DNA-Binding Specificity of the PAR Basic Leucine Zipper Protein VBP Partially Overlaps Those of the C/EBP and CREB/ATF Families and Is Influenced by Domains That Flank the Core Basic Region

NAOMI B. HAAS,¹ CARRIE A. CANTWELL,² PETER F. JOHNSON,² AND JOHN B. E. BURCH^{1*}

Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111,¹ and ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702²

Received 11 October 1994/Returned for modification 15 November 1994/Accepted 19 January 1995

The PAR subfamily of basic leucine zipper (bZIP) factors comprises three proteins (VBP/TEF, DBP, and HLF) that have conserved basic regions flanked by proline- and acidic-amino-acid-rich (PAR) domains and functionally compatible leucine zipper dimerization domains. We show that VBP preferentially binds to sequences that consist of abutted GTAAY half-sites (which we refer to as PAR sites) as well as to sequences that contain either a C/EBP half-site (GCAAT) or a CREB/ATF half-site (GTCAT) in place of one of the PAR half-sites. Since the sequences that we describe as PAR sites and PAR-CREB/ATF chimeric sites, respectively, were both previously described as high-affinity binding sites for the E4BP4 transcriptional repressor, we infer that these sequences may be targets for positive and negative regulation. Similarly, since the sequences that we describe as PAR-C/EBP and PAR-CREB/ATF chimeric sites are known to be high-affinity binding sites for C/EBP and CREB/ATF factors, respectively, we infer that these sites may each be targets for multiple subfamilies of bZIP factors. To gain insights regarding the molecular basis for the binding-site specificity of PAR factors, we also carried out an extensive mutational analysis of VBP. By substituting five amino acid residues that differ between the *Drosophila* giant bZIP factor and the vertebrate PAR bZIP factors, we show that the fork region, which bridges the basic and leucine zipper domains, contributes to half-site sequence specificity. In addition, we report that at least two domains amino terminal to the core basic region are required for VBP to bind to the full spectrum of PAR target sites. Thus, whereas direct base contacts may be restricted to basic-region residues (as indicated by GCN4-DNA crystal structures), several other domains also influence the DNA-binding specificity of PAR bZIP proteins.

Basic leucine zipper (bZIP) proteins are a class of eukaryotic factors that regulate transcription by binding as dimers to specific DNA sequences. These proteins perform a variety of regulatory functions and occur ubiquitously, having been found in organisms as diverse as fungi, plants, and mammals. Of the nearly 70 bZIP proteins that have been identified to date (10), most can be placed into one of five major subfamilies on the basis of their DNA recognition properties and amino acid sequence similarities (13). These bZIP subgroups include the AP-1, CREB/ATF, C/EBP, PAR, and plant G-box proteins. The proteins in each subfamily recognize highly similar or identical DNA sites whose consensus sequences are 9- or 10-bp palindromes composed of two 5-bp half-sites. Binding sites for the various classes of bZIP proteins may differ by either their half-site sequences or their half-site spacing properties. AP-1 proteins such as Fos, Jun, and GCN4 bind to a 9-bp pseudopalindromic sequence that can be viewed as two half-sites that overlap by a single base pair (22, 30), whereas the consensus binding sites for the other four subfamilies have directly abutted pairs of half-sites.

All bZIP domains consist of basic and leucine zipper regions that are separated by a short segment known as the fork. The basic region contains a set of highly conserved residues (the basic motif) that occurs in constant register with respect to the leucine repeats of the zipper (16, 34). Domain-switching experiments demonstrate that sequences amino terminal to the

leucine zipper, principally the basic region, dictate DNA-binding specificity (1, 17, 21, 26, 29). X-ray crystallographic structures of GCN4-DNA complexes (7, 15) reveal that the bZIP domain forms a continuous alpha helix when bound to DNA, as predicted by earlier theoretical models (23, 34). Each subunit of a dimer bound to DNA interacts with one of the half-sites in the recognition sequence. The basic region lies in the major groove and makes DNA contacts through amino acid side chains emanating from one side of the alpha helix. Five residues of the GCN4 basic region make base-specific contacts to DNA, while several basic-amino-acid-phosphate backbone interactions further stabilize the complex.

Paradoxically, the five base-contacting amino acids in GCN4 that presumably determine DNA-binding specificity correspond to residues that are highly conserved among bZIP proteins. Consequently, it has not been possible to deduce the basis for sequence discrimination by the various classes of bZIP factors from simple amino acid sequence comparisons and structural analogies to GCN4. Some information has been gained by constructing amino acid substitution mutations at specific basic-region residues and determining their effects upon binding specificity (13, 14, 32). Such studies have revealed that one of the highly conserved basic motif residues that makes a base-specific contact in the GCN4 crystal structure is critical for C/EBP specificity (13, 32). This position is occupied by alanine in most bZIP proteins but contains valine in the C/EBP proteins and, therefore, should be characterized as a variable residue. In addition, mutational studies have implicated the fork region in conferring the half-site spacing

* Corresponding author. Phone: (215) 728-3696. Fax: (215) 728-3574. Electronic mail address: jb_burch@fccc.edu.

FIG. 1. (A) The *Drosophila* giant bZIP transcription factor contains two residues in the basic region and three residues in the fork region that are distinct from those in the respective regions of vertebrate PAR bZIP transcription factors. (B) Schematic diagram of a bZIP dimer bound to DNA. Shown are the relative positions of the leucine zipper, fork, and basic regions; the arrows indicate the two half-sites of the palindromic binding-site sequence.

preferences of GCN4 and C/EBP (13, 14). For the most part, however, the principles that underlie differential sequence recognition by bZIP proteins remain obscure.

