Nucleic Acids Research, 1995, Vol. 23, No. 13 2531–2537

Determinants of half-site spacing preferences that distinguish AP-1 and ATF/CREB bZIP domains

Joon Kim+ and Kevin Struhl*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

Received February 28, 1995; Revised and Accepted April 20, 1995

ABSTRACT

The AP-1 and ATF/CREB families of eukaryotic transcription factors are dimeric DNA-binding proteins that contain the bZIP structural motif. The AP-1 and ATF/CREB proteins are structurally related and recognize identical half-sites (TGAC), but they differ in their requirements for half-site spacing. AP-1 proteins such as yeast GCN4 preferentially bind to sequences with overlapping half-sites, whereas ATF/CREB proteins bind exclusively to sequences with adjacent half-sites. Here we investigate the distinctions between AP-1 and ATF/CREB proteins by determining the DNA-binding properties of mutant and hybrid proteins. First, analysis of GCN4–ATF1 hybrid proteins indicates that a short surface spanning the basic and fork regions of the bZIP domain is the major determinant of half-site spacing. Replacement of two GCN4 residues on this surface (Ala244 and Leu247) by their ATF1 counterparts largely converts GCN4 into a protein with ATF/CREB specificity. Secondly, analysis of a Fos derivative containing the GCN4 leucine zipper indicates that Fos represents a novel intermediate between AP-1 and ATF/CREB proteins. Thirdly, we examine the effects of mutations in the invariant arginine residue of GCN4 (Arg243) that contacts the central base pair(s) of the target sites. While most mutations abolish DNA binding, substitution of a histidine residue results in a GCN4 derivative with ATF/CREB binding specificity. These results suggest that the AP-1 and ATF/CREB proteins differ in positioning a short surface that includes the invariant arginine and that AP-1 proteins may represent a subclass (and perhaps evolutionary offshoot) of ATF/CREB proteins that can tolerate overlapping half-sites.

INTRODUCTION

bZIP DNA-binding domains consist of two structurally and functionally distinct subdomains, the leucine zipper and the basic region, that are connected by a short fork (1). The C-terminal leucine zipper forms a coiled coil that mediates dimerization (2). This dimer interface and the connecting fork symmetrically position a divergent pair of basic region α-helices, which pass through the major groove of each DNA half-site (3–5). Upon binding to DNA, the previously unfolded basic region becomes α-helical (6–8), such that five conserved amino acid residues are positioned to contact specific base pairs in the target sites (9,10).

Two important families of eukaryotic transcriptional regulatory proteins, AP-1 and ATF/CREB, contain structurally related DNA-binding domains having the bZIP structural motif (Fig. 1). These families were initially defined by their preferred recognition sequences, termed AP-1 and ATF/CREB sites. The optimal AP-1 recognition sequence, TGA(C/G)TCA, consists of overlapping half-sites, TGAC and TGAG, that are non-equivalent due to the asymmetry imposed by the central C:G base pair (11–13). Biochemical and crystallographic analysis of a complex containing yeast GCN4 protein bound to the AP-1 site indicates that the optimal half-site is TGAC (9,13). High affinity ATF/CREB sites, TGACGTCAT, contain abutting TGAC half-sites that do not overlap at the central base pair. These observations suggest that AP-1 and ATF/CREB proteins recognize identical half-sites, but differ in their preferences for half-site spacing (13).

By definition, AP-1 proteins prefer to bind AP-1 sites, but they also bind ATF/CREB sites with only slightly lower affinity (13,14). Crystal structures of GCN4 bound to AP-1 (9) and ATF/CREB (10) sites reveal that the protein–DNA contacts are virtually identical except for the interaction of the invariant arginine in the basic region of bZIP proteins (Arg243 in GCN4) with the central base pair(s). When bound to the inherently asymmetric AP-1 site, Arg243 from one monomer (the left as conventionally drawn) forms two hydrogen bonds to the guanine at the central base pair (position 0); Arg243 of the right monomer contacts phosphates on the central base pair and the adjacent residue on the opposite strand (9). In contrast, GCN4 binds symmetrically to the ATF/CREB site with both Arg243 residues interacting with the equivalent guanine residues at the two central base pairs (10).

The ability of GCN4 to bind AP-1 and ATF/CREB sites, along with other information (15), indicates a surprising degree of flexibility at the protein–DNA interface. In particular, GCN4 must accommodate the 36° rotation and 3.4 displacement of the half-sites from one another that arises from the additional base pair in the ATF/CREB site. A comparison of the crystal structures

* To whom correspondence should be addressed

+Present address: Department of Biology, College of Science, Korea University, Seoul 136-701, Korea

