Article

DNA-Binding Specificity of PAR and C/EBP Leucine Zipper Proteins: A Single Amino Acid Substitution in the C/EBP DNA-Binding Domain Confers PAR-Like Specificity to C/EBP

FALVEY, Eileen Elisabeth, MARCACCI, Lysiane, SCHIBLER, Ulrich

Abstract

PAR and C/EBP family proteins are liver-enriched basic leucine zipper (bZip) transcription factors that bind to similar sites on the promoters of albumin and cholesterol 7 hydroxylase genes. However, C/EBP proteins have a more relaxed binding specificity than PAR proteins, in that they recognize many sites within promoter or randomly selected rat genomic DNA sequences that are ignored by PAR proteins. Thus, DNAse I protection experiments suggest that C/EBP recognizes a binding site with an affinity similar to the one of the cholesterol 7 hydroxylase gene promoter every 200 to 300 bp. The frequency of PAR protein binding sites with comparable affinities is about 20- fold lower in the rat genome. By using a PCR-based amplification assay we selected high affinity DMAbinding sites for C/EBPß and the PAR protein DBP from a pool of oligonucleotides. Both proteins indeed recognize similar sequences with the optimal core binding sequence 5'RTTAY.GTAAY3'. However, as expected, DBP, is considerably less tolerant to deviations from the consensus site. Here we have characterized a single amino acid substitution mutant of C/EBPß [...]

Reference

FALVEY, Eileen Elisabeth, MARCACCI, Lysiane, SCHIBLER, Ulrich. DNA-Binding Specificity of PAR and C/EBP Leucine Zipper Proteins: A Single Amino Acid Substitution in the C/EBP DNA-Binding Domain Confers PAR-Like Specificity to C/EBP. *Biological Chemistry Hoppe-Seyler*, 1996, vol. 377, no. 12, p. 797-810

DOI : 10.1515/bchm3.1996.377.12.797

Biol. Chem., Vol. 377, pp. 797 – 809, December 1996

DNA-Binding Specificity of PAR and C/EBP Leucine Zipper Proteins: A Single Amino Acid Substitution in the C/EBP DNA-Binding Domain Confers PAR-Like Specificity to C/EBP

Eileen Falvey, Lysiane Marcacci, and Ueli Schibler*

Department of Molecular Biology, University of Geneva, 30, Quai Ernest Ansermet, CH-1211 Geneva, Switzerland

* Corresponding author

PAR and C/EBP family proteins are liver-enriched basic leucine zipper (bZip) transcription factors that bind to similar sites on the promoters of albumin and cholesterol 7α hydroxylase genes. However, C/EBP proteins have a more relaxed binding specificity than PAR proteins, in that they recognize many sites within promoter or randomly selected rat genomic DNA sequences that are ignored by PAR proteins. Thus, DNAse I protection experiments suggest that C/EBP recognizes a binding site with an affinity similar to the one of the cholesterol 7α hydroxylase gene promoter every 200 to 300 bp. The frequency of PAR protein binding sites with comparable affinities is about 20-fold lower in the rat genome. By using a PCR-based amplification assay we selected high affinity DNA-binding sites for C/EBPβ and the PAR protein DBP from a pool of oligonucleotides. Both proteins indeed recognize similar sequences with the optimal core binding sequence 5′RTTAY.GTAAY3′. However, as expected, DBP, is considerably less tolerant to deviations from the consensus site. Here we have characterized a single amino acid substitution mutant of C/EBPβ that increases its target site specificity. This protein, C/EBP$\beta_{V>A}$, contains a valine to alanine substitution at position 13 of the basic domain (residue 216 of C/EBPβ). C/EBP$\beta_{V>A}$ selectively binds only the subset of C/EBP sites that are also DBP sites, both as oligonucleotides and within the natural contexts of the albumin and cholesterol hydroxylase promoters.

Key words: C/EBP / DBP / DNA binding specificity / HLF / PAR basic leucine zipper proteins / TEF.

Introduction

The superfamily of basic region leucine zipper (bZip) proteins comprises a wide range of transcriptional regulators that bind to their 9 or 10 bp DNA response elements as homo- and heterodimers (Cao *et al.*, 1991; Kouzarides and Ziff, 1988; Landschulz *et al.*, 1988; 1989; O'Shea *et al.*, 1992). Most bZip proteins can be classified into one of five families, AP-1, CREB/ATF, C/EBP, PAR, and the plant G-box proteins (Hurst, 1994; Johnson, 1993), based on their recognition site similarities, amino acid homologies, and the ability of family members to form heterodimers with each other. The leucine zipper motifs common among the bZip protein superfamily consist of amphipathic α-helical domains that form parallel coiled-coils that are stabilized via hydrophobic and polar interactions (Lumb and Kim, 1995; O'Shea *et al.*, 1992, and references therein). Upon binding to DNA, the previously unstructured basic regions of the yeast GCN4 protein dimers also become alpha-helical and interact through five amino acid-base contacts along the major groove at each half-site of their cognate DNA sequences. The protein:DNA complexes are further stabilized by several electrostatic contacts between positively charged amino acid side chains and negatively charged backbone phosphates. Thus, the two basic regions from the GCN4 dimer bind to a single turn of the DNA helix, making similar contacts with nucleotide bases at either side of the dyad-symmetrical recognition sequence. Similar contacts were found in both the GCN4:DNA and the cFos-cJun:DNA co-crystal structures, and it has therefore been suggested that other bZip proteins may also bind in an analogous fashion (Glover and Harrison, 1995; Kerppola and Curran, 1995).

The two liver-enriched transcription factors, DBP and C/EBPβ, belong to the PAR and C/EBP bZip subfamilies, respectively, of bZip transcription factors. The PAR subfamily, named for a common proline and acid rich peptide domain, also includes TEF/VBP, the thyrotroph embryonal factor and its chicken ortholog, the vitellogenin-gene binding protein (Drolet *et al.*, 1991; Iyer *et al.*, 1991), and HLF, the hepatic leukemia factor (Hunger *et al.*, 1992; Inaba *et al.*, 1992). In the liver of adult rats and mice, all three PAR bZip proteins accumulate according to a robust circadian rhythm (Falvey *et al.*, 1995; Fonjallaz *et al.*, 1996, Wuarin and Schibler, 1990) and are thus believed to regulate circadian gene expression in this tissue. Recently, PAR family members have also been suggested to play an important role in the regulation of apoptosis (Metzstein *et al.*, 1996; Inaba *et al.*, 1996). In fact, these transcription factors show extensive homology with the *Caenorhabditis elegans* protein CES-2, a bZip protein required for the programmed cell death of two serotoninergic neurons in *C. elegans*. The members of the C/EBP family appear to be involved in the differentiation of hepatocytes and adipocytes (Cao *et al.*, 1991; Descombes *et al.*, 1990; Descombes and Schibler, 1991; Lin and Lane, 1994; Wang

et al., 1995, Yeh *et al.*, 1995), in energy metabolism (McKnight *et al.*, 1989; Wang *et al.*, 1995), and in regulating acute-phase and immune responses (Screpanti *et al.*, 1995; and references therein). In fact, the first leucine zipper structure was uncovered in C/EBPα, and this protein can thus be considered as the founding member of the leucine zipper protein family (Landschulz *et al.*, 1988).

Although DBP and C/EBPβ belong to distinct bZip families and do not heterodimerize (Vinson *et al.*, 1993), the two transcription factors have much in common. Both are expressed in the terminally differentiated adult liver as well as a variety of other tissues, such as kidney and lung, and activate transcription from similar target sites in transfection experiments. For example, two potential DBP target promoters, albumin (Mueller *et al.*, 1990) and cholesterol hydroxylase (Lavery and Schibler, 1993), which contain PAR binding sites, are activated by DBP and C/EBPβ in cotransfection experiments, although higher concentrations of C/EBPβ than DBP are required for these purposes in nuclei of transfected cells. Although some DNA elements are recognized by both C/EBPβ and DBP, DNase I protection experiments with several promoter sequences revealed that DBP (and other PAR proteins) only recognize a subset of sites bound by C/EBPβ (and other C/EBP proteins).

In order to elucidate the DNA-binding site selectivity of PAR and C/EBP proteins, we have used a PCR-based selection protocol to define optimal C/EBPβ and DBP target sites, and have investigated the details of binding characteristics of recombinant proteins to these sites. We found that DBP and C/EBPβ preferentially bind to highly similar sequences, with the optimal binding site for both proteins being 5′RTTAY.GTAAY3′. C/EBPβ, however, tolerates more deviations from this consensus sequence and as a result binds many more sites than DBP on random DNA. For example, unlike DBP, C/EBPβ binds sequences with a T or a C at position +2 from the center of half-site symmetry equally well (i.e., the sequences RTTAY.GCAAY, RTTGY.GCAAY, and RTTGY.GTAAY). A mutation that changes a valine residue to an alanine in the basic region of C/EBPβ increases its specificity by allowing it to selectively bind DBP target sites. Interestingly, the alanine at this position is the site of a critical contact to the thymidine at position +1 in the co-crystal structures of GCN4 and cFos-cJun.

