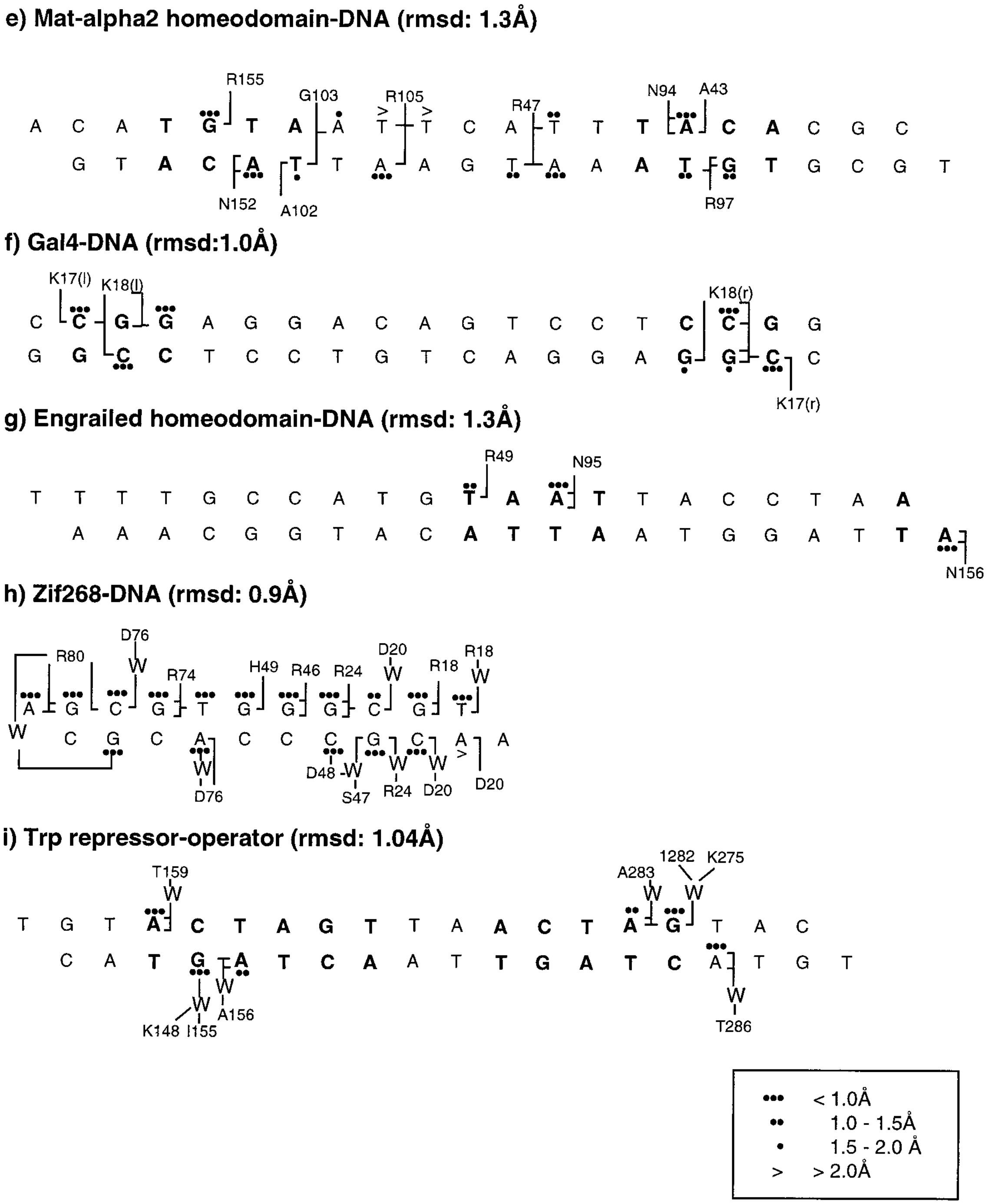

a) 434 repressor-operator (OR1; rmsd: 0.9Å)
R152
W
Q137
Q69
A A G T A C A A A C T T T C T T G T A T
T C A T G T T T G A A A G A A C A T A T
W
Q138
Q142
W
R83
Q68
b) 434 repressor-operator (OR2; rmsd: 0.9Å)
K78
W
Q68
Q136
Q132
A C A A A C A A G A T A C A T T G T A T
G T T T G T T C T A T G T A A C A T A T
Q69
Q73
W
R83 R146
Q131
c) 434 cro-operator (OR1; rmsd: 0.8Å)
R156
W
Q141
Q71
A A G T A C A A A C T T T C T T G T A T
T C A T G T T T G A A A G A A C A T A T
Q142
W
Q142
Q70
d) Lambda repressor-operator (rmsd: 1.3Å)
S85
Q84
W
W
N187
W
K135
K136
W - S177
A A T A C C A C T G G C G G T G A T A T
T A T G G T G A C C G C C A C T A T A T
N95
Q176
••• < 1.0Å
•• 1.0 - 1.5Å
• 1.5 - 2.0 Å
> > 2.0Å

FIGURE 4 The nine DNA sequences for which protein binding sites were correlated with the predicted base hydration sites of the DNA. The consensus