One strategy for identifying specificity determinants in bZIP proteins is to exploit the existence of naturally occurring amino acid variants within subfamilies. An example of such variation occurs in the PAR bZIP protein family, which was defined by three factors (VBP/TEF, DBP, and HLF) identified in vertebrate species (6, 9, 11, 12, 20). The PAR proteins display a high degree of sequence identity over their DNA-binding domains and contain a conserved proline- and acidic-amino-acid-rich (PAR) region at a location amino terminal to the basic region (6). The core basic regions of these three factors are identical in all but two positions, and the fork regions are perfectly conserved (Fig. 1). The fact that the two nonconserved core basic-region residues occupy positions analogous to those of two GCN4 residues that do not make base-specific or phosphate backbone contacts in the GCN4-DNA crystal structures (7, 15) suggested that PAR bZIP factors might have similar, if not identical, DNA-binding specificities. Evidence in support of this inference is provided by a comparison of the empirically determined optimal binding sites for VBP (this study) and HLF (8).

Considering that several of the known PAR factor binding sites are also recognized by C/EBP proteins (20), we were surprised to learn from a recent database search that the basic and fork regions of the PAR proteins are most similar to the *Drosophila* giant bZIP factor (4) rather than to C/EBP proteins. As shown in Fig. 1, only two and three residues, respectively, distinguish the basic and fork regions of giant from those of the PAR factors. These minimal differences notwithstanding,

the fact that giant has been reported to bind to CREB/ATF-like sites (4), whereas PAR factors have not, suggested that giant and the PAR proteins may have distinct target site specificities. To address the issue of whether one or more of the five giant-specific basic-region and fork region residues might be able to confer distinct binding properties when substituted into a PAR factor, we compared VBP and various VBP/giant chimeras in binding assays using previously identified VBP and giant target sites. We also used a PCR-based site selection protocol to determine the optimal binding sites for VBP and several VBP/giant chimeras. The results of these studies highlight several determinants of binding-site specificity that could not have been inferred by reference to the crystal structures of the yeast GCN4 bZIP factor. In addition, the resolution of high-affinity binding sites for VBP reveals that many target sites for PAR factors are also target sites for a previously described transcriptional repressor as well as positively acting transcription factors that belong to two other bZIP subfamilies.

MATERIALS AND METHODS

Construction of VBP and VBP/giant expression vectors. Oligonucleotides with methionine codons and flanking *NcoI* sites were used in conjunction with an *NcoI* site-containing antisense oligonucleotide directed against the 3' untranslated region to PCR amplify codons at positions 79 to 313, 228 to 313, or 243 to 313 from a previously described full-length VBP cDNA clone (12). These *NcoI* fragments were cloned in place of the beta-globin open reading frame segment in the pT7 β globin in vitro expression vector (generously provided by Tom Kadesch). The 235-, 86-, and 71-amino-acid proteins encoded by these three constructs are denoted full length, intermediate, and core VBP, respectively. VBP codons at positions 259, 263, 266, 269, and 270 (CGT, GCT, TTA, AAT, and CAG, which code for arginine, alanine, lysine, asparagine, and glutamine, respectively) were individually replaced with giant-specific codons (AAG, CGT, ATA, GAT, and GAG, which code for lysine, arginine, isoleucine, aspartic acid, and glutamic acid, respectively) in the context of the pT7VBP(79-313) plasmid by a two-step PCR method (35) that we have described in detail in a previous publication (12). After sequencing of each of the five resultant VBP/giant chimeric open reading frames to confirm that only the desired mutations were introduced, these five pT7VBP/giant plasmids were used as templates to program additional rounds of two-step PCRs in order to generate pT7VBP/giant plasmids with pairs of giant-specific codons. This process of two-step PCR mutagenesis and sequencing was serially repeated to introduce 12 permutations of giant-specific codon substitutions into the VBP open reading frame. A subset of these pT7VBP/giant plasmids (see below) was subsequently used as templates to obtain PCR fragments for the two more severely truncated open reading frame segments (codons at positions 228 to 313 and 243 to 313), which were each cloned back into the *NcoI* site of the pT7 expression vector.

In vitro expression of VBP and VBP/giant chimeras. The pT7VBP and pT7VBP/giant plasmids described in the preceding paragraph were used to direct in vitro transcription, and subsequent in vitro translation, reactions essentially as described in detail in a previous publication (12). In vitro translations were carried out in parallel with unlabeled methionine or [35 S]methionine by following protocols provided by the manufacturer (Promega). Labeled proteins were resolved on denaturing sodium dodecyl sulfate-polyacrylamide gels and exposed to film in order to confirm that the various transcripts were translated with equal efficiencies. Aliquots of unlabeled in vitro-translated proteins were stored at -70°C until further use.

Gel mobility shift assays. Equal amounts of in vitro-translated VBP and VBP/giant proteins, as determined by the analysis of ^{35}S -labeled proteins that were translated in parallel, were used for gel shift assays essentially as described in detail in a previous publication (12). The probes used for this analysis had identical 5' and 3' ends and differed only with respect to the nature of the 10-bp core sequence [5' AGGAGTG(N) $_{10}$ CTGATAA 3'] noted in each of the figures. To generate probes for these assays, 10-pmol aliquots of sense strand oligonucleotides were labeled with 50 pmol of [γ - ^{32}P]ATP and T4 polynucleotide kinase, annealed to 30 pmol of antisense strand oligonucleotides, and then resolved from the unincorporated [γ - ^{32}P]ATP by passage through Sephadex G-50. Probes (0.1 ng) were mixed with equivalent amounts of in vitro-translated protein (see above) in 30- μl (final volume) aliquots of binding buffer [250 mM KCl, 12 mM Tris (pH 7.8), 1 mM MgCl_2 , 0.1 mM dithiothreitol, 1 mM CaCl_2 , 5 mM NaCl, 0.1 mg of bovine serum albumin (BSA) per ml, 5% glycerol, 4 μg of poly(dI-dC)] and placed on ice for 1 h. The DNA and protein-DNA complexes were then resolved on 6% nondenaturing polyacrylamide gels that had been prerun for 30 min in a cold room. Gels were subjected to constant voltage (125 V), and the electrophoresis buffer (0.25 $\times$ TBE [19]) was recirculated manually every 30 min. In some cases, unlabeled double-stranded competitor binding sites were also

