CBF is a heterotrimeric protein that binds to DNA containing CCAAT motifs. Here we have analyzed interactions of recombinant CBF with DNA using hydroxyl radical footprinting and methylation interference assays. In the CBF-DNA complex, three separate DNA regions are protected from hydroxyl radical cleavage, one located over and immediately adjacent to the CCAAT motif itself and the other two located on both sides of the CCAAT motif. The methylation interference assay showed, however, that only in the CCAAT motif region methylation of bases was able to interfere with the formation of a CBF-DNA complex, suggesting that CBF makes sequence-specific contacts only in the CCAAT motif region. To further determine the specific DNA sequences necessary for CBF binding, we employed a polymerase chain reaction-mediated random binding site selection method. This analysis showed that CBF binding to DNA requires the CCAAT sequence and other specific sequences immediately flanking both ends of the CCAAT motif. We also showed that the nature of the flanking nucleotide sequences affects the affinity of CBF for DNA. Interestingly, most of the CCAAT motifs present in various higher eukaryotic promoters correspond to the CBF binding sites that were selected, consistent with the hypothesis that these motifs are binding sites for CBF and, hence, that CBF could regulate transcription of numerous eukaryotic genes.

In many higher eukaryotic class II promoters, CCAAT motifs (or ATTGG motifs in the opposite strand), are often found between −50 and −110 relative to the start of transcription (1–4). The precise location of these CCAAT motifs and the promoter sequences around the motif of a specific gene are highly conserved during evolution. Experiments in several laboratories have shown that in many eukaryotic promoters, including tissue-specific and inducible promoters, the mammalian heterotrimeric CCAAT-binding protein CBF (also called NF-Y and CFI) binds to these CCAAT motifs. Mutations of the CCAAT motif that cause loss of CBF binding also decrease the transcriptional activity of these promoters (5–15).

CBF consists of three different subunits, CBF-A, CBF-B, and CBF-C, that all are necessary to form a CBF-DNA complex (16–20). Specific amino acid segments in each CBF subunit are highly conserved during evolution. The conserved segments of CBF-A and CBF-C show amino acid sequence similarities with the histone-fold motifs of histone H2B and H2A, respectively (21–23). Mutational analyses of the three CBF subunits demonstrated that the histone-fold motifs of CBF-A and CBF-C are needed to form a CBF-A/CBF-C heterodimer; this heterodimer generates a hybrid surface, which then interacts with CBF-B to form the heterotrimeric CBF molecule (23–25). Our previous mutational study also showed that each CBF subunit contains a domain that interacts with DNA, indicating that in each of these CBF subunits specific amino acids are involved in formation of the CBF-DNA complex.

Previous studies showed that mutations of any nucleotide in the CCAAT sequence either decrease or abolish CBF binding, suggesting that CBF is a bona fide CCAAT-binding factor (1, 5, 26). However, it was also shown that some of the DNA sequences containing CCAAT do not bind CBF, indicating that specific nucleotide sequences other than CCAAT are necessary for CBF binding. Moreover, several other proteins that bind to CCAAT motifs of specific promoters have been isolated (2, 27, 28). Together, these studies raised the question whether the CCAAT motif in a promoter binds only to a specific protein such as CBF or to a heterogeneous set of DNA-binding proteins whose requirement of an intact CCAAT motif for binding seems less rigorous than for CBF. This in turn underscored the need to better define the sequences in the CBF binding sites that are required for CBF binding.

In the present study, we analyzed in more detail the interactions between the heterotrimeric CBF and DNA using hydroxyl radical footprinting and methylation interference assays. By utilizing a polymerase chain reaction-mediated random binding site selection method, we defined the specific DNA sequences necessary for CBF binding. Specific sequences flanking the CCAAT motif were identified that are involved in the formation of a CBF-DNA complex. Interestingly, the flanking sequences that we found favor CBF binding also flank CCAAT motifs in many eukaryotic promoters, consistent with the notion that CBF is the major CCAAT-motif binding protein which participates in the regulation of transcription of many eukaryotic genes.

MATERIALS AND METHODS

Recombinant CBF Subunits—Each of the three full-length CBF subunits was expressed in Escherichia coli and was purified as described by Coustry et al. (13). DNA binding of recombinant CBF was performed as described by Sinha et al. (20).

