Cell, Vol. 71, 1223–1237, December 24, 1992, Copyright © 1992 by Cell Press

The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted α Helices: Crystal Structure of the Protein–DNA Complex

Thomas E. Ellenberger,* Christopher J. Brandl,† Kevin Struhl,† and Stephen C. Harrison*‡
*Harvard University
Department of Biochemistry and Molecular Biology
Cambridge, Massachusetts 02138
†Harvard Medical School
Department of Biological Chemistry and Molecular Pharmacology
Boston, Massachusetts 02115
‡Howard Hughes Medical Institute
Harvard University
Cambridge, Massachusetts 02138

Summary

The yeast transcriptional activator GCN4 is 1 of over 30 identified eukaryotic proteins containing the basic region leucine zipper (bZIP) DNA-binding motif. We have determined the crystal structure of the GCN4 bZIP element complexed with DNA at 2.9 Å resolution. The bZIP dimer is a pair of continuous α helices that form a parallel coiled coil over their carboxy-terminal 30 residues and gradually diverge toward their amino termini to pass through the major groove of the DNA-binding site. The coiled-coil dimerization interface is oriented almost perpendicular to the DNA axis, giving the complex the appearance of the letter T. There are no kinks or sharp bends in either bZIP monomer. Numerous contacts to DNA bases and phosphate oxygens are made by basic region residues that are conserved in the bZIP protein family. The details of the bZIP dimer interaction with DNA can explain recognition of the AP-1 site by the GCN4 protein.

Introduction

DNA recognition by transcriptional regulatory proteins depends on "reading" the local DNA topography at specific binding sites. Sequence-specific variations in the functional groups of bases, the DNA helical geometry, and local deformability of DNA define a unique surface, which is complemented by the protein partner. The DNA-binding element of most transcription factors is a module that can function independently of the rest of the protein. In many instances this module is a stably folded domain, maintained by hydrophobic packing interactions or metal chelation (for reviews see Harrison and Aggarwal, 1990; Kaptein, 1991). In other cases, all or part of the recognition module is unfolded in solution and assumes a stable conformation when bound to DNA. Examples in which ordering is contingent upon association with DNA include the arm of λ repressor (Weiss et al., 1984; Clarke et al., 1991), the β-strand motif of Arc repressor (Bowie and Sauer, 1989), and the basic region leucine zipper (bZIP) motif of a number of eukaryotic transcription factors (Johnson and McKnight, 1989; Talanian et al., 1990; O'Neil et al., 1990, 1991; Weiss et al., 1990; Patel et al., 1990; Weiss, 1990; Saudek et al., 1991). The interaction with DNA stabilizes a protein conformation complementary to the binding site. In the case of the bZIP element at normal cellular concentrations, dimerization is also induced by DNA binding, and neither the basic region nor the leucine zipper are likely to be well ordered in the unbound protein (Johnson and McKnight, 1989; Talanian et al., 1990; O'Neil et al., 1990, 1991; Weiss et al., 1990; Patel et al., 1990; Weiss, 1990; Saudek et al., 1991). For this reason, we use the term element, rather than domain, to refer to this structure.

The bZIP DNA-binding element consists of 55 to 65 residues that determine the DNA binding affinity and specificity of more than 30 identified proteins from yeast, mammalian cells, and plants. DNA binding activity derives from two related properties, dimerization and direct interaction with DNA, which are functions of the leucine zipper and basic region, respectively (for review see Kerppola and Curran, 1991a). The leucine zipper contains a 4-3 heptad repeat of hydrophobic and nonpolar residues that pack together in a parallel α-helical coiled coil (O'Shea et al., 1989). The stability of the dimer results from the side-by-side packing of leucines and nonpolar residues in positions a and d of the heptad repeat, as well as a limited number of intra- and interhelical salt bridges, shown in a recent crystal structure of the GCN4 leucine zipper peptide (O'Shea et al., 1991). The structure of the basic region and the details of its interaction with DNA have awaited detailed structural analysis of the protein–DNA complex.

We report the crystal structure of the GCN4 bZIP element in complex with a specific DNA site. The bZIP element looks like an α-helical forceps gripping the major groove of DNA. Each chain of the homodimer forms a smoothly curving, continuous α helix. The carboxy-terminal parts of the monomers pack in a coiled coil, essentially identical in structure to the isolated leucine zipper. This dimer interface projects at nearly right angles to the axis of the DNA helix. The amino-terminal 25 residues pass through the major groove of each half site, making numerous contacts with the base pairs. The conservation of certain basic region residues in the bZIP family and the exact specification of the central 7 bp of the GCN4-binding site are consistent with the protein–DNA contacts observed in the crystal structure.

Results

Structure of the GCN4–DNA Complex

The bZIP element of the yeast transcriptional activator GCN4 (residues 226–281) was expressed in Escherichia coli, and crystals were obtained of the protein complexed with the 20-mer DNA shown in Figure 1. The X-ray structure of this complex was solved by multiple isomorphous replacement using phase information from 5-iododeoxyuridine (5I-dU)-substituted DNAs (Table 1; Figure 2; see Experimental Procedures). Each monomer of the GCN4–DNA-binding element forms a continuous α helix

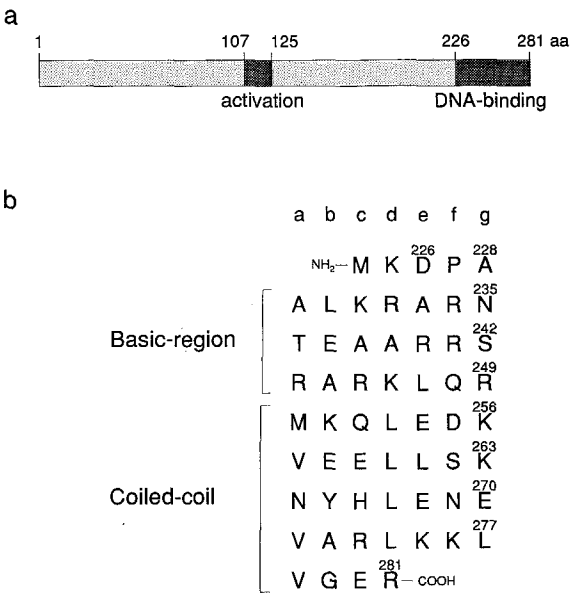

Figure 1. Sequence of the GCN4 bZIP Element and DNA Used in Crystallization

(a) DNA binding activity of the 281 residue yeast GCN4 protein is localized to the carboxy-terminal 56 amino acids. This DNA-binding element, which functions independently of the rest of the protein (Hope and Struhl, 1986), was used for crystallization of a complex with DNA.

(b) The sequence of the GCN4 bZIP element used in crystallization is displayed as a heptad repeat, with positions a and d corresponding to residues at the dimer interface of the coiled coil. The amino-terminal methionine and lysine were added for efficient expression in E. coli. GCN4 residues begin at Asp-226 and extend to the carboxyl terminus of GCN4, Arg-281.

(c) The 20-mer oligonucleotide in the crystallized complex contains an AP-1 site, which is a pseudopalindrome of overlapping half sites, centered on a guanine–cytosine base pair. Bases are numbered from the center of the binding site (position 0), with the designation of L or R for left or right half site and a prime for bases on the strand complimentary to the TGAC or TGAG sequences.

of 52 residues with no pronounced kinks or sharp bends (Figure 3). The helices pack tightly against one another at the carboxy-terminal dimer interface, forming an extensive van der Waals contact surface. This rigid coiled coil orients the basic region of each protomer over its respective half site so that a smooth bend of the α helix in the direction of the DNA axis allows it to traverse the major groove (Figure 3). The basic region α helix is anchored in the major groove by an array of positively charged and polar residues, which donate hydrogen bonds to unesterified oxygens of the phosphodiester backbone (see Figures 6b and 8). These backbone contacts precisely orient basic region side chains for a series of base-specific contacts described below. The amino-terminal end of the basic region continues as a straight α helix beyond the point of DNA contact, giving the protein the appearence of a pair of α-helical tweezers.

The Coiled-Coil Dimerization Motif

The GCN4 leucine zipper dimer interface is oriented almost perpendicular to the axis of the DNA (Figure 3). The structure of the leucine zipper in the DNA complex shares many features with that of a leucine zipper peptide determined by O'Shea et al. (1991). The crossing angle between monomers (about 18°) is the same in both instances, corresponding to Crick's knobs-into-holes packing scheme for two α helices (Crick, 1953). The main chain atoms of the carboxy-terminal 25 residues of the bZIP element (residues 256–279) superimpose with fair accuracy on those of the isolated leucine zipper (root mean square [RMS] error = 1.1 Å), and the orientations of side chains at positions a and d of leucine zipper heptad repeat are comparable. A buried hydrogen bond between Asn-264 side chains of both monomers (position a of the heptad repeat) is present in both structures. These conserved structural features reflect the importance of side chain–packing interactions of residues at positions a and d in stabilizing the coiled-coil dimer.

The surface of the leucine zipper is decorated with charged and polar residues, many of which participate in crystallographic contacts. A series of charged residues is located at positions e and g of the heptad repeat (see Figure 1b). These have the potential to form intermolecular electrostatic interactions, which, in principle, could stabilize the coiled-coil dimer. However, few salt bridges are present in the crystal structure. Salt bridges do link Glu-270 of either monomer to Lys-275 of reciprocal monomer. Glu-268 of the left monomer is bridged to Lys-263 of the right monomer; the analogous interaction with Glu-268 of the right monomer is blocked by a symmetry-related DNA in the crystals, which packs tightly against Lys-263 of the left monomer. All three of these salt bridges are located near the carboxy-terminal end of the leucine zipper. Several salt bridges join residues at positions e and g of succeeding turns of the helical repeat within each monomer. These intramolecular salt links include Glu-270–Arg-273 of both monomers and Glu-237–Arg-240 of the right monomer. Arg-240 of the left monomer hydrogen bonds to the phosphate of Cyt-0 instead of interacting with Glu-237. The small number of electrostatic interactions between GCN4 monomers suggests that the dimer is stabilized primarily by the van der Waals interactions at the core of the coiled coil. Electrostatic interactions may play a more important role in the stabilization of other bZIP dimers. For example, c-Fos has a predominance of acidic residues at positions e and g of the leucine zipper, unlike GCN4 (see Figure 1b). It has been proposed that the localized negative charge in c-Fos is destabilizing for an α-helical conformation and that the charge must be neutralized by complementary residues of Jun to form the Fos–Jun heterodimer (O'Shea et al., 1992).

