Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 1968–1972, March 1991
Genetics

A cDNA encoding a human CCAAT-binding protein cloned by functional complementation in yeast

(*Saccharomyces cerevisiae*/*Schizosaccharomyces pombe*/Hap2/CP1/NF-Y)

DANIEL M. BECKER, JOHN D. FIKES, AND LEONARD GUARENTE

Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Communicated by Gerald R. Fink, November 9, 1990

ABSTRACT **We constructed a comprehensive cDNA library from HeLa cell mRNA in a vector that directs expression of the cDNA in *Saccharomyces cerevisiae*. We used this library to clone the human counterpart of the *Sa. cerevisiae* CCAAT-binding transcription factor, Hap2, by functional complementation of a *hap2* mutation. The cDNA encoding the human Hap2 homolog encodes a protein of 257 amino acids that has a 62-amino acid carboxyl-terminal region 73% identical to the essential core region of Hap2. The amino terminus of the protein is highly enriched in glutamine residues, reminiscent of transcriptional activation domains of several other mammalian transcription factors. Analysis of human Hap2 expression reveals three major transcripts: a 4.1-kilobase species found in all cell types examined, a 7.0-kilobase species specific to B lymphocytes, and a 1.6-kilobase species that is expressed preferentially in HeLa cells and that likely corresponds to our cDNA clone. Thus, the human Hap2 homolog and related factors may play both a constitutive and cell type-specific role in gene expression. The general approach of cloning by complementation should allow the isolation of many human genes for which corresponding yeast mutations exist.**

The machinery governing the regulation of transcription has been remarkably well conserved through the evolution of eukaryotes. One example is provided by studies of the CCAAT-binding complex, Hap2/3/4, of the budding yeast *Saccharomyces cerevisiae*. The genes encoding Hap2, Hap3, and Hap4 were identified as genetic loci required for the expression of respiratory competence in *Sa. cerevisiae*—in particular, for expression of the principal isoform of cytochrome *c* (*CYC1*; for review, see ref. 1). The three proteins associate into a heteromeric complex that binds to an upstream activation sequence in the *CYC1* promoter (2, 3). The consensus sequence recognized by the Hap2/3/4 complex, CCAAT, is found in slightly varying forms in the promoters of other yeast genes that are required for respiration, permitting coordinate regulation (4).

CCAAT is an upstream sequence element that is also found in a multitude of higher eukaryotic promoters (5); it serves as the recognition sequence for a variety of mammalian transcription factors (6, 7). There are at least three such chromatographically separable CCAAT-binding factors in HeLa cells: CP1, CP2, and CTF/NF-1. These factors recognize overlapping, but distinct, subsets of known CCAAT-containing promoters and make distinguishable patterns of contacts with DNA in and around the CCAAT motif (8).

Of these three factors, CP1 bears the greatest resemblance to the yeast Hap complex. Its promoter preference *in vitro* and its contacts with DNA are identical to those of Hap2/3/4. Like Hap2/3/4, CP1 consists of a heteromeric association of at least two components, CP1A and CP1B, both of which are required for binding. Most strikingly, the subunits of CP1 and Hap2/3/4 can be interchanged *in vitro*. Yeast cell extracts from a strain bearing a *hap3* deletion will bind DNA in association with components provided by the CP1A chromatographic fraction; conversely, extracts from a strain lacking Hap2 expression can be complemented by components within the CP1B fraction (9). Thus, CP1 likely represents the human homolog of the yeast Hap complex, with the CP1B fraction containing the Hap2 homolog and the CP1A fraction containing a Hap3 equivalent.

The ability to substitute the human CP1 subunits for their yeast counterparts *in vitro* suggested that expression of the human proteins *in vivo* might well yield functional hybrid complexes in yeast strains bearing individual *hap* mutations. We report here the isolation of a HeLa cDNA whose expression in *Sa. cerevisiae* corrects the respiratory defect in a strain bearing a *hap2* deletion.* The general approach of cloning genes from heterologous species by complementation of yeast mutations has been previously reported (10–13).

MATERIALS AND METHODS

cDNA Synthesis. cDNA was synthesized from mRNA isolated from HeLa cells or from *Schizosaccharomyces pombe* cells by using RNase H^- Moloney murine leukemia virus reverse transcriptase (Superscript RT; Life Technologies/BRL) as described by the manufacturer.