Hydroxyl Radical Footprinting—DNA fragments corresponding to nucleotides −130 to −30 of the pro-α2(1) collagen promoter and nucleotides −140 to −45 of the pro-α1(1) collagen promoter were used for hydroxyl radical footprinting. The DNA fragments were labeled at their 3′ ends with [α-32P]dCTP using Klenow DNA polymerase. Hydroxyl radical footprinting reactions were performed according to the method of Dixon et al. (29). Briefly, the labeled DNA fragments (1–2 × 105 cpm) were incubated with pure recombinant CBF subunits (about 120 ng of total protein) in a 50-μl volume of 10 mM Hepes, pH 8.0, 100 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, and 1% glycerol for 20 min at room temperature. The DNA fragments without CBF were similarly incubated. To initiate hydroxyl radical-mediated DNA cleavage, 6 μl of a
mixture containing 2 μl of Fe(II)EDTA (125 mM Fe(II) and 250 mM EDTA), 2 μl of 29 mM sodium ascorbate, and 2 μl of 0.8% hydrogen peroxide was added to the DNA solutions. The reaction mixture was incubated for 2 min at room temperature and was stopped by adding 150 μl of stop solution containing 112 μl of TE (10 mM Tris, pH 8.0, 1 mM EDTA), 10 μl of 0.1 M thiourea, 25 μl of 3 M sodium acetate, 2 μl of 0.2 M EDTA, and 1 μl (5 mg/ml) of tRNA. The mixtures were extracted with phenol-chloroform, and DNAs were precipitated by ethanol. The DNA samples were separated on an 8% polyacrylamide-urea sequencing gel.

Methylation Interference Assay—Partial methylation of end-labeled DNA was performed as described by Baldwin et al. (30). One picomole of labeled DNA fragments was incubated in a 200-μl reaction volume containing 50 mM sodium cacodylate, pH 8.0, 1 mM EDTA, 1 μl of dimethyl sulfate at room temperature for 5 min and was stopped by adding 40 μl of dimethyl sulfate stop buffer containing 1.5 M sodium acetate, pH 7.0, 1 mM 2-mercaptoethanol, and 1 μl (5 mg/ml) of tRNA. DNA was precipitated by ethanol, dried, and resuspended in 10 mM TE.

The methylated DNA (6 × 10^6 cpm) was incubated with recombinant CBF subunits (100 ng) in a 20-μl DNA binding buffer containing 25 mM HEPES, pH 8.0, 75 mM KCl, 10% glycerol, 1 mM EDTA, 0.05% Nonidet P-40, 0.5 mM phenylmethylsulfonyl fluoride, 0.5 mM dithiothreitol, 100 ng of poly(dI-dC), and 1 μg of bovine serum albumin at room temperature for 20 min. After the reactions were fractionated on a 4% nondenaturating polyacrylamide gel, the CBF-bound DNA and unbound DNA were each eluted from the gel in a buffer containing TE and 0.3 M sodium acetate, purified by ElutipKit (Schleicher & Schuell), cleaved after incubation with piperidine, precipitated, resuspended, and separated on an 8% polyacrylamide-urea sequencing gel.

PCR1-mediated Random Binding Site Selection—Two different 67-base oligonucleotides, termed R1 and R2, were synthesized and used as primers in two independent binding site selection experiments. Both were flanked by sequences complementary to two 18-base oligonucleotides primers A and B containing XhoI and XbaI restriction sites, respectively, for cloning. The sequences of these four oligonucleotides are as follows: oligo R1, 5'-CTGAGTCCCCTCGAGCTAGCNNNNNNNNNNNNNNNNNNTTTGGTCTAGACCTCCGGCT-3'; oligo R2, 5'-CTGAGTCCCCTCGAGCTAGCNNNNNNNNNNNNNNNNNATGGNNNNNNNNNNNTTTGGTCTAGACCTCCGGCT-3'; primer A, 5'-CTGAGTCCCCCTCGAGCTCTCCGGCT-3'; primer B, 5'-ACCAGATCTGGAGGCCGA-3'.

Double-stranded radiolabeled oligonucleotides were prepared by annealing oligo R1 or R2 to 5-fold molar excess of primer B and extending with Taq DNA polymerase at 72 °C for 30 min in the presence of [α-32P]dCTP. The labeled DNA (50 fmol) was incubated in a 10-μl volume of DNA binding buffer together with the three recombinant CBF subunits (100 ng) or separately with either CBF-A/CBF-C (50 ng) or CBF-B (50 ng) alone. After reaction mixtures were fractionated by electrophoresis on a nondenaturing 4% polyacrylamide gel, the DNA-protein complex specifically formed in the presence of three CBF subunits was eluted from the gel and then amplified in a PCR using primers A and B in the presence of [α-32P]dCTP. The labeled DNA was used in a second round of CBF binding, followed by isolation of the DNA-protein complex and reamplification of the CBF-bound DNA. Af-

The abbreviations used are: PCR, polymerase chain reaction; oligo, oligonucleotide(s); MUT, mutant.
After four rounds of selection, the amplified DNA was digested with XhoI and XbaI restriction enzymes and subcloned into a pBluescript KS vector (Stratagene). The DNA sequence of inserts in individual clones was then determined. The plasmids containing the selected binding sites were digested with XhoI and NotI, labeled with [α-32P]dCTP using Klenow DNA polymerase, and assayed for binding with CBF.