FIG. 2. VBP preferentially binds to a natural VBP binding site, whereas a VBP/giant chimera preferentially binds to a natural giant binding site. Gel shift assays were performed with a natural VBP binding site and wild-type VBP (A) or with a natural giant binding site and a VBP/giant chimera that contains the five giant-specific residues substituted into the basic and fork regions of VBP (B). The distinct binding-site specificities of these two proteins are evident from the fact that VBP binding to the natural VBP site is effectively inhibited by competition by the natural VBP site but not by the natural giant site (compare lanes 2 and 3 with lanes 10 and 11 in panel A) whereas VBP/giant chimera binding to the natural giant site shows the converse binding-site preferences (compare lanes 10 and 11 with lanes 2 and 3 in panel B). Two other competitor binding sites that are closely related to the natural VBP site (TTTAC/GTAAA and GTTAC/GTAAT; lanes 6 and 7 and 8 and 9, respectively) also competed efficiently for binding to both VBP (A) and the VBP/giant chimera (B). Lanes 14 show that a 100-fold excess of the unrelated USF binding site (i.e., 10 ng) does not compete for either protein. All of the other competitor binding sites were present in either 30-fold excess (even lane numbers) or 100-fold excess (odd lane numbers) relative to the respective probes. We presume that the faster-mobility complexes in panel B are due to the binding of proteolytic cleavage products.

added, as indicated in the figure legends. An AMBIS radioanalytic imaging system was used to quantify the data from one such competition gel shift assay (see Fig. 4).

Binding-site selection experiments. Selection of optimal binding sites from a pool of random-sequence oligonucleotides was carried out by the method of Pollock and Treisman (25), with some modifications, most of which have been described previously (13). Epitope-tagged VBP proteins were constructed and expressed in *Escherichia coli* as follows. *Nco*I fragments containing wild-type and mutant VBP genes were inserted into pT5-Gtag, a derivative of the T7 expression vector pT5 (6a, 18). pT5-Gtag contains an expression cassette encoding the first 12 amino acids of GCN4 followed by an *Nco*I cloning site. Proteins expressed from this vector bear the sequence MSEYQPSLFALNA at their N termini and are therefore recognized by an antibody specific for the GCN4 N terminus (GCN4-N) (13). The VBP expression plasmids were introduced into the host strain BL21(DE3)(pLysS) (31), and the recombinant proteins were expressed and cell extracts were prepared as described previously (13). Protein levels were assessed by Western blotting (immunoblotting) with the GCN4-N antibody, and the concentrations of VBP protein in the extracts were equalized by diluting the more concentrated samples with control *E. coli* extract. The equalized extracts were diluted an additional 20-fold prior to use in the selection procedure.

A partly degenerate oligonucleotide [5' CATAGATGGATCCTCT(N)₁₈TCA GAATTCTCGAACC 3'] was made double stranded and labeled by annealing the reverse primer (5' GGTTCGAGAATTCCTGA 3') and filling in with Klenow DNA polymerase in a reaction mixture containing [³²P]dCTP. The double-stranded product was purified by polyacrylamide gel electrophoresis, eluted, and precipitated with ethanol. Approximately 0.8 pmol of labeled DNA probe was used in a binding reaction mixture containing the following components (final concentrations): 250 mM KCl, 16 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) (pH 7.9), 16% glycerol, 0.16 mM EDTA, 0.4 mM dithiothreitol, 0.4 mM phenylmethylsulfonyl fluoride, 0.1% Nonidet P-40, 100 µg of BSA per ml, 50 µg of poly(dI-dC) per ml, 1 µl of protein extract, and 1 µl of GCN4-N antiserum. After incubation for 30 min on ice, the reaction mixtures were transferred to tubes containing 10 µl of protein A-Sepharose beads and incubated another 90 min at 4°C with mixing. The beads were washed as described previously (13), except that the wash buffer contained 250 mM KCl. Bound DNA was eluted, precipitated with ethanol, and amplified by PCR as described elsewhere (13) by using the forward (5' CATAGATGGATCCTCT 3') and reverse primers. PCR products were extracted with chloroform and precip-

itated by adjusting the solution to 2.5 M ammonium acetate and adding 6 µg of glycogen carrier and 2 volumes of ethanol. Pellets were washed with 70% ethanol, counted, and resuspended in 10 mM Tris-1 mM EDTA, pH 8.0 (TE), at a concentration of 5,000 cpm/ml, and 5,000 cpm was used for the subsequent binding reaction. After five cycles of binding enrichment and amplification, the DNA was digested with *Bam*HI and *Eco*RI and ligated into Bluescript. The ligated DNA was digested with *Pst*I prior to transformation into DH5α cells to enrich for plasmids that contained inserts. Plasmid DNA was prepared from several independent clones from each binding-enriched pool, and the inserts were sequenced.

RESULTS

To assess whether the five giant-specific residues in the basic and fork regions alter the binding specificity of a PAR bZIP protein, we used PCR site-directed mutagenesis to introduce codons for these amino acids (KRIDE) at the corresponding positions in VBP (RALNQ) (Fig. 1). The resultant chimeric factor [which we will refer to as VBP/giant(KRIDE)] and wild-type VBP were translated in vitro, and the binding properties of these two proteins were analyzed by gel shift assays. The probe used to assay VBP binding corresponds to a sequence (TTTAC/ATAAA) from the chicken vitellogenin II promoter that was previously used to isolate VBP cDNA clones from an expression library (12). The probe for VBP/giant(KRIDE) corresponds to a giant binding site (ATGAC/GCAAG) from the *Drosophila* Kruppel promoter (4).

