Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 1624–1628, February 1995
Biochemistry

Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein–DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3

SATRAJIT SINHA*, SANKAR N. MAITY*†, JINGFANG LU, AND BENOIT DE CROMBRUGGHE†

Department of Molecular Genetics, The University of Texas, M. D. Anderson Cancer Center, Houston, TX 77030

Communicated by Sankar Adhya, National Institutes of Health, Bethesda, MD, November 10, 1994 (received for review August 22, 1994)

ABSTRACT **The CCAAT binding factor CBF is a heteromeric transcription factor, which binds to functional CCAAT motifs in many eukaryotic promoters. cDNAs for the A and B subunits of CBF (CBF-A and CBF-B) and for their yeast homologues HAP3 and HAP2 have been previously isolated, but the purified recombinant CBF-A and CBF-B together are unable to bind to CCAAT motifs in DNA. Here we report the isolation of a cDNA coding for rat CBF-C, demonstrate that recombinant CBF-C is required together with CBF-A and CBF-B to form a CBF–DNA complex, and show that CBF-C is present in this protein–DNA complex together with the other two subunits. We further show that CBF-C allows formation of a complex between the purified recombinant yeast HAP2 and HAP3 polypeptides and a CCAAT-containing DNA and is present in this complex, implying the existence of a CBF-C homologue in yeast. We show that CBF-A and CBF-C interact with each other to form a CBF-A–CBF-C complex and that CBF-B does not interact with CBF-A or CBF-C individually but that it associates with the CBF-A–CBF-C complex. Our results indicate that CBF is a unique evolutionarily conserved DNA binding protein.**

The CCAAT motif is a functional cis-acting element present in many eukaryotic promoters, including those of the genes for the two type I collagens, albumin, the major histocompatibility complex (MHC) class II, and other genes (1–3). In these genes this motif is a binding site for the heteromeric CCAAT binding transcription factor CBF, also called NF-Y and CP1 (3, 4). CBF subunits A and B (CBF-A and CBF-B) show considerable sequence identity in domains involved in DNA binding with similar functional segments in the yeast polypeptides HAP3 and HAP2 (refs. 5–12; K. Y. Sohn, S.S., S.N.M., and B.d.C., unpublished data). These polypeptides are components of a *Saccharomyces cerevisiae* CCAAT binding transcription factor that controls a series of nuclear genes active in mitochondrial function (13). A third yeast polypeptide, HAP4, has been found to be associated with HAP2–HAP3 but is not required for DNA binding (14, 15).

cDNAs for CBF-A and CBF-B have been isolated from several mammalian species (5–8), but the purified recombinant CBF-A and CBF-B together are unable to effect DNA binding (16). We have previously identified a rat liver polypeptide of apparent size 40 kDa, designated CBF-C, which allowed the recombinant CBF-A and CBF-B subunits (rCBF-A and rCBF-B) to bind to DNA (16). Here we report the isolation and characterization of a full-length cDNA clone for CBF-C and describe how the three subunits of CBF associate with each other and bind to DNA.‡

MATERIALS AND METHODS

Purification of CBF-C. The purification of CBF-C followed a scheme similar to that used for CBF-A (6). Briefly, nuclear extracts from 150 Sprague–Dawley rat livers were chromatographed twice through a DNA affinity column containing the DNA sequence of the mouse $\alpha 2$(I) collagen promoter from −105 to −64 covalently linked to CNBr-Sepharose. The 1 M NaCl eluate was fractionated by Mono S ion-exchange chromatography, which separated CBF-A and CBF-C from CBF-B. The Mono S flow-through was loaded over a Mono Q ion-exchange column and eluted with a salt gradient. The fractions active in DNA binding when complemented with CBF-B were fractionated by SDS/PAGE, which showed three predominant polypeptide species of 45, 40, and 32 kDa. The 40-kDa species contained the CBF-C activity after elution from the gel, whereas the 32-kDa species was identified as CBF-A (6).

Cloning of CBF-C cDNA. The 40-kDa CBF-C polypeptide was digested with trypsin, and the resulting peptides were fractionated by HPLC. Amino acid sequences of five tryptic peptides (Table 1) were obtained. Based on codon tables and the choice of either inosine residues or degenerate residues at certain places, sense and antisense oligonucleotides were synthesized corresponding to a portion of peptides 2 and 4 of Table 1. Rat liver poly(A)$^+$ RNA served to synthesize first-strand cDNAs, which were then used as templates for PCR by using the two different combinations of oligonucleotides. Major DNA bands obtained after 35 cycles of amplification were cloned by blunt-end ligation in Bluescript KS and sequenced. One of the PCR clones, obtained with the sense oligonucleotide of peptide 2 [5′-GTGCAGGAGCTGCC (C/A)(C/T)TIGC-3′] and the antisense oligonucleotide of peptide 4 [5′-CAGGGTCAGCTCGGTGATGAAGAT (T/C)TG (A/G)GC-3′] contained sequences adjacent to the primer sequences that encoded the expected amino acid sequences in peptides 2 and 4 of Table 1. It also contained a sequence coding for peptide 3 of Table 1. This 129-bp PCR cDNA clone was used to screen 1×10^6 plaques of a rat brain cDNA library constructed in Lambda ZAP (Stratagene). After rescue of the vector/insert, DNA sequencing showed that several cDNA clones contained the full-length coding sequence specifying a polypeptide of 334 amino acids.