**Determination of Dissociation Rate Constant (k_d)—** Labeled DNA fragments were incubated first with recombinant CBF in a 100-μl DNA binding buffer at room temperature for 20 min and then incubated with a 100-fold excess of a 42-base pair double-stranded oligonucleotide containing the CBF binding site of the pro-α2(1) collagen promoter over different time intervals. At each time point, 10 μl of the incubation mixture was loaded on a 4% polyacrylamide gel and run at 50 V. After the last aliquot of the incubation mixture was loaded, the gel was run at 150 V for 2 h. The amount of CBF-DNA complex remaining at each time point was measured by densitometry in a PhosphorImager (Molecular Dynamics). The equation ln(fraction CBF bound) = −k_d(t − time) was used to determine the dissociation rate constant k_d.

**In Vitro Transcription Assay—** In vitro transcription assay was performed using EL4 cell nuclear extracts depleted of CBF (12). DNA templates containing the wild-type pro-α2(1) collagen promoter and the pro-α1(III) collagen control promoter were used in the assay as described previously (5, 12). Mutations in the flanking sequences of CCAAT of the pro-α2(1) collagen promoter were generated by two-step PCR (24), and the mutations were verified by DNA sequencing. The following oligonucleotides used to generate the mutants are as follows:

- **oligo 1** (5'-CTCAACCAATCGAGGCCTAGG-3')
- **oligo 2** (5'-CTCGTAGGGTTGAGACGGT-3') for MUT1
- **oligo 3** (5'-CTGCCCCTACGATTGGAGCTAGG-3') and **oligo 4** (5'-CTCAGCTTAaGGGCGAGAGGTTG-3') for MUT2.

**RESULTS**

**Hydroxyl Radical Footprinting—** To identify the contacts in DNA that are produced by CBF binding, we used the hydroxyl radical footprinting method of Tullius and Dombroski (31). Hydroxyl radical is a small molecule cleavage agent that removes individual bases from a DNA fragment; the base cleavage is diminished, however, by the contacts produced by protein-DNA interaction. Thus, this footprinting method provides high resolution structural information on the DNA-protein contacts. Briefly, DNA fragment of the mouse pro-α2(1) collagen

**FIG. 2.** Summary of hydroxyl radical footprinting and methylation interference assay of recombinant CBF. Results of hydroxyl radical footprinting of recombinant CBF protein bound to pro-α2(1) collagen promoter DNA and to pro-α1(1) collagen promoter DNA are shown in Fig. 1. Results of methylation interference assay of CBF binding to pro-α2(1) collagen promoter DNA and to pro-α1(1) collagen promoter DNA are shown in Fig. 3.

**FIG. 3.** Methylation interference assay of CBF binding to pro-α2(1) collagen promoter DNA (A) and pro-α1(1) collagen promoter DNA (B). In both A and B, lanes 2 and 4 correspond to the DNAs that were bound with CBF (b), whereas lanes 1 and 4 correspond to the free DNAs (f). Nucleotide sequences of each DNA strand were determined by G+A sequencing reactions and run in the same sequencing gel. Asterisks represent bases that were absent or present at low levels in the CBF bound DNA.
A promoter containing a CCAAT motif was labeled at the 3′ end of either the upper or the lower strand. The labeled DNA fragments were then incubated with purified recombinant CBF subunits and subjected to hydroxyl radical cleavage. The concentration of reagents used in the reaction to generate hydroxyl radical footprints did not affect formation of the CBF-DNA complex, as examined by electrophoretic mobility shift assay. When the DNA binding reactions contained all three CBF subunits, three separate regions of the upper DNA strand and three separate regions of the lower strand were protected from hydroxyl radical cleavage (Fig. 1A, lanes 3 and 6). Each of the three protected regions on the lower strand was offset by about 4 to 5 residues in comparison with the protected regions on the upper strand. When the DNA binding reaction contained either CBF-B alone or the CBF-A/CBF-C heterodimer, no protection was seen (data not shown). Thus, the protection of three DNA regions from hydroxyl radical cleavage in each DNA strand was specifically due to interactions with the heterotrimeric CBF.

Among the three hydroxyl radical footprinting regions, one contained the CCAAT sequence, and the other two were located on either side of the CCAAT motif (Fig. 2). A previous comparison of the CBF binding sites in different promoters showed nucleotide sequence similarity only in the CCAAT motif and the bases immediately surrounding this motif and none in the more distal flanking regions that were protected from hydroxyl radical cleavage (1, 2). To determine the footprinting pattern of another CBF binding site, we performed the hydroxyl cleavage reaction using the DNA containing the CCAAT motif of the pro-$$\alpha$$1(1) collagen promoter. When this promoter DNA bound to CBF, three separate regions on both the upper and lower strand were protected from hydroxyl radical cleavage (Fig. 1B, lanes 3 and 6). The CBF footprints of the two promoters are summarized in Fig. 2. The results show that the footprinting patterns of the pro-$$\alpha$$1(1) and pro-$$\alpha$$2(1) collagen promoters were qualitatively very similar despite the absence of obvious sequence similarities other than in the CCAAT mo-