As shown in Fig. 2A, VBP bound to the VBP binding site as expected (lane 1) and this complex was effectively inhibited by competition by an excess of the unlabeled homologous binding site (lanes 2 and 3). In contrast, the giant binding site from the Kruppel promoter was a poor competitor (Fig. 2A, lanes 10 and 11), and the same was true of the related CREB/ATF

FIG. 4. The fork region of giant confers the ability to discriminate between two closely related palindromic sequences when assayed in the context of a VBP/giant chimera. Two competition gel shift assays were carried out in parallel with the ATTACGTAAT probe and various amounts of either the homologous unlabeled fragment (solid boxes) or a related unlabeled fragment (TTTACGTAAA) (striped boxes). Note that these two unlabeled sequences differ at only positions -5 and $+5$ (relative to the centers of the respective palindromes). The top panel shows the results obtained with VBP, whereas the bottom panel shows the results obtained with the VBP/giant(IDE) chimera, which contains the giant-specific fork region residues.

respectively, a significant number of the selected binding sites have an A at position $+1$ and, to a lesser extent, a T at position -1 . This finding was also not unexpected in light of the data shown in Fig. 2. Much to our surprise, however, a significant number of the selected binding sites contained sequences resembling a CREB/ATF half-site (GTCA, reading from $+1$ to $+4$). In particular, the three bases that abut the variable portion of the probe (TCA) match the $+2$ to $+4$ portion of a CREB/ATF half-site (GTCAT), and in seven of the selected binding sites this triplet was preceded by a G to yield a GTCA half-site. Since the right arms of these imperfect palindromes were not derived from degenerate bases, they were not included in the tally shown in Fig. 5. In addition, two other CREB/ATF half-sites were selected from the random portion of the cassette and are included in Fig. 5.

FIG. 5. Identification of high-affinity binding sites for VBP and VBP/giant chimeras. A PCR-based reiterative protocol (see Materials and Methods) was used to select high-affinity binding sites for wild-type VBP and VBP/giant chimeras containing giant-specific basic-region residues (KR), fork region residues (IDE), or basic-region and fork region residues (KRIDE). The sequences were aligned, and the base frequencies in each position were tabulated. Note that bases derived from the invariant portion of the amplification cassette were not included in this tally, as indicated by the fact that the sum of the four bases in some positions is less than the respective sample size.

The binding sites selected by VBP/giant(KR) and VBP/giant(KRIDE) chimeras are quite similar to those selected by wild-type VBP, and all three factors appeared to tolerate single CREB/ATF half-sites. In contrast, the lack of any CREB/ATF half-sites in the sample of binding sites selected by the VBP/giant(IDE) chimera suggests that the DNA-binding domain may be more severely constrained by giant-specific fork residues than by VBP-specific fork residues. The more pronounced preferences for G over A at position $+1$, and for A over G at position -2 , in the sites selected by VBP/giant(IDE) versus VBP appear to support this conclusion. Similar biases are seen when one compares the sites selected by VBP/giant(KRIDE) and VBP/giant(KR). It is also noteworthy that all of the binding sites that were obtained with the two chimeras with giant-specific fork region residues [VBP/giant(IDE) and VBP/giant(KRIDE)] derive from the random-sequence portion of the serially selected PCR cassette. In contrast, the proteins that contain VBP-specific fork region residues [VBP and VBP/giant(KR)] each selected numerous binding sites that overlapped with the fixed-sequence portion of this cassette. This is

FIG. 6. The similar optimal binding-site preferences for VBP and the VBP/giant(KRIDE) chimera are attributable to the combination of a more permissive giant basic region and a more restrictive giant fork region. A probe (ATTAC/GTAAT) that was identified as a high-affinity binding site for both VBP and the VBP/giant(KRIDE) chimera (Fig. 5) was used in gel shift assays with either VBP (A); the VBP/giant(KR) chimera, which contains the giant-specific basic-region residues (B); the VBP/giant(IDE) chimera, which contains the giant-specific fork region residues (C); or the VBP/giant(KRIDE) chimera, which contains the giant-specific basic-region and fork region residues (D). Note that a broader range of sites compete for binding to the VBP/giant(KR) chimera (B) than to any of the other proteins tested. On the other hand, the giant-specific fork region residues serve to restrict the binding-site specificity of the VBP/giant(KRIDE) chimera relative to that of the VBP/giant(KR) chimera (compare panels D and B). Note also that the presence of a CREB/ATF or C/EBP half-site (GTCAT and GCAAT, respectively) in conjunction with an optimal VBP half-site (GTAAT) only modestly decreases binding to either VBP or the VBP/giant(KRIDE) chimera (panels A and D, respectively), as predicted from the results of the optimal site selections (Fig. 5). We presume that the minor faster-mobility complexes in some of the lanes are due to the binding of proteolytic products.

indicated by the fact that the number of bases included in these compilations is less than the sample sizes for the +2 to +5 positions. As noted above, the compilations presented in Fig. 5 include bases from the degenerate portion of the original PCR cassette only.

Since the selection of binding sites that contain a CREB/ATF half-site was not anticipated, we carried out competition gel shift assays to test whether these sites might indeed be potential binding sites for VBP and VBP/giant chimeras. The results obtained with a high-affinity VBP binding site (ATTAC/GTAAT) probe and either VBP, VBP/giant(KR), VBP/giant(IDE), or VBP/giant(KRIDE) are shown in Fig. 6A through D, respectively. Considering first the data obtained with VBP (Fig. 6A), it is clear that GTTAC/GTCAT (which is referred to in the figure as a VBP/CREB site but which we will refer to henceforth in the text as a PAR-CREB/ATF chimeric site to reflect the mixed nature of this imperfect palindrome) was nearly as effective as two different VBP sites (ATTAC/GTAAT and GTTAC/GTAAT) in competing for VBP binding (Fig. 6A; compare lanes 4 and 5 with lanes 2 and 3 and lanes 12 and 13, respectively). A PAR-C/EBP chimeric site (GTTAC/GCAAT; referred to as a VBP/CEBP site in Fig. 6) was also an effective

competitor (Fig. 6A, lanes 8 and 9), whereas consensus CREB/ATF and C/EBP sites (ATGAC/GTCAT and ATTGC/GCAAT, respectively) were much poorer competitors (lanes 6 and 7 and lanes 10 and 11, respectively), albeit that each was more effective at competing than the USF binding site that we tested (lane 14). On the basis of these results, we conclude that PAR-CREB/ATF and PAR-C/EBP chimeric sites are relatively high-affinity binding sites for VBP.