sequences are bold. The interacting protein residues are indicated by one-letter amino acid codes; water-mediated protein-DNA contacts are labeled "W." The closeness of prediction is coded as in Fig. 1.



e) Mat-alpha2 homeodomain-DNA (rmsd: 1.3Å)

R155
G103
R105
R47
N94
A43
A C A T G T A A T T C A T T T A C A C G C
G T A C A T T A A G T A A A T G T G C G T
N152
A102
R97

f) Gal4-DNA (rmsd:1.0Å)

K17(l)
K18(l)
K18(r)
C C G G A G G A C A G T C C T C C G G
G G C C T C C T G T C A G G A G G C C
K17(r)

g) Engrailed homeodomain-DNA (rmsd: 1.3Å)

R49
N95
T T T T G C C A T G T A A T T A C C T A A
A A A C G G T A C A T T A A T G G A T T A
N156

h) Zif268-DNA (rmsd: 0.9Å)

D76
R80
W
R74
H49
R46
R24
D20
W
R18
R18
W
A G C G T G G G C G T
W C G C A C C C G C A A
W
D76
D48
W
S47
W
R24
W
D20
D20

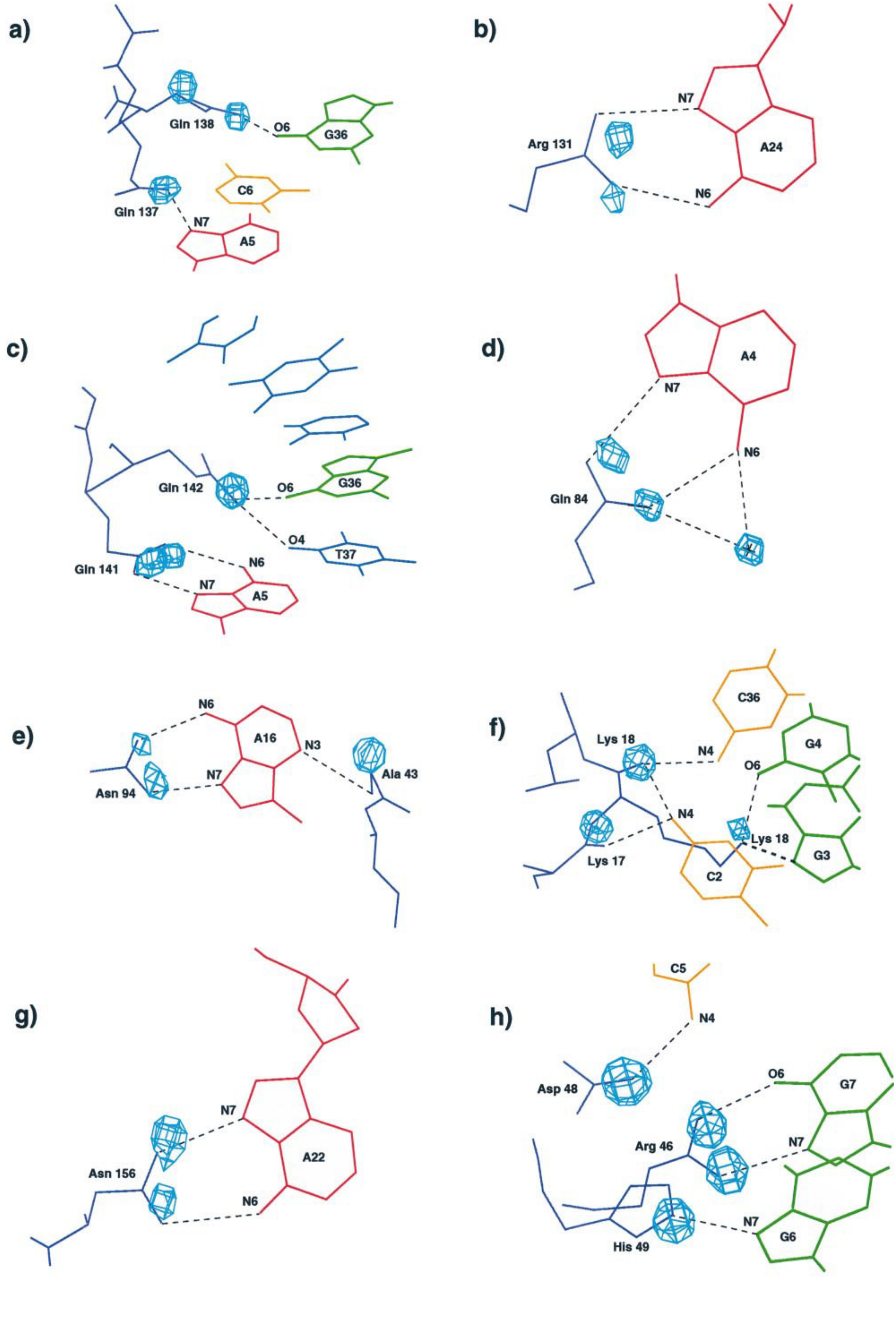
i) Trp repressor-operator (rmsd: 1.04Å)