Results

C/EBPβ and DBP Select Similar Binding Sequences

C/EBPβ, a member of the C/EBP family of transcription factors, and DBP, a member of the PAR family, were both originally isolated by their ability to bind to the D-site of the albumin promoter (Descombes *et al.*, 1990, Mueller *et al.*, 1990). These proteins have since been found to recognize similar sites in other potential target genes, such as the FP2 site of the C7αH promoter (Lavery and Schibler, 1993). In order to determine the optimal binding sites for these two proteins, we used a selective PCR-amplification and binding assay, similar to the one described by

C/EBPβ SELECTION

	Central dinucleotide	GATCTTGGTNNNNNNNNNNNNNNNTGTAAGATCT	Site Affinity C/EBPβ	DBP
1.	TG	GATCTTGGTTATGTAATGTTATACCTGTAAGATCT	+++	+++
2.		GATCTTGGTTATGTAATGTGTGGTATGTAAGATCT	+++	+++
3.		GATCTTGGTTATGTAATATGTCTAGTGTAAGATCT	+++	+++
4.		GATCTTGGTTATGTAATGTCCGTACTGTAAGATCT	+++	+++
5.		GATCTTGGTTATGTAATCTGTCTAGTGTAAGATCT	+++	+++
6.		GATCTTGGTTATGTAATTTCCGATGTGTAAGATCT	+++	+++
7.		GATCTTGGTTATGTAATATGTCTAGTGTAAGATCT	+++	+++
8.		GATCTTGGTTATGTAATACGCGGGGTGTAAGATCT	++	+++
9.		GATCTTGGTTATGTAATCTTGTTTGTGTAAGATCT	+++	++
10.		GATCTTGGTTATGTAACTTGCACTATGTAAGATCT	+++	+++
11.		GATCTTGGTTATGTAACATTGTATGTGTAAGATCT	+++	+
12.	CA	GATCTTGGTTACATAATCCCTCCTGTGTAAGATCT	+++	+++
13.		GATCTTGGTTACATAATTGCGACGGTGTAAGATCT	+++	+++
14.		GATCTTGGTTACATAATCCCCAGGGTGTAAGATCT	+++	+++
15.		GATCTTGGTTACATAATATCCTCCGTGTAAGATCT	+++	+++
16.		GATCTTGGTTACATAATCATCTCCATGTAAGATCT	+++	+++
17.		GATCTTGGTTACATAATCCCAACTATGTAAGATCT	+++	+++
18.		GATCTTGGTTACATAATCTCCCCTATGTAAGATCT	+++	+++
19.		GATCTTGGTTACATAACCTATACTGTGTAAGATCT	+++	+++
20.		GATCTTGGTTACATAACACTAGGGGTGTAAGATCT	++	+++
21.		GATCTTGGTGACATAACAATGCCCGTGTAAGATCT	+	++
22.		GATCTTGGTTACAAAATTAGCGCGCTGTAAGATCT	++	+/-
23.		GATCTTGGTGTACATAATCTTTGTATGTAAGATCT	+/-	+/-
24.		GATCTTGGTGTACATAATCTTTGTATGTAAGATCT	+	-
25.		GATCTTGGTGTACATAATTTCGCCCTGTAAGATCT	+/-	-
26.		GATCTTGGTGTACATAATCTTTGTATGTAAGATCT	+	-
27.		GATCTTGGTGTACATAACATTCATATGTAAGATCT	+/-	-
28.		GATCTTGGTATACATAACCCCCATATGTAAGATCT	+/-	+
29.		GATCTTGGTATACATGGCCCCCTCTGTGTAAGATCT	+/-	+/-
30.	CG	GATCTTGGTTACGTAATCTCCAATGTGTAAGATCT	++++	++++
31.		GATCTTGGTTACGTAATCCTCAGTATGTAAGATCT	+++	+++
32.	AC	GATCTTGGTTACTAACCTTCCTGTGTAAGATCT	+++	++
33.	TG	GATCTTGGTGTACACAACTCATCTATGTAAGATCT	++	-
34	TG/TG	GATCTTGGTTATGTAATTTGTGTTATGTAAGATCT	+++	+++
35.		GATCTTGGTTATGTAATCTCCATTATGTAAGATCT	+++	+++
36.		GATCTTGGTTATGTAATCACTATTATGTAAGATCT	+++	+++
37.		GATCTTGGTTATGTAATCTGTATTATGTAAGATCT	+++	+++
38.		GATCTTGGTTATGTAATCGCCATTATGTAAGATCT	+++	+++
39.		GATCTTGGTTATGTAATGCTGCTTGTGTAAGATCT	+++	+++
40.		GATCTTGGTTATGTAATCTTGTTTGTGTAAGATCT	+++	+++
41.		GATCTTGGTTATGTAAGTGTGGTTATGTAAGATCT	+++	+++
42.		GATCTTGGTTATGTAAGTGTGGTTATGTAAGATCT	+++	+++
43.		GATCTTGGTTATGTAATTGTTATGCTGTAAGATCT	+++	+++ (w)
44.		GATCTTGGTTATGTAACTTTGCCTGTGTAAGATCT	+++	+++ (w)
45.		GATCTTGGTTATGTAATCTGTATTATGTAAGATCT	+++	++
46.		GATCTTGGTTATGTAATCGCCATTATGTAAGATCT	+++	++
47.		GATCTTGGTTATGTATATTTCATTGTGTAAGATCT	+++	++
48.	CA/TG	GATCTTGGTTACACAATCTCTCTTATGTAAGATCT	+++	+++
49.		GATCTTGGTTACACAATCCTCTATATGTAAGATCT	+++	+++ (w)
50.		GATCTTGGTTACACAATCCCTGGGTATGTAAGATCT	+++	+++ (w)
51.		GATCTTGGTTACACAATCTTTTCTATGTAAGATCT	+++	+++ (w)
52.		GATCTTGGTATACATAACCCC CATATGTAAGATCT	+	-
53.		GATCTTGGTGTACATAACATT CATATGTAAGATCT	++	++
54.	CG/TG	GATCTTGGTTACGTAATTCTAGTGCTGTAAGATCT	+++	+++ (w)

Left-hand	G	T	T	A	C/T	A/G	T	A	A	T
A	0	0	0	42	0	19	1	42	40	1
C	0	0	0	0	21	0	4	0	1	6
G	42	0	0	0	0	22	0	0	0	2
T	0	42	42	0	21	1	37	0	1	33
Right-hand	N	T	T	A	T	G	T	A	A	T
A	8	1	0	13	0	0	0	19	19	0
C	3	2	0	2	0	0	0	0	0	0
G	5	1	2	4	0	19	0	0	0	0
T	3	15	17	0	19	0	19	0	0	19
CONSENSUS	G	T	T	A	C/T	A/G	T	A	A	T

Fig. 1 Alignment of Cloned DNA-Binding Sites Selected by C/EBPβ and the Consensus Sequence Derived from 54 Selected Clones.

Sites are grouped according to the sequence of the dinucleotide at the center of dyad symmetry (positions -1 and +1). Approximate relative binding affinities for each site with C/EBPβ and DBP were determined by EMSA (see Materials and Methods) and are shown on the right. A consensus binding site for C/EBPβ is deduced at the *bottom*. TA was been found at the center of dyad symmetry. Nucleotides of the left- and right-hand invariable oligonucleotide regions are indicated by asterices (*). (w) denotes the presence of a secondary binding site [indicated by (2)] with low affinity for DBP that was only observed with the higher of the two examined protein concentrations (50 nM).

Blackwell and Weintraub (1990) A 46-mer was synthesized containing 16 degenerate nucleotides flanked by a constant region to serve as a primer template for PCR amplification. Following ten rounds of amplification and C/EBPβ-binding selection (see Materials and Methods), a significant fraction of the PCR products formed complexes with recombinant C/EBPβ. The amplified DNA fragments thus obtained were cloned, and individual recombinant plasmids were isolated and their inserts sequenced. 54 distinct inserts were analyzed for binding affinity to recombinant C/EBPβ by EMSA, and the results are shown in Figure 1, with a C/EBPβ consensus sequence, -5G-4T-3T-2A-1Y + 1G + 2T + 3A + 4A + 5T. In the large majority of the selected sequences, the first two nucleotides (-5G and -4T) are derived from the invariable left-hand flank of the oligonucleotide. As shown below, oligonucleotides with an A at position -5 bind C/EBPβ and DBP equally well. We thus propose RTTA(C/T)(A/G)TAAY as a consensus sequence for C/EBP and PAR proteins (bottom of Figure 1; since AT was never observed as a central dinucleotide, we did not designate the central dinucleotides as YR).

To determine the consensus DBP binding site independently, a second binding site selection was performed. Eight rounds of PCR amplification and EMSA selection were performed using DBP recombinant protein to select for PCR-fragment binding in gel mobility-shift assays. The compilation of 43 DBP-selected binding sites is shown in Figure 2A. Every selected binding site employed the same position in the right-hand flank of the 46-mer to generate a DBP binding site, apparently because this sequence already contains four out of five nucleotides of a half-site. Purely based on the sequence comparison, we derived the consensus sequence -7Y-6Y-5R-4T-3T-2A-1T + 1G + 2T + 3A + 4A + 5G, where Y stands for pyrimidine and R stands for purine. While pyrimidines were somewhat more abundant than purines at positions -7 and -6, several strong binding sites contained a purine at one of these positions. We thus suspect that these nucleotides are not part of the core binding site. Moreover, as shown above (Figure 1) and below (Figure 2B) the G at +5 deviates from the selected DBP half-site 5′RTTAY3′. We thus propose 5′RTTATGTAAY3′ (bold letters in consensus sequence at the bottom of Figure 2) as the DBP core consensus sequence. This sequence is nearly identical to the optimal binding sites found for HLF, 5′GTTAC.GTAAT3′ (Hunger *et al.*, 1992; 1994; Inaba *et al.*, 1994) and TEF/VBP, 5′RTTAC.GTAAY3′ (Haas *et al.*, 1995), suggesting that all