Library Construction. (*i*) We constructed two *Sa. cerevisiae* expression plasmids, pDB10 and pDB20 (Fig. 1), by standard techniques. pDB10 contains the *CYC1* promoter, which directs low-level expression of the cloned cDNA, and pDB20 carries the *ADC1* (*ADH1*) promoter, which drives high levels of expression (14). Both vectors incorporate the high-efficiency inverted *Bst*XI cloning strategy of Aruffo and Seed (15). pDB20 plasmid was prepared for cDNA insertion by *Bst*XI digestion (New England Biolabs) followed by four rounds of spin dialysis in a Centricon-100 (Amicon) concentrator to remove quantitatively the 26-base-pair (bp) stuffer liberated by *Bst*XI digestion. (*ii*) cDNA was prepared for insertion into vector by ligation to phosphorylated, partially double-stranded adaptors (InVitrogen, San Diego). Excess adaptors and short cDNA products were removed after ligation by Sephacryl S-300 spin column chromatography (Pharmacia). (*iii*) Linkered cDNA was ligated to pDB20 at 16°C overnight. Ligation reactions were extracted with phenol/chloroform, precipitated, and resuspended in H_2O. A fraction of this ligation was transformed into *Escherichia coli* XL-1 Blue (Stratagene) by electroporation (16), and transformants were plated at a density of $0.5–1.5 \times 10^6$ colonies per 150-mm plate. Colonies were scraped from the surface of the plates and plasmid was recovered directly from the pooled bacteria, without additional amplification, by alkaline lysis followed by CsCl gradient centrifugation. The percentage and size distribution of inserts was determined for each

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.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. M59079).

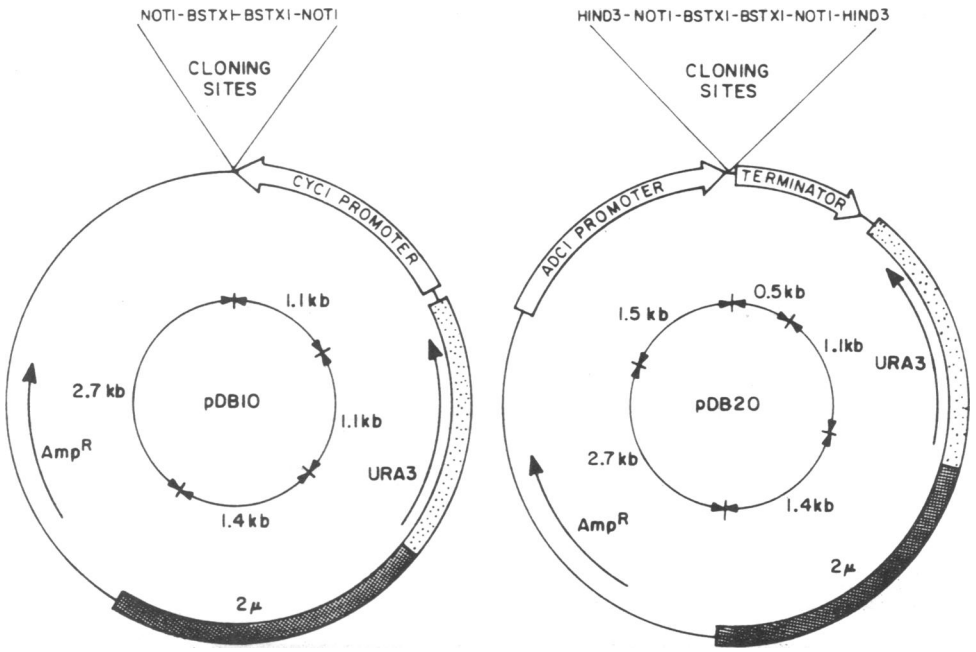

FIG. 1. cDNA expression vectors. Plasmids pDB10 and pDB20 are drawn approximately to scale. Yeast elements are drawn as thicker segments and include the origin of replication from the 2-μm circle for high-copy episomal maintenance in yeast, the *URA3* selectable marker, and one of two yeast promoters: *CYC1* or *ADC1* (*ADH1*). Direction of transcription from the yeast promoters is indicated. The pUC18 plasmid backbone is drawn as a single line. The cDNA insertion site of each vector consists of two identical, nonpalindromic *Bst*XI restriction sites in inverted orientation, separated by a 26-bp stuffer segment. *Not* I sites flank the cloning site to permit excision of the insert with an enzyme likely not to cut within the cDNA.

library by plasmid isolation from 12 colonies selected at random.

Yeast Strains and Media. *Sa. cerevisiae* strain JO1-1a (*MATa leu2-3,112 ura3-52 his4-519 adel-100 Δhap2*; ref. 17) was transformed (18) with plasmids from either the *Sc. pombe* or the HeLa cDNA libraries. Transformants were selected in 3% minimal top agar (0.67% yeast nitrogen base lacking amino acids/leucine, histidine, and adenine each at 40 mg per liter/0.1% Casamino acids/2% glucose/1 M sorbitol/3% agar) that was solidified on a bed of 2% agar of otherwise identical composition. After visible colony growth, transformants were recovered from the top agar and spread on rich medium (1% yeast extract/2% peptone/2% agar) containing 2% lactate (pH 6).

DNA Sequencing. Clones isolated from the HeLa cDNA library by complementation were subcloned into pBluescript SK(+) (Stratagene) and sequenced (Sequenase version 2, United States Biochemical) on both strands.

Northern (RNA) Analysis. RNA was resolved by electrophoresis through a 1% agarose gel containing 1.5% formaldehyde and blotted onto GeneScreen*Plus* membranes (DuPont/NEN) according to the manufacturer's instructions. Membranes were probed with a 600-bp coding region fragment isolated from clone HuHAP2.2. Transcript sizes were calculated from the migration in parallel of synthetic RNA species [0.24- to 9.5-kilobase (kb) RNA ladder, BRL] visualized by staining with ethidium bromide.