Expression of CBF-C and Two Other Subunits of CBF in *Escherichia coli*. An *Nde* I restriction site was generated in the initiating codon and an *Xho* I restriction site was generated following the stop codon in the cDNAs of each of the three CBF subunits by PCR, and these modified cDNAs were then cloned into the PET-23a vector (Novagen). The sequences of cDNAs obtained by PCR were verified by DNA sequencing. Synthesis of the recombinant polypeptides (rCBF-A, rCBF-B, and rCBF-C) was induced in the BL21(DE3)pLysS strain by

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MHC, major histocompatibility complex; r, recombinant; IPTG, isopropyl β-D-thiogalactopyranoside; GST, glutathione *S*-transferase.

*S.S. and S.N.M. contributed equally to this work.

†To whom reprint requests should be addressed.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. U17607).

Table 1. Tryptic peptide sequences of CBF-C protein

Peptide	Sequence
1	VMEEIR
2	(L)FRVQELPLAR
3	MISAEAPVLFAK
4	AAQIFITELTLR
5	FDQFDFLID(I)(V)P(R)

Single amino acids in parentheses represent probable amino acid residues.

adding 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 20 min at the midlogarithmic phase of bacterial growth. Bacteria were lysed by sonication in lysis buffer containing 50 mM Tris·HCl (pH 7.9), 100 mM NaCl, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 1 mM dithiothreitol, and 1% Triton X-100.

The recombinant subunits were purified over DEAE-cellulose followed by Mono S ion-exchange chromatography. rCBF-B was eluted from Mono S with 350 mM NaCl, whereas rCBF-A and rCBF-C were found in the flow-through fraction. rCBF-A and rCBF-C were further purified by chromatography over a Mono Q ion-exchange column: rCBF-C was present in the 100 mM NaCl flow-through, whereas rCBF-A was eluted with 250 mM NaCl.

For cloning glutathione *S*-transferase (GST)–CBF-C, an *Eco*RI restriction site was generated in the full-length CBF-C cDNA immediately preceding the initiating methionine and an *Xho* I site was generated after the stop codon by PCR. This modified CBF-C DNA was then subcloned in the pGEX-4T-3 (Pharmacia) vector and transformed in the HB101 strain. After induction with 1 mM IPTG for 20 min, the cells were pelleted, sonicated in lysis buffer, and centrifuged (16). The supernatant was loaded over glutathione-agarose beads, and GST–CBF-C was eluted by glutathione. Purification of GST–CBF-A and GST–CBF-B was as described (16).

Expression of HAP2 and HAP3 in *E. coli*. The HAP2 and HAP3 yeast polypeptides were expressed as GST fusion proteins. A *Bam*HI restriction site was generated at the initiating codon of the HAP2 and HAP3 DNAs and an *Eco*RI restriction site was generated after the stop codon. These modified DNAs were cloned in pGEX-2T vector (Pharmacia); subsequent IPTG induction in HB101 strain and purification of the rGST fusion polypeptides were done as described (16). HAP2 and HAP3 DNAs were isolated from yeast genomic DNA by PCR and the sequence of these PCR-generated clones was verified by DNA sequencing.

Generation of ^{35}S-Labeled CBF-C. The translation efficiency of RNA produced from the isolated CBF-C clone was very weak. To increase the efficiency of translation, we subcloned the open reading frame of CBF-C between the *Nde* I and *Xho* I restriction sites of the pCITE-2a vector (Novagen). The pCITE-2a containing CBF-C DNA (1 μg) was used for *in vitro* transcription by T7 RNA polymerase with an mCAP mRNA capping kit (Stratagene), and half of the RNA was translated using rabbit reticulocyte lysate in a total vol of 50 μl (Promega). CBF-C was labeled with [^{35}S]methionine during the translation reaction. Synthesis of the ^{35}S-labeled CBF-A and CBF-B was as described (5, 6).

Protein–Protein Interaction Assay. Five hundred nanograms of each GST–CBF subunit and 2 μl of each ^{35}S-labeled *in vitro* translated CBF subunit were mixed in separate reactions in a total vol of 20 μl of buffer I containing 50 mM Hepes, 100 mM KCl, 1 mM EDTA, 1% Nonidet P-40, 0.5 mM dithiothreitol, and 0.5 mM phenylmethylsulfonyl fluoride and incubated at room temperature for 20 min. After incubation for 10 min at room temperature with 2 μl (packed volume) of glutathione-agarose resin that was preequilibrated in buffer I, the GST–CBF subunits, which were bound to the glutathione-agarose resin, were precipitated by centrifugation. The precipitate was washed three times with 50 μl of buffer I and resuspended in 10 μl of SDS/PAGE sample buffer, which was loaded on a SDS/12% polyacrylamide gel. The ^{35}S-labeled CBF subunits, which were associated with the different GST–CBF subunit, were detected in the gel by autoradiography.