A

DBP SELECTION

GATCTTGGTNNNNNNNNNNNNNNNNTGTAAGATC	Site Affinity DBP	C/EBPβ
GATCTTGGTGGTTTC_TATGTGTTATGTAAGATC	+++	+++
GATCTTGGTGTCTTGAATGTCGTTATGTAAGATC	+++	+++
GATCTTGGTGTCCATACTACTGTTATGTAAGATC	+++	++
GATCTTGGTGGGTGACGCCTGGTTATGTAAGATC	+++	++
GATCTTGGTGTTCAATAAGTCGTTATGTAAGATC	+++	+++
GATCTTGGTGGTACTATTTTTGTTATGTAAGATC	+++	++
GATCTTGGTGGGAGGCGGTTCGTTATGTAAGATC	++	+
GATCTTGGTGTCTGTGTTTCCGTTATGTAAGATC	++	+
GATCTTGGTGGGATTGGTCCTGTTATGTAAGATC	+++	+++
GATCTTGGTGTTCGGTCACTCGTTATGTAAGATC	+++	+++
GATCTTGGTGGTCACTATCCTGTTATGTAAGATC	++	+++
GATCTTGGTGGCCAATTCATCGTTATGTAAGATC	+	+
GATCTTGGTGGTAAGCATGCCATTATGTAAGATC	+++	+++
GATCTTGGTGGGGGGAATTCCATTATGTAAGATC	+++	++
GATCTTGGTGGTTCGGGGGCCATTATGTAAGATC	+++	++
GATCTTGGTGGGGGCCCTATTATTATGTAAGATC	+++	++
GATCTTGGTGGTACGGTTTGTATTATGTAAGATC	+++	+++
GATCTTGGTGGTTTA_CCTCTATTATGTAAGATC	++	+++
GATCTTGGTGGACGGGTGGGCATTATGTAAGATC	++	+
GATCTTGGTGGGATG GATCGCTTATGTAAGATC	+++	+++
GATCTTGGTGGTGTAATACTAATTATGTAAGATC	++(2)	++(2)
GATCTTGGTGTCACTGTAATCGTTATGTAAGATC	++(2)	+++(2)
GATCTTGGTGTATGCACTGCCGTTATGTAAGATC	+++	+++(2)
GATCTTGGTGGTCTGCGTGGCGTTATGTAAGATC	+++	+++(2)
GATCTTGGTGTTCTGGACTAGGTTATGTAAGATC	+++	+++(2w)
GATCTTGGTGAATTCCTGCCTGTTATGTAAGATC	+++	+++(2w)
GATCTTGGTGTTTGGAGTTTAGTTATGTAAGATC	+++	+++(2w)
GATCTTGGTGGACTACAATCCGTTATGTAAGATC	+++	+++(2w)
GATCTTGGTGATACGCCGGTCGTTATGTAAGATC	+++	+++(2)
GATCTTGGTGGTCATAGAGCTGTTATGTAGATTC	+++	+++(2)
GATCTTGGTGTTGTGGACACTGTTATGTAAGATC	+++	+++(2)
GATCTTGGTGTCTGAACGTGTGTTATGTAAGATC	++	+++(2)
GATCTTGGTGGAATTATTACTGTTATGTAAGATC	++	++(2)
GATCTTGGTGTGTCCGTCTAGGTTATGTAAGATC	++	++
GATCTTGGTGTGTGTTGATTAGTTATGTAAGATC	++	++
GATCTTGGTGTAACTAGCACCATTATGTAAGATC	++	+(2)
GATC GTGGGTAAGGCCGCATTATGTAAGATC	+++	++(2w)
GATCTTGGTGGTCTCTATGCTATTATGTAAGATC	+++	+++(2)
GATCTTGGTGGGTTGTGTTCTATTATGTAAGATC	+++	+++(2w)
GATCTTGGTGGTCACAATCGCATTATGTAAGATC	+++	+++(2)
GATCTTGGTGGGTA GTGTCATTATGTAAGATC	++	++
GATCTTGGTGCCCCATGTAACAGTATGTAAGATC	++	++(2)
GATCTTGGTGTCTATTATCCGCTTATGTAAGATC	+++	+++(2)

	t	n	Y	Y	R	T	T	A	T	G	T	A	A	G	
G	7	11	7	5	26	1	0	0	0	43	0	0	1	42	G
A	9	7	3	3	15	0	0	43	0	0	0	43	42	1	A
T	20	15	13	16	0	42	43	0	43	0	43	0	0	0	T
C	7	10	20	19	2	0	0	0	0	0	0	0	0	0	C

	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	
Consensus	Y	Y	R	T	T	A	T	G	T	A	A	G	match
CYP2c6	C	A	A	T	T	A	T	G	C	A	A	T	9/10
alb D-site	T	G	A	T	T	T	T	G	T	A	A	T	9/10
C7αH	A	T	G	T	T	A	T	G	T	C	A	G	8/10

B

	A N N N T T A C / T A A N N N C A C C A A G	relative affinity
1.	TA CATTAC GTAAC CACACCAA	6
2.	AA GGTTAC GTAAC CACACCAA	6
4.	AG CATTAC GTAAC A GCACCAA	6
3.	ATAATTAC GTAATA GCACCAA	6
5.	AA GATTAC GTAACA CCACCAA	6
6.	AG C GTTAC ATAAC G TCACCAA	5
7.	AA GATTAC ATAATA CCACCAA	5
8.	AAA GTTAC ATAAT C GCACCAA	5
9.	AC T GTTAC ATAAC GG ACCAA	5
10.	AG G C TTAC ATAAT C G ACCAA	4
11.	AA C C TTAC GTAATA CCACCAA	4
12.	AG G C TTAC ATAAG CACACCAA	3
13.	AAA C TTAC GTAAA C C CACCAA	3
14.	AGA TTTAC ATAAA GG ACCAA	1
15.	ATAA TAC GTAAA GG ACCAA	1
16.	AA GATTA TAAA TACACCAA	0
17.	AGAATTA GTAAA GGCACCAA	0

C G
RTTA > TAAY
C A

Fig. 2 Binding of DBP to Selected DNA Sequences.
(A) Alignment of DNA-binding sites selected by DBP. A consensus sequence was derived from 43 selected clones. Since all sites selected a sequence within the right-hand invariable flank, these residues were considered significant and not neglected in determining consensus site. Relative binding affinities for DBP and C/EBPβ are shown on the right. As all high affinity sites contain the half-site RTTA(T) (see also Figures 1 and 2B), we propose the DBP consensus sequence RTTATGTAAY, although the invariable flank contains a G at position (+5). Several oligonucleotides contain two (2) C/EBP binding sites, two of which also bind two DBP dimers. The putative secondary DBP binding sites within these two oligonucleotides are written in italic fonts. Weak secondary C/EBP binding sites are indicated by (w). In these cases, a double shift could only be detected at the higher of the two protein concentrations (50 nM). DBP binding sites from the promoters of the cytochrome P_{450} CYP2C6 gene, the rat albumin gene, and the cholesterol hydroxylase (C7αH) gene are aligned underneath the consensus sequence. N, randomly mixed bases.
(B) 17 sites chosen from a pool of oligonucleotides containing either a CA or a CG central dinucleotide and mixed randomly at positions -7, -6, -5 and +5, +6, and +7. Relative affinities for DBP were scored using a scale of 1 to 6. N, randomly mixed base.

three PAR proteins bind analogous cognate recognition sequences. As shown at the bottom of Figure 2, three known DBP recognition sites, located in the promoters of the genes encoding albumin (Mueller *et al.*, 1990), the cytochrome p450, CYP2C6 (Yano *et al.*, 1992), and cholesterol 7α hydroxylase (Lavery and Schibler, 1994) are all permutations of the derived consensus core sequence. Although the same consensus sequence was selected by C/EBPβ and DBP, we noticed a striking difference between the outcomes of the two experiments. Sequences on the left side of the random 46-mer were used to generate most of the C/EBPβ sites, while, as mentioned above, the right-hand flank of the 46-mer was always selected by DBP. While several inserts contained two binding sites, only in one case did C/EBPβ select a single site from the right-hand flank. The different evolution of binding site selection by DBP and C/EBPβ is likely to reflect a more restricted sequence specificity by the former as compared to the latter protein. During the initial rounds of amplification, DBP may have selected only sites with a large number of matches with its optimal recognition sequence. Thus it used exclusively oligonucleotides with matches in addition to the five nucleotide match already present in the invariable right-hand flank. In contrast C/EBPβ may have bound to many different sequences adjacent to the left-hand flank in the first few selection rounds that were adapted to the optimal sequence only during later stages of amplification. Each of the above DBP binding sites was also analyzed for its affinity to C/EBPβ, to compare the complexes formed with recombinant DBP or C/EBPβ (see right-hand columns in Figure 2). In general, sites that bound DBP with high affinities were found to have similarly high affinities for C/EBPβ. However, many more sites resulted in complexes containing two protein dimers with recombinant C/EBPβ protein than with recombinant DBP (13 as compared to 2), again suggesting that there may be more sequence constraints on DBP-site recognition than on C/EBPβ-site recognition. In order to further analyze sequences that might be important for DBP recognition, an oligonucleotide with the sequence 5′ANNNTTAC(A/G)-TAANNNCA3′ at its core was used, containing either A or G at position 1 and random nucleotides at positions -7, -6, -5, +5, +6, and +7. A hierarchy of selected binding sites was determined by testing a set of these sites directly by EMSAs, and comparing their relative affinities, using a scale of 1 to 6. The results are shown in Figure 2B. An optimal binding site, 5′RTTAC.GTAAY3′, was derived by comparison of the highest affinity sites. Both DBP and C/EBPβ (data not shown) bind this site more avidly than any other site tested. This site differs from the consensus site derived in Figures 1 and 2A by containing a CG at the central dinucleotide. Curiously, this site never appeared in the DBP selection, and was found only twice in the C/EBPβ selection. A possible explanation for this apparent discrepancy can be seen upon inspection of the relative binding affinities in Figure 2B. Although 5′RTTAC.GTAAY3′ was indeed the highest affinity site, the purine at position -5 and the pyrimidine at position +5 were more important than the sequence of the central dinucleotide for recognition of poor sites, i.e., sites that deviate widely from the consensus, as can be seen in comparing the sequences of the lower-affinity binding sites. In these sites, the DBP-binding affinity to sites containing either a CG or a CA at the central positions (-1,+1) is similar. Thus, the CA may have been selected in the very early rounds of the selection-amplification since it is at least as strong (and perhaps even stronger) than the central CG in the context of imperfect sites. Conceivably, the very stringent selection in the initial rounds has resulted in an increased number of selected sites with CA at the central position of symmetry.

