J. Mol. Biol. (1995) **254**, 657–667

JMB

Crystal Structure of a bZIP/DNA Complex at 2.2 Å: Determinants of DNA Specific Recognition

Walter Keller, Peter König and Timothy J. Richmond*

Institut für Molekularbiologie und Biophysik
ETH-Hönggerberg, CH-8093 Zurich, Switzerland

The X-ray structure of the GCN4-bZIP protein bound to DNA containing the ATF/CREB recognition sequence has been refined at 2.2 Å. The water-mediated interactions between the basic domain and DNA are revealed, and combined with a more accurate description of the direct contacts, further clarify how binding specificity is achieved. Water molecules extend the interactions of both invariant basic domain residues, asparagine 235 and arginine 243, beyond their direct base contacts. The slight bending of the basic domain α-helix around the DNA facilitates the linking of arginine 241, 243 and 245 to main-chain carbonyl oxygen atoms *via* water molecules, apparently stabilizing interactions with the DNA.

Keywords: protein-DNA recognition; bZIP/GCN4; transcription factor; X-ray crystallography

**Corresponding author*

Introduction

The yeast transcription factor GCN4 activates genes containing the DNA sequence 5'-ATGACT-CAT-3' (Arndt & Fink, 1986; Hope & Struhl, 1987). The bZIP domain of GCN4 is sufficient to account for its dimerization and DNA binding properties, as is generally the case for the family of bZIP homology proteins. Two high-affinity DNA binding sites that show comparable affinities for the GCN4-bZIP were identified *in vitro*: the pseudosymmetric AP-1 site ATGA$^{C}/_{G}$TCAT and the 2-fold symmetric ATF/CREB site ATGAC|GTCAT (Weiss *et al*., 1990; Suckow *et al*., 1993). Full-length GCN4 can activate transcription from both types of sites *in vivo*, and the lower level of activity from ATF/CREB sites is likely due to competitive binding by the ACR-1 repressor (Sellers *et al*., 1990; Vincent & Struhl, 1992). The structures of the GCN4-bZIP domain in complex with AP-1 and ATF/CREB site DNAs were previously solved at 2.9 Å and 3.0 Å, respectively (Ellenberger *et al*., 1992; König & Richmond, 1993). A comparison of the two structures revealed that flexibility of the DNA, as well as of the protein, allows tight binding to both sequences despite a difference between them of one base-pair in the half-site spacing (König & Richmond, 1993). Although many aspects of DNA specific binding were explained by the direct contacts between protein and DNA seen in the lower-resolution structures, it is well recognized that water molecules can play a crucial role in the specificity of protein-DNA interaction (Joahimiak *et al*., 1991). Here, we present the refined structure of a bZIP/DNA complex at 2.2 Å and show that water molecules are an integral component of the protein-DNA interface, contributing significantly to specific recognition.

Results

The crystals studied contain the GCN4-bZIP dimer and a palindromic DNA molecule incorporating the ATF/CREB sequence (Figure 1). The C-terminal halves of each monomer form the leucine zipper, a parallel coiled coil of α-helices, and from a position over the major groove the coiled coil extends away from the DNA in a direction perpendicular to the overall DNA helix axis. The aligned 2-fold axes of the protein and DNA obey crystallographic symmetry. The N-terminal basic domains run to opposite sides of the DNA double helix as continuous extensions from the leucine zipper and are embedded in the DNA major groove over a 12 bp region. In addition to the direct base contacts made by five amino acid residues, Asn235, Thr236, Ala238, Ala239 and Arg243, and the hydrogen bonds to the phosphate backbone made by seven amino acid residues, Arg232, Arg234, Arg240, Arg241, Ser242, Arg243 and

Present address: P. König, MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK

0022–2836/95/490657–11 $12.00/0

Arg245 (Figure 2), extensive networks of solvent molecules partake in both forms of interaction. Overall, 46 water molecules were located by the structure refinement (Table 1).

Protein–DNA interactions

The two bZIP domain ''invariant'' residues, Asn235, and Arg243, as well as Lys246, make