The protein subunits in the GCN4–DNA complex diverge from the dimer axis in the segment comprising residues 245–255. This fork creates a smooth bend in each

Table 1. X-Ray Data Collection Statistics

Derivatives	Resolution	Reflections	Unique	Complete (%)	R_{sym}[a]	R_{iso}[b]	F_H/E[c]
Native (23°C)	16–3.1	15,670	4,488	89	0.064		
Native (−160°C)	8–2.9	21,237	5,289	96	0.099		
IdU-1[d]	16–3.5	11,811	3,450	98	0.080	0.147	1.27
IdU-2[d]	16–3.4	8,628	3,649	94	0.068	0.165	1.43
IdU-3[d]	16–3.4	17,428	3,600	94	0.081	0.156	1.66
IdU-4[d]	16–3.4	6,219	3,404	89	0.059	0.168	1.32

Refinement statistics (−160°C data)	
Resolution	8–2.9 Å
Reflections	5,289
Number of atoms	1,771
Solvent molecules	18
Crystallographic R factor	0.23
RMS deviation in bond lengths	0.023 Å
RMS deviation in bond angles	3.30°
RMS deviation thermal parameter	1.8 Å^2

[a] $R_{sym} = \Sigma_h\Sigma_i|I_{h,i} - I_h|/\Sigma_h\Sigma_i I_{h,1}$, where I_h is the mean intensity of i observations of reflection h.
[b] Mean isomorphous difference (10–3.4 Å resolution), $R_{iso} = \Sigma|F_{ph} - F_h|/\Sigma F_{ph}$, where F_p and F_{ph} are the structure factor amplitudes of the native and derivative data, respectively.
[c] Phasing power (10–3.4 Å resolution), $F_H/E = [(F_{H(calc)})^2/(F_{PH(obs)} - F_{PH(calc)})^2]^{1/2}$.
[d] Overall figure of merit = 0.50 (10–3.4 Å resolution).

subunit, which displaces its basic region of each monomer away from the dimer interface, in a direction parallel to the DNA axis (Figure 3). No contacts between protein monomers are evident amino-terminal to Met-250, the first residue in position a of the leucine zipper heptad repeat (see Figure 1b). In contrast, the subunits of the leucine zipper dimer lacking the basic region pack closely together at their amino-terminal residue, Arg-249 (O'Shea et al., 1991). This "closed" conformation of the leucine zipper may be the structure of the bZIP dimer in solution at high peptide concentrations, conditions that permit folding of the coiled coil but not the basic region (Weiss et al., 1990).

The α-helical conformation of the bZIP element is broken by Gly-279 less than one turn from the carboxyl terminus of the leucine zipper heptad repeat. A short strand, formed by Gly-279 and Glu-280, passes over the carboxyl terminus of each monomer. Electron density is present for Arg-281 of the left monomer only, where it contacts a symmetry-related DNA molecule, thereby stabilizing the orientation of the strand capping this monomer. The GCN4 leucine zipper consists of just four heptad repeats, ending with Leu-277 in the g position of the last repeat. This is one of the shortest leucine zippers in the bZIP family, and it may represent a nearly optimal packing of side chains at the dimer interface of the coiled coil.

Interaction of GCN4 with DNA

The DNA in the crystals contains an AP-1-binding site, identical to that bound by the mammalian Fos and Jun proteins (Bohmann et al., 1987; Rauscher et al., 1988; Franza et al., 1988; Risse et al., 1989). The AP-1 sequence is a pseudopalindrome of two 4 bp half sites that overlap on a central guanine–cytosine base pair (see Figure 1c). This is a symmetrized version of naturally occurring GCN4-binding sites, and it is bound by GCN4 with high affinity (Hill et al., 1986). Each half site of the AP-1 sequence is contacted by one subunit of the GCN4 bZIP dimer (Figure 3). Asn-235, a residue that is absolutely conserved in the bZIP protein family, is at the center of each monomer–half site interface. In the refined model, OD_1 of Asn-235 can accept a hydrogen bond from N_4 of Cyt-2, and ND_1 can donate a hydrogen bond to O_4 of Thy-3 (Figure 4). These contacts require that the basic region α helix lie deep in the major groove, and they specify two of the four base pairs in each half site. In addition, a water molecule appears to bridge between the ND_1 of Asn-235 and the N_6 and N_7 of Ade-4. Although this water appears as strong positive difference density in a 2Fo − Fc electron density map calculated from the final model and diffraction data collected at −160°C, this interpretation must be regarded as tentative owing to the limited resolution of the current structure. Lys-231 of the right half site monomer also projects into the major groove in this region; it does not appear to interact directly with a base pair. Lys-231 of the left monomer points away from the DNA.

Asn-235 lies in a hydrophobic pocket formed by the β carbons of Ala-238 and Ala-239, which are in van der Waals contact with the 5-methyl groups of Thy-3 and Thy-1, respectively (Figure 5a). This van der Waals surface specifies the thymines of base pairs 1 and 3 of each half site. The hydrophobic pocket is shielded from solvent by the side chains of Ser-242, Arg-240, and Arg-241, which cap the boundaries of the protein–DNA interface. The Arg-243 methylene groups of the left half site monomer further demarcate the medial boundary of the pocket.

The asymmetry of the AP-1 site causes GCN4 to contact the central base pair in an asymmetric manner. Arg-243 of the left half site monomer can donate hydrogen bonds from NH_1 and NH_2 to O_6 and to N_7 of Gua-0′ (Figure 6a). This bidentate pattern of hydrogen bonding between arginine and guanine was predicted as a simple mode of guanine recognition by proteins (Seeman et al., 1976), and it occurs at a conserved position in all three zinc fingers of the Zif268–DNA complex (Pavletich and Pabo, 1991).

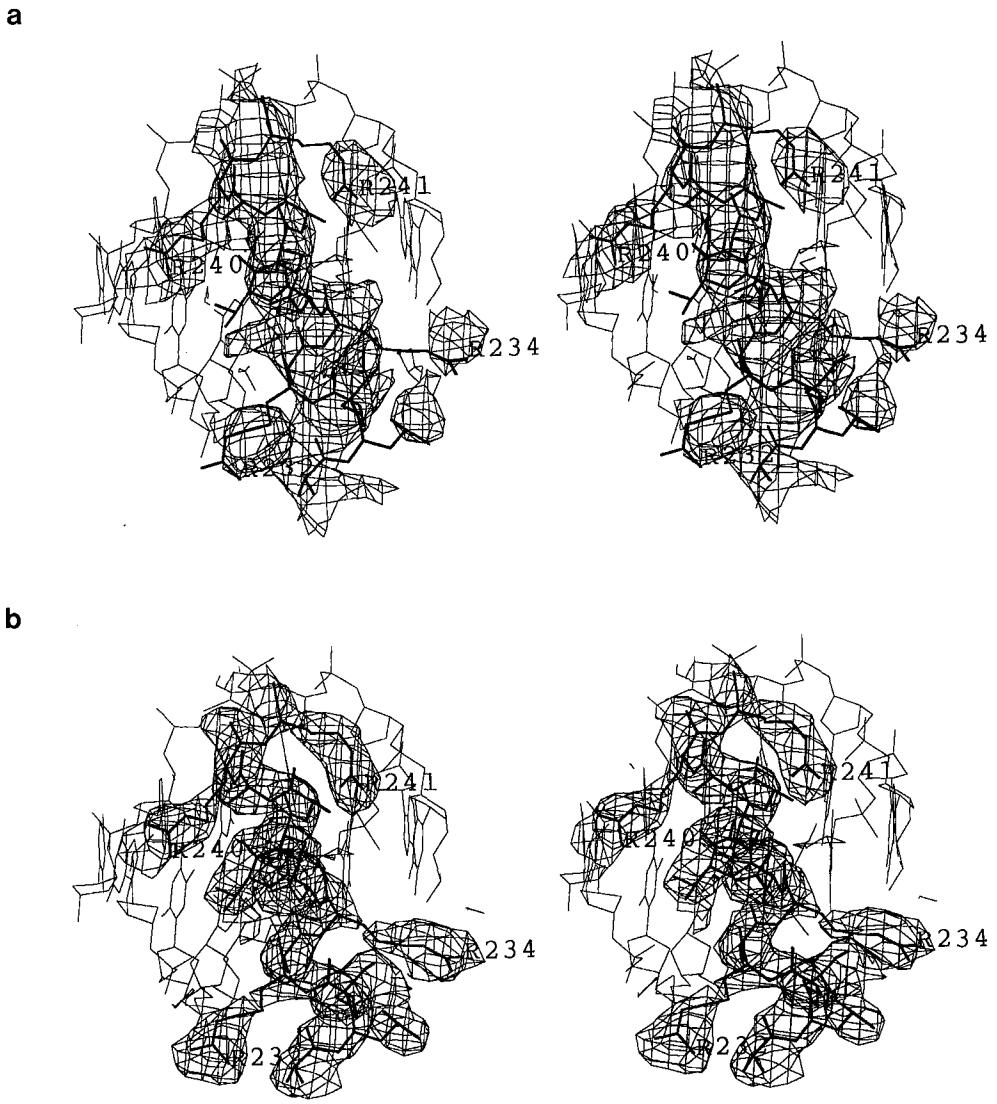

Figure 2. Electron Density of the Basic Region α Helix

(a) Stereo view of the basic region in the major groove of the binding site. Electron density calculated from the initial MIR phases after solvent flattening is shown superposed on the final model of the complex.

(b) Electron density of the final 2Fo − Fc map.

Arg-243 of the right half site monomer of GCN4 is, of course, unable to bond with Gua-0′. Instead, it donates hydrogen bonds to the unesterified phosphate oxygens of Cyt-0 and Ade-1L (Figures 6a and 8). One correlate of these asymmetric contacts is a slight displacement of the protein relative to the pseudodyad of the AP-1 site. The leucine zipper of the left half site monomer is drawn close to the DNA pseudodyad, and the dyad axis relating the main chain atoms of the two protomers is inclined about 3° from the DNA pseudodyad.

Despite the local asymmetry in protein orientation about the center of the binding site, the position of the basic region in the major groove and the protein contacts to base pairs 1, 2, and 3 are conserved between half sites. Superposition of basic region main chain atoms of the right and left monomers aligns the bases of the two half sites with reasonable precision (RMS deviation = 1.33 Å; Figure 7). Some differences in phosphodiester backbone geometry are present, but the exocyclic groups of bases contacted by the protein are nearly superimposable for the two halves of the binding site. This conserved orientation of each subunit on its half site suggests that the positioning of the basic region α helix is locally determined rather than being a consequence of the orientation of the leucine zipper dimer interface.