T159
W
A283
1282
K275
W W
T G T A C T A G T T A A C T A G T A C
C A T G A T C A A T T G A T C A T G T
W
W
K148
I155
A156
W
T286

•••	< 1.0Å
••	1.0 - 1.5Å
•	1.5 - 2.0 Å
>	> 2.0Å

Figure 4 Continued

good, such as the contacts involving serine 85 (Fig. 4 *d*). Mat-α2 contacts DNA in both conserved and nonconserved regions. The contact sites are well predicted in some cases and poorly predicted in others; for example, the interface between arginine R105 and basepairs 9 and 10 is poorly predicted, but the sites around adenine 16 are well predicted (Figs. 4 *e* and 5 *e*). The pattern of interaction around Gal-4 involves both the side chain and the backbone atoms of the protein. Despite this complexity, the interaction sites coincide well with the predicted sites (Figs. 4 *f* and 5 *f*). Of the proteins examined here, the engrailed homeodomain shows the poorest correlation between experimentally observed and predicted sites in terms of their rmsd (Fig. 4 *g*), but the hydration densities and contacting protein atoms still correspond closely (Fig. 5 *g*). Interestingly, this structure has the fewest interaction sites. In contrast, the zinc finger sites are uniformly very well predicted (Figs. 4 *h* and 5 *h*) and this interface also has the most interaction sites. Predictions


a)
Gln 138
O6
G36
C6
Gln 137
N7
A5
b)
N7
Arg 131
A24
N6
c)
Gln 142
O6
G36
O4
T37
Gln 141
N6
N7
A5
d)
A4
N7
N6
Gln 84
e)
N6
A16
N3
Ala 43
N7
Asn 94
f)
C36
Lys 18
N4
O6
G4
N4
Lys 18
Lys 17
C2
G3
g)
N7
A22
Asn 156
N6
h)
C5
N4
O6
Asp 48
G7
Arg 46
N7
N7
His 49
G6

for the trp repressor-operator, in which the protein-DNA interactions are mediated by water, are good (Fig. 4 *i*, not shown in Fig. 5).

A total of 85 protein atoms make direct contacts to the DNA bases in the nine analyzed non-CAP complexes, and 73 of these atoms are within 1.5 Å of the closest hydration site. Therefore, at this cutoff we can consider that 86% of the sites are marked and there are significantly more marked than unmarked positions at a 5% confidence level. A majority of protein atoms hydrogen-bonded to DNA bases are side chain atoms; of those 77, 68 are marked, and of the eight main chain atoms that hydrogen-bond to the DNA bases, five are marked by hydration sites.