Comparison of DBP and C/EBPβ Binding to the Optimal Recognition Site

Since DBP and C/EBPβ appear to select similar optimal sites, we reasoned that a difference in binding affinity of one factor for its cognate site may gain an advantage for the more avidly binding factor in a competition for a common recognition site. We therefore compared the binding affinities of DBP and C/EBPβ for the optimal binding site, 5′GTTAC.GTAAT3′ by using saturation binding assays in which a constant amount of protein was titrated with increasing amounts of labeled oligonucleotide. After incubation with recombinant protein at 25°C for 30 min, the free oligonucleotide was separated from protein-bound DNA complexes on polyacrylamide gels. After autoradiography to position the bound and free forms, the corresponding gel fragments were isolated and measured for radioactivity using a scintillation counter. The dissociation constants were determined by Scatchard plot analysis (Scatchard, 1949), as shown in Figure 3 A, to be 1.3×10^{-9} M and 6×10^{-10} M for DBP and C/EBPβ, respectively.

Scatchard plot analysis also allows determination of the total concentration of active protein dimers in the reaction (X-axis intercept). As seen in Figure 3A, these values are about 0.34×10^{-9} M and 0.39×10^{-9} M for DBP and C/EBPβ, respectively. If all the protein monomers (3×10^{-9} M) added to the reaction dimerized, the concentration of dimers in the binding assay would be 1.5×10^{-9} M. Thus, the fraction of active DBP and C/EBPβ dimers can be estimated to be about 23% and 26%, respectively, of the input protein.

The affinities for the optimal binding site of DBP and C/EBPβ were also compared to those of the two other PAR family members, TEF and HLF, by protein saturation analysis (Figure 3B). Using this method, increasing amounts of the recombinant proteins C/EBPβ, DBP, HLF, and TEF were added to a constant amount of labeled oligonucleotide, and the bound and free DNAs were quantified as described above. When protein dimers are in excess over DNA, the equilibrium dissociation constant is the concentration of protein dimers required to bind half of the input DNA (see Fonjallaz *et al.*, 1996). These protein dimer concentrations are estimated to be 7×10^{-9} M for C/EBPβ, 6.5×10^{-9} M for DBP, 9×10^{-9} M for HLF, and 10×10^{-9} M for TEF

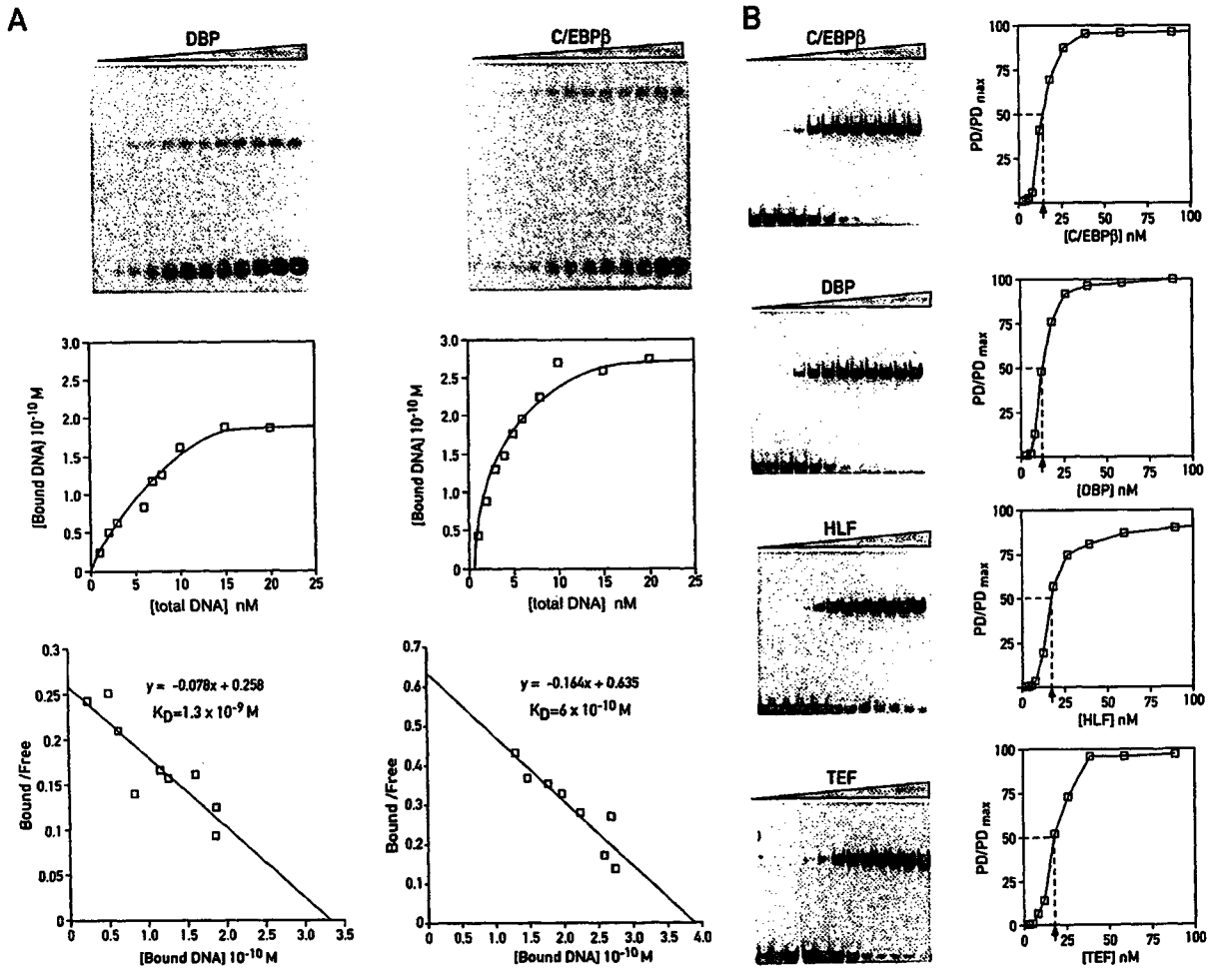


Fig. 3 Relative Affinities of C/EBPβ and the Three PAR Proteins, DBP, HLF, and TEF, for an Optimal Binding Site.
(A) Measurement of the relative DNA-binding affinities of DBP and C/EBPβ for an oligonucleotide containing an optimal DBP-binding site, 5′GTTAC.GTAAT3′, by Scatchard plot analysis. A constant amount of purified protein was added to increasing amounts of radioactively labeled probe, from 0 – 25 nM. Dissociation constants for DBP-DNA and C/EBPb-DNA complexes were determined by quantitative analysis of EMSAs, shown in the top panel. Autoradiographs of wet gels show the positions of bound and free DNA. Bound and free forms were excised from the gel and their radioactivity was quantitated by scintillation counting. The concentration of bound DNA was plotted against total input DNA to present the saturation binding profile of each reaction, center. These profiles were analyzed by Scatchard plots (Scatchard, 1949), below, to determine the dissociation constant, K_D, of each factor, whereby K_D = -1/slope of this plot.
(B). Protein saturation experiments to compare affinities of C/EBPβ and the three PAR proteins, DBP, HLF, and TEF. An increasing amount of purified protein was added to a constant amount of radio-labeled probe. Autoradiographs at left of gels that were analyzed as in(A). On the right, the ratio of protein-DNA complex (PD) in each reaction relative to the maximum protein-DNA complex (PD_{max}), expressed as a percentage, was plotted as a function of purified protein concentration. The concentration of protein at 1/2 maximum (as indicated by the dotted lines) gives an indication of the relative binding affinities of each transcription factor (see Results).

(see Figure 3B, note however that the protein monomer concentration is plotted in these graphs). Thus, this method yielded similar relative affinities for the four examined proteins. Given the high conservation of amino acids that are thought to establish base contacts within these four proteins (see Lavery and Schibler, 1994), this result is not surprising.

When corrected for the fraction of active protein dimers, the K_D values obtained for DBP in the binding assays with excess protein can be estimated to be 1.4×10^{-9} M ($0.23 \times 6.5 \times 10^{-9}$ M, see above). This is in excellent agreement with the 1.3×10^{-9} M determined in the binding assays with excess DNA (Figure 3A). For C/EBPβ, the corrected K_D value of 1.8×10^{-9} M ($0.26 \times 7 \times 10^{-9}$ M, see above) obtained in binding assays with excess protein is three times higher than the one estimated by Scatchard plotting.

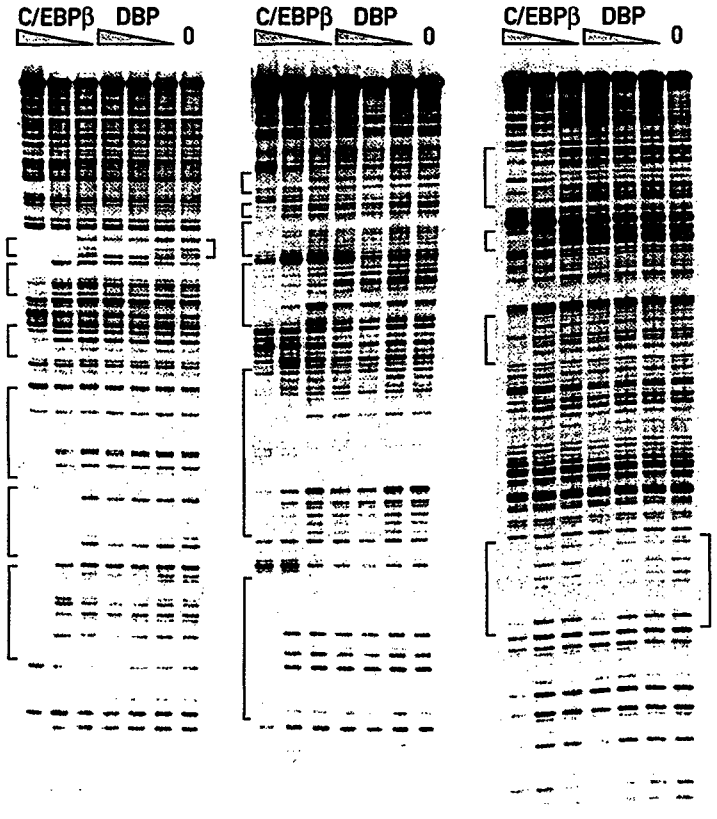
C/EBPβ Sites Occur Much More Frequently than DBP Sites in the Rat Genome

On the basis of quantitative immunoblot experiments (Descombes and Schibler, 1991; Fonjallaz *et al.*, 1996) we estimate that there are approximately 10^5 molecules of both DBP and C/EBPβ in homo- and heterodimers in the

8 p.m. rat liver nucleus. However, the effective concentration of a transcription factor, that is, its concentration required to occupy a particular cis-acting element, not only depends on its absolute concentration in the nucleus, but also on the abundance of all accessible binding sites within the genome.