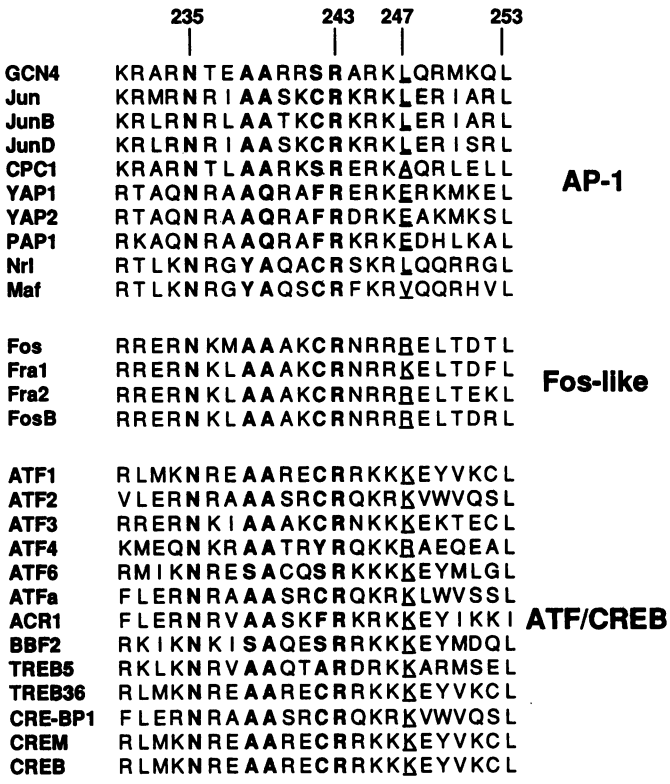

Figure 1. Basic regions of AP-1 and ATF/CREB proteins. Sequences for the indicated AP-1, ATF/CREB and Fos-like proteins as for the region corresponding to GCN4 residues 231–253. Proteins are defined as belonging to the AP-1 or ATF/CREB families according to their target sequences as reported in the literature; with rare exceptions, half-site spacing specificity has not been explicitly tested. Residues in bold directly contact base pairs in the crystal structures of GCN4 bound to the AP-1 and ATF/CREB sites, including Asn235 and Arg243, which are invariant in bZIP domains. The underlined residue at position 247 is positively charged in ATF/CREB proteins, but not in AP-1 proteins; Fos (and its relatives) also have a basic residue at this position, but appear to have intermediate half-site spacing specificity (see text). The leucine residue (except in the case of ACR1) at position 253 is the first leucine of the zipper dimerization region. References for the bZIP domains are as follows: GCN4 (28); Jun (29); JunB and Jun D (30,31); Cpc1 (32); Yap1 (33); Yap2 (34); Pap1 (35); Maf (36); Nrl (37); Fos (38); Fra1 (39); Fra2 (40); FosB (41); ATF1–4, 6 (17); ATF-a (42); ACR1 (18); BBF-2 (43); TREB5, 36 (44); CRE-BP1 (45); CREB (46); CREM (47).

strongly suggests that this flexibility is due to both DNA bending and protein conformation (10). The DNA in the GCN4–AP-1 site complex is straight (9), whereas it is bent by 20° towards the leucine zipper in the GCN4–ATF/CREB site complex (10). The distinct DNA conformations in the crystal structures are likely to reflect the differential intrinsic curvatures of the AP-1 and ATF/CREB sites in solution (16) and may not be significantly induced by protein binding. In addition to this effect on DNA structure, the basic regions in GCN4 diverge by an additional 5° in the ATF/CREB complex. Thus GCN4 (and presumably other AP-1 proteins) can adapt to the distinct curvatures of the AP-1 and ATF/CREB sites.

Although structures of ATF/CREB proteins bound to their cognate sites have yet to be described, it is reasonable to suppose that the protein–DNA contacts will be identical to those in the GCN4–ATF/CREB site complex. The basic regions of AP-1 and ATF/CREB proteins are very similar, including the five residues directly contacting base pairs in the GCN4 crystal structures (9,10). Despite these similarities, however, ATF/CREB proteins bind very poorly to AP-1 sites (14,17,18). Thus the different half-site spacing preferences of AP-1 and ATF/CREB proteins are likely to arise from the inability of ATF/CREB proteins to adapt to the overlapping half-sites in AP-1 target sequences.

The structural differences between AP-1 and ATF/CREB proteins that account for their distinct half-site spacing preferences are not understood. One consistent difference between these two classes of proteins occurs at a residue (247 in GCN4) in the fork between the leucine zipper and basic region (18). ATF/CREB proteins have a positively charged residue (nearly always lysine), whereas AP-1 proteins (with the apparent exception of Fos and its relatives; see below) typically do not (Fig. 1). Previously, we showed that Lys247 and Arg247 derivatives of GCN4 have DNA-binding properties that are intermediate between AP-1 and ATF/CREB proteins; in comparison with GCN4, these proteins bind with reduced affinity to the AP-1 site but with comparable affinity to the ATF/CREB site (15). Thus while the residue 247 can affect half-site spacing preferences, it does not fully account for the differences between AP-1 and ATF/CREB proteins. Moreover, the properties of the Lys247 and Arg247 derivatives of GCN4 suggest that there may be eukaryotic transcription factors that bind with a specificity intermediate between conventionally defined AP-1 and ATF/CREB proteins.

In this paper we investigate the distinctions between AP-1 and ATF/CREB proteins by determining the DNA-binding properties of mutant and hybrid proteins. Our results suggest that half-site spacing preferences reflect differences in positioning of the invariant arginine by a short region that spans the basic region and the fork. Further, they suggest that AP-1 proteins may represent a subclass of ATF/CREB proteins that can tolerate overlapping half-sites.