Whereas consensus CREB/ATF and C/EBP sites were relatively poor competitors for binding to VBP, these sites both competed efficiently for binding to VBP/giant(KR). In fact, as evidenced by the results shown in Fig. 6B, this chimeric factor bound to all of the competitors tested, with the exception of the unrelated USF binding site. This observation provides additional support for the idea that these giant-specific basic-region residues effect a broadening of the binding-site specificity when substituted in place of the analogous VBP residues. Conversely, sequences that deviated from the 8-bp core consensus sequence (TTAG/CTAA) competed poorly, if at all, for binding to VBP/giant(IDE), as shown by Fig. 6C. Again, the ability of giant-specific fork region residues to restrict DNA-binding specificity is evident not only from a comparison of the results obtained with VBP/giant(IDE) and VBP (Fig. 6C and A, respectively) but also from a comparison of the results obtained with VBP/giant(KRIDE) and VBP/giant(KR) (panels D and B, respectively).

Two considerations prompted us to investigate whether the repertoire of high-affinity binding sites for truncated proteins that contain only the core bZIP region might differ from the sets shown in Fig. 5. First, mutations in the extended basic region (EBR) that resides upstream of the core basic region were found to reduce the affinity of TEF (the mouse homolog of VBP) to a subset of the binding sites that were surveyed by Drolet et al. (6). And second, since both of the yeast GCN4-DNA crystal structures were derived by using severely truncated (i.e., essentially core bZIP) forms of GCN4 (7, 15), it was of interest to examine in a systematic fashion the possibility that additional domains might be relevant to the interactions of at least some bZIP factors with their respective binding sites. We thus carried out DNA-binding-site selections for truncated bZIP versions of VBP, as well as the three VBP/giant chimeras that we had analyzed as full-length proteins (Fig. 5). The choice of position for the truncation breakpoint was based on the fact that the VBP core bZIP region is encoded by a single exon (3) and the fact that this segment of VBP is essentially analogous to the segments of GCN4 used for crystal structure determinations. As shown in Fig. 7, each of these four truncated bZIP factors displayed a striking preference for the sequence (G/A)TTAC/GTAA(C/T). The fact that a larger number of deviations from this sequence was seen with full-length versions of these factors (Fig. 5) suggests that the truncation of sequences amino terminal to the core basic region effectively restricts the DNA-binding specificities of these factors. For example, whereas T and A at positions -1 and +1, respectively, were represented in the data sets obtained with full-length VBP and all of the VBP/giant chimeras (Fig. 5), these sequences were completely lacking in the data sets obtained with the truncated bZIP factors. Similarly, whereas CREB/ATF half-sites were well represented among the sequences selected with full-length VBP (Fig. 5), only two CREB/ATF half-sites were selected by the truncated bZIP version of this factor.

As noted above, the extended basic region (EBR) has been implicated as being an important determinant of DNA-binding specificity for the mouse homolog of VBP (6). To address whether this might be the only such determinant outside the

FIG. 7. Amino-terminal truncations that remove all but the core bZIP domain result in a general narrowing of binding-site specificities. A PCR method was used to generate VBP and three different VBP/giant chimeras that were truncated at the beginning of the (fourth) bZIP exon of VBP (see Materials and Methods). High-affinity binding sites for these core bZIP factors were identified by the protocol described in the legend to Fig. 5. A comparison of the results obtained by using these core bZIP factors with those presented in Fig. 5 for the respective full-length factors reveals that the truncated factors are biased toward selecting more restricted sets of binding sites. These results indicate that sequences amino terminal to the core bZIP domain contribute to the binding specificity of VBP as well as to the binding specificities of various artificial VBP/giant chimeras.

bZIP region of VBP, we expressed VBP and VBP/giant chimeras from a methionine residue that maps to a location immediately amino terminal to the EBR region and used gel shift assays to evaluate the binding properties of the resultant EBR-bZIP truncated proteins. The respective full-length and bZIP VBP and VBP/giant chimeras were also analyzed for comparison. The results of this analysis are shown in Fig. 8. As shown in Fig. 8A, the abilities of both VBP and VBP/giant(KR) to bind to the TTTAC/GTAAA probe are absolutely dependent on sequences that map to a location amino terminal to the EBR. This result is particularly striking considering that full-length VBP has nearly the same affinity for this site as it has for an optimal binding site (unpublished results; also Fig. 4). A similar dependence on sequences that reside at a location amino terminal to the EBR is also evident from the data presented in Fig. 8G (compare lanes 1 and 2). The second major conclusion from this analysis is that the EBR facilitates

the binding of VBP and various VBP/giant chimeric factors to nearly all of the probes tested. Interestingly, the one probe (ATTAC/GTAAT; Fig. 8C) bound by each of these four bZIP truncated factors corresponds to the common consensus optimal binding site that was deduced from the site selections that were carried out with these bZIP truncated proteins (Fig. 7). Curiously, the other probe (GTTAC/GTAAT; Fig. 8H) that might have been predicted to bind to all four of these bZIP truncated proteins on the basis of its match to the deduced consensus sequence [(G/A)TTAC/GTAA(C/T); Fig. 7] was a poor binding site for both VBP and the VBP/giant(IDE) chimera. Finally, in keeping with the evidence presented above for the relatively broad binding specificity of the VBP/giant(KR) chimera, the bZIP truncated version of this factor was found to bind to the largest number of probes tested.