To estimate the abundance of potential DBP and C/EBPβ binding sites in the rat genome, we cloned random genomic fragments approximately 500-700 bps in length, and identified DBP and C/EBPβ recognition sites on inserts from this 'mini-library' by DNase I footprinting. Ten random clones were selected and analyzed; three representative footprints are shown in Figure 4 A. The FP2 site of the C7αH promoter was used as a footprinting standard (data not shown): the amount of recombinant DBP or C/EBPβ required to achieve 50% protection of the C7αH FP2 site was defined as 1 footprinting unit (1 FU). The results from the 10 footprinting experiments are compiled in Figure 4 B. With one FU, a concentration which is likely to be physiologically relevant (Descombes *et al.*, 1990; Lavery and Schibler, 1993; Mueller *et al.*, 1990), a single DBP binding site and 24 C/EBPβ sites were found in 5000 bp. A threefold higher protein concentration (3 FU) yielded 3 and 32 footprints per 5000 bp for DBP and C/EBPβ, respectively. Thus, C/EBP binding sites with affinities close to those measured for cis-acting promoter elements appear to be approximately 10- to 20-fold more abundant than DBP binding sites in genomic DNA. If the randomly selected DNA fragments were representative for the entire genome, the diploid genome would contain approx. 1.2 to 3×10^6 DBP binding sites versus 3×10^7 C/EBPβ binding sites. Even if the numerous additional low-affinity binding sites that escape the footprinting analysis are neglected, the ratio of transcription factor dimers (about 5×10^4 for each DBP and C/EBPβ) per high-affinity binding site would be approximately 1/50 for DBP and 1/1200 for C/EBPβ. Although many of these binding sites may be inaccessible in native chromatin, the ability of C/EBPβ (and other C/EBP isoforms) to recognize a broader range of DNA sites could significantly reduce its effective concentration, and hence its ability to compete for the same binding site with DBP (and other PAR proteins).

A



B

	DBP			C/EBPβ		
	0.3	1.0	3.0	0.3	1.0	3.0
1	1	1	1	1	2	2
2	0	0	1	0	6	6
3	0	0	0	0	3	4
6	0	0	1	0	4	6
11	0	0	0	0	0	1
14	0	0	0	0	3	3
15	0	0	0	0	0	1
16	0	0	0	0	0	1
17	0	0	0	0	2	3
21	0	0	0	1	4	5
total	1	1	3	2	24	32
alb D-site	0	1	1	4	4	7
C7αH FP2	0	1	1	0	2	3

Fig. 4 DNase I Footprint Analysis of Random Rat Genomic Clones.
(A) Autoradiographs of three representative random footprints. End-labeled genomic fragments approximately 500 – 700 bps in length were incubated with 0, 0.3, 1.0, or 3.0 footprint units (as defined in text) of recombinant C/EBPβ or DBP and analyzed by DNase I protection. Observed C/EBPβ footprints are bracketed on the left and DBP footprints are bracketed on the right..
(B) Quantitative analysis of DNase I protection data from ten footprinting experiments on randomly selected clones to estimate the number of binding sites for C/EBPβ and DBP in the rat genome.

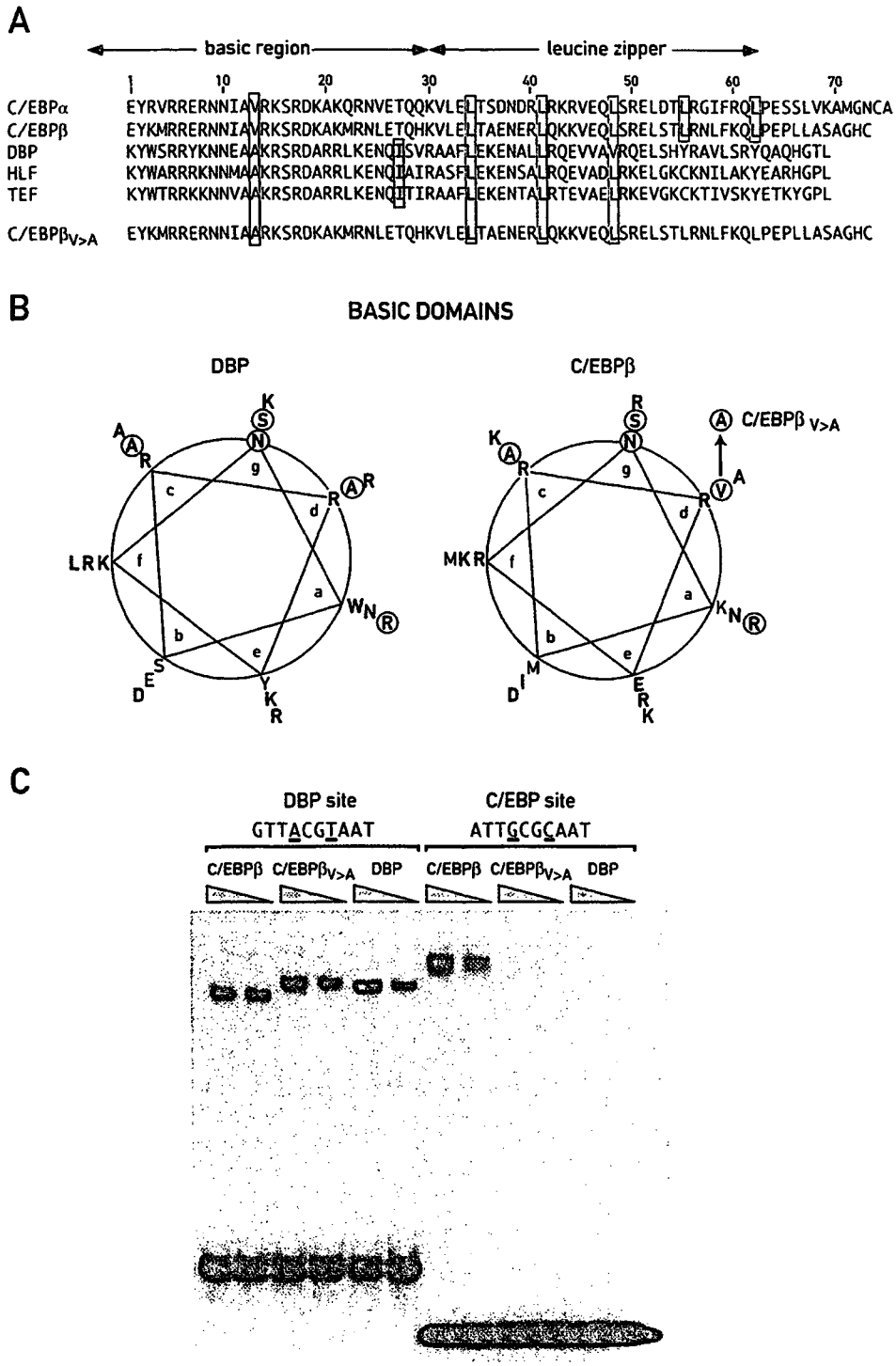


Fig. 5 Comparison of the DNA Binding Specificities of C/EBPβ, C/EBBβ$_{V>A}$, and DBP.
(A) The bZip regions of C/EBP family, C/EBPα and β, and PAR family proteins, DBP, HLF and TEF are aligned through their leucine zippers, in shaded boxes, to show homologies in the binding and dimerization domains. The valine residue in the binding domain of the C/EBP family that was changed to an alanine, is shown within the open box.
(B) A helical projection of amino acids in the α-helical DNA-binding basic-region amino acids of DBP, representing the PAR family, left, and C/EBPβ, representing the C/EBP family, right. Conserved amino acids within each family are shown in bold. The position of amino acids that make base contacts in the Fos-Jun and GCN4 bZip co-crystal structure are circled. The valine-alanine substitution at position 13 of the basic region in C/EBPβ is shown at right.
(C) EMSA analysis of binding reactions containing either the consensus DBP or C/EBP oligonucleotide with each of the three recombinant proteins, DBP, C/EBPβ$_{V>A}$, and C/EBPβ. Two concentrations of protein, 10 or 20 nM, were used for each set.

C/EBPβ Mutant Ala-216 Has an Increased Specificity for DBP-Binding Sites in the Albumin and C7αH Promoters

We compared the basic regions of the PAR and C/EBP families to try to identify residues that might account for the difference in their binding site specificity. The amino acid sequences of bZip domains of two C/EBP isoforms (C/EBPα and C/EBPβ) and three PAR proteins (DBP, TEF, HLF) are aligned in Figure 5A. Panel B depicts a helical wheel projection of the a-helical basic region amino acids of DBP, representing the PAR family, and C/EBPβ, representing the C/EBP family. One interesting candidate responsible for the broader sequence specificity of C/EBP proteins is val-216 (see Descombes and Schibler, 1991 for numbering of amino acids), located at position 13 of the basic region of C/EBPβ. Valine is conserved at this position within the C/EBP family, but members of the PAR family (as well as all other characterized bZip proteins) always have an alanine in this position. This alanine is also implicated as a specificity determinant both in genetic (Johnson, 1993, Kim *et al.*, 1993; Pu and Struhl, 1991; Suckow *et al.*, 1993a, b) and structural studies (Ellenberger *et al.*, 1992, Glover and Harrison, 1995; Kerppola and Curran, 1995; Kim *et al.*, 1993).