**Fig. 4.** DNA sequences of clones selected by using random oligonucleotide R1. A, sequence of the R1 oligonucleotide containing 27 random nucleotides. B, sequences of 33 clones selected by binding of CBF to DNA. The sequences of the 10 clones in Part 1 are the complementary strands of R1; each clone sequence contains an ATTG that is underlined at the 3′ end of the selected sequences. The defined sequence 3′ of ATTG in each clone is shown in lowercase letters. In Part 2, the sequences of 20 clones are aligned according to the position of ATTG. In Part 3 the sequences of three clones are shown; none contain ATTGG. C, summary of sequences selected at each position relative to ATTG in Parts 1 and 2. Each number indicates how many times a base appeared in a position.
tific. This suggests that the CBF contacts over the CCAAT motif are sequence-specific and those in the more distal flanking footprinting regions are not. In the three footprinting regions, each of the regions protected from hydroxyl radical cleavage on the complementary strand was shifted toward the 3'-boundary; this pattern was also found in the hydroxyl radical footprinting of the bacteriophage λ repressor and Cro proteins, characteristic of proteins that interact with DNA on one side of the helix only (31).

Previous mutational analyses of all three CBF subunits indicated that a DNA binding domain is present in each CBF subunit, suggesting that all three CBF subunits make contact with DNA in one of the three DNA regions. To determine the specificity of the contacts between CBF and DNA, we performed the hydroxyl radical cleavage reaction using DNA binding domain mutants of the CBF-A (A60) and CBF-C (C46/47) subunits. Each of these CBF mutants formed a heterotrimeric CBF that was, however, unable to bind DNA in an electrophoretic mobility shift assay (23, 25).

When the DNA binding reaction contained either a mixture of mutant A60, wild-type CBF-B, and wild-type CBF-C or a mixture of wild-type CBF-A, wild-type CBF-B, and mutant C46/47, none of the three DNA regions were protected from hydroxyl radical cleavage (data not shown). This indicated that none of the three DNA binding domains of the CBF subunits made contact separately in the CBF-DNA complex but rather that these domains must have interacted simultaneously with DNA to form a CBF-DNA complex.

**Methylation Interference**—To identify specific base contacts that contributed to CBF binding, we performed a methylation interference assay. In this assay, we treated the labeled DNA of the pro-α2(1) collagen promoter with dimethyl sulfate and then determined whether methylation of A and G bases interfered with CBF binding. Using this assay, we found that methylation
of two Gs in the upper strand and two Gs in the lower strand, all located in the CCAAT motif region, strongly interfered with CBF binding (Fig. 2 and Fig. 3A, lanes 2 and 4). In contrast, methylation of bases in the flanking regions of CCAAT that were protected from hydroxyl radical cleavage did not interfere with CBF binding. This indicated that the bases that formed specific contacts with CBF were only those of the CCAAT motif region. We also performed a methylation interference assay with the pro-α1(1) collagen promoter DNA. In this promoter, methylation of one G base in the upper strand and two A bases in the lower strand, all located within the CCAAT motif, interfered with CBF binding (Fig. 2 and Fig. 3B, lanes 2 and 4). In contrast, methylation of bases of the hydroxyl radical-footprinted regions flanking the CCAAT sequence did not interfere with CBF binding. Again this indicated that CBF made sequence-specific contacts with bases in the CCAAT motif region but not with bases flanking the CCAAT motif.

In the pro-α1(1) collagen promoter, methylation of both A and G interfered with CBF binding, suggesting that CBF contacted bases in both the major and minor grooves of DNA. Recently, Ronchi et al. (32) reported that several minor groove-binding drugs inhibited CBF binding to DNA, suggesting that CBF makes contact with minor groove DNA. However, this study did not rule out that CBF also makes contact with major groove DNA.

Selection of DNA Binding—To define the nucleotides that determine the specificity of CBF binding to DNA in an unbiased manner, we used PCR-mediated random binding site selection (33). For this the 67-base pair oligonucleotide R1 containing 27 random nucleotides flanked by 20 nucleotides of defined sequences at each end was used for the selection (Fig. 4A). The DNA binding assays were performed using radiolabeled double-stranded random oligonucleotides incubated with the three purified recombinant CBF subunits. For control experiments, the radiolabeled oligonucleotides were incubated in two separate reactions, one with CBF-B and the other with the CBF-A/CBF-C heterodimer. Formation of a protein-DNA complex was observed only in the reaction containing all three CBF subunits and not in control reactions containing either CBF-B alone or the CBF-A/CBF-C heterodimer alone; this indicated that the protein-DNA complex consisted of the heterotrimeric CBF-DNA complex.