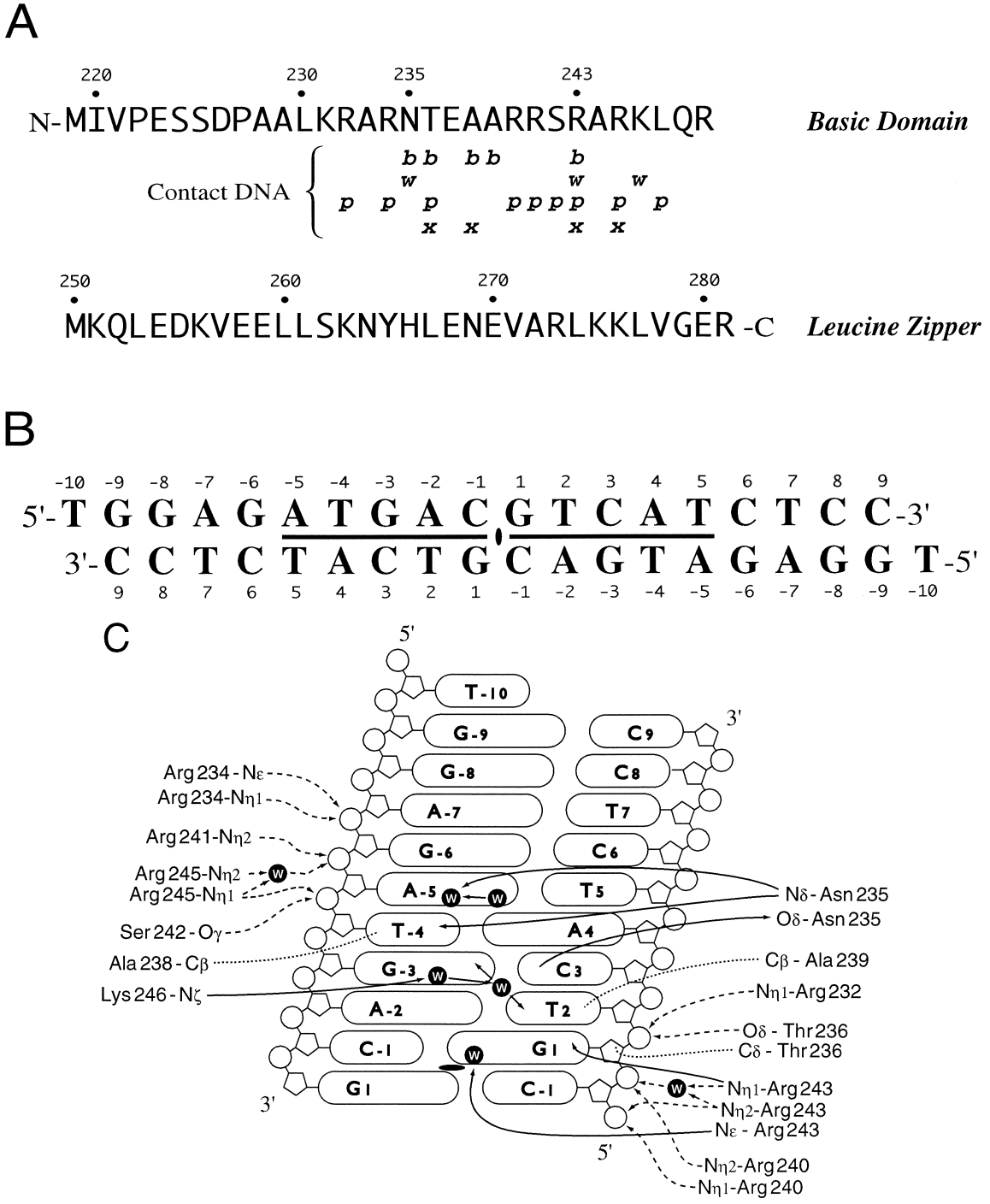


Figure 1. The protein and DNA sequences of the complex crystallized and interaction summary. A, GCN4-bZIP domain numbered according to the full-length sequence. Residues that make one or more DNA contacts are indicated: *b*, direct to base; *w*, *via* water to base; *p*, direct to phosphate; *x*, *via* water to phosphate. B, Self-complementary, 18 bp palindromic ATF/CREB DNA with terminal thymidine. A molecular and crystallographic 2-fold axis passes through the $C_{-1}G_1$ base step. The specific recognition sequence is marked with the bar. C, Protein-DNA interaction summary for one basic domain and DNA half-site. The location of the 2-fold axis is indicated. Direct and water (W) mediated interactions with bases are shown as continuous lines, interactions with the phosphate backbone as broken lines and hydrophobic interactions as dotted lines. Arrows indicate the direction of hydrogen bonds.

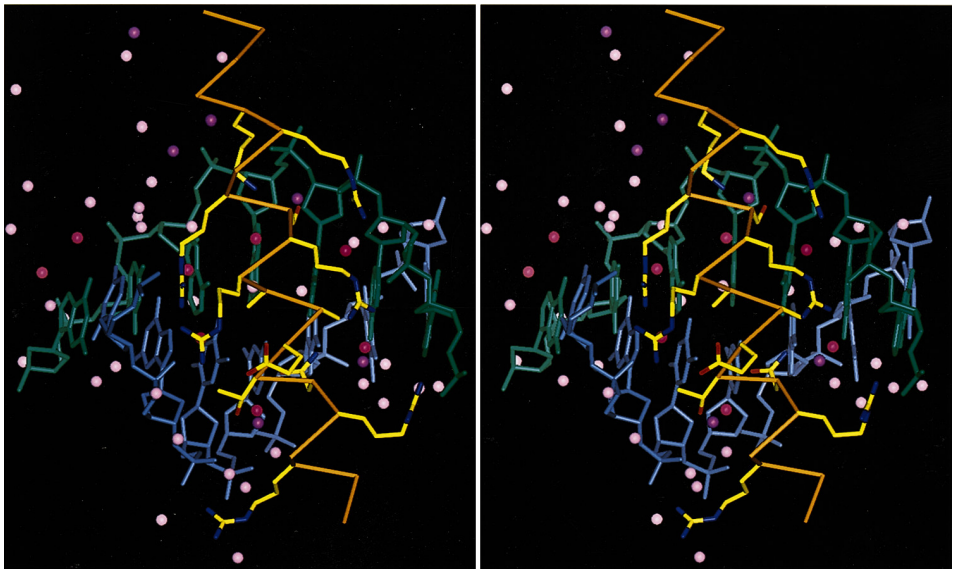


Figure 2. Water positions in the GCN4-bZIP/DNA complex. The stereodiagram shows a single basic domain α-helix (gold, from top to bottom, amino acid residues 250 to 230) and ATF/CREB DNA half-site (from left to right, bases G_1 to G_{-6} in green, bases C_{-1} to C_6 in blue). Solvent molecules identified at 46 sites in the first and second hydration shells are shown and classified: 1, water molecules involved in protein-DNA interactions (red); 2, water binding to basic domain carbonyl groups (purple), two of these in arginine ''helix bridge'' structures also belong to class 1; and 3, other sites (orchid).

water-mediated base contacts. The side-chain of Asn235 is pointing into the DNA major groove and hydrogen bonds to N4 of base C_3 and the O4 of base T_{-4}, determining the specificity towards base-pairs at positions 3 and 4. The propeller twist of these two base-pairs also allows a cross-chain hydrogen bond between these N4 and O4 atoms. The additional hydrogen-bonding functions of the asparagine carboxyl oxygen and amide nitrogen atoms are occupied by water molecules, the latter contacting N7 of A_{-5} and explaining the high preference of GCN4 for a purine in position –5 of ATF/CREB or position –4 of AP-1 target site sequences (Oliphant *et al*., 1989; Mavrothalassitis *et al*., 1990). The water molecule bound between Asn235 and A_{-5} is further stabilized by an additional hydrogen bond to a water molecule bonded to N6 of A_{-5} (Figure 3A). The seven oxygen and nitrogen atoms involved in

Table 1. Refinement statistics

A. *R-factor comparison of observed and calculated structure factors*[a]

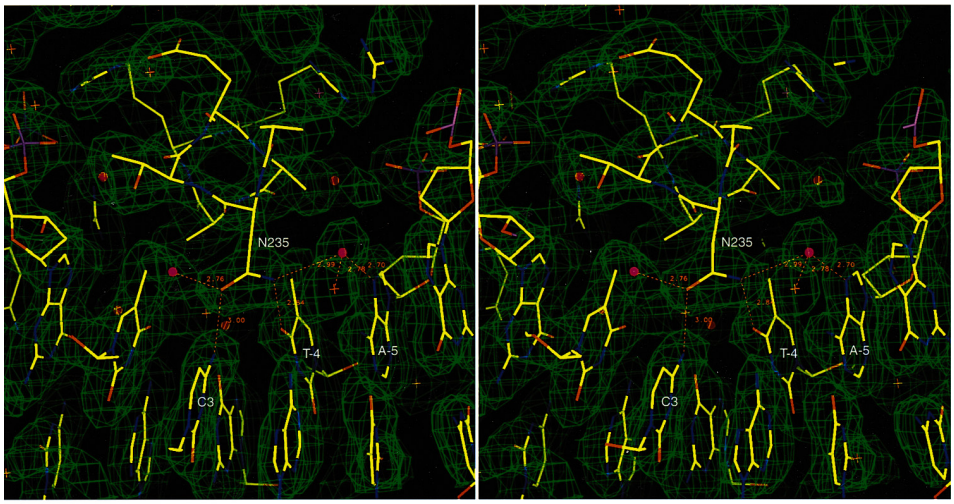
Resolution range (Å)		Working set		Test set		Completeness
Low	High	Number	*R*-factor	Number	*R*-factor	(%)
6.0	4.19	654	0.165	87	0.340	91.13
4.19	3.54	660	0.174	93	0.262	92.41
3.54	3.17	644	0.183	78	0.317	93.15
3.17	2.92	569	0.235	71	0.296	84.32
2.92	2.73	557	0.260	58	0.316	79.05
2.73	2.58	476	0.282	61	0.392	70.47
2.58	2.46	497	0.313	51	0.371	72.39
2.46	2.36	526	0.305	50	0.297	75.69
2.36	2.27	506	0.310	59	0.405	75.19
2.27	2.20	487	0.343	49	0.393	70.70
Overall		5576	0.214	657	0.316	80.7