MATERIALS AND METHODS

DNA manipulations

DNA molecules expressing full-length GCN4 and its derivative originate from YCp88-Sc4400, a centromeric vector with the *ura3* selectable marker that utilizes the SP6 promoter for protein expression *in vitro* (19). DNA molecules for expressing the isolated bZIP DNA-binding domains, GCNK58 and ATF1K60, derive from pGCNK58 (9) and pGEM3-ATF1 (17), which utilize the T7 promoter for protein expression *in vitro*. Molecules encoding GCN4–ATF1 hybrid proteins were made by using oligonucleotides encoding the junction sequences as primers for PCR on the parental molecules containing the isolated bZIP domains. Molecules in which patches of GCN4 sequences were substituted by their ATF1 counterparts were generated by using oligonucleotides converted to a clonable form by PCR or by mutually primed synthesis (20). The resulting fragments were cloned between the *Pst*I and *Alw*NI sites or *Alw*NI and *Xho*I sites (19) to generate full-length GCN4 derivatives. The DNA molecules that express a truncated version of Fos (150 residues) containing the entire bZIP domain and Fos^{G} (246 amino acids) have been described previously (21).

Residue 243 of GCN4 was randomized using sequential PCR. Complementary oligonucleotides contained a randomized codon 243 with 10 flanked bases on each side of the randomized area. PCR was first performed on YCp88-Sc4400 in two separate

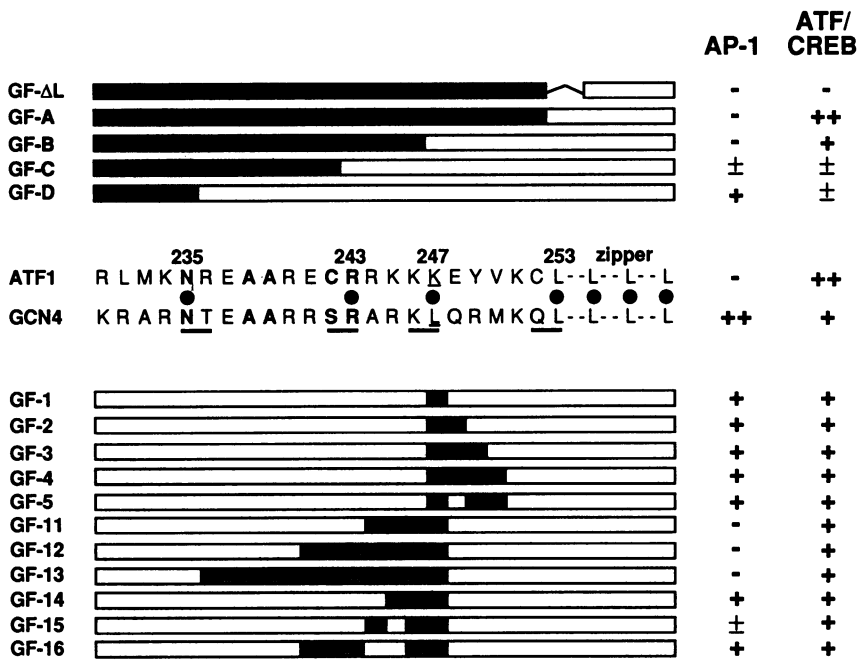

Figure 2. Structures of GCN4–ATF1 hybrid proteins. The sequences for GCN4 and ATF1 are shown in Figure 1 with the addition of conserved leucines in the zipper dimerization region. Shown above these sequences are ATF1–GCN4 chimeric proteins; shown below are GCN4 derivatives in which patches of ATF1 are substituted. For each derivative white bars represent GCN4 residues, black bars represent ATF1 residues and shaded bars represent conserved residues between GCN4 and ATF1 (bottom part only). The relative DNA binding activity of these proteins on AP-1 and ATF/CREB sites is indicated as follows: ++, binds with affinity comparable to that of GCN4 and/or ATF1; +, binds with reduced affinity in comparison with GCN4 and/or ATF1; ±, binds very weakly; –, binding not detected.

reactions using appropriate pairs of the randomized oligonucleotide and their respective outside primers. After amplification the resulting fragments, which are complementary at their 3′-ends except for the randomized area, were mixed and a second PCR was performed. The resulting mixture of fragments was cleaved with *Asp*718 and *Eco*RI and cloned into YCp88-Sc4400.

DNA binding experiments

GCN4 proteins were synthesized from the plasmid DNAs by transcription and translation *in vitro* using SP6 or T7 RNA polymerase and rabbit reticulocyte lysates. Equivalent amounts of ^{35}S-labeled proteins, determined by SDS–PAGE, were incubated with the appropriate target sequences (2 nM) and the resulting complexes were electrophoretically separated in native acrylamide gels. The binding sites were obtained either as 98 bp *Nde*I fragments or 65 bp fragments generated by PCR (22). The conditions for the DNA binding assay are such that the intensities of bands representing the protein–DNA complexes are roughly proportional to the binding constants (11,19,22,23).