Contacts that position the basic region on the DNA include the base-specific contacts discussed above and a number of basic and polar residues donating hydrogen bonds to the phosphate oxygens of the DNA backbone (see Figure 6b). Most contacts to the DNA backbone are conserved between protein monomers, with the exception of those near the center of the binding site (Figure 8). An

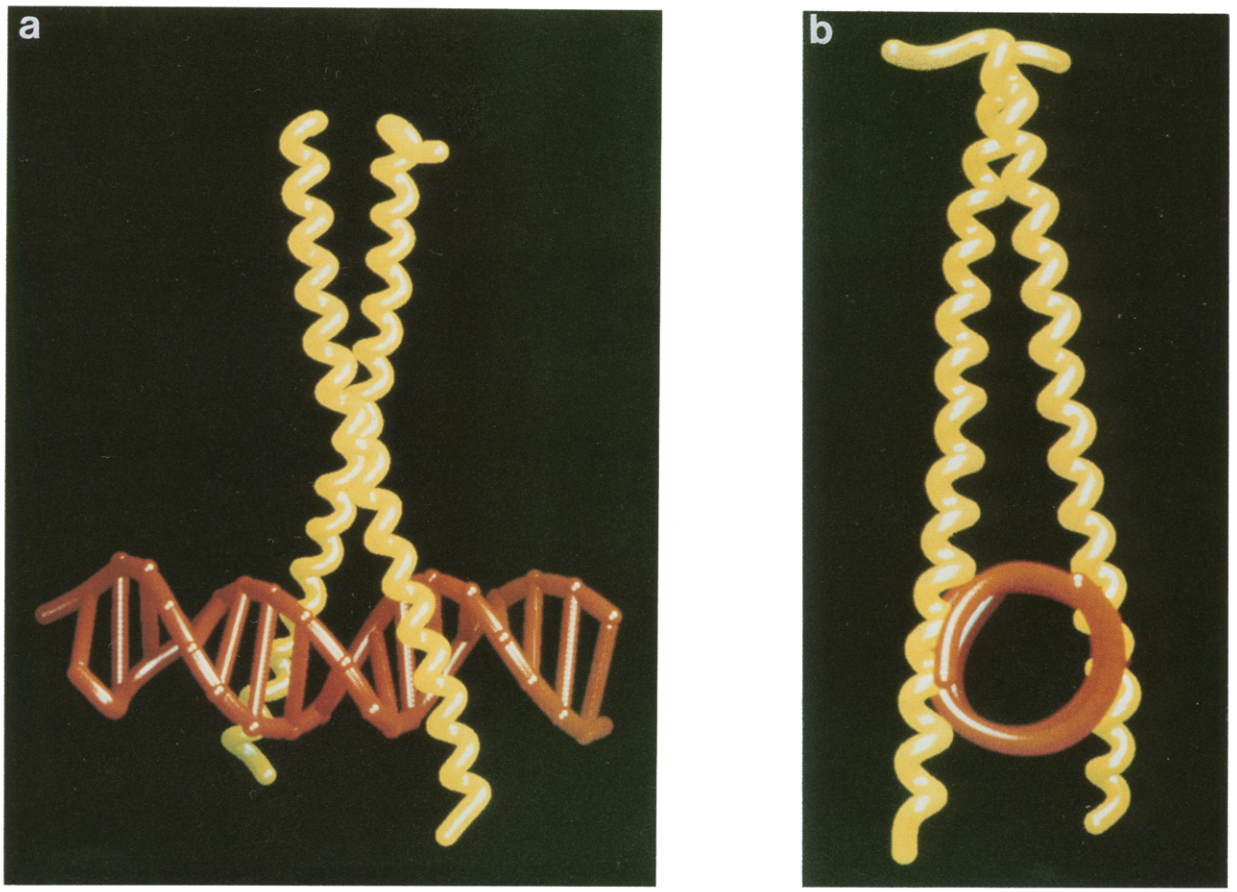

Figure 3. Structure of the GCN4 bZIP Complex

(a) The bZIP dimer (yellow) binds in the major groove of the DNA (red). Each bZIP protomer is a smoothly curved, continuous α helix. The carboxy-terminal residues of the monomers pack together as a coiled coil, which gradually diverges to allow the basic region residues to follow the major groove of either DNA half site. This divergence of the bZIP monomers corresponds to an unwinding of the coiled-coil super helix, with a slight righthanded rotation of basic region residues about the α-helical axis of each chain and a lateral displacement of each monomer along the helical axis of the DNA. The DNA in the complex is straight, and its conformation is in the B form across the region contacted by the protein.

(b) View down the DNA axis, with the DNA backbone represented as a red coil. The basic region residues amino-terminal to the point of DNA contact are in a straight, α-helical conformation. The amino-terminal residues of the basic region do not wrap around the back side of the binding site.

unesterified phosphate oxygen of Cyt-0 is contacted by Arg-243 of the right monomer, with the phosphate of Gua-0′ contacted by Arg-240 of the left monomer. The central asymmetry imposed by contacts to bases at position 0 is corrected by local contacts, which position both monomers over their respective half sites in an analogous way. This adjustment implies that the α-helical fork at the junction between the basic region and the leucine zipper coiled coil is somewhat flexible. Flexibility of GCN4 is likely to be a key component in its specific recognition of the activating transcription factor/cAMP response element–binding protein (ATF/CREB)-binding site (Sellers et al., 1990). The ATF/CREB site is a palindromic sequence homologous to the AP-1 sequence, but it contains an additional central guanine–cytosine base pair.

Structure of the DNA in the Complex

The AP-1 site in the GCN4 complex is straight, B-form DNA across the region contacted by the protein (see Figure 3). The central 15 bp of the AP-1 site DNA have an average helical rise and twist of 3.2 Å and 34°. No systematic variation in phosphodiester backbone or base pair geometry is evident across the binding site. DNAs in adjacent unit cells of the crystals pack end to end, forming a Watson–Crick base pair between the 5′ thymine of one complex and the 5′ adenine of the adjacent DNA (Figure 9a). The DNA is aligned along crystallographic unit cell axis c, which has a length (65.3 Å) commensurate with that of a B-form 20-mer oligonucleotide.

The DNA packing arrangement in the crystals changes markedly in response to crystal harvest conditions. Prior to collecting diffraction data from crystals at −160°C, the crystals were soaked in cryoprotectant to minimize crystalline ice formation during freezing (see Experimental Procedures). Following equilibration of the crystals in a cyroprotectant solution containing an elevated concentration of polyethylene glycol (molecular weight = 400) (PEG 400), the length of the crystallographic c axis is reduced by 6.1 Å (c = 59.2 Å). This shrinkage is due to a local rearrangement in base-stacking interactions that decreases

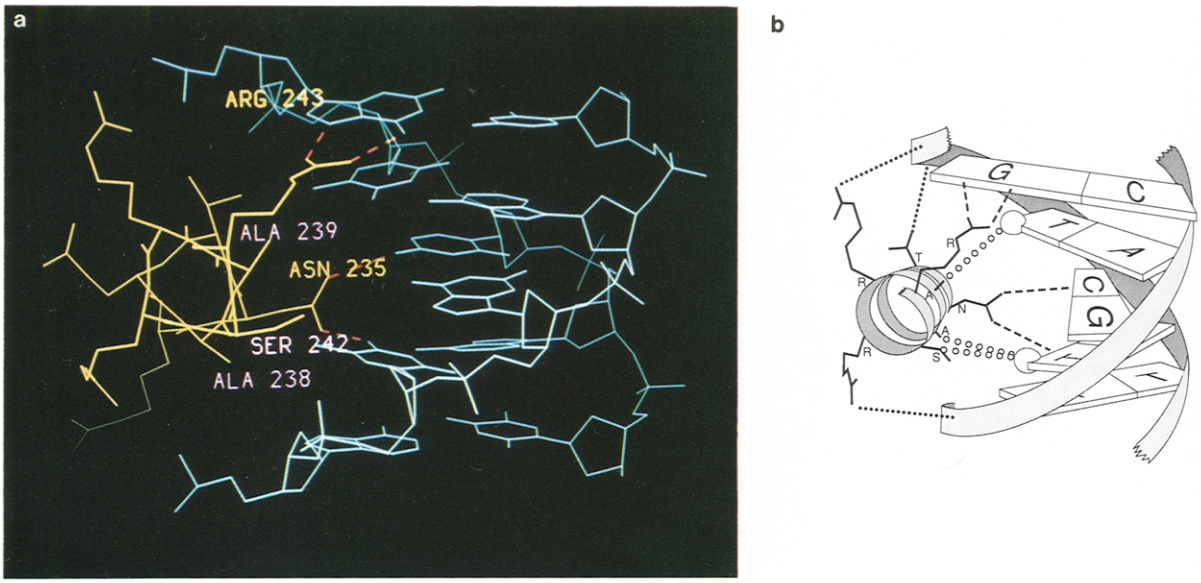

Figure 4.