RESULTS

Purification of CBF-C and Isolation of cDNA Clones. Since the 40-kDa CBF-C species copurifies with CBF-A through several chromatographic steps (16), we followed the same procedure to purify CBF-C to homogeneity previously used to purify CBF-A (6, 17). At the last step in this purification, CBF-C was identified by its binding to a CCAAT-containing DNA after its elution from SDS/polyacrylamide gel and complementation with rCBF-A and rCBF-B. Amino acid sequences were obtained for several tryptic peptides of CBF-C shown in Table 1. We selected two peptides, pep2 and pep4, and synthesized oligonucleotides corresponding to the sense and antisense strands, respectively. These oligonucleotides were used in PCR experiments in which a preparation of first-strand cDNA of rat liver poly(A)$^+$ RNA was used as template. In one PCR cDNA clone of 129 bp, the DNA sequences adjacent to the oligonucleotides encoded the expected amino acid sequences in pep2 and pep4 (Table 1). In addition, it also coded for the amino acid sequence of pep3 (Table 1). This 129-bp PCR cDNA was used as hybridization probe to obtain a larger cDNA of 1.2 kb. The sequence of this cDNA contains an open reading frame of 1002 bp encoding a protein of 334 amino acids (Fig. 1). The calculated molecular mass of the deduced polypeptide is 40 kDa, which is identical with the molecular mass of native CBF-C measured by SDS/PAGE. The open reading frame has no homology with any protein in the Swiss-Prot data base. The N-terminal amino acids are rich in charged amino acids, whereas the sequences

```
GGGGGATAGGAGTGACTTAAATTCTCCCTGTCCGAGGAGATATAAATAACTCCATGTAAA 60
ATTAATGCAGTTCTCGGTGTCAAAATGTCCACAGAAGGAGGGTTTGGCAGTACCAGCAGC
                          M  S  T  E  G  G  F  G  S  T  S  S  12
AGTGATGCCCAGCAAAGCCTCCAGTCCTTCTGGCCCAGAGTCATGGAAGAAATCCGAAAC 180
S  D  A  Q  Q  S  L  Q  S  F  W  P  R  V  M  E  E  I  R  N
TTAACAGTGAAAGATTTCCGAGTACAAGAACTACCACTGGCTCGTATTAAGAAGATTATG
L  T  V  K  D  F  R  V  Q  E  L  P  L  A  R  I  K  K  I  M  52
AAACTGGATGAAGATGTGAAGATGATCAGTGCAGAAGCCCCTGTGCTGTTTGCTAAGGCA 300
K  L  D  E  D  V  K  M  I  S  A  E  A  P  V  L  F  A  K  A
GCCCAGATCTTCATCACAGAGCTGACTCTTCGAGCCTGGATCCACACAGAGGATAACAAG
A  Q  I  F  I  T  E  L  T  L  R  A  W  I  H  T  E  D  N  K  92
CGTCGTACTCTTCAGAGGAATGATATTGCTATGGCAATTACAAAATTTGATCAGTTTGAC 420
R  R  T  L  Q  R  N  D  I  A  M  A  I  T  K  F  D  Q  F  D
TTTCTCATCGACATTGTTCCAAGAGATGAACTGAAACCTCCAAAGCGCCAGGAGGAGGTA
F  L  I  D  I  V  P  R  D  E  L  K  P  P  K  R  Q  E  E  V  132
CGCCAGTCTGTGACTCCCGCGGAGCCTGTCCAGTACTACTTCACGCTGGCTCAGCAGCCC 540
R  Q  S  V  T  P  A  E  P  V  Q  Y  Y  F  T  L  A  Q  Q  P
ACTGCTGTCCAGGTCCAGGGACAGCAGCAAGGCCAGCAGACCACCAGTTCTACGACCACC
T  A  V  Q  V  Q  G  Q  Q  Q  G  Q  Q  T  T  S  S  T  T  T  172
ATCCAGCCTGGCCAGATCATCATTGCGCAGCCTCAACAGGGTCAGACCACACCCGTGACC 660
I  Q  P  G  Q  I  I  I  A  Q  P  Q  Q  G  Q  T  T  P  V  T
ATGCAGGTTGGAGAAGGTCAGCAGGTGCAGATTGTGCAGGCCCAACCTCAGGGTCAGGCC
M  Q  V  G  E  G  Q  Q  V  Q  I  V  Q  A  Q  P  Q  G  Q  A  212
CAGCAGACCCAGAGTGGTACTGGACAGACCATGCAGGTGATGCAGCAGATCATTACCAAC 780
Q  Q  T  Q  S  G  T  G  Q  T  M  Q  V  M  Q  Q  I  I  T  N
ACAGGGGAGATCCAACAGATCCCGGTGCAGCTGAATGCCGGCCAGTTGCAGTATATCCGC
T  G  E  I  Q  Q  I  P  V  Q  L  N  A  G  Q  L  Q  Y  I  R  252
TTAGCCCAGCCTGTATCAGGCACCCAAGTTGTGCAGGGACAGATCCAGACCCTTGCTACC 900
L  A  Q  P  V  S  G  T  Q  V  V  Q  G  Q  I  Q  T  L  A  T
AATGCCCAACAGATCACACAGACAGAGGTCCAACAAGGACAGCAGCAGTTCAGCCAGTTC
N  A  Q  Q  I  T  Q  T  E  V  Q  Q  G  Q  Q  Q  Q  F  S  Q  F  292
ACAGACGGACAGCTGTACCAGATCCAGCAAGTCACCATGCCTGCAGGCCAAGACCTTGCC 1020
T  D  G  Q  L  Y  Q  I  Q  Q  V  T  M  P  A  G  Q  D  L  A
CAGCCCATGTTTATTCAGTCAGCCAACCAGCCCTCTGATGGGCAGACCCCCCAGGTGACT
Q  P  M  F  I  Q  S  A  N  Q  P  S  D  G  Q  T  P  Q  V  T  332
GGAGACTGAGGCCTGAGCTGCAAAGCCAAGGACCCCCAACACAATATTTGCCATAGAGCC 1140
G  D
CCCAGGCGATGGGGACACACCTTCCCTCACCAGAGGACCCGGGACTTCATTGCCTCCTGC
```