DISCUSSION

Most studies that have been directed toward understanding the molecular basis for the DNA-binding specificity of bZIP transcription factors have been focused on the core basic region, which has been shown to make base-specific and phosphate backbone contacts in GCN4-DNA crystal structures (7, 15). We have shown here that the fork region, which was previously found to be a determinant for preferential binding to either abutted or overlapping arrangements of half-sites for the GCN4 and C/EBP bZIP proteins (13, 14), is also involved in determining half-site specificity for a PAR bZIP factor. Additionally, we provide evidence that the DNA-binding specificity of VBP is modulated by at least two other domains in addition to the basic and fork regions. These auxiliary domains are located at a site amino terminal to the basic region and serve to broaden the binding specificity of full-length VBP relative to a truncated protein that contains only the core bZIP domain. Thus, although the primary DNA-binding determinants of bZIP proteins reside in the basic region, our findings demonstrate that more subtle deviations from the consensus binding specificity of PAR bZIP factors can be conferred by other regions of the protein.

The fork region. The analysis of C/EBP-GCN4 chimeras (13) and of amino acid substitution mutants of GCN4 (14) has led to the proposal that residues in the fork region dictate the ability to discriminate between abutted and overlapping configurations of half-sites (half-site spacing). On the basis of our analysis of VBP/giant chimeras, it is apparent that fork region residues can also influence discrimination between related palindromes such as TTTAC/GTAAA and ATTAC/GTAAT (Fig. 4). In general, the substitution of giant-specific fork region residues into VBP appears to increase half-site sequence specificity. For example, whereas PAR-CREB/ATF sites were represented in the set of high-affinity sites selected from a pool of degenerate sequences with full-length VBP, these sites were not selected with a full-length VBP/giant chimera that contained the fork residues from giant (Fig. 5). Our VBP/giant chimeras failed to bind to overlapping versions of any of the palindromes tested (data not shown), which is not surprising in view of the fact that the parental VBP and giant proteins bind to abutted palindromes.

How might these giant-specific fork region substitutions affect discrimination between related palindromes that differ solely at the +5 and -5 positions? We suggest that this may be due to the fork region of the alpha helix being rendered more rigid by intrahelical salt bridges (Fig. 9). Since this rigidity would be translated to the basic-region alpha-helical segment, compensatory adjustments might be required of target sequences for binding to occur. Thus, ATTAC/GTAAT and TT

FIG. 8. At least two domains that reside at a location amino terminal to the core bZIP domain of VBP help to stabilize binding to nonoptimal sites. Long, intermediate, and short forms (lanes 1, 2, and 3, respectively) of VBP and three VBP/giant chimeras were tested in gel shift assays for their abilities to bind to the indicated probes. The terms long and short refer to the full-length and core bZIP forms analyzed in preceding figures; the term intermediate denotes a form that includes the extended basic region which flanks the core bZIP domain (see Materials and Methods for details). The critical importance of at least two domains that reside at a location amino terminal to the core bZIP domain is evident from the fact that the binding-site specificity broadens in a stepwise fashion when the short, intermediate, and long versions of several of these factors are compared.

TAC/GTAAA may differ in their inherent flexibilities or curvatures such that only the former can stably contact the more rigidly positioned basic-region alpha-helical segment of VBP in the context of VBP/giant(fork region) chimeras. The importance of the intrinsic bend of a target sequence as a determinant of binding specificity has received some experimental support (24). However, we infer that the relative importance of

this parameter will have to be evaluated on a case-by-case basis.

Amino-terminal auxiliary domains. We have presented evidence that in addition to the basic and fork regions, at least two other domains that reside at a location amino terminal to the core basic region of VBP contribute to DNA-binding-site specificity. One of these domains likely maps to the so-called extended basic region of VBP. A cluster of basic amino acids [KKA(R/K)K] is conserved in the -41 to -37 region (relative to the beginning of the leucine zipper) for all three members of the PAR subfamily of bZIP factors, and mutations within this region have been shown to compromise the ability of TEF (the mammalian homolog of VBP) to bind to a subset of naturally occurring TEF binding sites (6). On the basis of data presented in this study, we infer that this extended basic region is critical for binding to a broad range of sites, with the exception being sites that were identified as high-affinity sites for the core bZIP fragment of VBP by a PCR-based site selection protocol. A similar conclusion was recently drawn on the basis of an analysis of HLF, which is another member of the PAR subfamily (8). It should be noted that analogous extended basic regions are not a feature common to all bZIP subfamilies, however. Indeed, we have shown that this dependence on the extended basic region for binding to a number of target sites can be mitigated to a large extent by altering residues within the core basic region of VBP (Fig. 7). Since site selections carried out with full-length and core bZIP fragments of VBP and VBP/giant chimeras yielded 10-bp consensus sequences in each case, residues outside the core basic domain probably make nonspecific phosphate backbone contacts with the DNA rather than base-specific contacts.

FIG. 9. Fork region intrahelical salt bridges (dashed lines) may be important determinants of binding-site specificity for bZIP transcription factors. The fork and first leucine zipper heptad repeats of VBP and giant are shown with the three distinct fork region residues highlighted in boldface type. As indicated in the figure, the aspartic acid at position -2 of giant (relative to the beginning of the leucine zipper) has the potential to form an intrahelical salt bridge with the arginine at position -6 and the glutamic acid at position -1 has the potential to form a salt bridge with the lysine at position -4 or the arginine at position +4 within the first heptad repeat. The substitutions of these fork region residues into VBP create analogous potential intrahelical salt bridges since the relevant charged amino acids are conserved between these two bZIP factors. Note that each of these fork region substitutions is sufficient to restrict the binding specificity of VBP (for example, see Fig. 3).

FIG. 10. Distinct subsets of the high-affinity binding sites for PAR factors may also be targets for members of the C/EBP or CREB/ATF subfamilies of bZIP factors as well as the E4BP4 transcriptional repressor. Consensus binding sites for C/EBP (ATTGC/GCAAT), PAR (RTTAC/GTAAAY), and CREB/ATF (ATGAC/GTCAT), as well as imperfect palindromes composed of either C/EBP and PAR or PAR and CREB/ATF half-sites, are represented in panel A. C/EBP, PAR, E4BP4, and CREB/ATF dimers are represented in panel B as indicated. The dashed vertical arrows represent predicted interactions between the various bZIP dimers and their respective consensus binding sites as well as the indicated imperfect palindromes. As discussed in the text, this diagram is based on results obtained in this study as well as data reported in the literature.