RESULTS

Cloning of Hap2 Homologs. cDNA libraries were constructed in the *Sa. cerevisiae* expression plasmid pDB20 (Fig. 1) from mRNA isolated from HeLa cells or the fission yeast *Sc. pombe*. For each library we were able to obtain $>2 \times 10^7$ primary *E. coli* transformants per μg of input mRNA, with >80% of the clones containing inserts.

To identify human and fission yeast homologs of Hap2, we used *Sa. cerevisiae* strain JO1-1a. Deletion in this strain of the gene encoding Hap2 disrupts expression of several genes required for respiratory competence; complementation restores the ability of the strain to grow on nonfermentable carbon sources such as lactate. Recent deletion analysis of the *Sa. cerevisiae* Hap2 protein demonstrates that only a portion of the yeast Hap2 protein is required to complement the chromosomal deletion in JO1-1a. This 65-amino acid region, termed the essential core, encodes both the DNA-binding activity and the subunit-interaction surfaces of the protein (17).

We transformed 12 μg of each of the cDNA libraries in parallel into JO1-1a. Yeast transformants were selected first for uracil prototrophy in the presence of glucose and then replated as a pool on rich medium containing lactate as the sole carbon source. Transformation with 12 μg of library yielded $>1 \times 10^6$ primary yeast transformants, and numerous colonies were obtained in each case after selection on lactate. The replating of transformants precludes an accurate calculation of the frequency with which independent complementing cDNAs occur in each library, but several clearly independent clones were examined in detail (see Fig. 2).

Two isolates of the human Hap2 homolog were characterized. The 1.64-kb HuHAP2.1 cDNA clone matches in size the principal message seen in HeLa cells by Northern analysis (see below). The second, HuHAP2.2, is truncated on both 5′ and 3′ ends. Sequencing the clones, as detailed in the following section, revealed a region within the human gene homologous to the Hap2 core; this region is indicated by denser shading in Fig. 2. HuHAP2.2, while truncated, retains this essential core. In addition, the lactate$^+$ phenotype it confers is indistinguishable from that of HuHAP2.1, evidence that a region of no more than 157 amino acids, encompassing the human Hap2 core, is sufficient for full complementation in yeast.

Three distinguishable classes of *Sc. pombe* Hap2 clones were characterized. SpHAP2.1 is a full-length 2.2-kb cDNA containing the entire 334 amino acid open reading frame also identified within a genomic clone (19). The position of its conserved core is indicated. SpHAP2.2 is also full length but is inverted relative to the *ADH* promoter of the vector. Yeast carrying this clone grew less well on lactate than colonies containing SpHAP2.1. This and additional experiments (data not shown) demonstrate that the vector contains within the *ADH* terminator a low-level cryptic promoter in opposing direction.

The third type of *Sc. pombe* clone confers a phenotype on lactate identical to that of SpHAP2.1 and is transcribed correctly from the *ADH* promoter. Like the human HuHAP2.2 clone, it contains a truncation of a substantial portion of the coding region while retaining the essential core. Surprisingly, there is a poly(A) tract at the apparent 5′ end of the clone. Fig. 2 *Lower* depicts the likely origin of this clone as a cDNA copy of an overlapping antisense transcript. The