FIG. 1. Nucleotide sequences of rat CBF-C cDNA and deduced amino acid sequence of rat CBF-C. Amino acid sequences (single-letter code) of five tryptic peptides (Table 1) are underlined. Initiating methionine codon is preceded by stop codons in all three frames. Comparison of amino acid sequence of CBF-C with the Swiss-Prot data base revealed no homology to any known protein.

from amino acid 150 to the C terminus is very hydrophobic and rich in glutamine residues. A similar, long glutamine-rich hydrophobic domain was also found in the amino acid sequence of CBF-B (5).

DNA Binding of rCBF-C. To test the function of rCBF-C in DNA binding, we expressed the CBF-C cDNA clone in *E. coli* by using two different expression vectors. In one expression vector, the full-length rCBF-C was synthesized in bacteria and the recombinant protein was partially purified by FPLC ion-exchange chromatography. In another expression vector, the full-length CBF-C was fused with the DNA sequences of GST and was synthesized as a fusion protein (GST–CBF-C). The rGST–CBF-C protein was purified by using glutathione affinity resin (see Fig. 4*A*). The two other subunits of CBF, CBF-A and CBF-B, were also synthesized by using the same two expression vectors and were also purified by similar procedures. No DNA binding occurred with any of the three combinations of two recombinant subunits (Fig. 2, lanes 2–4); however, when all three subunits were present, a CBF–DNA complex formed that had an electrophoretic mobility similar to that of the complex formed by the native CBF subunits and DNA (lane 5). To test whether all three CBF subunits participated in this complex, each of these subunits was replaced in three separate binding reactions by its GST homologue, which adds 26 kDa to the size of each subunit. In each case (lanes 6–8), a slower mobility complex was produced, indicating that all three CBF subunits were present in the CBF–DNA complex. The small differences in mobility between these complexes could have been due either to differences in conformation between the GST derivatives of the CBF subunits or to differences in the stoichiometry of the subunits within the CBF molecule. In another experiment, we also used CBF-C synthesized in rabbit reticulocyte lysate obtained after *in vitro* transcription and translation. In this system, the formation of a CBF–DNA complex was also dependent on the presence of all three CBF subunits (data not shown). Moreover, the SDS/PAGE mobility of 40 kDa for *in vitro* translated CBF-C was in excellent agreement with that of the native CBF-C polypeptide (data not shown) (16).