Although the more amino-terminal auxiliary domain of VBP has yet to be mapped, the functional significance of this domain is indicated by the fact that it is required to allow VBP to bind to the target site from the vitellogenin II promoter (TT-TAC/ATAAA) that was used to isolate cDNA clones for this factor (12) and that can be transactivated by VBP in cotransfection assays (3).

DNA-binding specificity of the PAR bZIP protein family. VBP/TEF, DBP, and HLF were originally classified as a distinct subfamily of bZIP factors because they have conserved PAR and basic-region domains and because they were shown to have functionally compatible leucine zipper dimerization domains. On the basis of the results of binding-site selection experiments carried out with VBP (this study) and HLF (8), it also appears that PAR factors have identical DNA-binding specificities. Considering that the consensus half-sites for CREB/ATF and C/EBP each differ in only a single position from the consensus PAR half-site (GTCAT and GCAAT versus GTAAAY, respectively), it is not surprising that PAR-CREB/ATF and PAR-C/EBP chimeric sites were represented among the high-affinity sites that were selected with full-length VBP (Fig. 5) and that both of these sequences were found to bind VBP effectively (Fig. 6). Conversely, a retrospective analysis of the sets of high-affinity binding sites that were selected with a C/EBP factor (13) and a CREB factor (2) revealed numerous examples of PAR-C/EBP chimeric sites and PAR-CREB/ATF chimeric sites, respectively. It should be noted, however, that these were previously described as imperfect C/EBP and CREB/ATF binding sites, respectively, since PAR half-sites had not been reported at the time these studies were published.

These two types of imperfect palindromes represent obvious targets for PAR factors and either C/EBP or CREB/ATF factors (Fig. 10). For example, sequences within the HNF3 β , hemopexin, and transthyretin gene regions that were previously described as imperfect C/EBP binding sites can now

more properly be described as PAR-C/EBP chimeric sites, and these sequences have recently been shown to be targets for PAR factors as well as C/EBP factors (28). Similarly, the two ATF binding sites from the adenovirus E4 promoter that confer E1A inducibility (27) can retrospectively be classified as PAR-CREB/ATF chimeric sites. Interestingly, in contrast to other ATF sites from adenovirus early promoters, these two particular ATF binding sites are also binding sites for an E1A-inducible binding activity termed E4F. Since cDNA clones have not yet been reported for E4F, the relationship between this factor and PAR factors is presently unknown. On the other hand, a factor that appears to be distinct from E4F was identified by screening a cDNA expression library with one of these atypical ATF sites. This factor, which has been named E4BP4, functions as a transcriptional repressor and has a binding specificity that is nearly identical to that of PAR factors (5). Thus, we suggest that E4BP4 may play a significant role in antagonizing PAR-mediated transcriptional activation. Moreover, on the basis of the interhelical salt bridge rules that have been proposed to govern leucine zipper interactions (33), it is possible that E4BP4 may be able to heterodimerize with PAR factors. This remains to be tested, however.

An emerging theme of bZIP DNA-binding specificity is that a conserved DNA-contacting surface, the basic region, can be altered by minor variations in basic-region residues or by adjacent structural domains to generate differences in DNA recognition (reference 14 and this study). It is clear that the fork region imposes constraints on the way in which the basic region is presented to the DNA, and, at least for VBP, two domains that reside at a location amino terminal to the basic region also contribute to binding-site specificity. These findings emphasize the difficulties of attempting to draw inferences about binding-site specificities from simple amino acid sequence comparisons and underscore the limitations of the crystal structure determined for the severely truncated GCN4 protein as a general paradigm for modeling the interactions between bZIP factors and their target DNA sequences. Further mutational studies and determination of physical structures for other bZIP factors should lead to a more universal understanding of bZIP protein-DNA specificity.

ACKNOWLEDGMENTS

We thank Marge Strobel, Kevin Ryder, and members of our laboratories for constructive criticisms regarding the manuscript. We also thank Robert Muhlhauser (Oligonucleotide Synthesis Facility, Fox Chase Cancer Center) for providing oligonucleotides and Marge Eienkel and Sarah Costello (Secretarial Services, Fox Chase Cancer Center) for help in preparing the manuscript.

This research was supported, in part, by National Institutes of Health grants 35535 (awarded to J.B.), CA-06927 (awarded to the Fox Chase Cancer Center), and T32 CA-09035-19 (awarded to the Fox Chase Cancer Center). Additional support was provided by an appropriation from the Commonwealth of Pennsylvania and by the National Cancer Institute under contract N01-CO-74101 with ABL.

REFERENCES

- Agre, P., P. F. Johnson, and S. L. McKnight. 1989. Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. *Science* **246**:922-925.
- Benbrook, D. M., and N. C. Jones. 1994. Different binding specificities and transactivation of variant CRE's by CREB complexes. *Nucleic Acids Res.* **22**:1463-1469.
- Burch, J. B. E., and D. L. Davis. 1994. Alternative promoter usage and splicing options result in the differential expression of mRNAs encoding for isoforms of chicken VBP, a member of the PAR subfamily of bZIP transcription factors. *Nucleic Acids Res.* **22**:4733-4741.
- Capovilla, M., E. D. Eldon, and V. Pirrotta. 1992. The *giant* gene of *Drosophila* encodes a b-ZIP DNA-binding protein that regulates the expression of other segmentation gap genes. *Development* **114**:99-112.