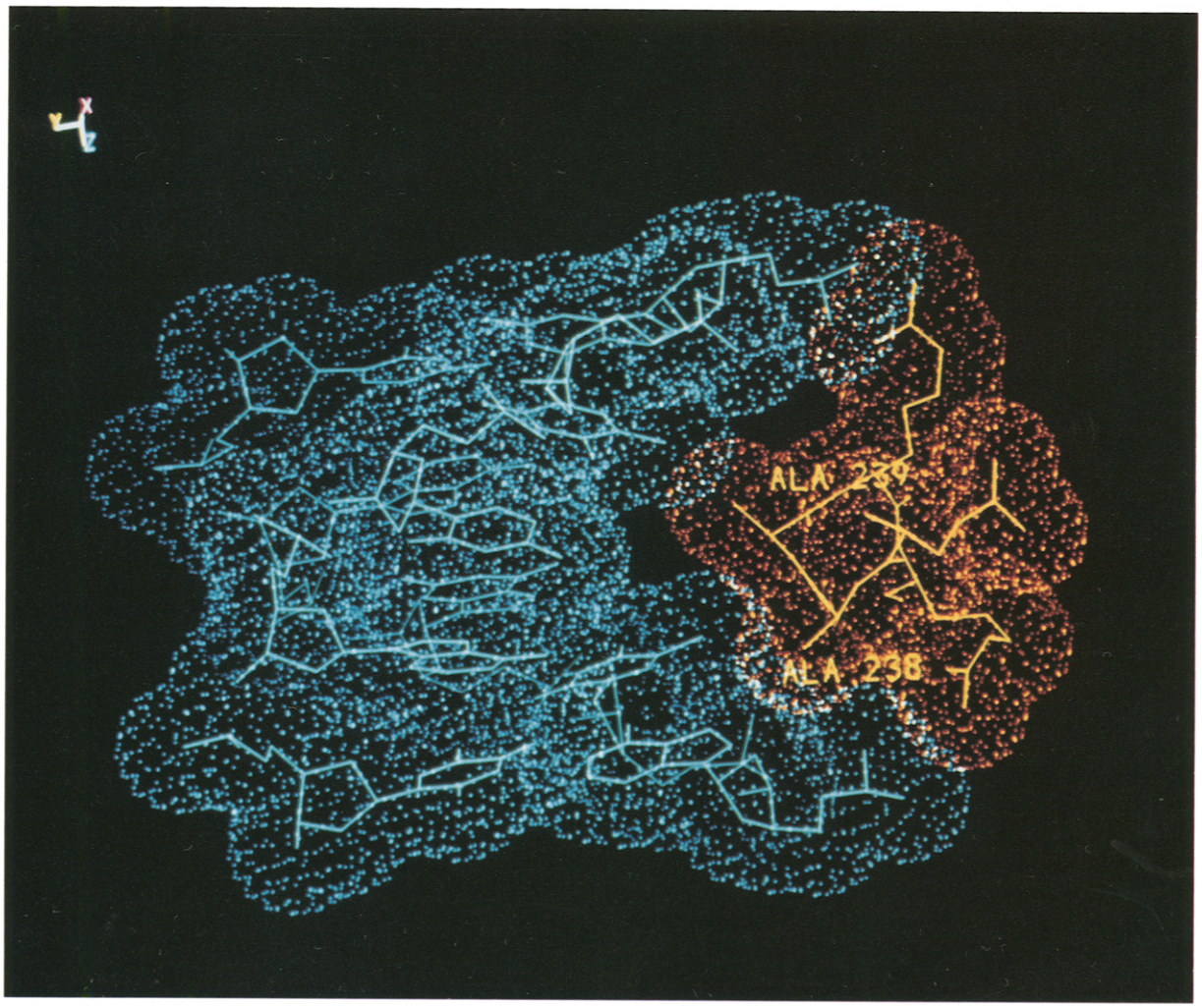

Figure 5.

the length of the DNA. Two adenines, which are located near the 5′ end of one strand, rotate out of the DNA helix, resulting in the collapse of this strand. The 5′ adenine of this strand now stacks against the fourth base in the strand, with the intervening two adenines forming a short loop (Figure 9b). This loop is stabilized by contacts to a symmetry-related molecule in the crystal. Electron density for the looped-out adenines is weak, suggesting some disorder.

The collapse of one DNA strand exposes the 3′ base (Thy-10′R) of the opposite strand for crystal-packing interactions with a symmetry-related DNA. The Watson–Crick base pair, which bridged DNA molecules in the crystals, is lost. Instead, two base triplets join adjacent DNAs in the new packing arrangement. Thy-9L stacks in the plane of base pair Ade-11R–Thy-9′R of the symmetry-related DNA, accepting a hydrogen bond from N_6 and donating a bond to N_7 of Ade-11R (Figure 9b). This base triplet stacks against a second triplet composed of Thy-10′R and Thy-8L–Ade-8′L. Thy-10′R O_2 accepts a hydrogen bond from Ade-8′L N_6. This rearrangement of the DNA conformation is restricted to bases at the ends of the DNA, which do not contact the protein, and it is presumably an artifact of soaking the crystals in an elevated concentration of PEG 400. The structure of the protein and the central 14 bp in the complex are the same before and after soaking the crystals in the cryoprotectant solution.

Discussion

Positioning the bZIP Element on DNA

The orientation of the basic region α helix in the major groove of DNA is crucial for making the base pair contacts that specify the AP-1 site. The basic region helical axis projects straight through the major groove. It is similar in this regard to helix 3 of the homeodomain, which traverses the major groove along a path that is almost parallel to the local direction of the phosphodiester backbone (Kissinger et al., 1990; Wolberger et al., 1991). This conserved orientation simply reflects the snug fit of an α helix in the major groove of B-form DNA. The rotational azimuth of the GCN4 basic region α helix is constrained by side chain–DNA interactions, and it differs from that of the engrailed and MATα2 homeodomains by about 20° (data not shown). Contacts from the GCN4 basic region include hydrogen bonds to the phosphodiester backbone and base-specific interactions, most of which are common to both half sites (Figure 8). This shared pattern of bonded and nonbonded interactions positions both monomers in nearly identical orientations with respect to their half sites.

Previous efforts at modeling the bZIP–DNA complex resulted in two distinct models of how an α-helical protein might bind a site extending over nearly one complete turn of the DNA helix. The scissors grip model proposed that the basic region consists of two α-helical segments joined by a kink, an arrangement that would permit the protein to wrap around the DNA and thereby facilitate contacts to the outside bases of the binding site (Vinson et al., 1989). It was suggested that the asparagine common to all bZIP proteins (Asn-235 of GCN4) would "cap" one helical segment of the basic region, breaking the basic region helix and permitting it to follow a helical path around the DNA. A second model, the induced helical fork, was proposed on the basis of experiments showing that the basic region of GCN4 was unfolded in solution and assumed a helical conformation only upon binding to DNA (O'Neil et al., 1990). The phenomonen of bZIP folding upon association with DNA has been observed for several bZIP elements (O'Neil et al., 1990; Weiss et al., 1990; Patel et al., 1990), as well as for full-length bZIP proteins (O'Neil et al., 1991). The authors of the induced helical fork model proposed that a partially structured basic region can interact with DNA and be stabilized in a relatively straight α-helical conformation by specific DNA contacts (O'Neil et al., 1990). The structure of the GCN4 bZIP complex bears out the central features of the induced helical fork model. Two relatively straight basic region α helices straddle the AP-1 site, directly contacting 4 bp of both half sites. The angle of protein approach to the major groove permits these DNA contacts to occur without sharply bending the basic region (Figures 3 and 4). The register of the basic region on DNA proposed in the induced helical fork model is approximately correct, although the details of the protein contacts to the DNA require amendment (Figures 4 and 8).

The DNA Footprint of GCN4

The positions of the GCN4 bZIP amino termini on DNA have been mapped by affinity DNA cleavage (Oakley and Dervan, 1990). GCN4 bZIP derivatives, with DNA scission agents covalently linked to their amino termini, produce patterns of DNA cleavage that are centered over each half site, 4–5 bp from the central Gua–Cyt-0 base pair. DNA

Figure 4. Detail of the Protein–DNA Interactions in the Left Half Site

(a) This view down the helical axis of the left half site protomer shows the conserved residue, Asn-235, in the center of the protein–DNA interface forming hydrogen bonds with O_4 of Thy-3 and N_4 of Cyt-2′. The basic region α helix is straight in the region of contact with the DNA. Asn-235 is surrounded by the β carbons of Ala-238 and Ala-239, which contact the thymine methyl groups of Thy-3 and Thy-1′, respectively. The β carbon of Ser-242 also contacts the Thy-3 methyl group. Arg-243, which is conserved in all bZIP proteins (Figure 10), can donate two hydrogen bonds to Gua-0′ at the center of the AP-1 site. Arg-243 of the right half site protomer contacts the phosphates of Cyt-0 and Ade-1L.

(b) Schematic of the DNA contacts shown in (a).

Figure 5. van der Waals Interactions between Ala-238 and Ala-239 and the Methyl Groups of Thy-3 and Thy-1

The van der Waals surfaces of two conserved alanines, Ala-238 and Ala-239, extend deep into the major groove of the binding site, where they contact thymine methyl groups at positions 3 and 1′, respectively.

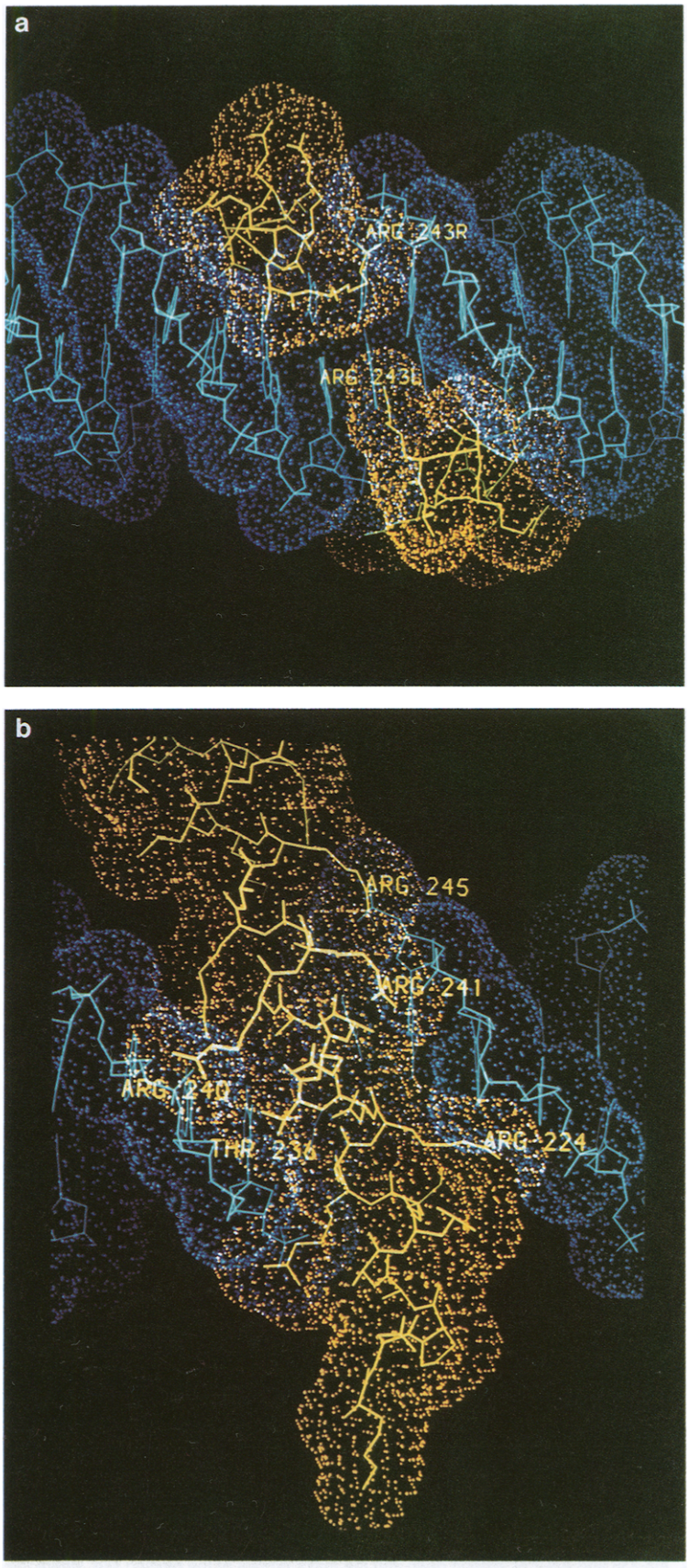

Figure 6. Asymmetric Protein Contacts to the Center of the AP-1 Site

(a) This view is from the perspective of the fork region of the protein, looking into the major groove at the center of the binding site. The two protein monomers lie in the major groove over either half site. Arg-243, which is absolutely conserved in the bZIP family (Figure 10), is in different conformations in the two monomers. Arg-243 of the left monomer (bottom of the figure) extends into the major groove and donates hydrogen bonds to N_7 and O_6 of Gua-0′. Arg-243 of the right monomer wraps back toward the DNA backbone, donating hydrogen bonds to the phosphates of Cyt-0 and Ade-1L.