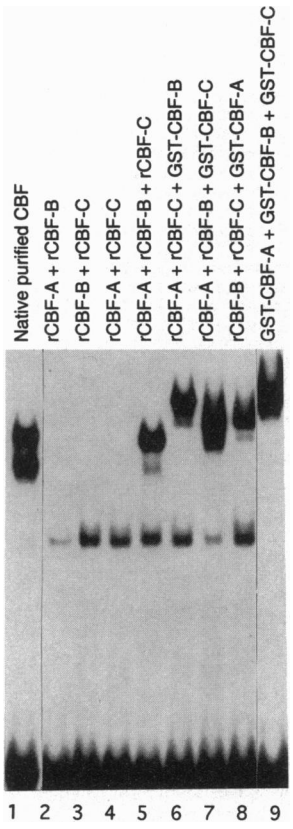

FIG. 2. DNA binding activity of CBF-C. Gel-shift assays using *E. coli* synthesized rCBF subunits. Bacteria transformed with PET-23a vectors containing the coding sequences of CBF-A, CBF-B, or CBF-C were treated with IPTG. Lanes 2–8, 2 μl of supernatant (10 μg of total protein) of appropriate *E. coli* lysates was used as source of rCBF-A, rCBF-B, and rCBF-C as indicated; lanes 6–8, 10 ng of GST–CBF-A, 20 ng of GST–CBF-B, and 20 ng of GST–CBF-C was used in the binding reactions. DNA binding reactions similar to those shown in lanes 2–5 were also performed with the semipurified rCBF subunits and gave results identical to those shown in lanes 2–5. A nonspecific retarded species appears in all lanes containing crude *E. coli* extracts; this nonspecific species disappeared when semipurified rCBF subunits were used. Reaction mixture shown in lane 1 contained 2 ng of purified rat liver CBF-B and 1 ng of a purified fraction containing both rat liver CBF-A and CBF-C. Reaction mixture shown in lane 9 contained 10 ng of each of the GST-derived subunits.

Binding of the three recombinant subunits of CBF to the CCAAT sequence in the mouse α2(I) collagen promoter was specific since no binding took place when the point mutation CCAAT → CCAAA was introduced into the motif (Fig. 3, lane 8). CBF-C was also required for the formation of CBF complexes with functional CCAAT motifs in other promoters, such as those of the α1(I) collagen, albumin, and MHC class II genes (1–3), to which cruder preparations of CBF had been previously shown to bind (lanes 1–6).

The yeast CCAAT-binding transcription factor has been previously reported to contain three different polypeptides—HAP2, HAP3, and HAP4 (14). This factor controls nuclear genes that encode mitochondrial proteins involved in growth on nonfermentable carbon sources. HAP2 is the yeast homologue of mammalian CBF-B and HAP3 is the homologue of CBF-A (5–7). The high degree of sequence conservation of HAP2 and CBF-B includes both a segment needed for DNA binding and an adjacent segment required for subunit interactions in the two polypeptides (11, 15). Similarly, in HAP3 and CBF-A the conserved segment is needed for DNA binding and presumably, subunit interactions (ref. 12; K. Y. Sohn, S.S., S.N.M., and B.d.C., unpublished data). Although HAP4 is present in the DNA–protein complex and provides a strong transcription activation function in yeast, it is not required for DNA binding (12, 15).

As could be expected from their sequence homologies with CBF-A and CBF-B, a mixture of purified rGST–HAP2 and rGST–HAP3 (Fig. 4*A*, lanes 1 and 2) was unable to bind to either the α2(I) collagen CCAAT sequence or the CCAAT sequence located in the promoter of the yeast *CYC1* gene, which is controlled by HAP2, HAP3, and HAP4 (Fig. 4*B*, lanes 1 and 5) (17). However, DNA binding took place when either rCBF-C (data not shown) or its purified GST derivative was added to purified GST–HAP2 and GST–HAP3 (lanes 2 and 6). In a separate experiment, rHAP2 and rHAP3 polypeptides, which were not GST fusion polypeptides, were also generated by *in vitro* transcription and translation in reticulocyte lysates. In gel-shift assays a new DNA–protein complex could be detected that migrated faster than the complex be-

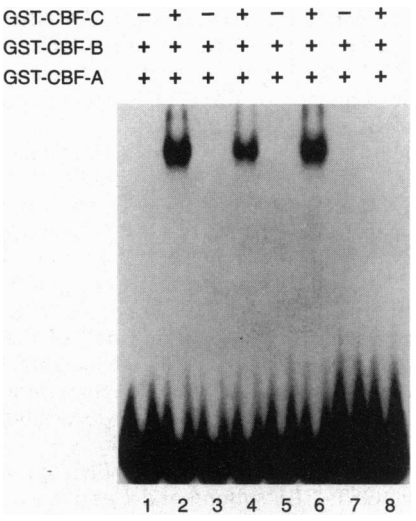

FIG. 3. CBF-C is needed for DNA binding to CCAAT-containing sequences in other promoters. Sequences of the oligonucleotides were from the mouse α1(I) collagen promoter (−120 to −77) (lanes 1 and 2), mouse albumin promoter (−106 to −67) (lanes 3 and 4), and mouse MHC class II promoter (−79 to −37) (lanes 5 and 6). Lanes 7 and 8, sequence of the mouse α2(I) collagen promoter (−105 to −64) containing a mutation in the CCAAT motif (CCAAT → CCAAA) was used. Amount of GST fusion proteins used in the binding reaction mixture was as described in Fig. 2 (lane 9).