B. *Model parameters*[b]

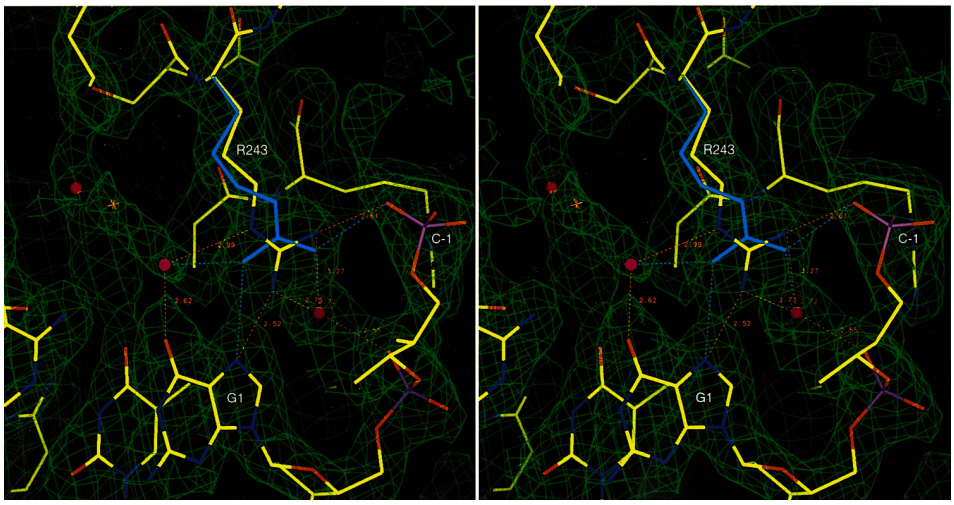
	Protein	DNA
Bond length rms deviation (Å)	0.006	0.022
Bond angle rms deviation (deg.)	1.13	3.64
Mean temperature factor (Å^2)	38.5	43.5

[a] The data in the working set were used in the structure refinement. The test set was used to calculate an unbiased measure of error (Brünger, 1992). *R*-factor = $\Sigma|F_o - F_c| / \Sigma F_o$, where F_o and F_c are the observed and calculated structure factor magnitudes. Completeness = % unique data measured.

[b] The asymmetric unit consists of Ala229 to Lys276, the 19 base ATF/CREB oligonucleotide, and 46 well-ordered solvent molecules.



A



B

Figure 3. The basic domain, invariant amino acids. A, Asn235 makes hydrogen bonds to three bases in the DNA major groove. The stereodiagram of electron density (green, 1.2σ level, $2F_o - F_c$ map) shows that the range of specific interaction of Asn235 extends beyond its direct contacts to N4 of C_3 and O4 of T_{-4} *via* a water molecule to N7 of A_{-5}. B, Arg243 occurs in two equally occupied conformations. A simulated-annealing OMIT map was calculated with Arg243 and all atoms in a sphere of 8 Å around C^{γ} omitted from the structure factor calculation (Hodel *et al*., 1992). Two independent Arg243 side-chain conformations were built in the electron density and refined to temperature factors of 19.6 $Å^2$ (blue side-chain and hydrogen bonds) and 32.6 $Å^2$ (CPK colored side-chain and red hydrogen bonds and distances). Parts of the phosphodiester chain are omitted for clarity.

the hydrogen-bonding network with Asn235 are essentially coplanar. Proper hydrogen-bonding geometry of this arrangement is preserved only when the plane of the amide group is inclined by approximately 35° with respect to the normal to the DNA helix axis.

Mutation of Asn235 can yield bZIP proteins with modified DNA binding specificity. In one case, substitution of this residue with histidine changes the requirement at position 3 from C to G (Suckow *et al*., 1994). The rearrangement suggested to account for this alteration was that the N^{ϵ} and N^{δ} of His235 form hydrogen bonds to the adjacent bases G_3 and A_4 of the same DNA strand. However, our model building indicates that this arrangement has an unlikely hydrogen-bonding geometry. Alternatively, when the plane of the histidine imidazole ring is oriented similarly to the asparagine amide group seen in the native complex, hydrogen bonds between N^{ϵ} and N7 of base G_3 directly and between N^{δ} and N7 of base A_{-5} *via* a water molecule can easily be constructed. This arrangement provides favorable hydrogen-bonding geometry without requiring any conformational rearrangements and simultaneously explains the necessity for purine in position –5. Natural variation in bZIP binding sites maintains the combination of a hydrogen bond donor and an acceptor in positions 3 and –4 to complement Asn235 (König & Richmond, 1993).

The side-chain of the other invariant amino acid residue, Arg243, is hydrogen bonded to G_1 at the

center of the binding site both directly from $N^{\eta 2}$ to N7 and indirectly from N^{ϵ} *via* a water molecule to O6 (Figure 3B). A second solvent molecule links this side-chain to the basic domain main-chain through the carbonyl oxygen atom of Thr236 (Figure 4B). In addition, the side-chain guanidinium group bridges between the phosphate groups of the two central nucleotides $C_{-1}G_1$ *via* direct and water-mediated interactions. As a consequence, the pucker of the C_1 ribose is clearly C3' *endo*, and the P-P distance of