Using site-directed mutagenesis, we converted the val-216 residue of C/EBPβ to an alanine, and the recombinant protein bearing this mutation was tested in binding assays with either the optimal DBP binding site, 5′GTTAC.GTAAT3′, or the previously proposed C/EBP site, 5′ATTGC.GCAAT3′ (Agre *et al.*, 1989; Shuman *et al.*, 1990). As shown in Figure 5C, all three proteins, DBP, C/EBPβ, and C/EBP$\beta_{V>A}$, formed complexes with the DBP site oligonucleotide. Using the same concentration of proteins, however, only C/EBP, but not DBP or C/EBP$\beta_{V>A}$, formed stable complexes with the C/EBP site 5′ATTGC.GCAAT3′ oligonucleotide. Thus, we conclude that the substitution of an alanine for a valine at position 13 of the C/EBP basic domain interfered with C/EBP-site binding.

The rat albumin and C7αH promoters contain binding sites for both DBP and C/EBPβ. There are four C/EBPβ binding sites on the albumin promoter, one of which is also a DBP binding site, and transcription from this promoter was stimulated by both DBP and C/EBPβ in cotransfection experiments (Descombes *et al.*, 1990; Mueller *et al.*, 1990). Likewise, the C7αH promoter contains one DBP site and four C/EBPβ sites and was also shown to be stimulated by both proteins in transfection experiments (Lavery and Schibler, 1993; and data not shown). To assay

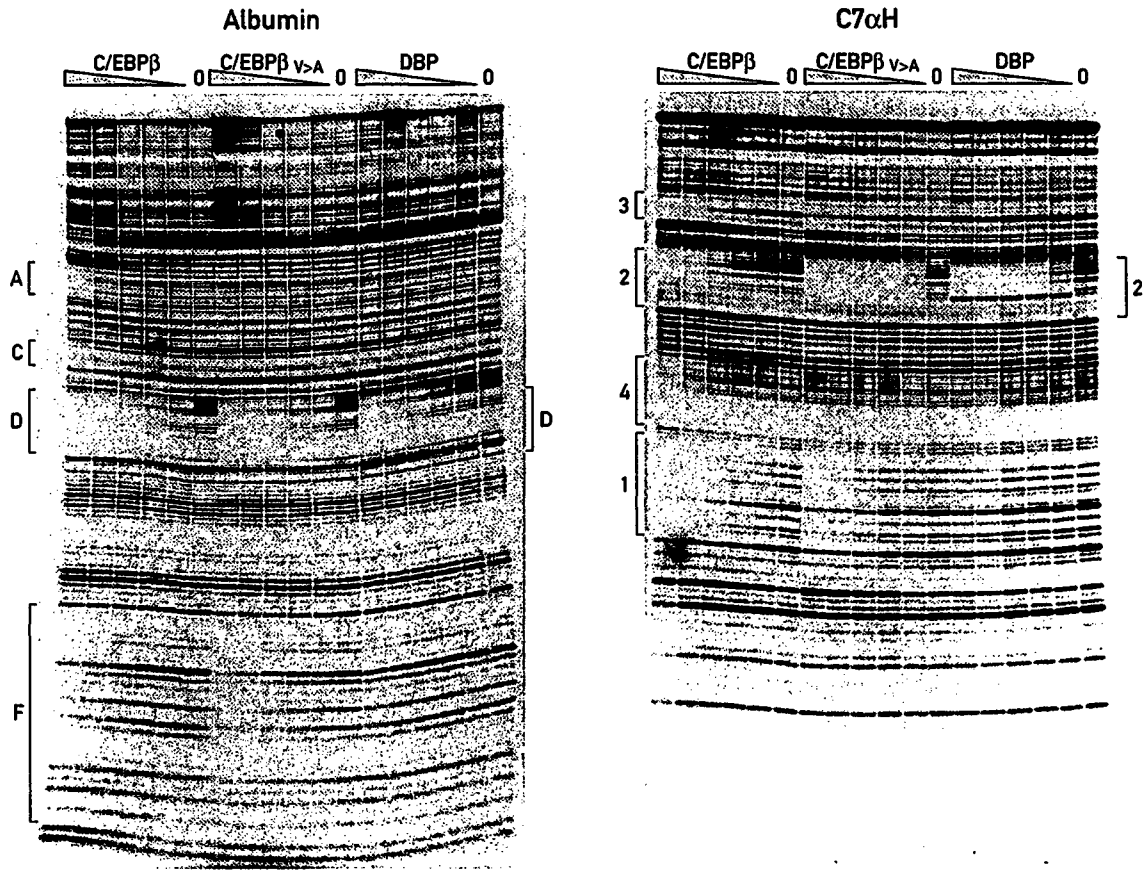


Fig. 6 DNase I Footprint Analysis of the Three Recombinant Proteins, DBP, C/EBP$\beta_{V>A}$, and C/EBPβ on the Albumin and the Cholesterol Hydroxylase C7αH) Promoters.
For each set, five steps of 2-fold dilutions and a control without protein is shown. C/EBPβ footprints are indicated by brackets on the left and DBP footprints are indicated on the right.

the affinity of the mutant and wild-type proteins to sites on the albumin and C7αH promoters, DNase I footprinting assays were performed using two-fold dilutions of recombinant proteins. As shown in Figure 6 (left panel), C/EBPβ initially protected the D site from DNase I digestion, followed by protection of sites A, C, and F, between 8 and 16 FUs (see definition, above). DBP protected mainly site D, with only very weak protection of site F at high protein concentrations (> 8 FUs). The mutant protein C/EBP$\beta_{V>A}$ initially protected site D, as well as site F between 8 and 16 FUs. However, C/EBP$\beta_{V>A}$, like DBP, did not protect sites A or C even at the highest protein concentration. Thus, the substitution mutant behaved more like DBP than C/EBPβ on the albumin promoter. A DNase I protection analysis performed on the C7αH promoter is shown on the right-hand panel of Figure 6. C/EBPβ protected sites FP1, FP2, and FP3 at 1 FU and an additional site, FP4, at 4 FUs. DBP, on the other hand, protected only FP2, even at the highest protein concentration, 16 FUs. C/EBP$\beta_{V>A}$, like DBP, initially protected only FP2, but protection of FP1 and FP2 is also seen at very high protein concentration (>16 FUs). Interestingly, a DBP-specific DNase I cleavage in FP2 is absent in the C/EBP$\beta_{V>A}$ footprint patterns. In summary, the valine to alanine mutant behaved more like DBP than C/EBP on both the albumin and C7αH promoters, although at high protein concentrations a weakly intermediate phenotype on the C7αH promoter was noticed.

Discussion

C/EBP Proteins Have a Broader DNA Sequence Specificity than PAR bZip Proteins

We have identified the optimal DNA-binding site for DBP to be the 10 bp symmetrical sequence, 5′RTTAC.GTAAY3′. This sequence is nearly identical to the optimal binding sites determined for the PAR proteins HLF, 5′GTTAC.GTAAT3′ (Hunger *et al.*, 1994; Inaba *et al.*, 1994) and TEF/VBP, 5′RTTAC.GTAAY3′ (Haas *et al.*, 1995). As we have shown here, DBP, HLF, and TEF/VBP bind this consensus site with similar equilibrium dissociation constants *in vitro*. The same sequence, 5′RTTAC.GTAAY3′, was also the optimum high affinity binding site for C/EBPβ. C/EBP proteins, however, are much less sensitive to permutations of this site and are thus considerably less discriminatory than PAR proteins in their choice of cognate DNA sequences. In contrast to PAR proteins, for example, C/EBPβ avidly binds to sites with deviations from the consensus at the ± 2 position such as RTTGY.GCAAY and RTTAY.GCAAY and, as a result, C/EBPβ protects more sites in random DNA footprints. Thus, C/EBPβ recognition sequences with physiologically relevant affinities (K_D about 10^{-8} M) can be found on average every 200 to 300 base pairs in random rat genomic DNA fragments, while PAR binding sites with similar affinities are about 10- to 20-fold less abundant in bulk genomic DNA. We have demonstrated here that this difference in selectivity is primarily due to a single amino acid in the basic region of C/EBPβ. When valine-216 of C/EBPβ is changed to alanine, the resultant protein, C/EBP$b_{V>A}$, specifically recognizes only the subset of C/EBβ binding sites that are also DBP sites. Thus, alanine-216 plays an important role in conferring higher selectivity to PAR proteins and related factors.

Although in cotransfection experiments target promoters harboring sites bound by C/EBP and PAR proteins are often activated by both, this may not be the case under more physiological conditions when the concentrations of these effector proteins are limiting. It is noteworthy in this context, that transcription rates of both the albumin gene and the cholesterol 7α hydroxylase gene are circadian and follow the circadian accumulation of PAR proteins (Wuarin *et al.*, 1990; Lavery and Schibler, 1993). Thus, while both C/EBP and PAR bZip proteins can bind to the same cis-acting elements, the latter proteins are more likely candidates for transcriptional regulators of these two genes. At least in part, PAR proteins may be dominant over C/EBP proteins since the concentration of free PAR proteins is expected to be much higher than the one of C/EBP factors. This may be true even if C/EBP proteins accumulate to higher nuclear levels than PAR proteins. Obviously, specific promoter- and/or enhancer-dependent protein:protein interactions are another important determinant of establishing the occupancy at cis-acting DNA elements by a given transcription factor. In fact, even C/EBPα and C/EBPβ, which have an identical DNA-binding domain and accumulate to very similar concentrations in hepatocytes, are not functionally redundant. While mice homozygous for a C/EBPβ mutant allele are viable and fertile (Screpanti *et al.*, 1995), disruption of the C/EBPα gene results in a lethal hypoglycemia a few hours after birth (Wang *et al.*, 1995). This suggests that these two C/EBP isoforms do not regulate the same target genes *in vivo*, at least not with the same efficiency. Given the identical *in vitro* DNA-binding specificity and affinity of C/EBPα and C/EBPβ for their target sequences (Cao *et al.*, 1991), the discrimination of cis-acting promoter and enhancer elements by these two C/EBP isoforms is likely to involve cooperative interactions with other nearby transcription factors. It may be useful to recall in this context that small changes in binding energy (ΔG_o), for example those gained by cooperative protein:protein interactions, result in large changes in the affinities ($\Delta G_o = RT\ln Kd$, where R and T are the ideal gas constant and the temperature in °K, respectively). Due to the more promiscuous binding selectivity of C/EBP proteins as compared to PAR proteins, such cooperative interactions are likely to be more important for C/EBP proteins than for PAR proteins in the recognition of functionally relevant binding sites.