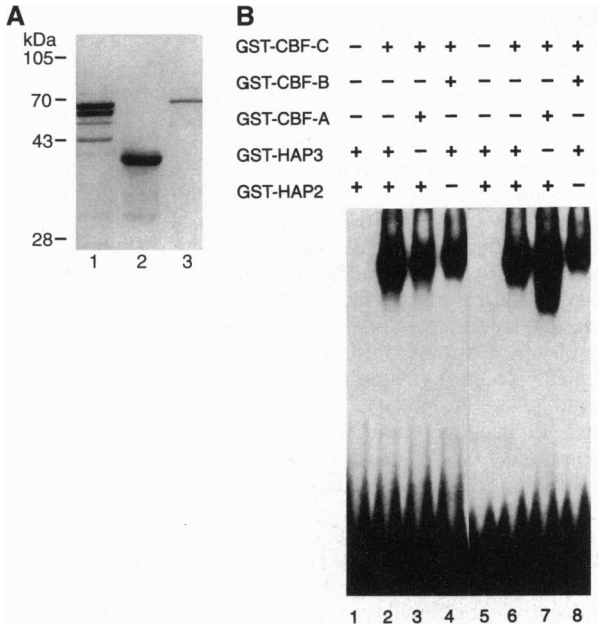

FIG. 4. CBF-C is needed for DNA binding of HAP2 and HAP3. (*A*) SDS/PAGE of purified GST–HAP2, GST–HAP3, and GST–CBF-C; 500 ng of GST–HAP2 (lane 1), 500 ng of GST–HAP3 (lane 2), and 100 ng of GST–CBF-C (lane 3) were fractionated on a SDS/12% polyacrylamide gel and the gel was stained with Coomassie blue. Molecular markers are indicated on the left. Migration of fusion proteins was as expected from their molecular mass. (*B*) Gel-shift assay using the mouse α2(I) collagen promoter sequence (lanes 1–4) and the yeast UAS2UP1 sequence (13) (−225 to −187) (lanes 5–8). Twenty nanograms of each of the indicated recombinant polypeptides was used in the binding reaction mixtures.

tween DNA and the endogenous CBF-like protein of the lysate only when lysate containing *in vitro* translated CBF-C was added to the lysates containing HAP2 and HAP3 polypeptides (data not shown); the faster mobility of this complex was due in part to the smaller size of HAP2 and HAP3 compared to CBF-B and CBF-A.

In the binding reactions of Fig. 4, GST–HAP3 could be replaced by GST–CBF-A (Fig. 4*B*, lanes 3 and 7) and GST–HAP2 could be replaced by GST–CBF-B (lanes 4 and 8), indicating that the DNA binding and subunit interaction functions of these polypeptides are interchangeable. These experiments suggest that these two functions of CBF are completely conserved between yeast and rodents. Similar implications had been made previously based on experiments in which crude preparations containing either yeast HAP2 or human CBF-B and other preparations described as containing either HAP3 or human CBF-A were used (18). The results of our *in vitro* biochemical experiment also imply the existence of a CBF-C homologue in yeast that together with HAP2 and HAP3 is needed for DNA binding and subsequent activation of the promoters controlled by these factors.

Interactions Between the Three Subunits of CBF. To determine how the three rCBF subunits interact with each other, we performed an *in vitro* protein–protein interaction assay. The three subunits were synthesized as ^{35}S-labeled polypeptides after transcription of the corresponding cDNA and translation of the respective RNA in a rabbit reticulocyte lysate. To examine whether a labeled CBF subunit specifically interacted with any of the GST–CBF subunits, each of the ^{35}S-labeled CBF subunits was incubated in separate reaction mixtures with each of the GST–CBF subunits. The reaction mixtures were further incubated with glutathione-agarose resin to precipitate GST–CBF subunits and complexes formed with these GST subunits. The precipitates were analyzed in a SDS/polyacrylamide gel to detect labeled CBF subunits in the complex. In this assay, the radiolabeled CBF-A was precipitated only with GST–CBF-C (Fig. 5*A*, lane 5) but not by GST alone, GST–CBF-B, or GST–CBF-A (Fig. 5*A*, lanes 2–4). Similarly, the radiolabeled CBF-C was precipitated only with GST–CBF-A (Fig. 5*B*, lane 3). The radiolabeled CBF-B could not be precipitated with either GST–CBF-A or GST–CBF-C alone (Fig. 5*C*, lanes 3–6), but it was precipitated only when both GST–CBF-A and GST–CBF-C were present together in the reaction mixture (Fig. 5*C*, lane 2). The absence of self-association between identical subunits suggested that, in this assay, none of the subunits of CBF formed homodimers. We conclude from this experiment that CBF-A and CBF-C associate with each other to form a CBF-A–CBF-C complex; this complex then associates with CBF-B to form a ternary (CBF-A–CBF-C) CBF-B complex, which is capable of binding to DNA (Fig. 6). This experiment also demonstrates that formation of this ternary complex can occur in the absence of DNA binding.