(b) The basic region α helix fits snugly into the major groove of each half site, and it is anchored in this position by base contacts (see [a]; Figure 4) as well as by an array of basic and polar residues that form hydrogen bonds and electrostatic interactions with the DNA phosphodiester backbone.

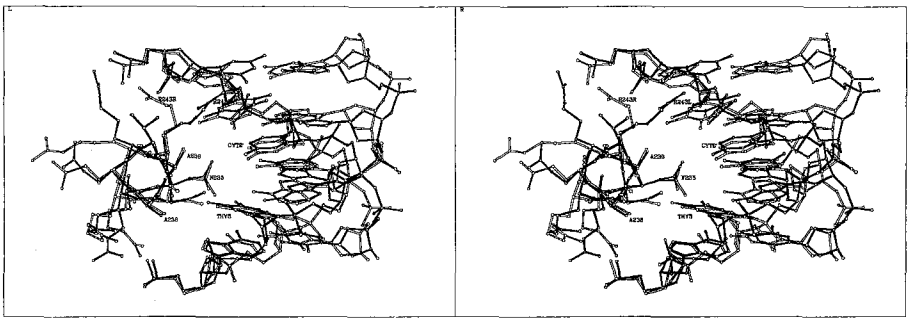

Figure 7. Superposition of the Protein Subunit–DNA Half Site Complexes

Superposition of the right half site–monomer complex onto the left, based on the best alignment of main chain coordinates for the two protomers. Although some variation in DNA backbone geometry is evident, the functional groups of bases contacted by the protein nearly superimpose. The RMS deviation between protomer–DNA complexes is 1.33 Å for all atoms shown, excluding the Gua–Cyt-0 base pair. Arg-243 of the left half site protomer contacts Gua-0 (at the top of the figure), whereas Arg-243 of the right monomer contacts the phosphate of Cyt-0 (see also Figure 6).

strand cleavage proceeds by a diffusible radical so that under conditions of limited cleavage the distribution of DNA strand breaks approximates the position of the scission agent. The cleavage pattern produced by a scission agent linked to Asp-226 (Oakley and Dervan, 1990) is in general agreement with the Asp-226–DNA–phosphate distances in the crystal structure. Asp-226 is three turns of the basic region helix amino-terminal to the residues that contact bases, and it is most closely apposed to phosphates 6 and 2′ of either half site. If the amino termini of the bZIP monomers wrapped around the back side of the DNA, as proposed in the scissors grip model (Vinson et al., 1990), Asp-226 would be located further away from the center of the binding site. The DNA affinity cleavage results are consistent with a straight, α-helical conformation of the basic region distal to the point of DNA contact.

The alkylation interference footprint of full-length GCN4 spans 11 to 14 bp of the AP-1 site (Gartenberg et al., 1990). Ethylation of phosphates at positions 1L to 5L, 1′L to 3′L, 0, 0′, 1R, 2R to 5R, 1′R, and 2′R of the AP-1 site strongly interferes with GCN4 binding (numbering according to Figure 1c). All of these phosphates are contacted by basic region residues, which contribute hydrogen bonds or electrostatic interactions (Figure 8). It is notable that phosphate 1R is not contacted by the bZIP element (Figure 8) and that ethylation of this phosphate does not strongly interfere with protein binding (Gartenberg et al., 1990). The symmetry-related phosphate 1L is contacted by Arg-243 of the right monomer. Arg-243 of the left monomer forms hydrogen bonds to Gua-0′ and therefore cannot contact phosphate 1R. The close correspondence between the phosphate contacts seen in the bZIP complex structure and the ethylation interference pattern for full-length GCN4 suggests that residues outside of the bZIP region contribute little, if at all, to DNA binding. Consistent with this suggestion, the bZIP element of GCN4 binds DNA with specificity and affinity indistinguishable from the full-length protein (Weiss et al., 1990).

Methylation of guanines at positions 5L, 2L, 0, and 2R of the AP-1 site disrupts GCN4 binding to DNA (Gartenberg et al., 1990). With the exception of position 5L, all of these bases are in intimate contact with the bZIP element in the crystal structure. Gua-2L and Gua-2R are located at the center of each half site, and Asn-235 accepts a hydrogen bond from Cyt-2′, the base pairing partner of these guanines (Figures 4 and 8). Gua-0′, at the center of

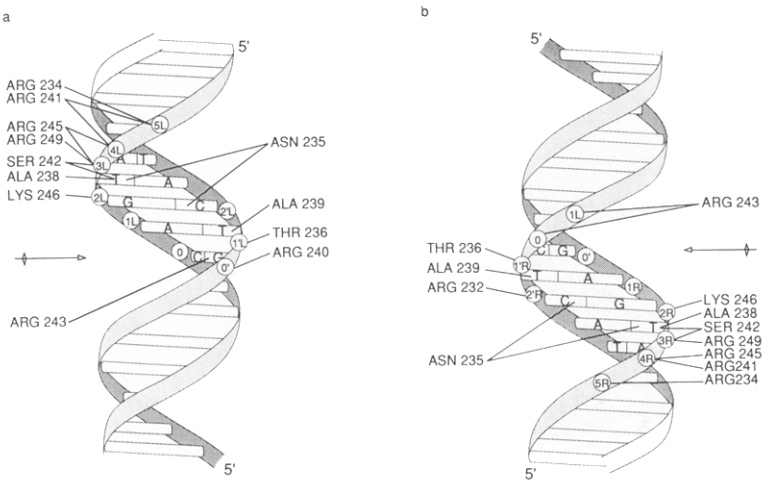

Figure 8. Summary of Protein–DNA Contacts in the GCN4 bZIP Complex

(a) Contacts made by the left half site monomer. The arrow shows the orientation of the DNA pseudodyad, pointing from the direction of the leucine zipper dimer interface.

(b) Contacts made by the right half site monomer. This view is related to that in (a) by a rotation about the DNA helical axis of approximately 180°.

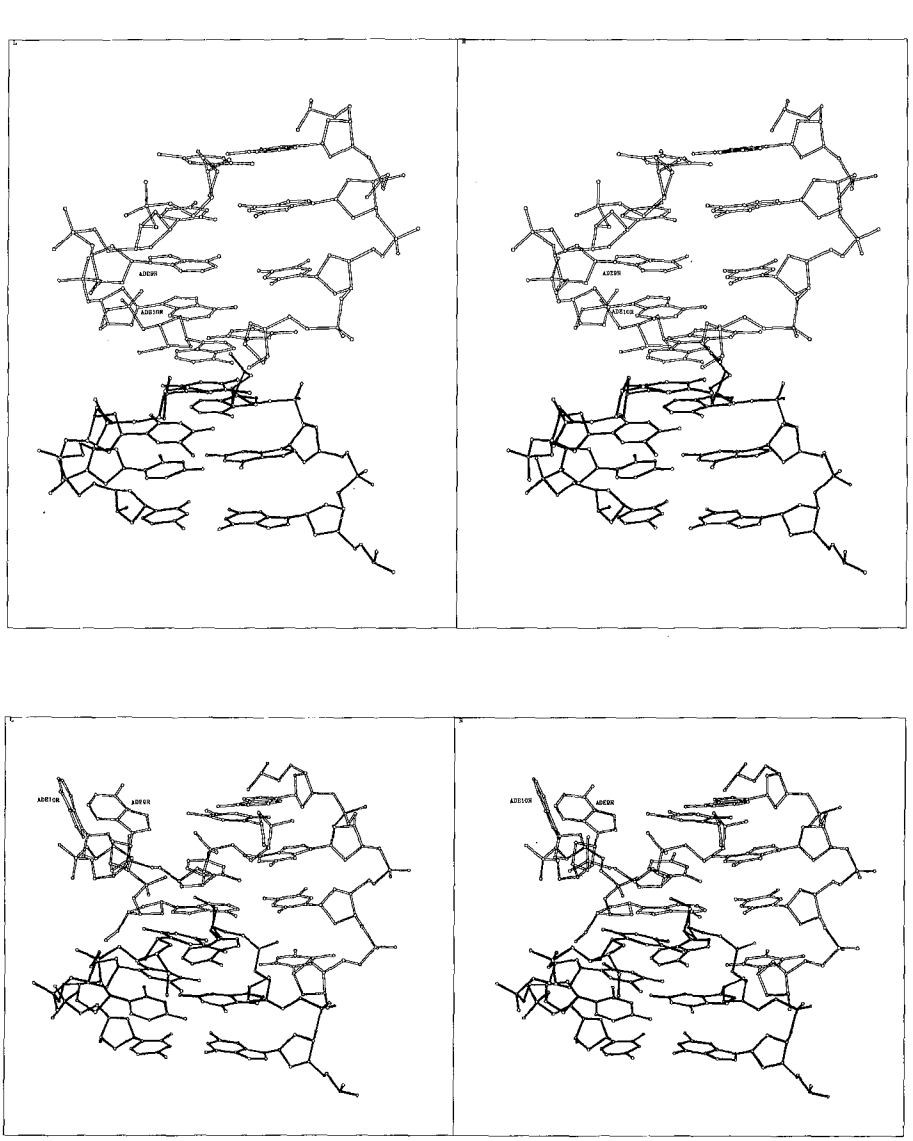

Figure 9. Alternative DNA Packing Contacts in the GCN4 Complex Crystals

(a) Stereo view of the crystallographic contacts between symmetry-related DNAs in the bZIP complex crystals. Each DNA complex contains 5′ complementary ends (Figure 1c), which base pair in the crystals, forming a pseudocontinuous double helix.