Water-mediated contacts are present in the lambda repressor complex, the phage 434/OR1 complex, the phage 434/OR1 Cro complex, the phage 434/OR2 complex, and the Zif268 complex. These are the only type of contacts in the trp repressor-operator complex. Quite significantly, of the 25 water-mediated contacts in the protein-DNA complexes examined, 24 of the sites were predicted within 1.5 Å.

Interactions in the conserved regions of the complexes

Conserved regions are those parts of the DNA sequence that are labeled by binding assays as crucial for specific protein-DNA binding. Of the 11 complexes studied, 10 contained regions of conserved bases, and for those the predictions of base-protein contacts based on the hydration model were compared in the conserved and nonconserved regions to determine whether the ratio of successful predictions is the same or different in both regions. Outside of the conserved regions, 54% of the bases that made protein contacts had those binding protein atoms marked by hydration sites. However, within the conserved regions, 86% of the bases that made protein contacts had at least one of the binding protein atoms marked by predicted hydration sites and 81% of the interacting bases in the conserved regions had all of the protein atoms involved in those contacts marked. Using the zI test at a 5% confidence level, there are significantly more marked contacts to bases within the conserved regions. On the other side, in the nonconserved regions, there is no significant difference between the number of marked and unmarked protein positions.

DISCUSSION

The concept of conserved hydration sites in water-mediated protein-DNA complexes has been explored by Shakked et al. (1994) who showed that the hydration sites in the recognition DNA sequence in crystal structures of both free DNA and of DNA complexed with trp repressor/operator (Otwinowski et al., 1988) are the same. In that case, the conserved water molecules mediate the protein-DNA contacts.

In the analysis presented here, we show that in direct protein-DNA complexes, the protein atoms that form hydrogen bonds to DNA reside close to the hydration sites predicted for free DNA. Analysis of 11 protein-nucleic acid complexes shows that the positions of the protein atoms are "marked" by these predicted hydration sites.

The CAP-DNA structure demonstrates that models of base and phosphate hydration shells mark the interacting protein atoms with approximately the same accuracy; 70% of the protein-base interactions are marked within 1.5 Å, and 78% of the protein-phosphate interactions are marked within 2.0 Å. The first hydration shells around phosphates and bases are independent of each other, as is revealed by the fact that there are no hydration sites common to both phosphates and bases, and there are no contacts made by the same protein atoms to both base and phosphate.

The average rmsd between the crystallographically determined water positions and the predicted hydration sites is 1.0 Å. Thus, the experimental water positions are accurately predicted by the hydration sites. This result reinforces evidence from previous studies that the hydration sites represent the position of the actual water molecules bound to the DNA bases or phosphates.

The algorithm used here to predict DNA hydration sites in protein-DNA complexes can be used to predict hydration sites around any DNA sequence with either experimentally determined or modeled 3-D structure. Therefore, hydration sites predicted around a sequence with, for instance, a known regulatory function yet with unknown protein-DNA structure can then be helpful in understanding potential binding of the regulatory protein.

The percentage of binding protein atoms that fit in the hydration sites is larger in the conserved regions than in the nonconserved regions. This result may be a reflection of the small number of protein-DNA base interactions outside of the conserved region. However, since interactions in the conserved regions are critical for high affinity protein binding, there might be another explanation. Protein-DNA interactions are energetically driven by interactions within the conserved region, and these interactions are therefore optimized. A good correlation between observed and predicted contacts in the conserved regions is then perhaps a corollary of the fact that the hydration sites represent the most probable and energetically most favorable binding positions for hydrophilic DNA binders. Binding outside these regions is energetically less optimized and correlation with the hydration sites is worse. Further analyses of DNA-protein com-

FIGURE 5 Examples of the protein-DNA interfaces. The pseudoelectron densities are the predicted sites. The actual observed protein and DNA residues are shown. (*a*) OR1-DNA (Aggarwal et al., 1988); (*b*) OR2-DNA (Shimon and Harrison, 1993); (*c*) Cro-DNA (Harrison et al., 1988); (*d*) lambda-DNA (Beamer and Pabo, 1992); (*e*) mat α2-DNA (Wolberger et al., 1991); (*f*) gal4-DNA (Marmorstein et al. 1992); (*g*) engrailed homeodomain-DNA (Kissinger et al., 1990); (*h*) ZIF-DNA (Pavletich and Pabo, 1991).

plexes with different binding affinities are required to confirm this idea.