After four rounds of selection, the amplified DNAs were cloned, and 33 different clones were sequenced. The sequences are compiled and are divided into three groups in Fig. 4B. In the first group, each of 10 different clones contained the 5′-ATTGG-3′ sequence at the 3′ end of the selected sequences. The next nucleotide 3′ of the ATTGG sequence was a G in the flanking sequence of the starting oligonucleotide. In the second group of 20 different clones, all sequences contained the ATTGG sequence but at different positions on either strand. In the third group none of three different clones contained the ATTGG sequence. In subsequent experiments the DNAs in the third group bound very weakly to CBF. Thus most of the selected sequences contained ATTGG (or CCAAT in the complementary strand), indicating that CBF bound preferentially to DNAs containing the CCAAT motif. To determine whether the flanking sequences of the ATTGG motif had any specific role in CBF binding, we aligned the 30 selected clones with respect to ATTGG (Fig. 4C). Although no single nucleotide was selected in every clone at any position either 5′ or 3′ of ATTGG, specific nucleotides were selected in several positions with high frequency. Most notably, at the 5′ end of ATTGG, G residues in position 1 and T and C residues in position 2 were preferentially selected. At the 3′ end of ATTGG, T and C residues in positions 1 and 2 and A and G residues in position 3 were preferentially selected. These results indicated that sequences immediately flanking ATTGG played a role in CBF binding and that (T/C) ATTGG (T/C)/(T/C)/(A/G) was the sequence that favored CBF binding.

To confirm the role of the flanking sequences, we repeated the PCR-mediated binding site selection using oligonucleotide R2 (Fig. 5A). The difference between R1 and R2 is that instead of 27 random nucleotides, R2 contains an ATTGG core sequence flanked by 12 5′ random nucleotides and 10 3′ random nucleotides. After four rounds of DNA binding and PCR amplification of the bound DNA, the amplified DNAs were cloned, and 30 different clones were sequenced. The sequences are compiled in Fig. 5B, and the frequency of the selected nucleotides in each position is shown in Fig. 5C. The results indicated the following; at the 5′ end of ATTGG, G and T residues were preferentially selected in positions 1 and 2, respectively; at the

**Table I**

DNA binding affinity of recombinant CBF for selected binding sites

| Site     | Sequence* | Binding affinity* |
|----------|-----------|-------------------|
| Group 1  | TG        | (CT)(CT)(A/G)     |
| A11      | AGACGTATGG | TTAATCTCTC       |
| A12      | TCGTGCCTCTG | CTTGTGTTAAT     |
| A13      | AGATGTGGGGTG | CTAATTTACA     |
| A14      | AGGAGAATGCTG | TTANAGACCT     |
| Group 2  | TG        |                   |
| A11      | AGAGAATGCTG | TgGAGCAAGG       |
| A12      | GGAGGTTTCTTG | CGTGTATATCA     |
| Group 3  | --        | (CT)(CT)(A/G)     |
| A31      | GGCCGGATTCTT | TTTGAGGATGC     |
| A32      | AGATGGATCTCgG | CTATCTGATTG     |
| A33      | GTCTGGCTCCtT | TTTGCGCTGCT     |
| A34      | AGAGGAATCCTta | TGTTACTGTT     |
| A35      | GAGGGTTCCCggG | CTTGAACTTGG     |
| Group 4  | AG        |                   |
| A41      | TAGTCACTCggG | CCAGGGATT       |
| A42      | ATGGGTGTGGggG | TTGTGACGCT    |
| A43      | ACCGGTTCGCCctT | aTGAGATCTG     |
| A44      | AAGCCTGGGCggG | CatTTATATT     |

*The sequences of selected clones are divided into four groups according to the derived consensus sequences for CBF binding present in group 1. Dashed lines above groups 2, 3, and 4 indicate which side the flanking sequences do not match the consensus CBF binding sites; lowercase letters in the DNA sequences indicate non-consensus flanking nucleotides.

*To determine CBF binding affinity, the DNA of each selected clone was labeled with the same specific activity and was used in a DNA binding assay with recombinant CBF as shown in Fig. 6. The relative DNA binding affinity was determined by visual estimation of CBF-bound DNA and was scaled from +++++ (strong) to +/− (weak).
FIG. 7. Dissociation rate of CBF-DNA complexes formed with selected clones containing different flanking sequences. A, dissociation of CBF-DNA complexes with individual clones at different times after addition of competitor DNA. B, amounts of DNA-protein complexes at different times as measured by densitometry.
TABLE II
Summary of CBF binding efficiency to selected sites from the four groups

| Site     | Sequence       | Binding affinity | \( k_d \) |
|----------|----------------|------------------|------------|
| Group 1  | TG ATGG (C/T)(C/T)(A/G) | ++ + + | -0.0214 |
| A11      | TG TTA          | ++ + + | 0.0193 |
| Group 2  | TG ATGG          | ++ + + | -0.0246 |
| A21      | TG TGC          | ++ + + | -0.0402 |
| A22      | TG TCT          | ++ + + | -0.0914 |
| Group 3  | ATGG (C/T)(C/T)(A/G) | ++ + + | -0.0468 |
| A31      | Ta TTG          | ++ + + | -0.0846 |
| A32      | Tt TTG          | ++ + + | -0.0193 |
| Group 4  | ATGG             | ++ + + | -0.0914 |
| A41      | TG CTA          | ++ + + | -0.0914 |
| A42      | gG CCc          | ++ + + | -0.0914 |