RESULTS

Analysis of GCN4–ATF1 hybrid proteins

To initially localize the region responsible for the half-site spacing distinction between AP-1 and ATF/CREB proteins, we generated a set of hybrid bZIP domains. N-Terminal portions of the human ATF1 were fused to C-terminal portions of yeast GCN4 at residues conserved between both proteins (Fig. 2). The resulting proteins were synthesized *in vitro* and tested for their ability to bind AP-1 and ATF/CREB sites (Fig. 3).

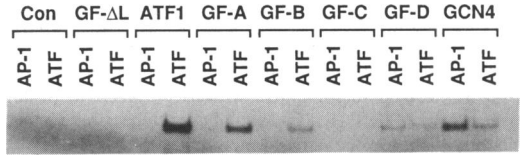

Figure 3. DNA binding of GCN4–ATF1 chimeras. Protein–DNA complexes formed by incubating equivalent amounts of *in vitro* synthesized ^{35}S-labeled proteins (see Fig. 2) with DNA fragments containing ATF/CREB or AP-1 target sequences. All proteins except GF-ΔL were generated as isolated bZIP domains.

Although the hybrid bZIP domains (and particularly GF-C) bind less avidly to the target sequences than either the ATF1 or GCN4 bZIP domains, they display half-site specificity characteristic of AP-1 or ATF/CREB proteins. GF-A and GF-B bind very poorly to the AP-1 site, indicating that the ATF1 bZIP domain retains its half-site spacing preferences even when the leucine zipper and fork region are replaced by the corresponding segment of GCN4. Conversely, GF-D, which contains additional GCN4 residues from the central portion of the basic region, shows AP-1 specificity. As expected, a hybrid protein containing only three heptad repeats of the GCN4 leucine zipper (GF-ΔL) fails to bind DNA, presumably because it cannot dimerize (5,24). These results implicate residues in the basic region as playing an important role in half-site spacing preferences. However, our previous observation that the Arg247 and Lys247 derivatives of GCN4 show reduced binding to the AP-1 site but normal binding to the ATF/CREB site suggests that residues in the fork region are also important (15).

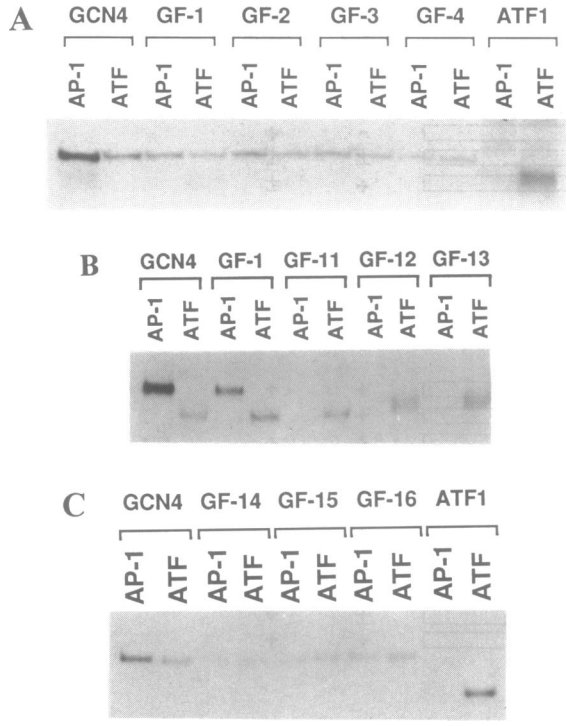

Figure 4. DNA binding of GCN4 derivatives containing short patches of ATF1. Protein–DNA complexes formed by incubating equivalent amounts of *in vitro* synthesized ^{35}S-labeled proteins (see Fig. 2) with DNA fragments containing ATF/CREB or AP-1 sites. (**A**) Mobility differences reflect the fact that GCN4 and the various GF derivatives were synthesized as full-length proteins, whereas ATF1 was synthesized as a 60 residue bZIP domain. Binding sites were 64 and 65 bp DNA fragments generated by PCR. (**B**) Proteins were synthesized as full-length derivatives. Mobility differences of the complexes arise because the AP-1 site was a 98 bp *NdeI* fragment, whereas the ATF/CREB site was a 65 bp fragment generated by PCR. (**C**) Mobility differences reflect the fact that GCN4 and the various GF derivatives were synthesized as full-length proteins, whereas ATF1 was synthesized as a 60 residue bZIP domain. Binding sites were 64 and 65 bp DNA fragments generated by PCR.

To map the distinctions between AP-1 and ATF/CREB proteins more precisely, we examined a set of hybrid bZIP domains in which short patches of GCN4 were replaced by the corresponding regions of ATF1 (Figs 2 and 4). Residues 248–250 do not appear to play a significant role, because hybrid proteins GF-2–GF-5 behave similarly to GF-1. However, GCN4 can be converted to a protein with ATF/CREB specificity by replacing the region spanning residues 244–247 with the corresponding ATF1 residues (GF-11 and the more extensively substituted derivatives GF-12 and GF-13). Further subdivision of this short region indicates that residues 244 and 247 are particularly important; GF-15, a GCN4 derivative containing only these two substitutions from ATF1, shows a clear preference for ATF/CREB sites over the AP-1 site. In contrast, GF-14 and GF-16 bind with comparable affinity to the two target sequences, indicating that residues 241, 242 and 245 are not significantly involved in the discrimination of AP-1 and ATF/CREB sites. Residues 243 and 246 are conserved between GCN4 and ATF1 and hence cannot be examined by the hybrid protein approach. Interestingly, the crucial residues identified by this approach, which correspond to 244 and 247 of GCN4, do not contact DNA in the GCN4 complexes with the AP-1 and ATF/CREB sites (9,10).