(b) In preparation for freezing, the crystals were equilibrated in a higher concentration of PEG 400, which causes a rearrangement of the DNA ends in the crystals. Two adenines (Ade-9R and Ade-10R) loop out of the DNA stack, resulting in the collapse of one strand and the formation of two base triplets between symmetry-related DNA.

the binding site, is directly contacted by Arg-243 of the left monomer (Figures 6 and 8). No direct contacts are seen to base 5L, which is a thymine in the complex studied here (Figure 1c). The identities of base pairs 5L and 5R have negligible influence upon the binding affinity of GCN4 (Hill et al., 1986), and the disruptive effect of guanine methylation at position 5 may therefore be indirect. Methylation of adenine N_3 at positions 3′ and 1 of each half site also interferes with GCN4 binding (Hill et al., 1986). This result is unexpected because this modification lies in the minor groove. The GCN4 complex does not yield a detectable hydroxy-radical footprint (Hill et al., 1986), and no minor groove contacts are made by the protein in the crystal structure (Figure 3). Adenine methylation may interfere with GCN4 complex formation owing to the enhanced rate of depurination of N_3-methyladenine or the delocalized charge on the purine ring of N_3-methyladenine, which could interfere with protein binding in the major groove.

Conformation of the AP-1 DNA in the GCN4 Complex

The DNA in the GCN4 bZIP complex is straight, and its conformation is in the B form across the region contacted by the protein (Figure 3). A complex of full-length GCN4 with DNA has an electrophoretic mobility in gels that is dependent upon the position of the binding site with re-

spect to the DNA ends, results similar to those obtained with proteins that bend DNA (Gartenberg et al., 1990). However, the positional dependence of the complex mobility is not influenced by the addition of a static bend in the DNA, suggesting that the aberrent mobility of the complex does not result from protein-induced bending of the DNA. It may instead be due to the unusual shape of the protein, which creates a T- or L-shaped complex, depending upon the position of the binding site with respect to the DNA ends, or it may be due to increased DNA flexibility induced by the protein (Gartenberg et al., 1990). The Fos–Jun heterodimer complex with DNA also has an electrophoretic mobility that is dependent upon binding site position, but, in this case, the mobility is influenced by a static bend in the DNA (Kerppola and Curran, 1991b, 1991c). These results are consistent with bending of the AP-1 DNA by the Fos–Jun heterodimer, but the GCN4 bZIP complex structure does not lend itself to a simple explanation of how a bZIP element might stabilize DNA in a bent conformation. The DNA in the GCN4 crystal structure is extensively contacted by both protomers without distortion of B-form DNA geometry. For some proteins, it is possible that residues outside the bZIP element influence the conformation of the bound DNA.

Base-Specific Interactions with the AP-1 Site

GCN4 appears to specify the core of the AP-1 sequence through contacts to the central 7 bp. Base pair substitutions at any of these positions markedly reduce GCN4 binding affinity (Hill et al., 1986; Oliphant et al., 1989; Sellers et al., 1990; Pu and Struhl, 1992). Five amino acids from one monomer and four from the other make contacts with the bases (Figure 8). Arg-243 of the left half site monomer anchors the protein to the center of the binding site, donating two hydrogen bonds to Gua-0′ (Figure 6a). Asn-235 bridges the DNA strands, hydrogen bonding with bases at positions 3 and 2′ of each half site (Figure 4). This conserved asparagine could in principle specify the binding site sequence at positions 3 and 2 as either 5′-Thy-Gua-3′ or 5′-Cyt-Ade-3′ by alternative choices of rotation about the Cβ–Cα bond, but Ala-238, together with Ser-242, appears to select the half site sequence as 5′-Thy-Gua-3′ by van der Waals contacts with the methyl group of Thy-3 (Figure 5). Moreover, close apposition of base pair 2 and the Asn-235 side chain excludes a thymine methyl group from positions 2′L and 2′R of the AP-1 site. A van der Waals interaction between Ala-239 and a thymine methyl group specifies position 1 of either half site (Figures 5 and 8). In this way, Asn-235, Ala-238, Ala-239, Ser-242, and Arg-243 can specify the identities of three contiguous base pairs of each half site and the central guanine–cytosine base pair. GCN4 exhibits some sequence specificity for bases at position 4, flanking the AP-1 core, with purines being preferred over pyrimidines (Hill et al., 1986; Oliphant et al., 1989). The degree of sequence preference is somewhat greater at position 4R than at position 4L. We see two residues that could contribute to specificity at base pair 4, Lys-231 and Asn-235. In our model, the side chain of Lys-231 lies near base pair 4 on the right half site, and a water molecule bridges between Asn-235 and Ade-4 on both half sites.

The significance of the basic region van der Waals contacts to thymine methyl groups in the AP-1 site has been demonstrated by chemical methods. Substitution of 5-deoxyuridine for thymine at position 3 or 1 would eliminate contacts with Ala-238 and Ala-239, and, indeed, it markedly reduces binding affinity for GCN4 (Pu and Struhl, 1992). Similar results have been obtained for the Fos–Jun heterodimeric complex with the AP-1 site (Risse et al., 1989). A contact between Ala-238 and Thy-3 has also been shown by ultraviolet-induced cross-linking of Ala-238 to a 5-bromo-deoxyuridine-substituted binding site (Blatter et al., 1992). Genetic selection for GCN4 mutants with DNA binding activity on altered binding sites has yielded mutants with large, hydrophobic side chain substitutions at positions 235, 238, 239, and 242 (Tzamarias et al., 1992; J. Kim, D. Tzamarias, and K. S., unpublished data), which are located on the DNA-contacting surface of the basic region (Figure 4). In general, these mutants have gained binding activity to sites containing adenine–thymine base pair substitutions or transversions. The results are consistent with the view that thymine methyl groups in the major groove strongly influence GCN4–DNA binding specificity.

AP-1 Site Recognition by Fos and Jun

The AP-1 site in the GCN4 complex (Figure 1c) is the same as that bound specifically by the proto-oncogenes *fos* and *jun*. Fos and Jun preferentially heterodimerize, and it is dimer stability that determines the relative DNA binding affinities of hetero- and homodimeric pairings of these proteins (Kouzarides and Ziff, 1988; Smeal et al., 1989; Turner and Tjian, 1989; Gentz et al., 1989; Cohen and Curren, 1990; Schuermann et al., 1991; O'Shea et al., 1992). The basic region residues of Fos and Jun family members are homologous to those of GCN4, with absolute conservation of all the residues forming base-specific contacts in the GCN4 complex (Figure 10). In fact, the Jun basic region can be substituted for that of GCN4, creating a protein with wild-type GCN4 activity in vivo (Struhl, 1987). The Fos basic region also binds the AP-1 site as a homodimer, providing that the GCN4 leucine zipper is substituted for that of Fos (Sellers and Struhl, 1989) or that one of the protomers is supplied with a Jun leucine zipper (Risse et al., 1989). It is therefore likely that the Fos–Jun heterodimer contacts the AP-1 site in a manner similar to the GCN4 homodimer. This scheme predicts that either Fos or Jun contacts Gua-0′ at the center of the AP-1 pseudopalindrome. A preferential interacton of either Fos or Jun with Gua-0′ would orient the heterodimer with respect to the asymmetric AP-1 site. The relative probability of either Fos or Jun contacting Gua-0′ might be influenced by residues flanking the conserved arginine (GCN4 residue 243), which could influence the bending angle of the basic region fork in the complex or alter the stability of the partially folded basic region prior to binding DNA.

Recognition of the ATF/CREB Site

GCN4 also binds specifically, although with slightly reduced affinity, to the consensus sequence recognized by

```
gcn4     DPAALKRARNTEAARRSRARKLQRMKQL
acr1     RKRKEFLERNRVAASKFRKRKKEYIKKI
atf-1    LKREIRLMKNREAARECRRKKKEYVKCL
atf-2    EKRRKVLERNRAAASRCRQKRKVWVQSL
atf-3    ERKKRRRERNKIAAAKCRNKKKEKTECL
atf-4    DKKLKKMEQNKRAATRYRQKKRAEQEAL
atf-5    REKENPKERNKMAAAKCRNRRRELTDTL
atf-6    LRRQQRMIKNRESACQSRKKKKEYMLGL
c/ebp    NEYRVRRERNNIAVRKSRDKAKQRNVET
cpc-1    DVVAMKRARNTLAARKSRERKAQRLELL
creb     RKREVRLMKNREAARECRRKKKEYVKCL
crem     RKRELRLMKNREAARECRRKKKEYVKCL
cys-3    AAEEDKRKRNTAASARFRIKKKQREQAL
embp-1   LKRERRKQSNRESARRSRLRKQQECEEL
fos-c    EKRRIRRERNKMAAAKCRNRRRELTDTL
fra-1    ERRRVRRERNKLAAAKCRNRRKELTDFL
hbp-1    ELKAQKRLSNRESARRSRLRKQAECEEL
jun-c    IKAERKRMRNRIAASKCRKRKLERIARL
junB     IKVERKRLRNRLAATKCRKRKLERIARL
junD     IKAERKRLRNRIAASKCRKRKLERISRL
lap      DEYKMRRERNNIAVRKSRDKAKMRNLET
taf-1     KREKRKQSNRESARRSRLRKQAEAEEL
opaque2  ERVRKRKESNRESARRSRYRKAAHLKEL
pap1     EPSSKRKAQNRAAQRAFRKRKEDHLKAL
tga-1    EKVLRRLAQNREAARKSRLRKKAYVQQL
yap-1    ETKQKRTAQNRAAQRAFRERKERKMKEL
```

Figure 10. Sequence Alignment of the Basic Regions of bZIP Proteins

This partial listing of bZIP protein basic region sequences highlights the residues contacting DNA bases in the GCN4 complex (large type). Those that are common to most (or all) bZIP family members are in bold.

the ATF/CREB family of proteins (Sellers et al., 1990). The ATF/CREB site is similar to the AP-1 site, but it has an additional central base pair that creates a true palindrome of 5′-TGAC-3′ half sites. Therefore, the ATF/CREB site can be thought of as an expanded AP-1 site in which the half sites are rotated by 36° and displaced from one another by 3.4 Å to admit the additional cytosine–guanine base pair. To accomodate this extra base pair, it is necessary either to increase the angle of monomer divergence or to bend the segment between Arg-243 and Met-250. It is probable that Arg-243 from each monomer contacts the appropriate guanine at the center of the ATF/CREB site, since GCN4 fails to bind a palindrome of two 5′-TGAG-3′ half sites, where the central guanines are located on the opposite strand (Sellers et al., 1990). An additional arginine–guanine interaction might stabilize a bend in the α helix at the fork region of the protein. Alternatively, it might permit a gradual divergence of the monomers over the amino-teminal segment of the leucine zipper, compensating for some loss in dimerization free energy. In either case, recognition of the AP-1 and ATF/CREB sites clearly requires some flexibility in the fork segment of GCN4.