Comparison with the GCN4 and Fos-Jun Complexes

The recent structural analyses of GCN4 and Fos-Jun cocrystals with the AP-1 site indicate a single hydrophobic contact between the methyl group of the thymine residue at the ±1 position in the AP-1 site, ATGACTCAT and the alanine residue at position 13 of the basic region (Ellen-

berger *et al.*, 1992, Glover and Harrison, 1995; König and Richmond, 1993). Unlike three of the other four residues in the basic region that make contacts at the AP-1 site, this alanine-thymidine contact is not complicated by additional interactions with other bases or amino acids (see also Johnson, 1993). Interestingly, this alanine residue is highly conserved throughout the bZip family of proteins, but the equivalent position contains a valine in all C/EBP family members. By analogy to GCN4 and Fos-Jun, this C/EBP valine should contact the ±2 position of the C/EBP consensus site (the 5 bp AP-1 half-sites are overlapping and the central base pair of the 9 bp AP-1 site is designated as 0, whereas in the 10 bp symmetrical C/EBPβ site the central dinucleotide pair is designated as ±1. Thus, the ±1 position of the AP-1 site is equivalent to ±2 of the C/EBPβ site). Our results are consistent with an interaction between the valine residue of C/EBPβ and either C or T at ±2 of the C/EBPβ consensus site. Indeed, changing the valine of C/EBPβ to an alanine weakens the binding of the mutant protein, C/EBP$\beta_{V>A}$, to a consensus site with C at ±2, but does not alter its affinity to a DBP consensus site, containing T at ±2. This suggests that the bulkier valine residue might establish a hydrophobic contact to the methyl group of the thymine residue equivalent to the alanine-methyl group contact in the AP-1/GCN4 and Fos-Jun crystal structures. Furthermore, since C/EBPβ binds similarly well to sites with C at ± 2, additional interactions may compensate for the loss of the alanine-methyl group contact of the thymine residue. For example, the second methyl group of valine might serve to stabilize contacts to neighboring amino acids or nucleotides. The precise nature of these contacts must await further physical studies of C/EBP and PAR proteins with their binding sites.

Genetic Evidence for the Role of Alanine at Position 13 of the Basic Region of bZip Proteins

A number studies aimed at identifying specificity determinants of bZip proteins have investigated the role of the highly conserved alanine at position 13 (Johnson, 1993; Kim *et al.*, 1993; Pu and Struhl, 1991; Suckow *et al.*, 1993a, b). Kim and coworkers (1993) found that an alanine to valine substitution at position 13 of GCN4, Val-239, bound as well to AP-1 sites with an A at position ±1 as they do to the wild-type AP-1 site with a T at ±1. However, in addition to this effect at ±1, Val-239 displayed altered specificity at the ±2 positions of the AP-1 site, ATGACTCAT. They have suggested, in the case of GCN4, that the protein would require conformational adjustment to relieve the steric crowding of the invariant Arg-243 contact with the central base-pair at the AP-1 site. Thus, adaptability in the local conformation and positioning of the basic region may be an important aspect of sequence recognition by bZip proteins. Steric hindrance of the Arg-243 side chain that contacts the central base pair might be relieved in the C/EBPβ interaction with its recognition site, where the C/EBPβ site is symmetrical rather than overlapping at the central base pair and the ±3 position base pair is an A-T rather than a C-G.

Suckow and coworkers have tested the effect of mutants at the position 13 residue of a GCN4 peptide containing the bZip domain (1993b). A number of position 13 mutants were found to alter and broaden the specificity for position ±1 of GCN4 target sites. The GCN4 peptide containing a valine at basic region position 13, however, had similar binding properties to the wild type peptide. In this case, their pool of GCN4 binding sites with single base changes did not contain any potential C/EBP sites since the C/EBP consensus site, 5′ATTGC.GCAAT3′, varies from their GCN4 site, ATGAC.GTCAT, in two half-site positions. However, together with the results of Kim and coworkers discussed above, this is consistent with the notion that the A at ±3 in the C/EBPβ site, which is required for C/EBPβ recognition (Johnson, 1993, this study), may be important for sterically accommodating the bulky valine residue.

DNA-Binding Specificity of the PAR and C/EBP Protein Families

It has recently been demonstrated that the region between the basic region and the zipper, the so-called fork region, may contribute to altering the binding specificity of bZip proteins by altering the conformation of the basic motif as it interacts with the DNA recognition site (Johnson, 1993). Indeed, Haas and coworkers (1995) have found residues in this region of the PAR protein VBP to be important for its target site specificity. However, in the case of the C/EBP and PAR proteins, the fork region is unlikely to be important in the common specificity of these two classes of proteins. Although the PAR proteins possess a high conservation of amino acids throughout their extended basic regions, the adjoining fork regions, and their leucine zippers, the bZip domains of C/EBP proteins, on the other hand, share little similarity with the PAR proteins in the fork region and leucine zippers.

The amino acid residues involved in DNA target site selection for the PAR and C/EBP protein families probably reside in the amino acids residues shared between DBP and the mutant protein, C/EBP$\beta_{V>A}$, since these proteins recognize identical target sites. These common amino acids lie within the basic region, including the counterparts of the quintet of amino acids that make base contacts at the AP-1 site in the GCN4 and cFos-cJun crystal structures, as well as four neighboring residues. The fact that these five amino acids are well conserved among all bZip proteins, though many have different target sites, has made it difficult until now to understand the basis for sequence specificity in this class of proteins. As we have shown here, however, minor differences in the amino acid composition at these critical positions can indeed account for significant differences in target site spectrum.

Materials and Methods

PCR Mediated Binding Site Selection

A modification of the selective amplification and binding (SAAB) method (Blackwell and Weintraub, 1990) was used to determine

the C/EBPβ and DBP consensus sequences. A 46-mer oligonucleotide, 5′GATCGAGATCTTACA(N)$_{16}$ACCAAGATCTGGTGC3, containing random nucleotides at 16 positions, was made double-stranded by annealing it to a reverse primer, 5′GCACCAGATCTTGGT3′, complementary to the 3′ arm of the 46-mer, and filling with Klenow polymerase and nucleotides. The double-stranded product was labeled with [γ-^{32}P]ATP and kinase. 10 ng were used in binding reactions for 30 min at 37°C with diluted recombinant C/EBPβ protein, from 0.1 to 25 ng/ml, and free oligonucleotides were separated from protein-complexes by electromobility shift assays (EMSA), as described previously (Lichtsteiner *et al.*, 1987). The gel was exposed for several hours, at which time the C/EBPβ-complex could be seen in the lane with highest protein concentrations, but not in the lane with the lowest concentrations. An area from the lane containing the lowest C/EBPβ concentration (supposedly containing trace amounts of protein:DNA complexes) was excised from the gel and the DNA was eluted for 1 hour at 50 °C in TEN (10 mM Tris-HCl pH 7.5, 1mM EDTA, 100 mM NaCl). 1% of this product was used as a template in PCR reactions (100 μl) with 50 pmol forward primer, 5′GATCGAGATCTTACA3′, homologous to the 5′ arm of the 46-mer, and 50 pmol of reverse primer (above) in 10 mM TrisHCl pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl, 1 mg/ml BSA, 10% DMSO, 2.5 mM dNTP's. After 30 cycles of PCR amplification with an annealing temperature of 55 °C, the PCR product was purified by polyacrylamide gel electrophoresis and a further selection for C/EBPβ binding sites was carried out as described above. Following ten rounds of selection and PCR amplification, the amplified DNA product was cut with Bgl II, and cloned into either an M13 or pKS vector and individual isolates were sequenced. The selection for DBP sites was carried out with the same method, using the same degenerate oligonucleotides and primers, with two modifications: a higher concentration of recombinant DBP, approximately 1 ng/ml, was used for the binding selection, and eight rounds of PCR amplification and selection were performed prior to cloning and sequencing of the PCR product. Relative binding of the cloned sequences was evaluated by electrophoretic mobility shift assays. To this end, the recombinant plasmids were cleaved with Bgl II and the resulting fragments were radioactively labeled by filling in the ends with Klenow polymerase and [α-^{32}P]dATP, dTTP, dGTP, and dCTP. The labeled plasmid fragments (vector and insert, approx. 100000 cpm/assay) were directly used for electrophoretic mobility shift assays (EMSA) with two concentrations (10 nM and 50 nM) of DBP and C/EBPβ. The high molecular weight vector fragments migrate more slowly than the protein:DNA complexes and did thus not interfere with analysis of the EMSA (data not shown).

In order to further analyze the sequence requirements of DBP, an oligonucleotide, 5′GATCGAGATCTTACAGATCT**NNNTTAC-(A/G)TAANNNC**ACAAGATCTGGTGC3′, was synthesized and cloned directly into a pKS vector, and individual random clones were isolated and sequenced. 17 individual isolates were tested directly for binding with recombinant DBP and C/EBPβ, as described above.