* Complete sequences of the selected clones are shown in Table I. Lowercase letters in the DNA sequences indicate non-consensus flanking nucleotides.

* Relative binding affinity of CBF is as described in Table I.

* Dissociation rate constants (\( k_d \)) were determined by the experiment described in Fig. 7.

3’ end, pyrimidine nucleotides were preferentially selected in positions 1 and 2 and purine nucleotides in position 3. The results also indicated that the favorable sequence for CBF binding was TG ATGG (T/C)(T/C)(A/G). Therefore, very similar sequences flanking the ATTGG motif were preferentially selected in the two different experiments.

Analysis of the Selected Binding Sites—To analyze in more detail CBF binding with respect to different flanking sequences of ATGG, we divided the selected sequences into four groups (Table I). Most sequences in the first group contained consensus flanking sequences on both sides of the ATTGG motif, whereas sequences in groups 2 and 3 contained consensus flanking sequences only on the 5’ side and 3’ side, respectively. Group 4 contained no consensus flanking sequences at either end.

Individual DNA fragments from each group were labeled and tested in binding assays with recombinant CBF subunits (Fig. 6). The relative binding efficiencies of the different DNAs are summarized in Table I. All four DNAs in group 1 bound to CBF with high affinity. In contrast, all of the DNAs in groups 2 and 3 bound to CBF with moderate affinity, except for A31 in group 3 which bound to CBF with high affinity. The DNAs in group 4 bound weakly to CBF.

To investigate the biochemical details of CBF binding to the different DNA sequences, we also determined the dissociation rate constants (\( k_d \)) of the CBF-DNA complexes (Fig. 7, A and B). The \( k_d \) measurements are summarized in Table II. The \( k_d \) of the CBF-DNA complex formed with A11 in group 1 was very similar to the \( k_d \) of the complexes formed with A21 and A22 in group 2 and A32 in group 3, although the sequences in these two groups bound to CBF more weakly than those in group 1. In contrast, the \( k_d \) of A31 in group 3, whose affinity for CBF was similar to A11’s, was 1.9-fold higher than the \( k_d \) of A11, indicating that the CBF complex formed with A31 was less stable than the complex formed with A11. The \( k_d \) values of A34 of group 3 and A41 of group 4, which bound to CBF very weakly, were about 4-fold higher than that of A11. These results indicated that the DNA sequences in group 1 bound to CBF with high affinity and formed stable CBF-DNA complexes, thus accounting for the high number of selected DNA sequences in this group. These results also indicated that modifying one or two nucleotides of the flanking consensus sequences (groups 2 and 3) modestly decreased CBF binding, whereas modifying three or more nucleotides of the flanking sequences (group 4) decreased or abolished CBF binding. Thus, we concluded that high affinity CBF binding sites contained TG and (T/C)(T/C)(A/G) as 5’- and 3’-flanking sequences of the pentanucleotide ATTGG, respectively, and that CBF could not bind to DNA containing an ATTGG motif flanked by (A/G)(A/C) on its 5’ side and (A/G)(A/G)(T/C) on its 3’ side.