GCN4 binds to the ATF/CREB site with 2- to 5-fold lower affinity than to the AP-1 site, whereas members of the ATF and CREB families of bZIP proteins show the opposite preference, with many binding weakly to the AP-1 site (Hai et al., 1988, 1989). ATF and CREB proteins show conservation of residues at the positions of GCN4 residues Asn-235, Ala-238, Ala-239, and Arg-243 (Figure 10), implicating these residues as important DNA-binding determinants of the ATF and CREB proteins. Relative binding preferences for the AP-1 and ATF/CREB sites may derive from residues in the fork segment, which stabilize the orientation of the basic region helices in a conformation suitable for either the palindromic ATF/CREB site or the pseudopalindromic AP-1 site. Recent mutational analyses of GCN4 have shown that the identity of residue 247 is particularly relevent in this regard (Figure 11; J. Kim, D. Tzamarias, and K. S., unpublished data).

DNA Recognition by Other bZIP Proteins

The basic regions of the more than 30 identified bZIP proteins are closely related, especially at those residues that contact the DNA in the GCN4 complex (Figure 10). The identities of Asn-235 and Arg-243 are absolutely conserved in the bZIP family. Ala-238 and Ala-239 are found in most bZIP proteins; sequence variants contain valine, serine, or glutamine at one of these positions and alanine at the other. Ser-242, which forms a van der Waals interaction with the methyl group of Thy-3, is typically serine or cysteine in other bZIP proteins, although aromatic residues are present in a few examples. A slight reorientation of the Ser-242 side chain would allow it to donate a hydrogen bond to the phosphate of Thy-3; however, this bond is not found in the current refined model. Many of the basic residues contacting the DNA phosphates in the GCN4 complex are conserved in the bZIP family. The high degree of sequence conservation in the bZIP basic region, particularly at residues making DNA contacts in the GCN4 complex, would suggest that the basic region of these proteins contacts DNA with approximately the same register and orientation as the GCN4 basic region.

If we assume that the conformation of the bZIP element is approximately the same for all bZIP proteins, what can be concluded about the interactions of other bZIP proteins with their cognate binding sites? Most of the reported binding sites for other bZIP proteins show little homology with either the AP-1 site or the ATF/CREB site. However, a common feature of many of these sites is one or more central cytosine–guanine base pairs, which define the dyad axis of an approximate palindrome. On this basis, these sites can be classified as AP-1 like, with overlapping half sites, or ATF/CREB like, with abutting half sites. The sequences of half sites so defined show a strong preference for thymine at position 3 and thymine or guanine at position 2. These sequences would permit the invariant basic region asparagine (Asn-235 of GCN4) to form hydrogen bonds with O_4 of thymine at position 3 and N_4 of cytosine or N_6 of adenine at position 2′. The invariant arginine (Arg-243 of GCN4) of one or both monomers could interact with the central guanine(s). Conservation of alanine at the position of GCN4 Ala-238 is consistent with the frequent occurrence of thymine at position 3 of the binding sites. Thymine is often present at position 1′ as well, and it could interact with the homolog of GCN4 Ala-239, which is a hydrophobic residue in many other bZIP proteins (Figure 10). It should be emphasized that these assignments are speculative and based on the DNA contacts made by the

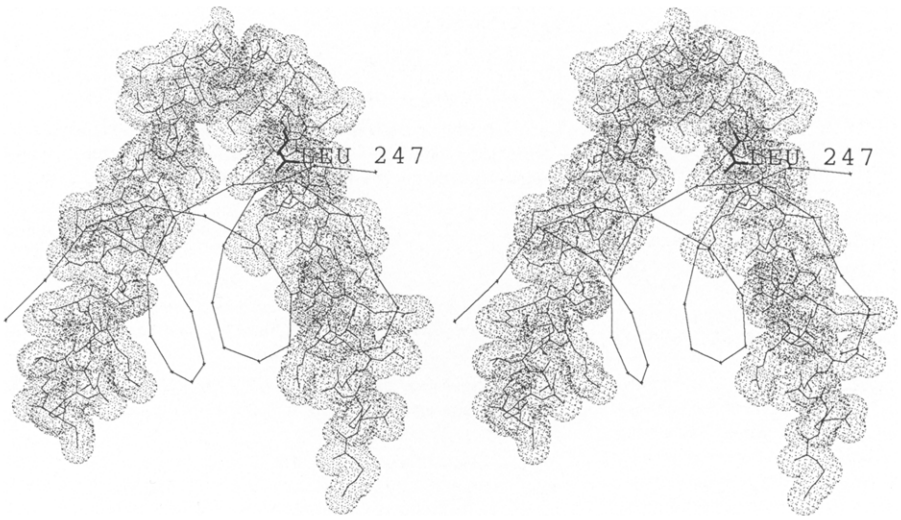

Figure 11. Detail of the Fork Region of the bZIP Dimer

The GCN4 subunits diverge at the amino-terminal end of the coiled-coil dimerization interface, allowing each basic region α helix to traverse the major groove. In this stereo diagram, the DNA phosphodiester backbone is represented by a line trace through the phosphate positions. While GCN4 prefers the pseudopalindromic AP-1-binding site over the palindromic ATF/CREB sequence, mutation of Leu-247 to lysine or arginine changes this relative preference in favor of the ATF/CREB half site spacing.

GCN4 protein. While it is possible to model interactions between the conserved Asn . . . Ala–Ala . . . Ser–Arg quintet and the binding sites of other bZIP proteins, their binding specificities cannot be fully accounted for by these conserved residues oriented as they are in the GCN4 complex. Clearly other factors, such as subtle differences in the basic region orientation on DNA or DNA contacts from nonconserved residues, influence the DNA binding specificities of various bZIP family members.

Experimental Procedures

Crystallization of the bZIP-DNA Complex

The pGCNK58 peptide, which comprises the carboxy-terminal 56 residues of GCN4, with the residues methionine–lysine appended to the amino terminus, was expressed in E. coli and purified to near homogeneity by ion exchange chromatography as previously described (Weiss et al., 1990). For crystallization trials, pGCNK58 was complexed with synthetic AP-1 DNA oligomers ranging in length from 16 to 22 bases. DNAs were synthesized by standard phosphoramidite chemistry and purified by reverse-phase high performance liquid chromatography and/or gel electrophoresis in 20% acrylamide per 7 M urea gels. The best complex crystals were obtained by vapor diffusion at 22°C from a solution of 0.95 mM pGCNK58 monomer, 0.57 mM DNA (a duplex of 5′-TTCCTATGACTCATCCAGTT-3′ and 5′-AAACTGGATGAGTCATAGGA-3′), 0.15 M NaCl, 0.03 M $MgCl_2$, 0.025 M MES buffer (pH 5.75), 1 mM spermine–HCL, and 12% (v/v) PEG 400. This hanging drop was equilibrated over a well solution containing twice these concentrations of salts and PEG 400. Plate-like crystals grew over the course of 2 weeks to final dimensions of approximately 0.7 mm × 0.3 mm × 0.1 mm. These crystals belong to the space group $P2_12_12_1$ and have unit cell dimensions of a = 49.3 Å, b = 90.3 Å, c = 65.4 Å (22°C), with one protein–DNA complex in the asymmetric unit (62% solvent content). Isomorphous derivatives of the pGCNK58–DNA complex were prepared with 5I-dU-substituted DNAs, using crystallization conditions identical to those for the native complex. Four isomorphous derivatives, representing five independent atomic positions, were utilized in the structure determination. These include substitutions at thymines 5L, 8L, 4′L, 7R, and 4′R (see Figure 1c for numbering scheme). Crystals failed to grow from mixtures of pGCNK58 with DNAs containing 5I-dU substitutions at position 1′ or 3 of the binding site (see Figure 1 for numbering of the AP-1 site). These thymine methyl groups are contacted by the protein (Figure 5).

X-Ray Data Collection and Processing

X-ray diffraction data were collected from native and 5I-dU derivative crystals with a Nicolet–Siemens area detector mounted on an Elliott GX-13 X-ray source. Reflection intensities were integrated and scaled with the programs XDS and XSCALE (Kabsch, 1988). The native crystals diffract to a maximum resolution of 2.8 Å. However, diffraction anisotropy and the limited crystal lifetime in the X-ray beam restricted data collection at 22°C to 3.1 Å resolution (Table 1). The diffraction quality of the derivative crystals was dependent upon the sites of heavy atom substitution, with maximum resolutions ranging from 3.5 to 3.1 Å for various derivatives (Table 1). Diffraction data was also collected from frozen native crystals at −160°C. Prior to freezing, crystals were soaked for 45 min in harvest buffer containing 35% PEG 400 as cryoprotectant (standard harvest conditions included 24% PEG 400). Following this soak, the crystal was suspended in a thin film of harvest buffer contained in a 0.7 mm diameter nichrome wire loop and immediately frozen in a chilled nitrogen stream. Freezing extended the crystal lifetime in the X-ray beam, allowing the collection of a complete data set to 2.9 Å resolution from one crystal.

Structure Determination

The iodine positions in the 5I-dU derivative crystals were located by difference Patterson methods and refined against origin-removed Patterson coefficients, using the program HEAVY (Terwilliger and Eisenberg, 1983). These positions were confirmed by difference Fourier analysis, and heavy atom parameters for five independent sites from four derivatives were simultaneously refined against the most probable phase. The results of heavy atom refinement are shown in Table 1.

The initial multiple isomorphic replacement (MIR) electron density map showed clear density corresponding to the DNA and both α-helical protein monomers. A B-form DNA model and the refined GCN4 coiled-coil model (GCN4 residues 249–279; O'Shea et al., 1991) were positioned in the electron density. The basic region of each monomer was modeled as a straight polyalanine helix. This initial model guided the fitting of a generous envelope around protein and DNA density. Electron density outside the envelope was flattened by cycles of phase calculation from the enveloped density and combination with MIR phases (Bricogne, 1976; Figure 2a). Prior to each cycle of phase combination, electron density in a 3.5 Å sphere around each heavy atom site was set to zero to remove artifacts around the sites of heavy atom

substitution. Phases calculated from the flattened map converged after several cycles, and these were used in a second round of heavy atom refinement. This method of refinement avoids the correlated refinement of heavy atom sites common to several derivatives (Rould et al., 1989).