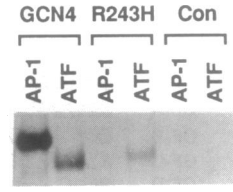

Figure 5. DNA binding of the His243 derivative of GCN4. Protein–DNA complexes formed by incubating equivalent amounts of *in vitro* synthesized ^{35}S-labeled proteins (full-length) with the AP-1 site. The AP-1 site was a 98 bp *NdeI* fragment containing the AP-1 site or a 65 bp PCR-generated fragment containing the ATF/CREB site.

Mutational analysis of the invariant arginine (Arg243) in the GCN4 basic region

We carried out this analysis for two reasons. First, Arg243 is the only residue that contacts the central base pair(s) that distinguishes AP-1 and ATF/CREB sites and that interacts differentially in the GCN4–AP-1 and GCN4–ATF/CREB site complexes (9,10). Although this residue is invariant in bZIP proteins and hence cannot be the critical determinant that distinguishes AP-1 and ATF/CREB sites, it is likely to play an important role in half-site spacing. Secondly, bZIP proteins that bind exclusively to AP-1 sites have not been described and such derivatives would be useful for analyzing the physiological roles of AP-1 and ATF/CREB binding activities. We considered the possibility that mutations at Arg243 would be more deleterious for ATF/CREB binding activity than for AP-1 binding activity, because they would eliminate contacts to two bases of the ATF/CREB site but only to one base of the AP-1 site.

Upon randomizing Arg243 of GCN4 we obtained 12 mutant proteins which contained substitutions of Met, Lys, Phe, Ser, Asn, Gln, Leu, His, Asp, Cys, Val and Trp. As expected from the invariance of this arginine in bZIP proteins and the importance of the central base pair(s) in efficient binding (11), nearly all of these mutant proteins failed to bind either the AP-1 or ATF/CREB sites. Surprisingly, however, the His243 protein bound to the ATF/CREB site but not the AP-1 site (Fig. 5). The complex of the His243 protein with the ATF/CREB site appears to migrate more slowly than the corresponding complex with GCN4, suggesting an altered conformation of the protein–DNA complex. Thus, in contrast to our initial hypothesis, a mutation in Arg243 can convert GCN4 into a protein with ATF/CREB binding specificity.

Half-site spacing preferences of the Fos oncoprotein

Fos was initially described, and is typically considered, as belonging to the AP-1 family, because Fos–Jun heterodimers display AP-1 binding specificity (25). However, Fos was also identified in a search for ATF/CREB proteins (17) and it can form heterodimers with other ATF/CREB proteins that display ATF/CREB specificity (26). Moreover, at the residue corresponding to position 247 in GCN4, Fos contains a lysine, which is typical of ATF/CREB proteins (18). Thus it is unclear whether Fos should be characterized as an AP-1 or ATF/CREB protein. This question is complicated by the fact that Fos cannot bind to DNA as a homodimer; hence, its apparent DNA binding properties might reflect the contributions of its heterodimeric partner.

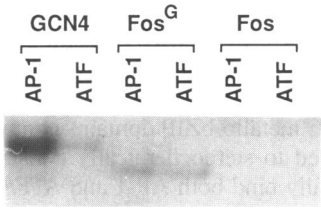

Figure 6. Half-site spacing specificity of Fos. Protein–DNA complexes formed by incubating equivalent amounts of *in vitro* synthesized ^{35}S-labeled full-length GCN4, Fos^{G} (246 residues) and a truncated version of Fos (150 residues) containing the intact bZIP domain with 65 bp DNA fragments containing the AP-1 or ATFCREB binding sites generated by PCR.

To circumvent this problem, we analyzed the DNA binding specificity of Fos^{G}, a hybrid containing the basic and fork regions of Fos and the GCN4 leucine zipper dimerization element (21). As the leucine zipper does not affect DNA binding specificity (Fig. 3), the specificity of Fos^{G} should define the inherent sequence recognition properties of Fos. As shown in Figure 6, Fos^{G} binds to the AP-1 and ATF/CREB sites with equal affinity, indicating that Fos has DNA binding characteristics that are intermediate between those of canonical AP-1 and ATF/CREB proteins.

DISCUSSION

Determinants of half-site preferences in AP-1 and ATF/CREB proteins

Our results indicate that the distinction between AP-1 and ATF/CREB proteins is due primarily to a short region corresponding to GCN4 residues 243–247. GCN4 can be converted to a protein with ATF/CREB specificity either by changing residues 244 and 247 to the corresponding amino acids from ATF1 or by changing Arg243 to His. Conversely, the presence of ATF1 residues at GCN4 positions C-terminal to 248 (GF-4) or at positions 241 and 242 (GF-16) does not significantly affect half-site spacing specificity. In addition, the similar properties of GF-1 and GF-14 suggest that residue 245 does not play a significant role.