In our previous studies, we used the pro-\( \alpha2(1) \) collagen promoter fragment as a CBF binding site (5, 12, 13, 26). In this promoter, the sequence of the CBF binding site is 5’-CCATTGTTG-3’. The 5’-flanking sequence does not conform, however, to the consensus 5’-flanking sequence, and only two of three nucleotides on the 3’-flanking side match the consensus 3’-flanking sequence. Thus, this sequence belongs to group 4 of the selected sequences. We have analyzed CBF binding to this natural promoter site versus the selected consensus site. The binding reactions were performed using different concentrations of recombinant CBF protein, and the amounts of CBF-DNA complex formed in different reactions were plotted relative to CBF concentrations (Fig. 8, A and B). In this assay, the CBF concentration required for half-maximal binding to the pro-\( \alpha2(1) \) collagen promoter sequence, to A31, and to A11 were 2.3, 0.48, and 0.18 μg/ml, respectively. The results indicated that the binding affinity of the pro-\( \alpha2(1) \) collagen promoter DNA to CBF was more than 10 times lower than that of A11, a consensus CBF binding site. To confirm that the low affinity CBF binding observed for the pro-\( \alpha2(1) \) collagen promoter sequence was due to the presence of nonconsensus sequences flanking ATTGG, the flanking sequences in the collagen promoter were mutated to full consensus flanking sequence (MUT1) and also to a complete nonconsensus flanking sequence (MUT2) (Fig. 9A). As a result, MUT1 DNA bound to CBF with higher affinity than did wild-type DNA; whereas MUT2 DNA did not bind CBF (Fig. 9B). These results were in agreement with the binding selection data and supported our previous conclusions regarding the role of specific ATTGG-flanking sequences in the formation of CBF-DNA complexes.
Transcriptional Activation of CBF with Respect to Binding Affinity—Recently, we demonstrated in an in vitro reconstitution assay that recombinant CBF activates transcription of promoters containing CBF binding sites including the pro-α2(1) collagen promoter (12, 13). Using this assay, we tested the transcriptional activation of the mutant pro-α2(1) collagen promoters containing mutations in the flanking sequences of ATTGG. In this assay, CBF activated transcription of the MUT1 promoter containing consensus flanking sequences slightly more than it activated transcription of the wild-type promoter. In contrast, CBF did not activate transcription of the MUT2 promoter containing the nonconsensus flanking sequences of ATTGG (Fig. 9, C and D).

The ATTGG or CCAAT Motif in Various Eukaryotic Promoters—Computer analysis of 502 unrelated RNA polymerase II promoter regions showed that approximately 30% of the promoters contained a CCAAT sequence (or ATTGG sequence on the complementary strand) and that in a large number of vertebrate promoters the CCAAT motif was located around nucleotide −80 upstream of the transcription start site (4). Interestingly, in most of these promoters the flanking sequences of ATTGG were TG on the 5′ side and (T/C)(T/C)(A/G) on the 3′ side, identical to the flanking sequences selected as high affinity CBF binding sites. Moreover, in less than 1% of...
the promoter sequences, the ATTGG motif was flanked by AA at its 5’ end and by G(A/G) at its 3’ end, sequences that inhibit formation of a CBF-DNA complex. Hence we concluded that the vast majority of CCAAT motifs present in eukaryotic promoters should be able to bind CBF.

DISCUSSION

Our hydroxyl radical footprinting analysis of the CBF-DNA complex provides clear evidence that CBF makes contact with three separate segments of DNA; one of these segments contains the CCAAT sequence, whereas the other two, located on each side of the CCAAT motif, contain no specific sequences. In contrast, only in the CCAAT motif region methylation of bases is able to inhibit CBF binding. Together, these data provide strong evidence that the contacts CBF makes over the CCAAT motif and immediately adjacent to the CCAAT motif are sequence-specific, whereas the contacts in the more distal regions on both sides of the CCAAT motif are not. Previous studies had suggested that, in the CBF-DNA complex, the CBF-B subunit makes sequence-specific contacts. This conclusion was based on the observation that an allele-specific mutation in the DNA binding domain of HAP2, the yeast homolog of CBF-B, was defective in forming a protein-DNA complex with wild-type DNA. In contrast, the mutant favored formation of a HAP-DNA complex with a mutant DNA containing CCAAC instead of CCAAT (34). Our recent studies showed that the DNA binding domains of CBF-A and CBF-C are located within their histone-fold motifs (23, 25). Interestingly, the histone-fold motif is a three-dimensional structural motif present in all four core histones, H2A, H2B, H3, and H4, although the primary sequence of these histones only showed a low degree of homology (22). In each of these histones, the histone-fold motif provides a surface on which protein-protein interactions can occur to form a histone octamer; through multiple histone-fold motifs, the octamer also makes contact with DNA in a sequence-independent manner (35). Based on this property, we hypothesize that the histone-fold motifs of the CBF-A/CBF-C heterodimer can make nonspecific contact with DNA, possibly in the two contact regions identified in the hydroxyl radical footprinting experiments on both sides of the CCAAT motif.

Since our binding site selection experiments identified specific nucleotide sequences only in the CCAAT motif region (the CCAAT motif and immediate flanking sequence), this provides further evidence that CBF makes sequence-specific contact only with the CCAAT motif region. The sequences selected from a completely random oligonucleotide contained the CCAAT sequence, indicating that the CCAAT pentanucleotide is necessary for CBF binding. This is consistent with previous observations that mutations in the CCAAT motif decreased or abolished CBF binding. Together with CCAAT, specific sequence flanking the pentanucleotide were also selected, indicating that in addition to the contacts between CBF and the CCAAT motif, other interactions are needed to form a stable CBF-DNA complex. The flanking sequences that were selected were grouped according to their affinity to CBF. Our study showed that changes in one or two nucleotides of the consensus flanking sequences decreased CBF binding somewhat, whereas changes in more than two nucleotides strongly decreased or abolished it. This suggests that CBF binding sites present in various eukaryotic promoters may have different affinities for CBF. Moreover, mutations in the flanking sequences of the CCAAT motif that abolished CBF binding also blocked transcriptional activation by recombinant CBF. However, within the context of the pro-α2(1) collagen promoter, mutations in the CBF-flanking sequences that increased CBF binding did not significantly increase transcriptional activation. It is possible that relatively weak binding of CBF to the pro-α2(1) collagen promoter is stabilized by another DNA-binding protein that binds near the CBF binding site. Indeed, recent studies showed that in the pro-α2(1) collagen promoter, binding sites for the SP1 transcription factor are located near the CBF binding site (38). Recently, Wright et al. (36) demonstrated that, in the major histocompatibility complex class II-associated invariant gene, binding of CBF to a weak CBF binding site is stabilized by the binding of SP1 near the CBF binding site.