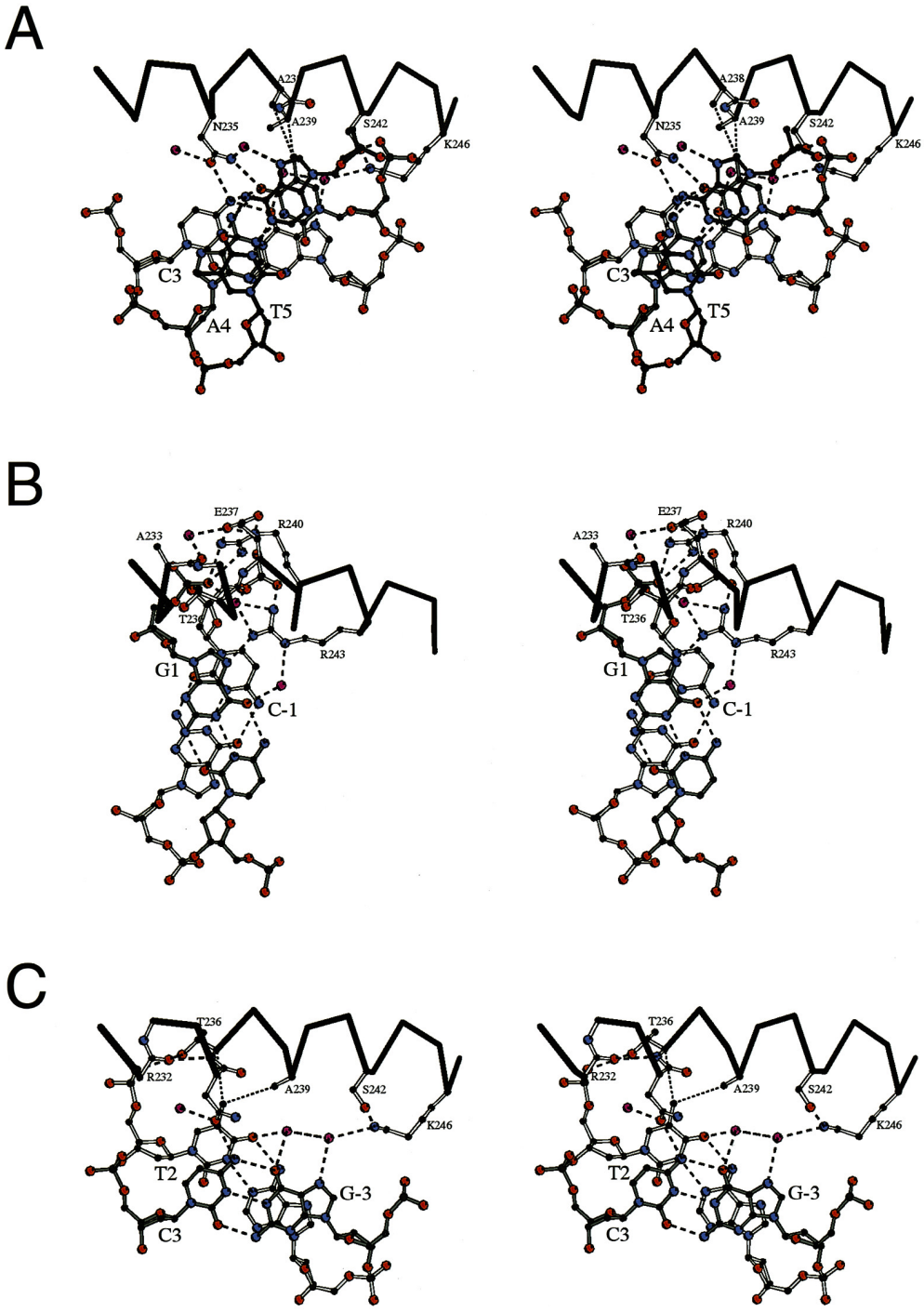


Figure 4. Stereo diagrams of the base-specific interactions. The views are down the overall DNA helix axis, and hydrogen bonds (broken) and hydrophobic contacts (dotted) are indicated. A, Interaction of Asp235 with three consecutive base-pairs and the hydrophobic contacts of the T_{-4} methyl group. B, Interactions at the central CG base-pairs. Arg243 forms an arginine helix bridge with the carbonyl group of Thr235 while making multiple interactions with the DNA (27). Arg240 bridges between the C_{-1} and G_1 phosphate groups of the same DNA strand. The C_{-1} deoxyribose moeity has the C3' *endo* conformation with a pucker angle of 53.4°. C, Water-mediated interaction of Lys246 with base G_{-3} and the hydrophobic contacts of the T_2 methyl group.

$C_{-1}G_1$ is 5.9 Å, both features typical of *A*-form DNA. This alteration from *B*-form conformation appears to be the predominant cause of the bend toward the leucine zipper in the center of the DNA, and is associated with the high positive roll angle of the central C/G base step and the large inclination of base-pairs in the binding site as noted earlier (König & Richmond, 1993). A similar *A*-like deformation seen in the E2/DNA complex can be associated with Arg342 linking two phosphate groups (Hedge *et al*., 1992). Difference electron density from a $2F_o - F_c$ map and a simulated-annealing OMIT map, indicated that the Arg243 side-chain adopts two different conformations, each of which was refined independently to an occupancy of 0.5 (Figure 3B). However, the alternative conformation with the guanidinium group in the opposite orientation occupies the same volume as the first and maintains the same hydrogen-bonding scheme. Neither orientation resembles either of the conformations seen in the AP-1 complex (Ellenberger *et al*., 1992). In comparison, the same residue in the AP-1 complex shows a different conformation in each half-site, one where Arg243 contacts N7 and O6 the central base-pair G_0 directly, and the other bridging the adjacent phosphate groups of C_0 and A_{-1}. The interactions made by Arg243 in the ATF/CREB complex are a combination of the two extremes seen in the AP-1 complex.

The conformation of the central base-pairs is reinforced by Arg240, which in addition to Arg243, hydrogen bonds directly to the phosphate groups of C_{-1} and G_1. The orientation of its guanidinium group is stabilized by a double hydrogen bond from Glu237 on the previous turn of the α-helix, whose side-chain takes up an unusual conformation ($\chi_1 = -74°$, $\chi_2 = -46°$) in order to make this interaction and to indirectly hydrogen bond to the carbonyl oxygen atom of Ala233 *via* a water molecule (Figure 4B). Glu237 has been shown to be important for specific recognition of the ATF/CREB site *versus* the binding site of the bZIP C/EBP, as mutation E237I in combination with T236N and A239V switch the specificity completely (Suckow *et al*., 1993). Whereas the mutation T236N mainly enhances binding to the C/EBP site, the additional mutation E237I appears to abolish binding to the ATF/CREB site. Direct interaction of Glu237 with Arg240 is not apparent in the lower-resolution AP-1 complex structure, and in most other bZIP proteins that preferentially bind the AP-1 site, the carboxylic acid side-chain is replaced with a hydrophobic group. The Glu237/Arg240 combination may make a significant contribution to the ability of the GCN4 bZIP to bind the ATF/CREB and AP-1 sites equally well.

Lys246 lies in the tunnel between the fork of the bZIP α-helices and the DNA major groove. A water molecule is hydrogen bonded to its main-chain carbonyl group and in van der Waals contact with C^{ϵ} of Met250 on the opposite coil. The side-chain of Lys246 is involved in a hydrogen-bonding network in the DNA major groove that links it to N7 of base G_{-3} *via* a water molecule (Figure 4C). A second water molecule, connected to the first, forms a cross-strand bridge between O6 of G_{-3} and O4 of base T_2, and thereby enhances the negative propeller twist observed for both base-pairs. Lys246 also hydrogen bonds to O^{γ} of Ser242, which in turn is bound to the T_{-4} phosphate group. A double mutant that includes K246Q permits binding to sequences containing G or T at position –2 in the AP-1 site (Kim *et al*., 1993), which corresponds to ATF/CREB position –3.

Mutagenesis studies have shown that Thr236 is important for the specificity of DNA recognition at base-pair positions 2 and 3 (Suckow *et al*., 1993). Since the side-chain has no direct or water-mediated interactions with bases, the importance of this residue can be understood only in the context of the surrounding base-specific contacts. The C^{γ} and C^{α} atoms span between the G_1 deoxyribose group C2' and C3' atoms and the T_2 methyl group, respectively, apparently stabilizing the *A*-like DNA conformation in this region by hydrophobic interaction (Figure 4B and C). This particular side-chain conformation is favored by a hydrogen bond from O^{γ} to the main-chain carbonyl oxygen atom of Arg232. The interdependence of the basic domain protein-DNA interactions is exemplified by Thr236. The peptide plane between the C^{α} atoms of Thr236 and Glu237 is tipped by 17° relative to the α-helix axis due to the Thr236 side-chain-Arg232 main-chain bridge, and due to the arginine helix bridge from Arg243 to the Thr236 carbonyl group. Therefore, Thr236 directly effects the orientation of Glu237, Arg232 and Arg243.