Although residues 243, 244 and 247 clearly play the major role, we cannot exclude the possibility that other residues might influence half-site preferences. First, most of the hybrid and mutant proteins bind with lower affinity to the AP-1 or ATF/CREB sites than the native GCN4 and ATF1 bZIP domains. While this reduced affinity probably reflects a non-specific perturbation of the bZIP domains due to the chimeric junction, the possibility of altered DNA binding specificity cannot be excluded. Secondly, the GCN4 derivatives with ATF/CREB specificity (GF-11 and GF-15) show a residual level of AP-1 binding activity; thus additional residues are likely to be required to fully convert GCN4 into an ATF/CREB protein. Thirdly, with the exception of Arg243, which was specifically analyzed, our approach cannot assess the importance of residues conserved between GCN4 and ATF1. In this regard, Arg240 is of interest, because it lies on the same α-helical surface as residues 243, 244 and 247.

Structural interpretations of the mutant proteins

Comparison of the GCN4 co-crystal structures indicates that binding to the ATF/CREB site is accompanied by increased divergence of the basic region and increased curvature of the target DNA relative to that observed when GCN4 is bound to the AP-1 site (9,10). As a consequence, the GCN4 basic region is, in relative terms, closer to DNA in the ATF/CREB site than in the AP-1 site. This difference in positioning is likely to underlie the distinct half-site preferences of AP-1 and ATF/CREB proteins and it provides a plausible basis to account for the roles of residues 243, 244 and 247.

Arg243 directly contacts guanine residues at the central base pair(s) and adjacent phosphate(s) of the AP-1 and ATF/CREB sites and these contacts are the only interactions that differ in the respective GCN4 complexes (9,10). The ATF/CREB specificity of the His243 derivative of GCN4 is likely due to a direct interaction between the histidine residue and position ±0 of the target site. Consistent with this interpretation, all other substitutions tested at this position abolish binding to either site. Molecular modeling using the GCN4 co-crystal structures suggests that His243, which is considerably smaller than Arg243, is too far away to make significant contacts with the AP-1 site. In contrast, His243 is probably close enough to contact the central guanines and adjacent phosphates in the ATF/CREB site.

While the crystal structure of the ATF1–DNA complex has yet to be solved, modeling the corresponding ATF1 residues into the GCN4 co-crystal structures provides some clues about the basis of half-site spacing specificity. Leu247 of GCN4 does not contact base pairs or phosphates of the AP-1 site (9) and molecular modeling suggests that lysine or arginine substitutions are unlikely to have significant interactions. However, at the ATF/CREB site lysine and arginine residues at position 247 are likely to contact phosphates at ±1A and (to a lesser extent) ±0C on the same strand but opposite half-site of the ±0G residue that interacts with Arg243 of the same monomer. Conversely, large hydrophobic residues at position 247 (such as leucine in GCN4 and some other AP-1 proteins) might interfere with the ionic interaction of Arg243 with the ±1A phosphate in the ATF/CREB site but not the AP-1 site. These considerations suggest that positively charged residues at position 247 confer a relative increase in the interaction with ATF/CREB sites, whereas large hydrophobic residues might cause a relative decrease.

Residue 244 is adjacent to the invariant asparagine that contacts the central base pair(s), and residues 247, 244 and 243 lie on the same surface of the α-helix; Arg240, which is present in both GCN4 and ATF1, also lies on this surface. Ala244 of GCN4 does not contact DNA in either the AP-1 or ATF/CREB complexes (9,10), but an arginine residue (such as occurs in ATF1 and GF-15) would probably contact the same ±1 phosphate that can interact with positively charged residues at position 247. Thus ATF1 and the GCN4 hybrid proteins with ATF/CREB specificity (GF-11 and GF-15) contain four positively charged residues (Arg240, Arg243, Arg244 and Lys247) on this α-helical surface, whereas GCN4 contains only two. At the ATF/CREB site these four positively charged residues are well positioned to make phosphate contacts to three adjacent nucleotides on the same strand. However, at the AP-1 site only two phosphate residues are available for contacts (due to the overlapping half-sites) and, more importantly, there is insufficient space to accommodate all four positively charged residues without electrostatic repulsion.

Thus a plausible hypothesis for half-site spacing preferences is that a positively charged surface at positions 240, 243, 244 and 247 cannot tolerate the constraints imposed by the overlapping half-sites of the AP-1 sequence

Fos is an intermediate between the AP-1 and ATF/CREB families

Unlike conventional AP-1 and ATF/CREB proteins that respectively show moderate or strong preferences for their cognate sites, Fos^G, a homodimeric form of Fos (21), binds these sequences with equal affinity. This intermediate specificity of Fos^G almost certainly represents the half-site preferences of Fos; it is unlikely to be an artifact of the hybrid protein because comparable fusions at the first leucine of the zipper dimerization region do not affect DNA-binding specificity (GF-A, Fig. 3; 3). These intermediate half-site preferences nicely account for the ability of Fos to display conventional AP-1 or ATF/CREB specificity upon heterodimeric association with AP-1 or ATF/CREB family members (26). Moreover, they suggest that Fos (and perhaps the related Fra-1, Fra-2 and FosB proteins, which also contain a positively charged residue at position 247; Fig. 1) represents a functional and perhaps evolutionary link between the conventionally defined AP-1 and ATF/CREB families.