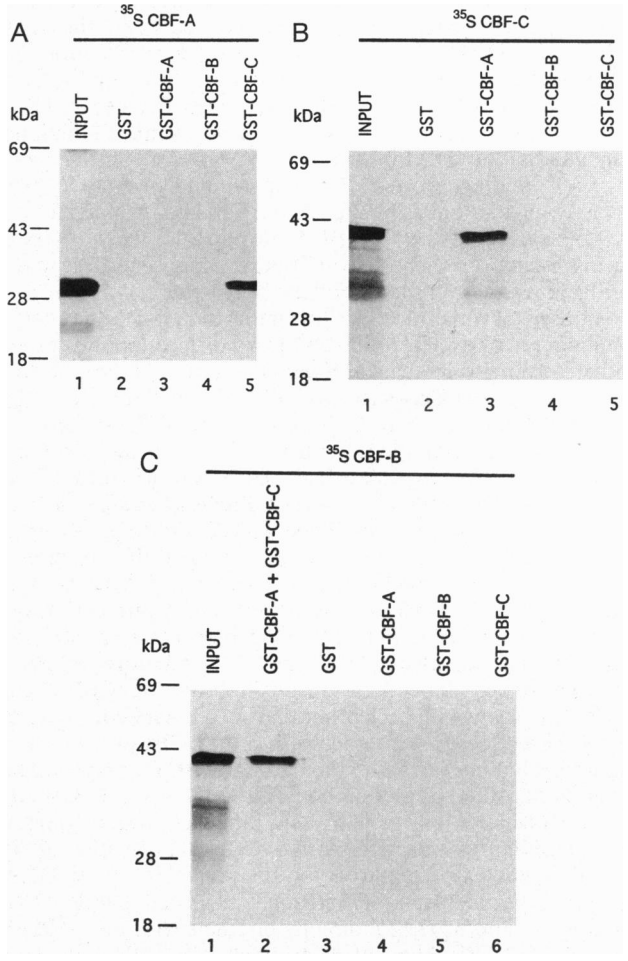

FIG. 5. CBF-C interacts with CBF-A, and CBF-B interacts with the CBF-A–CBF-C complex. (*A*) *In vitro* translated ^{35}S-labeled CBF-A (2 μl) was incubated in separate reaction mixtures with 500 ng of each GST–CBF subunit or 500 ng of GST as indicated. Reaction mixtures were then incubated with 5 μl (packed volume) of glutathione-agarose resin to precipitate GST or GST fusion proteins. Proteins bound to the glutathione-agarose resin were solubilized by 10 μl of SDS/PAGE sample buffer and fractionated in a SDS/12% polyacrylamide gel. The bound labeled CBF-A was detected by autoradiography. (*B* and *C*) Similar experiments were performed by using radiolabeled ^{35}S-CBF-C and ^{35}S-CBF-B, respectively. (*C*) Lane 2, a mixture of 250 ng of GST–CBF-A and 250 ng of GST–CBF-C was incubated with ^{35}S-CBF-B.

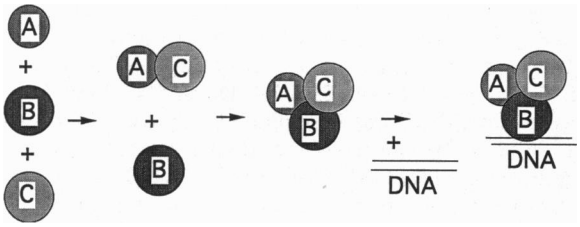

FIG. 6. Pathway for assembly of CBF subunits. The first two steps in the assembly of the CBF subunits are based on the results of Fig. 5. Evidence exists for interactions between the DNA binding domain of CBF-B and the CCAAT sequence (ref. 12; S.N.M., unpublished observation). CBF footprints on DNA cover at least 20 bp in which the CCAAT motif is acentric, suggesting that additional interactions occur possibly through CBF-A and/or CBF-C.

DISCUSSION

We have used rCBF-C to demonstrate that the three different subunits of CBF are needed for binding to CCAAT motifs in DNA and that all three polypeptides are present in the CBF–DNA complex. Furthermore, our studies also establish that CBF-C allows formation of a complex between the yeast HAP2 and HAP3 polypeptides and a CCAAT-containing DNA and show that CBF-C is present in this complex. Both CBF-A and CBF-B show in the sequences involved in DNA binding a high degree of conservation with similar functional domains in the HAP2 and HAP3 subunits of the yeast CCAAT binding protein. The experiments presented here strongly suggest that as in CBF-A and CBF-B, a segment of the amino acid sequence of CBF-C should also be conserved during evolution and that this putative conserved segment fulfills an essential function of the polypeptide.

Many eukaryotic transcription factors bind to DNA either as homodimers or as heterodimers of two different polypeptides and are often grouped according to the nature of their protein dimerization interfaces such as leucine zippers, or coiled coils, and helix–loop–helix motifs (19–21). None of the subunits of CBF shows any homology with the known protein–protein interaction motifs, suggesting that CBF may represent a unique heteromeric DNA binding protein. Previous studies of both CBF-B and its yeast homologue HAP2 indicated that the subunits of CBF associate with each other in the absence of DNA and that an evolutionarily conserved, 21-amino acid sequence in CBF-B represented the subunit interaction domain of this subunit (11, 22). Our present study strongly suggests that the assembly of the CBF subunits follows a specific pathway. Indeed, our results indicate that CBF-A and CBF-C associate with each other to form a binary complex and that CBF-B does not interact with either CBF-A or CBF-C separately but interacts with the CBF-A–CBF-C complex. This suggests that the interaction between CBF-A and CBF-C results in formation of a new protein interface, which interacts with the subunit interaction motif of CBF-B. Formation of this ternary complex is required for binding of CBF to DNA. Genetic experiments in yeast strongly suggested that the DNA binding domain of HAP2 directly interacts with the CCAAT sequence (12). Biochemical experiments with CBF-B also support this notion (S.N.M., unpublished observations). Based on CBF DNA footprints of at least 20 bp in which the CCAAT sequence is acentric, we believe that additional interactions with DNA occur, possibly through CBF-A and/or CBF-C, but the precise domains of these interactions within the CBF subunits still need to be defined.