Protein conformation

Three of the five arginine side-chains in the basic domain that hydrogen bond to non-esterified phosphate group oxygen atoms, Arg241, Arg243 and Arg245, form a bridge with a water molecule to the $i-7$ main-chain carbonyl oxygen atom (i.e. Arg243 in Figure 4B as residue i). The guanidinium moiety of each arginine side-chain lies parallel and in van der Waals contact with the peptide connecting residues $i-3$ and $i-4$, and donates a hydrogen bond from each of its N^{η} to a single water molecule. This water molecule anchors the arginine side-chain to the $i-7$ carbonyl atom. In this conformation, a N^{η} atom of each arginine hydrogen bonds to a phosphate oxygen atom. As there is a considerable entropic cost of approximately 1.2 to 2 kcal/mol for localizing an arginine side-chain to a single conformation (Koehl & Delarue, 1994; Sternberg & Chickos, 1994), it is likely that the basic domain adopts these structures when it binds DNA and takes on an α-helical structure. Presumably, the slight bending of the basic domain α-helix seen in the ATF/CREB complex facilitates these additional hydrogen bonds made to the main-chain carbonyl groups, as has been observed for α-helices (Barlow & Thornton, 1988). In the basic domain α-helix, eight of a total of 17 main-chain carbonyl oxygen

atoms, those on the outward-facing convex surface, are bound to solvent molecules (18), some of which in turn are hydrogen bonded to side-chains or phosphate oxygen atoms. The water molecules that extend the reach of the Arg243 (Figure 4B) and Arg245 side-chains are fully coordinated; the fourth hydrogen bond is made to an adjacent DNA phosphate oxygen atom. We have observed a highly similar ''arginine helix bridge'' conformation in the protein-DNA interface of other reported high-resolution complex structures, such as Arg344 in the E2/DNA complex (Hedge *et al*., 1992). In the estrogen receptor complex with two dimers per crystal asymmetric unit (Schwabe *et al*., 1993), all four Arg33 side-chains form a bridge with a hydrogen bond to one of the N^{η} atoms, while the other N^{η} atom makes a base contact. In both structures, the bridging water molecules donate hydrogen bonds to the phosphate backbone.

The structure of the leucine zipper in the GCN4-ATF/CREB complex is highly similar to the α-helical coiled coil seen in the crystal structures of the AP-1 complex (rmsd 0.5 Å for Met250 to Lys275 C^{α} positions; Ellenberger *et al*., 1992); and the GCN4 leucine zipper peptide (rmsd 0.7 Å for equivalent C^{α} positions; O'Shea *et al*., 1991). However, in contrast to the latter two structures, the molecular 2-fold axis of the ATF/CREB complex is coincident with a crystallographic 2-fold axis. The side-chains of Met250 and Asn264, which occupy **a** positions in the first and third heptade repeat of the coiled coil, are disordered and were modeled in two complementary orientations in order to avoid the steric overlap that would occur if strict 2-fold symmetry were imposed. For Asn264, these orientations resemble the two conformations observed for the leucine zipper peptide and AP-1 complex, and crystallographic symmetry is apparently maintained because of dynamic interchange or statistical superposition of conformers throughout the crystal. Met250 and Asn264 are not involved in crystal packing interactions in the ATF/CREB complex. NMR results suggest that the leucine zipper in solution is either 2-fold symmetric or, as consistent with the results here, has asymmetric conformations in rapid interchange (Oas *et al*., 1990). Although the deviation of superposition of main-chain C^{α} atoms of the three X-ray structures over the leucine zipper is generally within experimental error, the N-terminal regions show significant differences (Figure 5A). The 2-fold related C^{α} distances between residues in the **a** and **d** positions of the first heptade repeat are larger for the DNA complexes: both 7.5 Å in the ATF/CREB complex and 7.2 and 7.0 Å in the AP-1 complex as compared with both 6.3 Å for the coiled coil alone.

DNA conformation

Although the overall conformation of the DNA in the ATF/CREB complex is predominately *B*-form in appearance, the central region that is in intimate contact with the protein shows in particular some features reminiscent of *A*-DNA. The central C_{-1} deoxyribose has C3'-*endo* conformation and a short 5.9 Å P-P distance to the neighboring G_1, characteristic of *A*-DNA. In addition, the average base-pair displacement over the entire oligonucleotide is $-1.4(\pm 0.2)$ Å with phosphate groups rotated towards the major groove. The major and minor groove widths are clearly correlated with DNA bending: the center of the DNA is bent toward the major groove resulting in a narrowed and deepened major groove and a widened minor groove as compared with straight *B*-DNA. Overall, the conformation is intermediate between *A* and *B*-DNA, as is typical of other protein/DNA complexes (Nekludova & Pabo, 1994).

The spine of hydration seen in the *B*-form dodecamer structure (Chuprina *et al*., 1991) occurs in the ATF/CREB complex in the minor groove along base-pairs 4 to 7, but the water molecules in the widened minor groove of base-pairs −3 to 3 do not span between DNA strands with hydrogen bonds. The 20° bend in the DNA toward the leucine zipper is a summation of the 9.5° roll angle of the central base-pair step and the bending into the minor groove that occurs symmetrically about the molecular dyad at the base-pair step A_4T_5 (Table 2, Figure 5A). The analysis of DNA bending by the circular permutation assay suggests that at least the bent aspect of this conformation occurs for the free DNA (Paolella *et al*., 1994). In the case of a helix-turn-helix motif protein, *trp*-repressor, a comparison of both the free and bound operator structures revealed that the bound form has significant *A*-like character while the unbound form does not (Shakked *et al*., 1994). With respect to the AP-1 complex DNA, the bound ATF/CREB site has greater *A*-like character as measured by the base-pair tilt/inclination values (Figure 5C).