A second solvent-flattened MIR map was calculated, incorporating phases from the newly refined heavy atom parameters. This MIR map showed clear density for a number of protein side chains that were not evident in the initial map, and it confirmed that the coiled-coil model was in the correct register with respect to electron density. Side chains were added as appropriate, and the model was rebuilt using the solvent-flattened density as a guide. This partial model was subjected to energy minimization using the program X-PLOR (Brünger, 1990). Dynamic restraints were applied to the DNA dihedral angles during all stages of X-PLOR refinement to maintain proper deoxyribose ring geometry.

After several cycles of energy minimization refinement and manual rebuilding, the model (crystallographic R factor = 0.28, for reflections with F/SIGF >1.5) was used for the molecular replacement solution of the −160°C diffraction data set, which extended to higher resolution. Since the unit cell axis parallel to the DNA axis was significantly shorter at −160°C, bases at the ends of the DNA model were removed prior to use in molecular replacement. The 3′ base and two 5′ bases were removed from each DNA strand. A unique solution was identified with a Patterson space rotation search and translational correlation search (Brünger, 1990). Energy minimization refinement of the partial model against the −160°C data reduced the R factor from 0.49 to 0.32. A difference Fourier synthesis, using phases calculated from the partially refined model, showed the positions of the bases omitted from the ends of the DNA. These bases were added, and the model was manually rebuilt before further refinement. Cycles of simulated annealing and energy minimization brought the crystallographic R factor to 0.23 on all reflections (8 Å–2.9 Å). The positions of selected segments of the model were periodically checked by examining electron density difference maps calculated following cycles of simulated annealing and energy minimization refinement with these residues omitted (Brünger, 1990). Refinement was completed by conventional least squares refinement without torsional restraints, using the program TNT (Tronrud et al., 1987). The crystallographic R factor of the final model is 22.8%, with RMS deviation in bond lengths and bond angles of 0.023 Å and 3.3°, respectively (Table 1).

Acknowledgments

We thank Drs. Tom Alber and Erin O'Shea for providing coordinates of the GCN4 leucine zipper peptide prior to publication and members of the Harrison and Struhl laboratories for discussions and suggestions. This work was funded in part by National Institutes of Health grants to S. C. H. and to K. S. and by the Lucille P. Markey Charitable Trust. C. J. B. was supported by a fellowship from the Medical Research Council of Canada. T. E. E. is a Lucille P. Markey Scholar. Coordinates of the GCN4 complex are being deposited in the Brookhaven Data Bank.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received September 21, 1992; revised October 23, 1992.

References

Blatter, E. E., Ebright, Y. W., and Ebright, R. H. (1992). Identification of an amino acid–base pair contact in the GCN4–DNA complex by bromouracil-mediated photocrosslinking. Nature *359*, 650–652.

Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K., and Tjian, R. (1987). Human proto-oncogene c-*jun* encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. Science *238*, 1386–1392.

Bowie, J. U., and Sauer, R. T. (1989). Equilibrium dissociation and unfolding of the arc repressor dimer. Biochemistry *28*, 7139–7143.

Bricogne, G. (1976). Methods and programs for direct-space exploitation of geometric redundancies. Acta Cryst. *32*, 832–846.

Brünger, A. T. (1990). X-PLOR v.2.1 Manual (New Haven, Connecticut: Yale University Press).

Clarke, N. D., Beamer, L. J., Goldberg, H. R., Berkower, C., and Pabo, C. O. (1991). The DNA binding arm of λ repressor: critical contacts from a flexible region. Science *254*, 267–270.

Cohen, D. R., and Curran, T. (1990). Analysis of dimerization and DNA-binding functions in Fos and Jun by domain-swapping: involvement of residues outside the leucine zipper/basic region. Oncogene *5*, 929–939.

Crick, F. H. C. (1953). The packing of α-helices: simple coiled-coils. Acta Cryst. *6*, 689–697.

Franza, B. R., Rauscher, F. J., III, Josephs, S. F., and Curran, T. (1988). The Fos complex and Fos-related antigens recognize sequence elements that contain AP-1 binding sites. Science *239*, 1150–1153.

Gartenberg, M. R., Amphe, C., Steitz, T. A., and Crothers, D. M. (1990). Molecular characterization of the GCN4–DNA complex. Proc. Natl. Acad. Sci. USA *87*, 6034–6038.

Gentz, R., Rauscher, F. J., III, Abate, C., and Curran, T. (1989). Parallel association of Fos and Jun leucine zippers juxtaposes DNA-binding domains. Science *243*, 1695–1699.

Hai, T., Liu, F., Allegretto, E. A., Karin, M., and Green, M. R. (1988). A family of immunologically related transcription factors that includes multiple forms of ATF and AP-1. Genes Dev. *2*, 1216–1226.

Hai, T., Liu, F., Allegretto, E. A., Karin, M., and Green, M. R. (1989). Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. Genes Dev. *3*, 2083–2090.

Harrison, S. C., and Aggarwal, A. K. (1990). DNA recognition by proteins with the helix-turn-helix motif. Annu. Rev. Biochem. *59*, 933–969.

Hill, D. E., Hope, I. A., Macke, J. P., and Struhl, K. (1986). Saturation mutagenesis of the yeast *his3* regulatory site: requirements for transcriptional induction and for binding by GCN4 activation protein. Science *234*, 451–457.

Hope, I. A., and Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell *46*, 885–894.

Johnson, P. F., and McKnight, S. L. (1989). Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. *58*, 799–839.

Kabsch, W. (1988). Evaluation of single crystal X-ray diffraction data from a position-sensitive detector. J. Appl. Cryst. *21*, 916–924.

Kaptein, R. (1991). Zinc-finger structures. Curr. Opin. Struct. Biol. *2*, 109–115.

Kerppola, T. K., and Curran, T. (1991a). Transcription factor interactions: basics on zippers. Curr. Opin. Struct. Biol. *1*, 71–79.

Kerppola, T. K., and Curran, T. (1991b). Fos–Jun heterodimers and Jun homodimers bend DNA in opposite orientations: implications for transcription factor cooperativity. Cell *66*, 317–326.

Kerppola, T. K., and Curran, T. (1991c). DNA bending by Fos and Jun: the flexible hinge model. Science *254*, 1210–1214.

Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (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.

Kouzarides, T. and Ziff, E. (1988). The role of the leucine zipper in the Fos–Jun interaction. Nature *336*, 646–651.

Oakley, M. G., and Dervan, P. B. (1990). Structural motif of the GCN4 DNA binding domain characterized by affinity cleaving. Science *248*, 847–850.

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

O'Neil, K. T., Hoess, R. H., and DeGrado, W. F. (1990). Design of DNA binding peptides based on the leucine zipper motif. Science *249*, 774–778.

O'Neil, K. T., Shuman, J. D., Ampe, C., and DeGrado, W. F. (1991). DNA-induced incease in the α-helical content of C/EBP and GCN4. Biochemistry *30*, 9030–9034.

O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989). Evidence that the leucine zipper is a coiled coil. Science *243*, 538–542.

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

O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1992). Mechanism of specificity in the Fos–Jun oncoprotein heterodimer. Cell *68*, 699–708.

Patel, L., Abate, C., and Curran, T. (1990). Altered protein conformation on DNA binding by Fos and Jun. Nature *347*, 572–574.

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

Pu, W. T., and Struhl, K. (1992). Uracil interference, a rapid and general method for defining protein–DNA interactions involving the 5-methyl group of thymines: the GCN4–DNA complex. Nucl. Acids Res. *20*, 771–775.

Rauscher, F. J., III, Sambucetti, L. C., Curran, T., Distel, R. J., and Spiegelman, B. M. (1988). Common DNA binding site for Fos protein complexes and transcription factor AP-1. Cell *52*, 471–480.

Risse, G., Jooss, K., Neuberg, M., Bruller, H.-J., and Müller, R. (1989). Asymmetrical recognition of the palindromic AP1 binding site (TRE) by Fos protein complexes. EMBO J. *8*, 3825–3832.

Rould, M. A., Perona, J. J., Söll, D., and Steitz, T. A. (1989). Structure of *E. coli* glutaminyl–tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. Science *246*, 1135–1142.

Saudek, V., Pasley, H. S., Gibson, T., Gausepohl, H., Frank, R., and Pastore, A. (1991). Solution structure of the basic region from the transcriptional activator GCN4. Biochemistry *30*, 1310–1317.

Schuermann, M., Hunter, J. B., Hennig, G., and Müller, R. (1991). Non-leucine residues in the leucine repeats of Fos and Jun contribute to the stability and determine the specificity of dimerization. Nucl. Acids Res. *19*, 739–746.

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

Sellers, J. W., and Struhl, K. (1989). Changing Fos oncoprotein to a Jun-independent DNA-binding protein with GCN4 dimerization specificity by swapping leucine zippers. Nature *341*, 74–76.

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

Smeal, T., Angel, P., Meek, J., and Karin, M. (1989). Different requirements for formation of Jun:Jun and Jun:Fos complexes. Genes Dev. *3*, 2091–2100.

Struhl, K. (1987). The DNA-bind domains of the jun oncoprotein and the yeast GCN4 transcriptional activator protein are functionally homologous. Cell *50*, 841–846.

Talanian, R. V., McKnight, C. J., and Kim, P. S. (1990). Sequence-specific DNA binding by a short peptide dimer. Science *249*, 769–773.

Terwilliger, T. C., and Eisenberg, D. (1983). Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions. Acta Cryst. *39*, 813–817.

Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. Acta Cryst. *43*, 489–501.

Turner, R., and Tjian, R. (1989). Leucine repeats and an adjacent DNA-binding domain mediate the formation of functional c-fos–c-jun heterodimers. Science *243*, 1689–1694.

Tzamarias, D., Pu, W. T., and Struhl, K. (1992). Mutation in the bZIP domain of yeast GCN4 that alter DNA-binding specificity. Proc. Natl. Acad. Sci. USA *89*, 2007–2011.

Vinson, C. R., Sigler, P. B., and McKnight, S. L. (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science *246*, 911–916.

Weiss, M. A. (1990). Thermal unfolding studies of a leucine zipper domain and its specific DNA complex: implications for scissors' grip recognition. Biochemistry *29*, 8020–8024.

Weiss, M. A., Elison, J. L., and States, D. J. (1984). Dynamic filtering by two-dimensional ^1H NMR with application to phage λ repressor. Proc. Natl. Acad. Sci. USA *81*, 6019–6023.

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

Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991). Crystal structure of a MAT α2 homeodomain–operator complex suggests a general model for homeodomain–DNA interactions. Cell *67*, 517–528.