Because this pathway of subunit interactions is unique among DNA binding proteins and because of the high degree of conservation of analogous functional domains in the subunits of CBF and those of the yeast CCAAT binding protein, one can postulate that CBF could have a unique role in the transcription activation process of eukaryotic promoters. Given that CBF is a ubiquitous protein and given that in various higher eukaryotic promoters the functional CBF binding sites are often located around −80 at a relatively close distance of the TATA motif (23), one can postulate that CBF interacts both with a variety of upstream sequence-specific DNA binding transcription factors and with some of the proteins involved in the formation of preinitiation transcription complexes. Although indirect evidence for cooperativity between upstream activators and CBF has previously been reported (24), the specific roles of CBF in the transcription activation process remain to be established. The availability of recombinant CBF-C together with recombinant CBF-A and CBF-B should help in studies aimed at better understanding these functions.

We thank Dr. Ken Williams (W. M. Keck Foundation, Yale University, New Haven, CT) for performing amino acid sequence analysis. We thank Michele Sawadogo and Randy Legerski for discussions. We are grateful to Patricia McCauley for editorial assistance. This work was supported by Grants CA49515 and CA16672 from the National Institutes of Health. S.N.M. was supported by an Arthritis Foundation Investigator Award.

1. Maity, S. N., Golumbek, P. T., Karsenty, G. & de Crombrugghe, B. (1988) *Science* **241,** 582–585.
2. Maire, P., Wuarin, J. & Schibler, U. (1989) *Science* **246,** 343–346.
3. Dorn, A., Bollekens, J., Staub, A., Benoist, C. & Mathis, D. (1987) *Cell* **50,** 863–872.
4. Chodosh, L. A., Baldwin, A. S., Carthew, R. W. & Sharp, P. A. (1988) *Cell* **53,** 11–24.
5. Maity, S. N., Vuorio, T. & de Crombrugghe, B. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 5378–5382.
6. Vuorio, T., Maity, S. N. & de Crombrugghe, B. (1990) *J. Biol. Chem.* **265,** 22480–22486.
7. van Huijsduijnen, R. H., Li, X. Y., Black, D., Matthes, H., Benoist, C. & Mathis, D. (1990) *EMBO J.* **9,** 3119–3127.
8. Becker, D. M., Fikes, J. D. & Guarente, L. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1968–1972.
9. Hahn, S. J., Pinkham, R. W., Miller, R. & Guarente, L. (1988) *Mol. Cell. Biol.* **8,** 655–663.
10. Pinkham, J. L., Olesen, J. T. & Guarente, L. (1987) *Mol. Cell. Biol.* **7,** 575–578.
11. Maity, S. N. & de Crombrugghe, B. (1992) *J. Biol. Chem.* **267,** 8286–8292.
12. Xing, Y., Fikes, J. D. & Guarente, L. (1993) *EMBO J.* **12,** 4647–4655.
13. Olesen, J., Hahn, S. & Guarente, L. (1987) *Cell* **51,** 953–961.
14. Forsberg, S. L. & Guarente, L. (1989) *Genes Dev.* **3,** 1166–1178.
15. Olesen, J. T. & Guarente, L. (1990) *Genes Dev.* **4,** 1714–1729.
16. Maity, S. N., Sinha, S., Ruteshouser, C. E. & de Crombrugghe, B. (1992) *J. Biol. Chem.* **267,** 16574–16580.
17. Forsberg, S. L. & Guarente, L. (1988) *Mol. Cell. Biol.* **8,** 647–654.
18. Chodosh, L. A., Olsen, J., Hahn, S., Baldwin, A. S., Guarente, L. & Sharp, P. A. (1988) *Cell* **53,** 25–35.
19. Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. & McKnight, S. L. (1988) *Genes Dev.* **2,** 786–800.
20. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56,** 777–783.
21. O'Shea, E. K., Rutrowski, R. & Kim, P. S. (1989) *Science* **243,** 538–542.
22. Xing, Y., Zang, S., Olesen, J. T., Rich, A. & Guarente, L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3009–3013.
23. Bucher, P. (1990) *J. Mol. Biol.* **212,** 563–578.
24. Milos, P. & Zaret, K. S. (1992) *Genes Dev.* **6,** 991–1004.