Discussion

ATF/CREB-bZIP target site specificity

The basic domain of the bZIP α-helix is architecturally the simplest known specific DNA-binding motif. Nevertheless, different members of the bZIP family of transcription factors are able to recognize a selection of DNA sites with substantial diversity exemplified by those for the ATF/CREB (ATGACGTCAT) and the C/EBP (GATTGCGCAATC) proteins. Sequence-specific DNA binding is achieved by three means: (1) direct interaction of the base-specific hydrogen bonding and hydrophobic groups in the DNA major groove; (2) indirect, water-mediated hydrogen bonds to the bases in the major groove; and (3) a general fit to the local and global sequence-dependent conformational features of the DNA, which may be at least in part induced by the protein. How direct interactions might completely specify a particular DNA site is relatively easy to understand given a sufficiently large binding site. However, given the logistical difficulty of arranging side-chains on one face of a

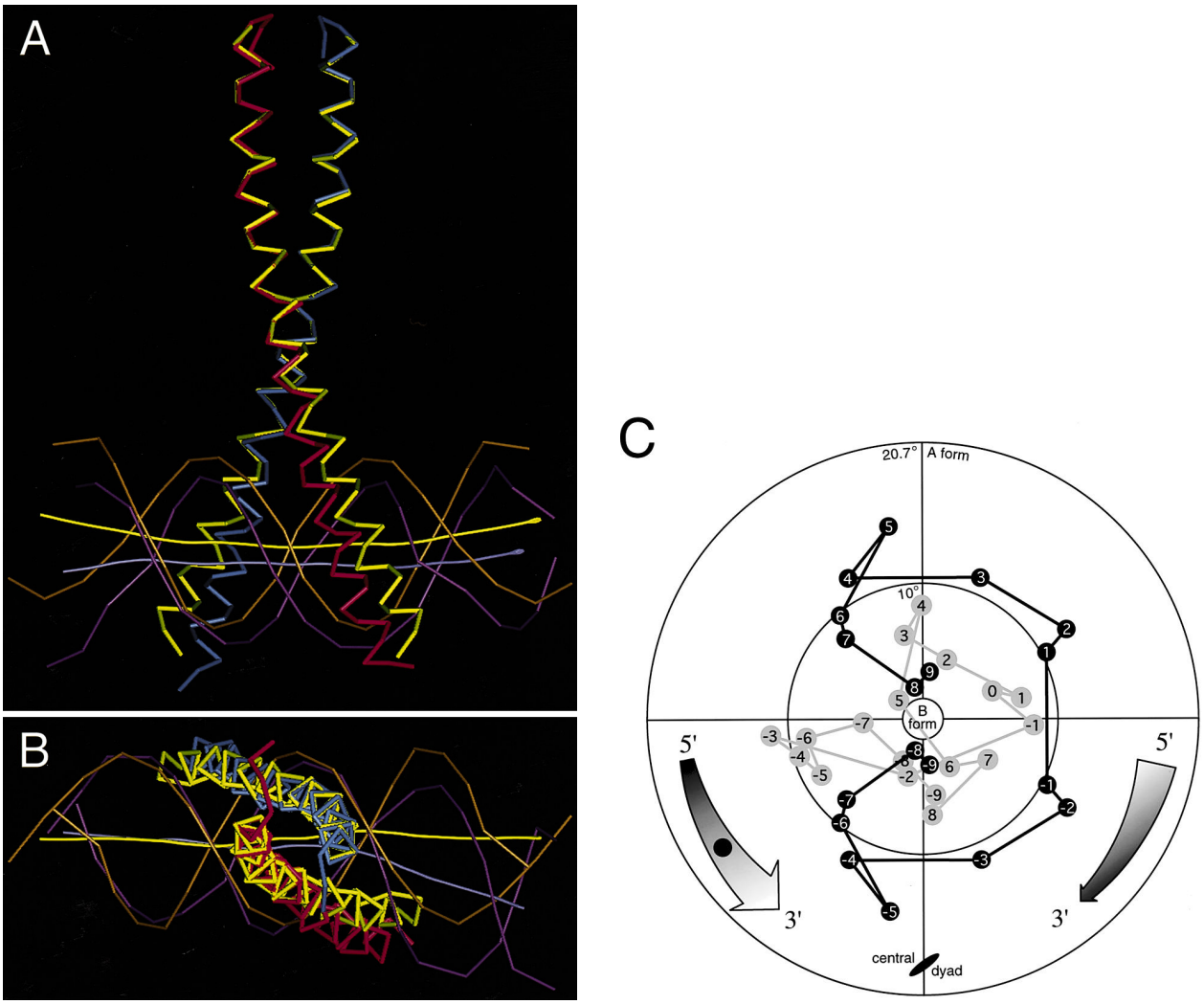


Figure 5. Comparison of the ATF/CREB and AP-1 (Ellenberger *et al*., 1992) complexes. The leucine zipper structures were superimposed based on the C^{α} atoms from Met250 to Lys275 (rmsd = 0.48 Å) to exhibit the differences between the locations of the basic domains and the conformations of the DNA sites. The C^{α} and DNA backbone traces are shown in yellow and brown for the ATF/CREB complex and in red and blue, and purple for the AP-1 complex, respectively. The DNA helix axes calculated with the program CURVES (Lavery & Sklenar, 1989) are shown. A, The ATF/CREB DNA is shifted toward the widened fork of the bZIP protein and bent 20° overall toward the coiled coil as compared with the straight AP-1 DNA. B, The view down the coiled coil shows that CTCAT AP-1 half-site is bent by 10° relative to the ATF/CREB half site. C, Polar projection of the unit vectors normal to the base pairs as viewed down the overall DNA helix axis (Otwinowski *et al*., 1988). The vectors positions (+ direction is indicated by the arrow marked with a black dot) for both the ATF/CREB (black) and AP-1 (gray) are shown and confirm the unusual conformation of the ATF/CREB site DNA. The angular deviation displayed is a combination of base-pair tilt and inclination, is near zero for *B*-form DNA and 20.7° for *A*-form (Arnott *et al*., 1980), and is dominated by large inclination values for the central 12 base-pairs of the ATF/CREB DNA (König & Richmond, 1993).

five turn α-helix to directly and uniquely readout each base of a 5 bp DNA half-site, it is apparent that indirect readout mechanisms will come into play. For example, the presence of the invariant Asn235 in the center of the DNA half-site is capable of rejecting ten of the possible two base-pair combinations dependent on direct hydrogen-bonding interactions alone, but is still compatible with eight different base-pair steps: CA, CC, AA, AC, TG, TT, GG and GT. The interaction between Asn235 and the purine A_{-5} *via* a water interaction demonstrates how the fixed asparagine side-chain extends the specificity determination for this residue to three base-pairs. The additional requirement of a purine in position –5 reduces the number of base-pair triples at positions 3 to 5 from 64 to 16. Direct hydrophobic contact by the C^{β} atom of Ala238 with the 5-methyl group of T_{-4} should be sufficient to reduce this number to four combinations since position 4 must be A. The water-mediated hydrogen bond made by Lys246 to N7 of G_{-3}, and the hydrogen-bonding network that this side-chain participates in is apparently sufficient to discriminate against T at this site, although since other

complicated hydrogen-bonding schemes may be imaginable, protein mutagenesis data are essential to verify the importance of Lys246. Again assuming the absolute specification of T by hydrophobic interaction, T_2 is determined by its 5-methyl group-contact with Ala239. The invariant residue Arg243 hydrogen bonds only with the N7 atom of G_1 and would thus also be compatible with A at this site. In the AP-1 complex, Arg243 is apparently capable of absolute specification of this G since it hydrogen bonds to both the O6 and N7 atoms. For the ATF/CREB complex, we are left after this simple analysis using direct and water-mediated interactions, in principle, with the sequences GTCAT, GTCAC, ATCAT and ATCAC as potential specific recognition half-sites from the 1024 possible.