Based on the structural considerations discussed above, we have suggested that a positive charge at residue 247 and along the α-helical surface defined by residues 240, 243, 244 and 247 is important for discrimination between AP-1 and ATF/CREB proteins. The intermediate half-site preferences of Fos are in good accord with this hypothesis; although Fos and its family members contain arginine (an ATF/CREB-type residue) at position 247, residues 240 (alanine) and 244 (asparagine) are neutral. In this regard the sequence and binding properties of Fos^G are similar to those of the Lys247 and Arg247 derivatives of GCN4 (15). More generally, our hypothesis about positive charge is fairly good at distinguishing AP-1 and ATF/CREB proteins from their primary sequences, but there are some ambiguities and exceptions. While these apparent discrepancies are very likely due to the inadequacies of an oversimplified hypothesis, they may also reflect the fact that half-site spacing preferences for most AP-1 and ATF/CREB proteins have not been determined explicitly.

AP-1 proteins may be a subclass and evolutionary offshoot of ATF/CREB proteins

Although the AP-1 and ATF/CREB families of bZIP domains have distinct properties, many observations indicate that they belong to the same superfamily. AP-1 and ATF/CREB proteins are structurally and immunologically related, they can interact across family lines to form DNA-binding heterodimers, they recognize identical half-sites and there are natural members of the superfamily (Fos and perhaps the related Fra-1, Fra-2 and FosB) that have intermediate half-site-spacing preferences. Thus it is very likely that AP-1 and ATF/CREB proteins arose from a common ancestor.

ATF/CREB proteins appear to represent the functional ground state of the superfamily. First, all proteins of the superfamily recognize ATF/CREB sites, whereas only some of them bind AP-1 sites. Secondly, the half-site spacing specificity mutant proteins described here and elsewhere (15) retain normal activity on the ATF/CREB sequence, while losing the ability to bind the AP-1 site. Thirdly, we were unable to generate a protein that exclusively recognizes AP-1 sites by mutating Arg243 of GCN4, which recognizes the central base pair(s) of AP-1 and ATF/CREB sites; in fact, the His243 derivative has ATF/CREB specificity. Fourthly, although metallo-bZIP domains containing the GCN4 basic region joined to stereochemically constrained metal ion complexes generally bind both AP-1 and ATF/CREB sites, one derivative strongly discriminates against the AP-1 site (27). The selective inactivation of AP-1 binding activity in these natural, mutated and artificial examples suggest that AP-1 proteins might represent a subclass of ATF/CREB proteins that tolerates overlapping half-sites.

There are two contrasting views of the evolutionary relationship between AP-1 and ATF/CREB proteins. As AP-1 proteins have the additional function of tolerating overlapping half-sites, it can be imagined that ATF/CREB proteins represent the original ancestor of the superfamily, from which AP-1 proteins subsequently evolved. Alternatively, it could be argued that AP-1 proteins are the original ancestor, because they are relatively more promiscuous in their DNA binding specificity and can be easily converted into proteins with ATF/CREB specificity. It is difficult to distinguish between these two views solely by functional criteria; a systematic evolutationary analysis of the superfamily might be informative.

ACKNOWLEDGEMENTS

We are indebted to Tom Ellenberger for modeling the mutant proteins in terms of the X-ray crystal structures and for his crucial insights into the differences between the GCN4–AP-1 and GCN4–ATF/CREB complexes. We thank Steve Beverly and an unknown reviewer for discussion of the evolutionary relationships between the AP-1 and ATF/CREB proteins, Susanna Wurgler and Peter Vaillincourt for helpful discussions and Tom Ellenberger for comments on the manuscript. This work was supported by a postdoctoral fellowship from the Damon Runyon–Walter Winchell Cancer Research Foundation (J.K.), a research grant from the Ministry of Education, Korea (J.K.) and research grants from the National Institutes of Health (GM 30186 and GM 46555) to K.S.

REFERENCES

1 Landschulz,W.H., Johnson,P. F. and McKnight,S.L. (1988) *Science*, **240**, 1759–1764.
2 O'Shea,E.K., Rutkowski,R. and Kim,P.S. (1989) *Science*, **243**, 538–542.
3 Agre,P., Johnson,P.F. and McKnight,S.L. (1989) *Science*, **246**, 922–926.
4 Talanian,R.V., McKnight,C.J. and Kim,P.S. (1990) *Science*, **249**, 769–771.
5 Pu,W.T. and Struhl,K. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 6901–6905.
6 O'Neil,K.T., Hoess,R.H. and DeGrado,W.F. (1990) *Science*, **249**, 774–778.
7 Weiss,M.A., Ellenberger,T., Wobbe,C.R., Lee,J.P., Harrison,S.C. and Struhl,K. (1990) *Nature*, **347**, 575–578.
8 Patel,L., Abate,C. and Curran,T. (1990) *Nature*, **347**, 572–575.
9 Ellenberger,T.E., Brandl,C.J., Struhl,K. and Harrison,S.C. (1992) *Cell*, **71**, 1223–1237.
10 Konig,P. and Richmond,T.J. (1993) *J. Mol. Biol.*, **233**, 139–154.
11 Hill,D.E., Hope,I.A., Macke,J.P. and Struhl,K. (1986) *Science*, **234**, 451–457.
12 Oliphant,A.R., Brandl,C.J. and Struhl,K. (1989) *Mol. Cell. Biol.*, **9**, 2944–2949.
13 Sellers,J.W., Vincent,A.C. and Struhl,K. (1990) *Mol. Cell. Biol.*, **10**, 5077–5086.
14 Hai,T., Liu,F., Allegretto,E.A., Karin,M. and Green,M.R. (1988) *Genes Dev.*, **2**, 1216–1226.