Most importantly, the CCAAT-flanking sequences selected in our experiments are also found around the CCAAT motifs in most eukaryotic promoters harboring a CCAAT sequence in these proximal promoters. It is possible that the binding of CBF to these promoters is regulated in vivo both by the affinity of CBF binding and by the presence of other transcription factors that bind nearby. Our results are consistent with the notion that CBF is the protein that binds to CCAAT motifs present in many eukaryotic gene promoters and hence that it could regulate transcription of many of these genes.

Acknowledgments—We thank Heidi Eberspaecher for oligonucleotide synthesis. We are also grateful to Jude Richard for editing the manuscript.

REFERENCES

1. Dorn, A., Bolleken, J., Staib, A., Benoist, C., and Mathis, D. (1987) Cell 50, 863–872
2. Chodosh, L. A., Baldwin, A. S., Carthew, R. W., and Sharp, P. A. (1988) Cell 53, 11–24
3. Hatamochi, A., Golumbek, P. T., Van Schaftingen, E., and de Crombrugghe, B. (1988) J. Biol. Chem. 263, 5940–5947
4. Golumbek, P. T., Romani, A., and de Crombrugghe, B. (1988) J. Biol. Chem. 263, 5940–5947
5. Sinha, S., Maity, S. N., Lu, J., and de Crombrugghe, B. (1995) J. Biol. Chem. 270, 21445–21448
6. Roy, B., and Lee, A. S. (1995) Mol. Cell. Biol. 15, 2269–2274
7. Mathis, D. (1990) J. Cell. Physiol. 139, 218–224
8. Mantovani, R., Li, X. Y., Knapp, A., Pasquali, J., Benoist, C., and Mathis, D. (1992) EMBO J. 11, 3151–3155
9. Wright, K. L., Vilen, B. J., Itoh-Lindstrom, Y., Moore, T. L., Li, G., Criscitiello, M., Cogswell, P., and Ting, J.-F. (1994) EMBO J. 13, 4042–4053
10. Kim, I.-S., and de Crombrugghe, B. (1995) J. Biol. Chem. 270, 668–675
11. Ronchi, A., Bellorini, M., Mongelli, N., and Mantovani, R. (1995) Nucleic Acids Res. 23, 12815–12820
12. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
13. Roy, B., and Lee, A. S. (1995) Mol. Cell. Biol. 15, 2269–2274
14. Mathis, D. (1990) J. Cell. Physiol. 139, 218–224
15. Roy, B., and Lee, A. S. (1995) Mol. Cell. Biol. 15, 2269–2274
16. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
17. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
18. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
19. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
20. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
21. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
22. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
23. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
24. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
25. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
26. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
27. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
28. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
29. Kim, I.-S., Sinha, S., and de Crombrugghe, B. (1999) J. Biol. Chem. 274, 21893–21898
30. Baldwin, A. S., Oettinger, M., and Struhl, K. (1994) in Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) Suppl. 25, pp. 12.1.1–12.5.7, Wiley-Liss, Inc., New York
31. Tullius, T. D., and Dombroski, B. A. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 5469–5473
32. Ronchi, A., Bellorini, M., Mongelli, N., and Mantovani, R. (1995) Nucleic Acids Res.
26572

DNA Binding Specificity of CBF/NF-Y

33. Blackwell, T. K., and Weintraub, H. (1990) Science 250, 1104–1110
34. Xing, Y., Fikes, J. D., and Guarente, L. (1993) EMBO J. 12, 4647–4655
35. Arents, G., and Moudrianakis, E. N. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 11170–11174
36. Wright, K. L., Moore, T. L., Vilen, B. J., Brown, A. M., and Ting, J. P.-Y. (1995) J. Biol. Chem. 270, 20978–20986
37. Maxam, A., and Gilbert, W. (1980) Methods Enzymol. 65, 497–559
38. Hasegawa, T., Zhou, X., Garret, L. A., Ruteshouser, E. C., Maity, S. N., and de Crombrugghe, B. (1996) Nucleic Acids Res. 24, 3253–3260