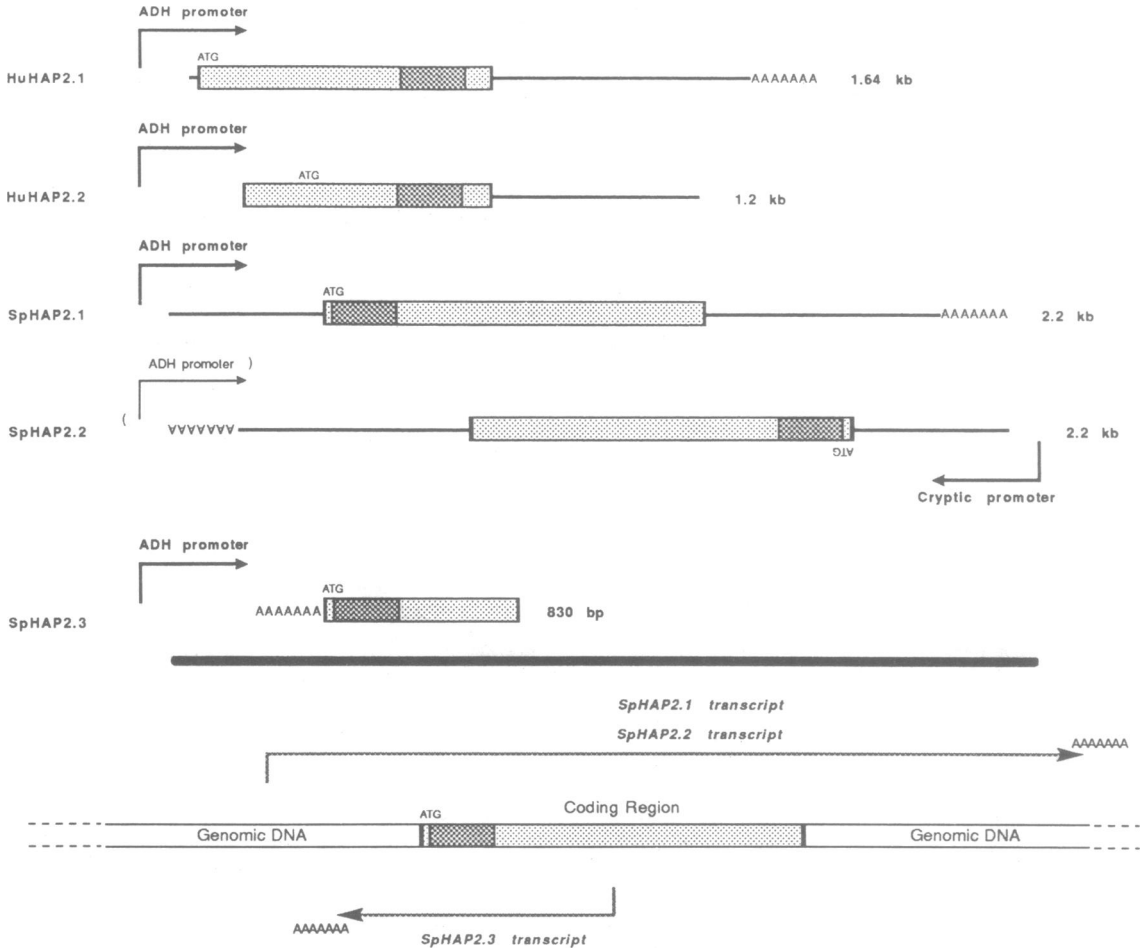

FIG. 2. Human and *Sc. pombe* Hap2 homologs. Twelve micrograms of each cDNA expression library was introduced in parallel into yeast strain JO1-1a (*MATa leu2-3,112 ura3-52 his4-519 adel-100 Δhap2*). Yeast transformants were selected first for uracil prototrophy in the presence of glucose and then replated on rich medium containing 2% lactate (pH 6). Plasmids conferring respiratory competence were transformed into *E. coli* and characterized by restriction mapping and sequencing. (*Upper*) Two isolates of the human Hap2 homolog (clones HuHAP2.1 and HuHAP2.2) and three isolates of the *Sc. pombe* Hap2 homolog (SpHAP2.1, SpHAP2.2, SpHAP2.3). Inserts are drawn to scale, and their orientation relative to the *ADH* promoter of the vector is shown. Coding regions are indicated as rectangles; the conserved core region is more heavily shaded. Location of ATG codons presumed to direct translation initiation is marked. (*Lower*) Genomic organization of the *Sc. pombe HAP2* gene. The separate transcripts that probably give rise to the various *Sc. pombe* cDNA clones are shown.

presence of such a transcript is not unprecedented: the *Sa. cerevisiae HAP3* gene is coincident with an overlapping antisense transcription unit of unknown function (20).

Structure of the Human Hap2 Homolog. We sequenced 1347 nucleotides from the 5′ end of clone HuHAP2.1 (Fig. 3). A single open reading frame of 257 amino acids was found, predicting a protein of 27.7 kDa, consistent with the 25-kDa molecular mass measured by sedimentation of partially purified CP1B activity (8). Two major features of the predicted structure are noteworthy.

The amino terminus of the protein contains a marked concentration of glutamine residues: glutamine constitutes 32 of the first 100 residues and fully 40% of the residues from amino acids 10 to 70. Clustering of the glutamine residues suggests that this amino-terminal region represents a discrete functional domain. Glutamine-rich regions in a number of mammalian transcription factors are potent activators of transcription in higher eukaryotic cells (21–23). Whatever the function of this domain in mammalian cells, however, it is dispensable in yeast; clone HuHAP2.2 lacks this region but fully complements the *hap2* null mutation. In yeast, therefore, the principal activation function of the hybrid complex is probably contributed by the acidic *Sa. cerevisiae* Hap4 subunit, as in the native Hap heteromer (24).

Near the carboxyl terminus is a region bearing striking similarity to the essential core of Hap2. In this basic region there is 73% amino acid identity between human and *Sa. cerevisiae* proteins (Fig. 4). Of note is a short region within this domain at which the evolutionary constraints appear less stringent—amino acids 197–207 of the human protein. The divergence here highlights the border previously mapped within the *Sa. cerevisiae* Hap2 core between the DNA recognition and the subunit association surfaces (17).

A search of the GenBank (version 64.0) and European Molecular Biology Laboratory (version 23.0) databases reveals no significant sequence relationship of the human homolog to proteins other than Hap2. Direct comparison with the CCAAT-binding factors CTF/NF1 (25) and C/EBP (26) identifies no significant similarity.