At this stage, the conformational dependence of the DNA on its sequence must play a role in the discrimination between binding sites. It is difficult to make a conclusive analysis of these effects because although some general rules for sequence-dependent DNA structure are emerging (e.g. Dickerson *et al*., 1994) for free DNA, the problem is substantially more complicated here, since it must also take into account the sequence-dependent propensity for DNA to adopt a particular structure induced by the binding protein. A site-specific protein may use seemingly non-specific interactions to sense and induce the structure of the DNA seen in the complex as well as to adjust the overall binding energy to meet functional requirements. Several examples of this category of interaction are apparent in the bZIP-ATF/CREB complex. Thr236 has been shown to be a DNA sequence discriminator even though its hydrophobic interaction with the G_1 deoxyribose would at first appear to be non-specific. The Thr236 interaction either aids the induction of the special structure of the DNA seen at the center of the site and/or is itself affected by it such that it helps specify the orientation of the surrounding residues including Arg243. Our high-resolution, bZIP complex structure shows the apparent importance of the precise orientation of several of the arginine side-chains that make phosphate contacts. Because they have much less conformational freedom due to the helix bridge structure than would have been expected *a priori*, they may act as precision sensors of, for example, major groove width, a parameter known to vary with sequence (Nekludova & Pabo, 1994). The stabilizing interaction of Glu237 with Arg240 also supports this role for the arginine side-chains. The two CG base-pair step at the center of the 10 bp binding site takes on an *A*-like conformation that is most obviously detected by the protein by positioning Arg240 to measure the short distance between the phosphate groups of G_1 and C_{-1}, as well as using the guanidinium group of Arg243 to fill the space between the N7 atom of G_1 and the phosphate group of C_{-1}. Taken together with the direct and water-mediated recognition of base-specific chemical groups described, recognition of sequence-dependent conformational features of the target site yield the observed specificity for the GTCAT and GTCAC half-site sequences. Mutagenesis and *in vitro* binding experiments combined with further high-resolution structural information will establish quantitatively the importance of these interactions to site-specific DNA recognition.

Half-site spacing—comparison of ATF/CREB and AP-1 site complexes

The difference of 6 bp in half-site spacing center to center for the ATF/CREB site as compared with the 5 bp separation for the AP-1 site is accommodated by a 20° bend toward the leucine zipper and larger base-pair inclination in the ATF/CREB complex (König & Richmond, 1993). These conformational differences in the DNA sites are the principle reason that the same direct base contacts can be made in both complexes, notwithstanding the differences at the central GC base-pairs noted above. Because the basic domain helices must reach further around the ATF/CREB DNA helix to make the equivalent contacts seen in the AP-1 structure,

Table 2. DNA conformational parameters for the ATF/CREB half site

Base-pair		C G	G C	T A	C G	A T	T A	C G	T A	C G	C G	*B* DNA[a]	*A* DNA[a]
x-Disp	(Å)	–1.44	–1.44	–1.40	–1.56	–1.32	–1.33	–1.35	–1.36	–1.23	–1.35	0.0	–5.28
Incl.	(deg.)	5.5	5.5	6.1	6.8	4.5	3.6	7.1	6.2	7.1	7.0	1.5	20.7
Tip	(deg.)	–1.9	1.9	–1.8	–1.2	–1.2	–3.3	–0.9	–1.2	0.0	–2.2	0	0
Buckle	(deg.)	–5.2	5.2	–3.9	4.1	–1.8	1.8	0.6	–0.9	–3.9	1.2	0	0
Prop.	(deg.)	–13.0	–13.0	–9.3	–10.4	–10.1	–11.8	–13.7	–11.0	1.9	–7.1	–13.3	–7.5
Open	(deg.)	2.24	2.24	2.22	–0.12	4.21	8.60	2.97	2.52	–4.57	0.87	0	0
Twist	(deg.)	31.4	34.2	32.0	41.7	25.0	43.0	30.3	38.8	30.7		36.0	30.7
Roll	(deg.)	9.5	–3.4	3.5	3.5	–6.1	1.3	1.3	0.8	–3.3		0.9	11.4
Tilt	(deg.)	0.0	3.1	–0.7	–2.8	1.8	5.6	–2.3	–0.9	–0.4		0	0
Slide	(Å)	0.66	–0.64	0.0	0.46	–0.62	0.08	–0.17	0.23	0.15		0.1	–1.9
P–P	(Å)	5.94	7.12	6.43	6.71	6.77	6.97	6.72	6.32	6.08			
		5.94	6.96	6.95	6.52	6.57	7.13	6.66	6.32	7.55		7.0	5.9

The program CURVES was used to calculate helical parameters (Lavery & Sklenar, 1989).
[a] Parameters from fiber diffraction data (Arnott *et al*., 1980).

the whole DNA site is shifted toward the fork of the coiled coil by about 1.5 Å (Figure 5A). As a consequence of this shift, the side-chain of Leu247 sterically clashes with the phosphate group of A_{-2}, and thereby may be the major determinant of the approximately fivefold reduction of binding affinity for the ATF/CREB site compared with the AP-1 site as suggested by Kim *et al*. (1993). CREB proteins have lysine or arginine at this position to facilitate the interaction with the DNA backbone of the ATF/CREB site. When the ATF/CREB and AP-1 structures are aligned based on their leucine zippers alone, the shift of the DNA helix results in a difference in the trajectories of the DNA axes of the non-homologous half-sites (GTCAT *versus* CTCAT) of approximately 10° as viewed from the leucine zipper (Figure 5B). The coupling between half-site separation and rotation of the DNA helix axis around the projected axis of the leucine zipper may be of consequence for higher assemblies, which include factors that may recognize features of both the basic domain/DNA and leucine zipper.

Materials and Methods

Crystallization, data collection and processing

The polypeptide C62GCN4 and the oligonucleotide containing the ATF/CREB site were prepared as described (König & Richmond, 1993). A 0.2 mM solution of 1:1 protein/DNA complex and 5 mM $MgCl_2$ was mixed with twice the volume of 10% PEG 6000, 30 mM sodium citrate (pH 4.6), 50 mM sodium acetate, 5 mM $MgCl_2$ and 1 mM NaN_3 in the wells of a FALCON Micro Test III assay plate (96 wells) and covered with mineral oil (Sigma). Crystals appeared within 24 hours and grew to their final size (0.2 to 0.3 mm) in five to seven days. Crystal parameters were similar to those reported previously (König & Richmond, 1993): $a = b = 58.66$ Å, $c = 86.88$ Å, space group $P4_12_12$, and one monomer per asymmetric unit. X-ray data were recorded at 4°C from two crystals on beam line X11 at the DESY/EMBL Outstation using $\lambda = 0.92$ Å, a MAR image plate system, and 1.5° rotation per exposure. Local programs were used for data processing (T.J.R., unpublished). The merged data yielded 6597 unique reflections between 15 and 2.2 Å with $R_{merge} = 0.09$ and a mean multiplicity of 5.