We thank Gary Parkinson for discussions about the CAP-DNA interface.

This work was funded by the National Institutes of Health Grant GM 21589 and by the grant agency of the Czech Republic (to B.S.). J.W. was a Henry Rutgers Scholar.

REFERENCES

Aggarwal, A. K., D. W. Rodgers, M. Drottar, M. Ptashne, and S. C. Harrison. 1988. Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science*. 242:899–907.

Beamer, L. J., and C. O. Pabo. 1992. Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. Mol. Biol*. 227:177–196.

Berman, H. M. 1991. Hydration of DNA. *Curr. Opin. Struct. Biol*. 1:423–427.

Berman, H. M. 1994. Hydration of DNA: take 2. *Curr. Opin. Struct. Biol*. 4:345–350.

Carrell, H. L. 1979. BANG: a computer program for the rapid calculation of bond lengths and angles. The Institute of Cancer Research, Fox Chase Cancer Center, Philadelphia.

Cohen, D., K. Vadaparty, and B. Dickinson. 1995. Efficient algorithms for geometric queries in macromolecular structure databases. University of Pittsburgh, Pittsburgh.

Franklin, R. E., and R. G. Gosling. 1953. Molecular configuration in sodium thymonucleate. *Nature*. 171:740–741.

Harrison, S. C., J. E. Anderson, G. B. Koudelka, A. Mondragon, S. Subbiah, R. P. Wharton, C. Wolberger, and M. Ptashne. 1988. Recognition of DNA sequences by the repressor of bacteriophage 434. *Biophys. Chem*. 29:31–37.

Kissinger, C. R., B. S. Liu, E. Martinblanco, T. B. Kornberg, and C. O. Pabo. 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell*. 63:579–590.

Langley, R. 1971. Practical Statistics. Dover Publications, New York.

Marmorstein, R., M. Carey, M. Ptashne, and S. C. Harrison. 1992. DNA recognition by Gal4: structure of a protein-DNA complex. *Nature*. 356:408–414.

Otwinowski, Z., R. W. Schevitz, R.-G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, and P. B. Sigler. 1988. Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature*. 335:321–329.

Parkinson, G., A. Gunasekera, J. Vojtechovsky, X. Zhang, T. A. Kunkel, H. Berman, and R. H. Ebright. 1996b. Aromatic hydrogen bond in sequence-specific protein-DNA recognition. *Nature Struct. Biol*. 3:837–841.

Parkinson, G., C. Wilson, A. Gunasekera, Y. Ebright, R. H. Ebright, and H. M. Berman. 1996a. Structure of the CAP-DNA complex at 2.5 angstrom resolution: a complete picture of the protein-DNA interface. *J. Mol. Biol*. 259:395–408.

Pavletich, N. P., and C. O. Pabo. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*. 252: 809–817.

Sack, J. S. 1988. CHAIN: a crystallographic modeling program. *J. Mol. Graphics*. 6:224–225.

Schneider, B., and H. M. Berman. 1995. Hydration of the DNA bases is local. *Biophys. J*. 69:2661–2669.

Schneider, B., D. M. Cohen, L. Schleifer, A. R. Srinivasan, W. K. Olson, and H. M. Berman. 1993. A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. *Biophys. J*. 65:2291–2303.

Schneider, B., K. Patel, and H. M. Berman. 1998. Hydration of the phosphate group in double helical DNA. *Biophys. J*. 75:2422–2434.

Seeman, N. C., J. M. Rosenberg, and A. Rich. 1976. Sequence specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA*. 73:804–808.

Shakked, Z., G. Guzikevich-Guerstein, F. Frolow, D. Rabinovich, A. Joachimiak, and P. B. Sigler. 1994. Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature*. 368:469–473.

Shimon, L. J. W., and S. C. Harrison. 1993. The phage 434 O_R2/R1–69 complex at 2.5 Å resolution. *J. Mol. Biol*. 232:826–838.

Westhof, E. 1993. Structural Water Bridges in Nucleic Acids. Water and Biological Macromolecules. CRC Press, Boca Raton.

Wolberger, C., A. K. Vershon, B. Liu, A. D. Johnson, and C. O. Pabo. 1991. Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell*. 67:517–528.