Expression of the Human Hap2 Homolog. We used the coding region of clone HuHAP2.2 to probe a variety of primate tissues and several human cell lines for expression of the human Hap2 homolog. The Northern blot shown in Fig. 5 demonstrates constitutive expression of a 4.1-kb mRNA. This is the predominant transcript in heart, brain, small bowel, kidney, liver, and spleen, as well as in the T-lymphoblastoid Jurkat cell line and the Raji B-lymphoma cell line.

The 4.1-kb transcript is not, however, the principal transcript in HeLa cells. Lane 1 reveals the overwhelming preponderance in HeLa cells of a 1.6-kb species, which in other cells represents a minor species relative to the 4.1-kb mRNA. Although the 1.64-kb HeLa HuHAP2.1 cDNA clone

```
1    CAGGCTGTCCCTGGTGGACAAGGTCAAACCATCATGCAAGTACCTGTTTCTGGAACACAG
                                    M  Q  V  P  V  S  G  T  Q
61   GGTTTGCAGCAAATACAGTTGGTCCCACCTGGACAGATCCAGATCCAGGGTGGACAGGCT
10    G  L  Q  Q  I  Q  L  V  P  P  G  Q  I  Q  I  Q  G  G  Q  A
121  GTGCAGGTGCAGGGCCAGCAGGGCCAGACCCAGCAGATCATCATCCAGCAGCCCCAGACG
30    V  Q  V  Q  G  Q  Q  G  Q  T  Q  Q  I  I  I  Q  Q  P  Q  T
181  GCTGTCACTGCTGGCCAGACTCAGACACAGCAGCAGATTGCTGTCCAGGGACAGCAAGTG
50    A  V  T  A  G  Q  T  Q  T  Q  Q  Q  I  A  V  Q  G  Q  Q  V
241  GCACAGACTGCTGAAGGGCAGACCATCGTCTATCAACCAGTTAATGCAGATGGCACCATT
70    A  Q  T  A  E  G  Q  T  I  V  Y  Q  P  V  N  A  D  G  T  I
301  CTCCAGCAAGTTACAGTCCCTGTTTCAGGCATGATCACTATCCCAGCAGCCAGTTTGGCA
90    L  Q  Q  V  T  V  P  V  S  G  M  I  T  I  P  A  A  S  L  A
361  GGAGCACAGATTGTTCAAACAGGAGCCAATACCAACACAACCAGCAGTGGGCAAGGGACT
110   G  A  Q  I  V  Q  T  G  A  N  T  N  T  T  S  S  G  Q  G  T
421  GTCACTGTGACACTACCAGTGGCAGGCAATGTGGTCAATTCAGGAGGGATGGTCATGATG
130   V  T  V  T  L  P  V  A  G  N  V  V  N  S  G  G  M  V  M  M
481  GTTCCTGGGGCTGGCTCTGTGCCTGCTATCCAAAGAATCCCTCTACCTGGAGCAGAGATG
150   V  P  G  A  G  S  V  P  A  I  Q  R  I  P  L  P  G  A  E  M
541  CTTGAAGAAGAGCCTCTCTACGTGAATGCCAAACAATACCACCGTATTCTTAAGAGGAGG
170   L  E  E  E  P  L  Y  V  N  A  K  Q  Y  H  R  I  L  K  R  R
601  CAAGCCCGAGCTAAACTAGAGGCAGAAGGGAAAATTCCAAAGGAGAGAAGGAAATACCTG
190   Q  A  R  A  K  L  E  A  E  G  K  I  P  K  E  R  R  K  Y  L
661  CATGAGTCTCGGCACCGTCATGCCATGGCACGGAAGCGTGGTGAAGGTGGACGATTTTTC
210   H  E  S  R  H  R  H  A  M  A  R  K  R  G  E  G  G  R  F  F
721  TCTCCAAAGGAAAAGGATAGTCCCCATATGCAGGATCCAAACCAAGCCGATGAAGAAGCA
230   S  P  K  E  K  D  S  P  H  M  Q  D  P  N  Q  A  D  E  E  A
781  ATGACACAGATCATCCGAGTGTCCTAA
250   M  T  Q  I  I  R  V  S  *
                                    CCCCACGCCATGTGATGGAGCTGATCAAGGTCA
841  TGTTTCTCACTGTTCCAGGAAATTGATCAACTCTTCCAATGGGACATTGATGATCACATT
901  CTGCCCTTTACTACAGGACAGAAACCACTTAGTTTTTAATAAGTGGCTCAGTATTACAAT
961  TGCAGCGATGCCTGGCAAATTGAAGTTGCAAGCCTTTGTCTTACTCTTTCAGTCAGGACC
1021 TATTTCAGACTGTTGATGACATTGACATTTCATGGATTCGGATGATGCAAGGAGACACAT
1081 GCCACCCCAGCAGTACACAAAGCACTTACCTTGACTGGTGAAGTCAGCCAGCCATCTCAG
1141 CAGGGAAGAACAACCTTCTCAACCAAAACTGCAATCAGGAAAGCAGCAGCCCACTCCTTC
1201 CCATGGAATCTAGATAGCATTTATCCCTCTGGTGGGGGAATCTCACACTGACTAACTGAA
1261 TGGACGCTTTGTATTCAGAGATGGCTTCCAGGCAAGGAAGTTCTATCTTGTGATACGGAA
1321 ACAATAAATTCCTGAGATACATTCTTC
```