Structure refinement

Structure refinement was carried out with XPLOR (Brünger, 1988) using the 3 Å structure as the starting model (König & Richmond, 1993). Six refinement cycles were performed, where each cycle began with slow-cooling simulated annealing, followed by conjugate gradient minimization and temperature factor optimization, and ended with model rebuilding using the program O (Jones *et al*., 1991). Data from 6 to 2.2 Å were used for refinement and yielded an overall *R*-factor of 0.214 and R_{free} of 0.316 (Table 1). Restraints for base planarity were used in XPLOR, but the ribose pucker was not constrained. The four side-chains Arg243, Met250, Asp264 and Arg273 had clearly more than one conformation and were modeled in the two most obvious orientations, each with occupancy factor 0.5, and were constrained not to interact with each other during refinement. Simulated annealing OMIT maps were calculated for all residues in the contact region between protein and DNA, for the five C and N-terminal residues, and for side-chains where density appeared weak in $2F_o - F_c$ map. The final model consists of the DNA 19-mer, amino acid residues Ala229 to Leu277 and 46 well-ordered solvent molecules. For the two C-terminal amino acid residues, Lys276 and Leu277, only the main-chain atoms and C^β could be assigned unambiguously. The DNA shows generally higher *B*-factors in the flanking regions than in the central protein binding region. The different sets of atomic force constants for protein (Engh & Huber, 1991) and DNA (PARAM11.DNA from Brünger, 1992) gave greater variation in the bonding parameters for DNA than for protein.

Acknowledgements

We thank J. Schwabe for the coordinates of the estrogen receptor complex. This research was supported in part by the Swiss National Fond and the Krebs Liga Zürich. Atomic coordinates are deposited in the Brookhaven Databank as entry 2dgc.

References

Arndt, K. & Fink, G. R. (1986). GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5' TGACTC 3' sequences. *Proc. Natl Acad. Sci. USA*, **83**, 8516–8520.

Arnott, S., Chandrasekaran, R., Birdsall, D. L. L., A. G. W. & Ratliff, R. L. (1980). Left-handed DNA helices. *Nature*, **283**, 743–746.

Barlow, D. J. & Thornton, J. M. (1988). Helix geometry in proteins. *J. Mol. Biol.* **201**, 601–619.

Brünger, A. T. (1988). Crystallographic refinement by simulated annealing. *J. Mol. Biol.* **203**, 803–816.

Brünger, A. T. (1992). *XPLOR V3.1*. Yale University Press, New Haven and London.

Chuprina, V. P., Heinemann, U., Nurislamov, A. A., Zielenkiewicz, P., Dickerson, R. E. & Saenger, W. (1991). Molecular dynamics simulation of the hydration shell of a *B*-DNA decamer reveals two main types of minor-groove hydration depending on groove width. *Proc. Natl Acad. Sci. USA*, **88**, 593–597.

Dickerson, R. E., Goodsell, D. S. & Neidle, S. (1994). . . . the tyranny of the lattice . . . , *Proc. Natl Acad. Sci. USA*, **91**, 3579–3583.

Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α-helices: crystal structure of the protein-DNA complex. *Cell*, **71**, 1223–1237.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Crystallog. sect. A*, **47**, 392–400.

Hedge, R. S., Grossman, S. R., Laimins, L. A. & Sigler, P. B. (1992). Crystal structure at 1.7 Å of the bovine papillovirus-1 E2 DNA-binding domain bound to its DNA target. *Nature*, **359**, 505–512.

Hodel, A., Kim, S.-H. & Brünger, A. T. (1992). Model bias in macromolecular crystal structures. *Acta Crystallog. sect. A*, **48**, 851–858.

Hope, I. A. & Struhl, K. (1987). GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *EMBO J.* **6**, 2781–2784.

Joahimiak, A., Haran, T. E. & Sigler, P. B. (1991). Mutagenesis supports water mediated recognition in

the trp repressor-operator system. *EMBO J.* **13**, 367–372.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & M., K. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.

Kim, J., Tzamarias, D., Ellenberger, T., Harrison, S. C. & Struhl, K. (1993). Adaptability at the protein-DNA interface is an important aspect of sequence recognition by bZIP proteins. *Proc. Natl Acad. Sci. USA*, **90**, 4513–4517.

Koehl, P. & Delarue, M. (1994). Application of a self-consistent mean field theory to predict side-chain conformation and estimate their conformational energy. *J. Mol. Biol.* **239**, 249–275.

König, K. P. & Richmond, T. J. (1993). The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* **233**, 139–154.

Lavery, R. & Sklenar, H. (1989). Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dynam.* **6**, 655–667.

Mavrothalassitis, G., Beal, G. & Papas, T. S. (1990). Defining target sequences of DNA-binding proteins by random selection and PCR: determination of the GCN4 binding sequence repertoire. *DNA Cell Biol.* **9**, 783–788.

Nekludova, L. & Pabo, C. O. (1994). Distinctive DNA conformation with enlarged major groove is found in Zn-finger-DNA and other protein DNA complexes. *Proc. Natl Acad. Sci. USA*, **91**, 6948–6952.

Oas, T. G., McIntosh, L. P., O'Shea, E. K., Dahlquist, F. W. & Kim, P. S. (1990). Secondary structure of a leucine zipper determined by nuclear magnetic resonance spectroscopy. *Biochemistry*, **29**, 2891–2894.

Oliphant, A. R., Brandl, C. J. & Struhl, K. (1989). Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Mol. Cell. Biol.* **9**, 2944–2949.

O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science*, **254**, 539–544.

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

Paolella, D. N., Palmer, C. R. & Schepartz, A. (1994). DNA targets for certain bZIP proteins distinguished by an intrinsic bend. *Science*, **264**, 1130–1133.

Schwabe, J. W. R., Chapman, L., Finch, J. T. & Rhodes, D. (1993). The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell*, **75**, 567–578.

Sellers, J. W., Vincent, A. C. & Struhl, K. (1990). Mutations that define the optimal half-site for binding yeast GCN4 activator protein and identify an ATF/CREB-like repressor that recognizes similar DNA sites. *Mol. Cell. Biol.* **10**, 5077–5086.

Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinowich, 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.

Sternberg, M. J. E. & Chickos, J. S. (1994). Protein side-chain conformational entropy derived from fusion data-comparison with other empirical scales. *Protein Eng.* **34**, 149–155.

Suckow, M., von Wilcken-Bergmann, B. & Müller-Hill, B. (1993). Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP and TAF-1 that are involved in specific DNA binding. *EMBO J.* **12**, 1193–1200.

Suckow, M., Schwamborn, K., Kisters-Woike, B., von Wilcken-Bergmann, B. & Müller-Hill, B. (1994). Replacement of invariant bZip residues with the basic region of the yeast transcriptional activator GCN4 can change its DNA binding specificity. *Nucl. Acids Res.* **22**, 4395–4404.

Vincent, A. C. & Struhl, K. (1992). ACR1, a yeast ATF/CREB repressor. *Mol. Cell. Biol.* **12**, 5394–5405.

Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. & Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature*, **347**, 575–578.

Edited by A. Klug

(Received 20 July 1995; accepted 26 September 1995)