5. Cowell, I. G., A. Skinner, and H. C. Hurst. 1992. Transcriptional repression by a novel member of the bZIP family of transcription factors. *Mol. Cell. Biol.* **12**:3070–3077.
6. Drolet, D. W., K. M. Scully, D. M. Simmons, M. Wegner, K. Chu, L. W. Swanson, and M. G. Rosenfeld. 1991. TEF, a transcription factor expressed specifically in the anterior pituitary during embryogenesis, defines a new class of leucine zipper proteins. *Genes Dev.* **5**:1739–1753.
- 6a. Eisenberg, S. Unpublished data.
7. Ellenberger, T. E., C. J. Brandl, K. Struhl, and S. C. Harrison. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: crystal structure of the protein-DNA complex. *Cell* **71**:1223–1237.
8. Hunger, S. P., R. Brown, and M. L. Cleary. 1994. DNA-binding and transcriptional regulatory properties of hepatic leukemia factor (HLF) and the t(17;19) acute lymphoblastic leukemia chimera E2A-HLF. *Mol. Cell. Biol.* **14**:5986–5996.
9. Hunger, S. P., K. Ohyashiki, K. Toyoma, and M. L. Cleary. 1992. HLF, a novel hepatic bZIP protein, shows altered DNA-binding properties following fusion to E2A in t(17;19) acute lymphoblastic leukemia. *Genes Dev.* **6**:1608–1620.
10. Hurst, H. C. 1994. Transcription factors. 1. bZIP proteins. *Protein Profiles* **1**:123–168.
11. Inaba, T., W. M. Roberts, L. H. Shapiro, K. W. Jolly, S. C. Raimondi, S. D. Smith, and A. T. Look. 1992. Fusion of the leucine zipper gene HLF to the E2a gene in human acute B-lineage leukemia. *Science* **257**:531–534.
12. Iyer, S. V., D. L. Davis, S. N. Seal, and J. B. E. Burch. 1991. Chicken vitellogenin gene-binding protein, a leucine zipper transcription factor that binds to an important control element in the chicken vitellogenin II promoter, is related to rat DBP. *Mol. Cell. Biol.* **11**:4863–4875.
13. Johnson, P. F. 1993. Identification of C/EBP basic region residues involved in DNA sequence recognition and half-site spacing preference. *Mol. Cell. Biol.* **13**:6919–6930.
14. Kim, J., D. Tzamarias, T. Ellenberger, S. C. Harrison, and K. Struhl. 1993. Adaptability at the protein-DNA interface is an important aspect of sequence recognition by bZIP proteins. *Proc. Natl. Acad. Sci. USA* **90**:4513–4517.
15. Konig, P., and T. J. Richmond. 1993. The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* **233**:139–154.
16. Kouzarides, T., and E. Ziff. 1988. The role of the leucine zipper in the fos-jun interaction. *Nature (London)* **336**:646–651.
17. Kouzarides, T., and E. Ziff. 1989. Leucine zippers of *fos*, *jun* and GCN4 dictate dimerization specificity and thereby control DNA binding. *Nature (London)* **340**:568–571.
18. Landschulz, W. H., P. F. Johnson, and S. L. McKnight. 1989. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science* **243**:1681–1688.
19. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. Mueller, C. R., P. Maire, and U. Schibler. 1990. DBP, a liver-enriched transcriptional activator, is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. *Cell* **61**:279–291.
21. Nakabeppu, Y., and D. Nathans. 1989. The basic region of *Fos* mediates specific DNA binding. *EMBO J.* **8**:3833–3841.
22. Oliphant, A. R., C. J. Brandl, and K. Struhl. 1989. Defining sequence specificity DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of the yeast GCN4 protein. *Mol. Cell. Biol.* **9**:2944–2949.
23. O'Neil, K. T., R. H. Hoess, and W. F. DeGrado. 1990. Design of DNA-binding peptides based on the leucine zipper motif. *Science* **249**:774–778.
24. Paolella, D. N., C. R. Palmer, and A. Schepartz. 1994. DNA targets for certain bZIP proteins distinguished by an intrinsic bend. *Science* **264**:1130–1133.
25. Pollock, R., and R. Treisman. 1990. A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res.* **18**:6197–6204.
26. Ransone, L. J., P. Wamsley, K. L. Morley, and I. M. Verma. 1990. Domain swapping reveals the modular nature of Fos, Jun, and CREB proteins. *Mol. Cell. Biol.* **10**:4565–4573.
27. Rooney, R. J., P. Raychaudhuri, and J. R. Nevins. 1990. E4F and ATF, two transcription factors that recognize the same site, can be distinguished both physically and functionally: a role for E4F in E1A *trans* activation. *Mol. Cell. Biol.* **10**:5138–5149.
28. Samadani, U., A. Porcella, L. Pani, P. F. Johnson, J. Burch, R. Pine, and R. H. Costa. Cytokine regulation of the liver transcription factor HNF-3b is mediated by the C/EBP family and interferon regulatory factor 1. Submitted for publication.
29. Sellers, J. W., and K. Struhl. 1989. Changing *Fos* oncoprotein to a *Jun*-independent DNA-binding protein with GCN4 dimerization specificity by swapping "leucine zippers." *Nature (London)* **341**:74–76.
30. Sellers, J. W., A. C. Vincent, and K. Struhl. 1990. Mutations that define the optimal half-site for binding yeast GCN4 activator protein and identify an ATF/CREB-like repressor that recognizes similar DNA sites. *Mol. Cell. Biol.* **10**:5077–5086.
31. Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**:60–89.
32. Suckow, M., B. von Wilcken-Bergmann, and B. Muller-Hill. 1993. Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP, and TAF-1 that are involved in specific DNA binding. *EMBO J.* **12**:1193–1200.
33. Vinson, C. R., T. Hai, and S. M. Boyd. 1993. Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes Dev.* **7**:1047–1058.
34. Vinson, C. R., P. B. Sigler, and S. L. McKnight. 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **246**:911–916.
35. Zaret, K. S., J.-K. Liu, and C. M. DiPersio. 1990. Site-directed mutagenesis reveals a liver transcription factor essential for the albumin enhancer. *Proc. Natl. Acad. Sci. USA* **87**:5469–5473.