FIG. 3. Nucleotide sequence and predicted amino acid sequence of the cDNA encoding the human Hap2 homolog. The nucleotide sequence of the 5′ end of clone HuHAP2.1 was determined on both strands. The deduced amino acid sequence of a 257-amino acid open reading frame is shown with translation beginning at the first methionine residue. Glutamine residues are boxed, and the region of similarity to the *Sa. cerevisiae* Hap2 protein is underlined.

appears likely to correspond to this predominant 1.6-kb transcript, we cannot rule out the possibility that it represents a truncated copy of the 4.1-kb message. There are several additional discrete transcripts between 1.6 and 4.1 kb in HeLa cells, but these are expressed at levels even lower than that of the 4.1-kb message.

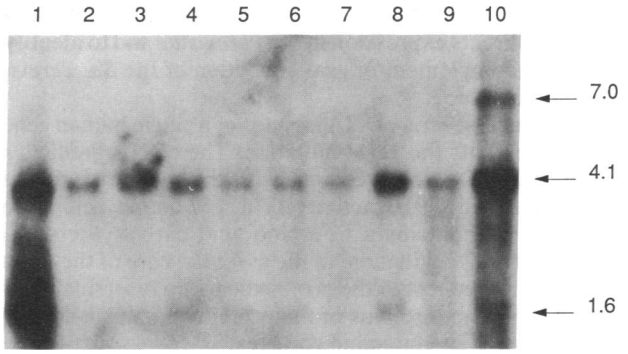

FIG. 5. The human Hap2 homolog has a complex pattern of expression. Tissues and cell lines were assayed for expression of the human Hap2 homolog by Northern analysis; the coding region of clone HuHAP2.2 was used as probe. Lane 1 contains 10 μg of poly(A)$^+$ RNA obtained from HeLa cells after two rounds of selection over oligo(dT)-cellulose. Lanes 2–8 each contain 10 μg of total RNA isolated from baboon tissues. Lanes: 2, heart; 3, brain; 4, ileum; 5, jejunum; 6, kidney; 7, liver; 8, spleen. Lane 9 contains 10 μg of total RNA from the human Jurkat T-cell line. Lane 10 contains 10 μg of once-selected poly(A)$^+$ RNA from the human Raji B-cell line. Sizes of the principal transcripts are indicated at right. Note that the disproportionately large signal in the HeLa lane (lane 1) and the Raji lane (lane 10) reflects enrichment of mRNA in these lanes due to oligo(dT) selection. The autoradiogram was exposed for 36 hr at −80°C with two intensifying screens.

The Raji B-lymphoma line expresses, in addition, a 7-kb transcript that appears to be B-cell specific (lane 10). Serial dilution (data not shown) of Raji RNA to normalize the signal from the 4.1-kb species does not extinguish this band; conversely, the substantial 4.1-kb signal in the HeLa lane (lane 1) is not accompanied by evidence of a larger transcript. The 7-kb transcript is not visible in a T-lymphocyte line (Jurkat, lane 7) but is visible on the original exposure in normal spleen RNA (lane 8).

DISCUSSION

The human CCAAT-binding complex CP1 contains subunits functionally related to the yeast Hap2 and Hap3 CCAAT-binding proteins. We report here the cloning of a cDNA that encodes a subunit of a human CCAAT-binding complex that is probably identical to CP1B. Our approach was to construct