DNase Footprinting of Randomly Selected Rat Genomic DNA Fragments

To determine the frequency of DBP binding sites versus C/EBPβ sites in the rat genome, we cloned random genomic fragments, approximately 500 – 700 bps long, into a pKS vector and footprinted this random library with either recombinant DBP or C/EBPβ. Total rat DNA was digested with Sau3A or Hae III and fragments within a 300 – 700 bp size range were isolated, gel-purified, and cloned into pKS vectors digested with Bam H1 or Sma 1, respectively. DNase I footprinting was carried out as described previously (Lichtsteiner *et al.*, 1987), using 3 dilutions of recombinant DBP or C/EBPβ protein. As the intermediate protein concentration we chose DBP and C/EBPβ concentrations that yielded half-saturation of the FP2 site of the cholesterol 7α hydroxylase promoter (Lavery and Schibler, 1993) under identical conditions. The lower and higher concentrations were threefold lower and higher, respectively, than this concentration.

Electrophoretic Mobility-Shift Assays (EMSA) and Determination of Dissociation Constants

Saturation binding analysis was carried out using a double-stranded oligonucleotide containing the DBP optimal binding site. For Scatchard plot analysis, 3×10^{-9} M of recombinant DBP or C/EBPβ protein was added to a set of binding reactions containing twofold increments from 0 to 25×10^{-9} M unlabeled oligonucleotide containing a small amount of ^{32}P-labeled oligonucleotide as a tracer. Binding reactions were incubated for 30 min at 25 °C and analyzed by EMSA. After autoradiography, bound and free products were excised from wet gels and their radioactivity quantitated by scintillation counting. Reaction conditions and EMSAs have been described previously (Lichtsteiner *et al.*, 1987). For protein saturation experiments, 0.5 ng of labeled oligonucleotide was added to reactions containing twofold increments of recombinant protein, between 1 and 100×10^{-9} M, and binding reactions were analyzed as above.

Construction of the C/EBPβ$_{V>A}$ Mutant and Expression of Recombinant Proteins

Site-directed oligonucleotide mutagenesis was used to convert val-216 of C/EBPβ to an alanine residue. An oligonucleotide complementary to C/EBPβ but containing two base changes was synthesized and mutagenesis was performed using the T7-directed Sculptor mutagenesis kit (Amersham), according to the manufacturer's instructions. After verification of the mutation by sequencing, a restriction fragment was subcloned into a pET vector bearing the C/EBPβ gene and the protein was expressed using the T7 expression system of Studier *et al.* (1990) and purified from bacterial lysates with heparin agarose chromatography as previously described (Descombes *et al.*, 1990). All other procedures were performed using standard molecular biology techniques (Sambrook *et al.*, 1989).

Acknowledgements

We thank D. Drolet and G. Rosenfeld for providing a TEF cDNA plasmid, A. Look for an HLF cDNA plasmid, D. Lavery and E. Schmidt for their critical reading of the manuscript, and N. Roggli for expert preparation of the Figures.
This work was supported by the Swiss National Science Foundation and by the Canton of Geneva.

References

Agre, P., Johnson, P.F., and S.L. McKnight. (1989). Cognate DNA-binding specificity retained after leucine zipper exchange between GCN-4 and C/EBP. Science *246*, 922 – 926.

Blackwell, T.K., and H. Weintraub. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science *250*, 1104 – 1110.

Cao, Z., R. M. Umek, and S. L. McKnight. (1991). Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. Genes Dev. *5*, 1538 – 1552.

Descombes, P., M. Chojkier, S. Lichtsteiner, E. Falvey, and U. Schibler. (1990). Lap, a novel member of the C/EBP gene family, encodes a liver-enriched transcriptional activator protein. Genes Dev. *4*, 1541 – 1551.

Descombes, P. and U. Schibler. (1991). A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP are translated from the same mRNA. Cell *67*, 569 – 579.

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.

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 a helices: crystal structure of the protein-DNA complex. Cell *71*, 1223 – 1237.

Falvey, E., F. Fleury-Olela, and U. Schibler. (1995). The rat hepatic leukemia factor (HLF) gene encodes two transcriptional activators with distinct circadian rhythms, tissue distribution and target preferences. EMBO J. *14*, 4307 – 4317.

Fonjallaz, P., V. Ossipow, G. Wanner, and U. Schibler. (1996). The two PAR leucine zipper proteins TEF and DBP display similar circadian and tissue-specific expression, but have different target promoter preferences. EMBO J.*15*, 351 – 362.

Glover, J. N. M., and S. C. Harrison. (1995). Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. Nature *373*, 257 – 261.

Haas, N. B., C. A. Cantwell, P. F. Johnson, and J. B. Burch. (1995). 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. Mol. Cell. Biol. *15*, 1923 – 1932.

Hunger, S. P., K. Ohyashiki, K. Toyama, 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.

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. Mol. Cell. Biol. *14*, 5986 – 5996.

Hurst, H. C. (1994). Transcription factors 1, bZIP proteins. Protein Profiles *1*, 123 – 168.

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.

Inaba, T., L. H. Shapiro, T. Funabiki, A. E. Sinclair, B. G. Jones, R. A. Ashmun, and A.T. Look. (1994). DNA-binding specificity and trans-activating potential of the leukemia-associated E2A-hepatic leukemia factor fusion protein. Mol. Cell. Biol. *14*, 3403 – 3413.

Inaba, T., Inukai, T., Yoshihara, T., Seyschab, H., Ashmun, R.A., Canman, E., Laken, S.E., Kastan, M.B., and Look, A.T. (1996). Reversal of apoptosis by the leukemia-associated E2A-HLF chimaeric transcription factor. Nature *382*, 541 – 544.

Iyer, S. V., D. L. Davis, S. N. Seal, and J. B. 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.

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.

Kerppola, T., and T. Curran. (1995). Transcription: Zen and the art of Fos and Jun. Nature *373*, 199 – 200.

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.

König, 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.

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

Landschulz, W. H., P. F. Johnson, and S. L. McKnight. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA-binding proteins. Science *240*, 1759 – 1764.

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.

Lavery, D. J., and U. Schibler. (1993). Circadian transcription of the cholesterol 7α hydroxylase gene may involve the liver-enriched bZIP protein DBP. Genes Dev. *7*, 1871 – 1884.

Lavery, D. J., and U. Schibler. (1994). DBP and related transcription factors of the PAR family. In: Liver Gene Expression. F. Tronche and M. Yaniv, eds (Austin, TX, USA; R.G. Landes Co.), pp. 259 – 275.

Lichtsteiner, S., J. Wuarin, and U. Schibler. (1987). The interplay of DNA-binding proteins on the promoter of the mouse albumin gene. Cell *51*, 963 – 973.

Lin, F. T., and M. D. Lane. (1994). CCAAT/enhancer binding protein alpha is sufficient to initiate the 3T3-L1 adipocyte differentiation program. Proc. Natl. Acas. Sci. USA *91*, 8757 – 8761.

Lumb, K. J., and P.S. Kim. (1995). Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper. Science *268*, 436 – 439.

McKnight, S. L., M. D. Lane, and S. Glueckson-Waelsch. (1989). Is CCAAT/enhancer-binding protein a central regulator of energymetabolism? Genes. Dev. *3*, 2021 – 2024.

Metzstein, M.M., Hengartner, M.O., Tsung, N., Ellis, R.E., and Horvitz, H.R. (1996). Transcriptional regulator of programmed cell death encoded by Caenorhabditis elegans gene ces-2. Nature *383*, 545 – 547.

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.

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

Pu, W.T., and K. Struhl. (1991). Highly conserved residues in the bZip domain of yeast GCN4 are not essential for DNA-binding. Mol. Cell. Biol. *11*, 4918 – 4926.

Sambrook, J., T. Maniatis, and E. F. Fritsch. (1989). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Scatchard, G. (1949). The attractions of proteins for small molecules and ions. Ann. N.Y. Acad. Sci. *51*, 660 – 672.

Screpanti, I., L. Romani, P. Musiani, A. Modesti, E. Fattori, D. Lazzaro, C. Sellitto, S. Scarpa, D. Bellavia, G. Lattanzio, F. Bistoni, L. Frati, R. Cortese, A. Gulino, G. Ciliberto, F. Constantini, and V. Poli. (1995). Lymphoproliferative disorder and imbalanced T helper response in C/EBPβ-deficient mice. EMBO J. *14*, 1932 – 1941.

Shuman,J.D., C.R. Vinson, and S.L. McKnight. (1990). Evidence of changes in protease sensitivity and subunit exchange on DNA-binding by C/EBP. Science *29*, 771.

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.

Suckow, M., B. von Wilcken-Bergmann, and B. Muller-Hill. (1993a). Identification of three residues in the basic regions of

bZip proteins GCN4, C/EBP, and TAF-1 that are involved in specific DNA-binding. EMBO J. *12*, 1193 – 1200.

Suckow, M., B. von Wilcken-Bergmann, and B. Muller-Hill. (1993b). The DNA-binding specificity of the basic region of the yeast transcriptional activator GCN4 can be changed by substitution of a single amino acid. Nucleic Acids Res. *21*, 2081 – 2086.

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.

Wang, N. D., M.J. Finegold, A. Bradley, C. N. Ou, S. V. Abdelsayed, M. D. Wilde, L. R. Taylor, D. R. Wilson, and G. Darlington. (1995). Impaired energy homeostasis in C/EBPα knockout mice. Science *269*, 1108 – 1112.

Wuarin J., and U. Schibler. (1990). Expression of the liver-enriched transcriptional activator protein DBP follows a stringent circadian rhythm. Cell *63*, 1257 – 1266

Yano, M., E. Falvey, and F. J. Gonzalez. (1992). Role of the liver-enriched transcription factor DBP in expression of the cytochrome P_{450} CYP2C6 gene. Mol. Cell. Biol. *12*, 2847 – 2854.

Yeh, W. C., Z. Cao, M. Classon, and S. L. McKnight. (1995). Cascade regulation of terminal adipocyte differentiation by three members of the C/EBP family of leucine zipper proteins. Genes Dev. *7*, 168 – 181.

Received September 25, 1996; accepted October 11, 1996