15 Kim,J., Tzamarias,D., Ellenberger,T.E., Harrison,S.C. and Struhl,K. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 4513–4517.
16 Paolella,D.N., Palmer,C.R. and Schepartz,A. (1994) *Science*, **264**, 1130–1133.
17 Hai,T., Liu,F., Coukos,W.J. and Green,M.R. (1989) *Genes Dev.*, **3**, 2083–2090.
18 Vincent,A.C. and Struhl,K. (1992) *Mol. Cell. Biol.*, **12**, 5394–5405.
19 Pu,W.T. and Struhl,K. (1991) *Mol. Cell. Biol.*, **11**, 4918–4926.
20 Oliphant,A.R., Nussbaum,A.L. and Struhl,K. (1986) *Gene*, **44**, 177–183.
21 Sellers,J.W. and Struhl,K. (1989) *Nature*, **341**, 74–76.
22 Tzamarias,D., Pu,W.T. and Struhl,K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 2007–2011.
23 Hope,I.A. and Struhl,K. (1985) *Cell*, **43**, 177–188.
24 van Heeckeren,W.J., Sellers,J.W. and Struhl,K. (1992) *Nucleic Acids Res.*, **20**, 3721–3724.
25 Curran,T. and Franza,B.J. (1988) *Cell*, **55**, 395–397.
26 Hai,T. and Curran,T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3720–3724.
27 Cuenoud,B. and Schepartz,A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 1154–1159.
28 Hope,I.A. and Struhl,K. (1986) *Cell*, **46**, 885–894.
29 Bohmann,D., Bos,T.J., Admon,A., Nishimura,T., Vogt,P.K. and Tjian,R. (1987) *Science*, **238**, 1386–1392.
30 Ryder,K., Lanahan,A., Perez-Albuerne,E. and Nathans,D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1500–1503.
31 Hirai,S., Ryseck,R., Mechta,F., Bravo,R. and Yaniv,M. (1989) *EMBO J.*, **8**, 1433–1439.
32 Paluh,J.L. and Yanofsky,C. (1991) *Mol. Cell. Biol.*, **11**, 935–944.
33 Moye-Rowley,W.S., Harshman,K.D. and Parker,C.S. (1989) *Genes Dev.*, **3**, 283–292.
34 Bossier,P., Fernandes,L., Rocha,D. and Rodrigues-Pousada,C. (1993) *J. Biol. Chem.*, **268**, 23640–23645.
35 Toda,T., Shimanuki,M. and Yanagida,M. (1991) *Genes* Dev., **5**, 60–73.
36 Kataoka,K., Noda,M. and Nishizawa,M. (1994) *Mol. Cell. Biol.*, **14**, 700–712.
37 Kerppola,T.K. and Curran,T. (1994) *Oncogene*, **9**, 675–684.
38 Distel,R., Ro,H.S., Rosen,B.S., Groves,D. and Spiegelman,B.M. (1987) *Cell*, **49**, 835–844.
39 Cohen,D., Ferreira,P., Gentz,R., Franza,B.J. and Curran,T. (1989) *Genes Dev.*, **3**, 173–184.
40 Nishina,H., Sato,H., Suzuki,T., Sato,M. and Iba,H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3619–3623.
41 Zerial,M., Toschi,L., Ryseck,R., Schuermann,M., Muller,R. and Bravo,R. (1989) *EMBO J.*, **8**, 805–813.
42 Gaire,M., Chatton,B. and Kedinger,C. (1990) *Nucleic Acids Res.*, **18**, 3467–3473.
43 Abel,T., Bhatt,R. and Maniatis,T. (1992) *Genes Dev.*, **6**, 466–480.
44 Yoshimura,T., Fujisawa,J.-I. and Yoshida,M. (1990) *EMBO J.*, **9**, 2537–2542.
45 Maekawa,T., Sakura,H., Kanei-Ishii,C., Sudo,T., Yoshimura,T., Fujisawa,J.-I., Yoshida,M. and Ishii,S. (1989) *EMBO J.*, **8**, 2023–2028.
46 Hoeffler,J.P., Meyer,T.E., Yun,Y., Jameson,J.L. and Haebner,J.F. (1988) *Science*, **242**, 1430–1433.
47 Foulkes,N.S., Borrelli,E. and Sassone-Corsi,P. (1991) *Cell*, **64**, 739–749.