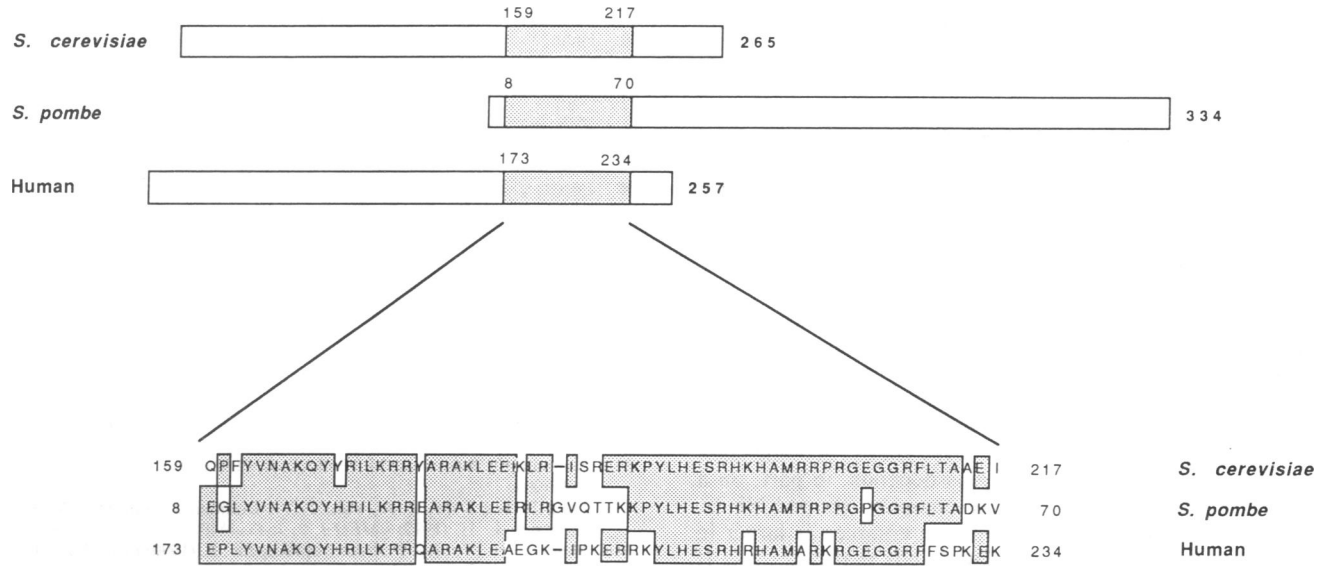

FIG. 4. Sequence conservation of the Hap2 core. Regions of amino acid sequence similarity between the *Sa. cerevisiae* Hap2 protein and homologs from *Sc. pombe* (19) and HeLa cells are shown schematically above and in detail below. Only identities are boxed.

a comprehensive cDNA library from HeLa cell mRNA in a vector that directs expression in *Sa. cerevisiae* and to identify clones that complement *in vivo* a deletion of the *Sa. cerevisiae* homolog Hap2.

We obtained several cDNA isolates of a single human gene that complements the yeast mutation. The cDNA predicts a protein of 27 kDa, in agreement with the measured size of partially purified CP1B activity. The 257-amino acid open reading frame contains a 62-amino acid carboxyl-terminal domain that is 73% identical to the essential core of the yeast Hap2 protein; the rest of the protein bears no resemblance to Hap2. The amino terminus of the protein contains a domain highly enriched in glutamine residues.

The Hap activation complex in yeast orchestrates the expression of nuclear genes required for mitochondrial oxidative function. The range of promoters containing the CCAAT recognition site in mammalian cells (5), however, suggests that the human Hap homolog normally functions in a role far more general than that of the yeast complex. There are two precedents of other yeast transcription factors with homologs in mammalian cells that serve regulatory roles different from that of the cognate yeast protein. The yeast activator GCN4 is related to the mammalian cJUN transcription factor, and both bind to the same DNA sequence (27, 28). In yeast, GCN4 regulates transcription of amino acid biosynthetic genes in response to starvation; the mammalian cJUN mediates cell proliferation in response to extracellular growth factors. The yeast Mcm1 protein regulates genes specifying the mating type of cells, whereas the homologous serum response factor regulates the immediate response of mammalian cells to serum stimulation (29–32). A divergence in the cellular role of conserved transcription factors may prove commonplace, facilitated by the modular nature of these proteins.

Dissociation of *CP1* from carbon control may have been achieved by elimination, through evolution, of the Hap4 subunit. Hap4 confers on the yeast complex its responsiveness to the cell's catabolic state (P. Sugiono and L.G., unpublished data), and carries the acidic transcription activation domain of the heteromer (24). The requirement for Hap4 in the yeast complex can be partially bypassed by fusion of an acidic activation domain directly to Hap2 (17). The glutamine-rich amino terminus of the human Hap2 may represent the evolutionary equivalent of such a fusion; glutamine-rich regions of the mammalian transcription factors Sp1 and OTF-2 (Oct-2) have been shown to be potent activators of transcription (21–23). This stoichiometry would be consistent with chromatographic evidence suggesting that HeLa CP1 consists of two, not three, components (8).

Analysis of human Hap2 expression reveals the presence of several highly regulated transcripts in addition to a constitutive 4.1-kb mRNA. HeLa cells express predominantly a 1.6-kb mRNA that is present as only a minor species in other cell types. This 1.6-kb mRNA probably corresponds to our HuHAP2.1 cDNA clone. The Raji B-cell lymphoma line expresses a 7-kb transcript that is absent from other major cell types.

We do not yet understand the structural relationships among the different RNA species or their significance to the regulation of gene expression. They may serve, however, in lieu of a separable Hap4 component to couple CP1 to different cell type-specific regulatory programs. For example, expression in B lymphocytes of the genes encoding class II major histocompatibility antigens depends, in part, on a CCAAT box and a Hap2/3/4-like complex, termed NF-Y. Data suggest that the CCAAT-binding complex is larger in class II-positive cells than in related cells that lack class II expression (33). This increase in size might well reflect the increased contribution of a Hap2 subunit that is encoded in the class II-positive cells by the larger, B cell-specific 7-kb transcript.

As this work was prepared for publication, Maity *et al.* (34) and Hooft van Huijsduijnen *et al.* (35) reported the cloning by standard techniques of the genes encoding the 40-kDa subunit of the corresponding CCAAT-binding complex from the rat and mouse. These clones differ in two respects from the human Hap2 clone reported here. (*i*) Both rat and mouse clones extend the reading frame 90 residues further to the amino terminus. This extension probably represents protein sequence encoded in the 4.1-kb transcript (Fig. 5) that is absent from the 1.6-kb HeLa mRNA that we cloned. (*ii*) The rat clone is missing a 6-amino acid sequence found in both human and mouse clones, Val-Thr-Val-Pro-Val-Ser (Hu-HAP2 residues 93–98), a difference that may result from different mRNA splicing.

We thank David Housman, Phil Sharp, and Laurie Glimcher for RNA and Alex Tzagaloff for plasmid YEP352. HeLa cells were obtained from the Massachusetts Institute of Technology Cell Culture Center. This work was supported by grants from the National Institutes of Health (L.G.) and postdoctoral fellowships from the American Cancer Society (D.M.B. and J.D.F.). D.M.B. is currently a Special Fellow of the Leukemia Society of America.

1. Forsburg, S. L. & Guarente, L. (1989) *Annu. Rev. Cell Biol.* **5**, 153–180.
2. Olesen, J., Hahn, S. & Guarente, L. (1987) *Cell* **51**, 953–961.
3. Hahn, S. & Guarente, L. (1988) *Science* **240**, 317–321.
4. Keng, T. & Guarente, L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 9113–9117.
5. Bucher, P. (1990) *J. Mol. Biol.* **212**, 563–578.
6. Raymondjean, M., Cereghini, S. & Yaniv, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 757–761.
7. Dorn, A., Bollekens, J., Staub, A., Benoist, C. & Mathis, D. (1987) *Cell* **50**, 863–872.
8. Chodosh, L. A., Baldwin, A. S., Carthew, R. W. & Sharp, P. A. (1988) *Cell* **53**, 11–24.
9. Chodosh, L. A., Olesen, J., Hahn, S., Baldwin, A. S., Guarente, L. & Sharp, P. A. (1988) *Cell* **53**, 25–35.
10. Lee, M. G. & Nurse, P. (1987) *Nature (London)* **327**, 31–35.
11. Schild, D., Brake, A. J., Kiefer, M. C., Young, D. & Barr, P. J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2916–2920.
12. Fikes, J. D., Becker, D. M., Winston, F. & Guarente, L. (1990) *Nature (London)* **346**, 291–294.
13. Colicelli, J., Birchmeier, C., Michaeli, T., O'Neill, K., Riggs, M. & Wigler, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 3599–3603.
14. Ammerer, G. (1983) *Methods Enzymol.* **101**, 192–201.
15. Aruffo, A. & Seed, B. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8573–8577.
16. Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988) *Nucleic Acids Res.* **16**, 6127–6145.
17. Olesen, J. & Guarente, L. (1990) *Genes Dev.* **4**, 1714–1729.
18. Sherman, F., Fink, G. R. & Hicks, J. B. (1986) *Methods in Yeast Genetics* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 117–119.
19. Olesen, J., Fikes, J. D. & Guarente, L. (1991) *Mol. Cell. Biol.*, in press.
20. Hahn, S., Pinkham, J., Wei, R., Miller, R. & Guarente, L. (1988) *Mol. Cell. Biol.* **8**, 655–663.
21. Courey, A. J. & Tjian, R. (1988) *Cell* **55**, 887–898.
22. Gerster, T., Balmaceda, C. G. & Roeder, R. G. (1990) *EMBO J.* **9**, 1635–1643.
23. Muller-Immergluck, M. M., Schaffner, W. & Matthias, P. (1990) *EMBO J.* **9**, 1625–1634.
24. Forsburg, S. L. & Guarente, L. (1989) *Genes Dev.* **3**, 1166–1178.
25. Santoro, C., Mermod, N., Andrews, P. C. & Tjian, R. (1988) *Nature (London)* **334**, 218–224.
26. Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. & McKnight, S. L. (1988) *Genes Dev.* **2**, 786–800.
27. Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. & Tjian, R. (1988) *Science* **238**, 1386–1392.
28. Struhl, K. (1988) *Nature (London)* **332**, 649–650.
29. Passmore, S., Elble, R. & Tye, B.-K. (1989) *Genes Dev.* **3**, 921–935.
30. Keleher, C. A., Passmore, S. & Johnson, A. D. (1989) *Mol. Cell. Biol.* **9**, 5528–5230.
31. Jarvis, E. E., Clark, K. L. & Sprague, G. F. (1989) *Genes Dev.* **3**, 936–945.
32. Norman, C., Runswick, M., Pollock, R. & Treisman, R. (1988) *Cell* **55**, 989–1003.
33. Finn, P. W., Kara, C. J., Van, T. T., Douhan, J., Boothby, M. R. & Glimcher, L. H. (1990) *EMBO J.* **9**, 1543–1549.
34. Maity, S. N., Vuorio, T. & de Crombrugghe, B. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5378–5382.
35. Hooft van Huijsduijnen, R., Li, X. Y., Black, D., Matthes, H., Benoist, C. & Mathis, D. (1990) *EMBO J.* **9**, 3119–3127.