Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 945–949, February 1987
Biochemistry

Amino acid sequence of the mRNA cap-binding protein from human tissues

(translational initiation/peptide sequencing/cDNA cloning/oligodeoxynucleotide probes)

WOJCIECH RYCHLIK, LESLIE L. DOMIER, PAUL R. GARDNER*, GARY M. HELLMANN, AND ROBERT E. RHOADS†

Department of Biochemistry, University of Kentucky College of Medicine, Lexington, KY 40536

Communicated by Sidney Udenfriend, October 31, 1986 (received for review September 23, 1986)

ABSTRACT **The 25-kDa mRNA cap-binding protein (CBP) involved in translation was purified by affinity chromatography from human erythrocytes and rabbit reticulocytes. The sequences of eight human and seven rabbit tryptic and V8 proteolytic peptides were determined. Based on the peptide sequence data, oligodeoxynucleotide probes were synthesized and used to screen human fibroblast and lymphocyte λ cDNA libraries. The DNA sequence obtained from recombinant λ phage inserts was found to code for all but one peptide. A 23-base oligonucleotide was synthesized based on the DNA sequence and used to prime synthesis of cDNA from human placental mRNA to construct a third library in λgt10. Screening with a 22-base oligonucleotide, whose sequence was upstream from the 23-base primer, yielded numerous recombinant phages with ≈250-base inserts. The 1900-base-pair cDNA sequence compiled from all phage inserts appeared to represent the entire primary sequence of CBP (M_r 25,117). Blot analysis of human placental and HeLa mRNA revealed multiple CBP mRNA species ranging from 1925 to 2250 bases. The amino acid sequence of CBP showed homology to the cap-binding PB2 protein of influenza virus.**

The 25-kDa mRNA cap-binding protein (CBP) recognizes and binds to the 7-methylguanosine-containing mRNA "cap" during an early step in the initiation of protein synthesis (1–4). Valuable information on the highly specific interaction between mRNA caps and CBP as well as the interactions of CBP with eukaryotic initiation factors 4A, 4B, and P220 (5–8) could potentially be derived from the primary structure of CBP. A potential regulatory role for CBP, which may act at the rate-limiting step for initiation (9–12), could also be investigated using cDNA probes to follow changes in the cellular level of CBP mRNA. At another level, CBP activity may be modulated by phosphorylation/dephosphorylation, because CBP is isolated as a mixture of phosphorylated and nonphosphorylated forms (13, 14) and because the dephosphorylated form has been correlated with a decrease in protein synthesis (15). Sequence information could be used to define and characterize the site of phosphorylation. Finally, determination of the entire coding region of the CBP mRNA could be used to establish unequivocally the size of the initial translation product, in view of reports of CBPs of higher molecular weight (16–20).

In this report we describe the sequence analysis of tryptic peptides of rabbit and human CBP, the construction of oligonucleotide probes, and the selection and sequence analysis of recombinant λ phages from human fibroblast, lymphocyte, and placental cDNA libraries that establish the primary structure of CBP. The mRNA coding for CBP was also detected by hybridization to cloned probes.

MATERIALS AND METHODS

Materials. Trypsin treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone was purchased from Cooper Biomedical (Malvern, PA). Protease V8 was obtained from Miles. All solvents used for HPLC were obtained from Fisher and were HPLC grade. Enzymes used for the cloning techniques were purchased from New England Biolabs. Rabbit reticulocyte lysate was obtained from Hazelton Dutchland (Denver, PA). Human erythrocyte lysate was obtained as described (13). m^7GTP-Sepharose (21) is a product of Pharmacia P-L Biochemicals.

Phage Libraries. A λgt11 cDNA library, prepared from WI-38 human fibroblast mRNA (22), was kindly provided by Ernest Beutler (Scripps Clinic). A λgt10 cDNA library was prepared from human IM9 lymphocytes by a modification of the method of Gubler and Hoffman (23) and generously provided by Jonathan Whittaker (University of Chicago).

Preparation of λgt10 Library Enriched in CBP Sequences. Human placenta poly(A)$^+$ mRNA was purified by the methods of Le Meur *et al.* (24) and Desrosiers *et al.* (25). cDNA was prepared by the method of Gubler and Hoffman (23). The first cDNA strand was synthesized using a 23-base oligodeoxynucleotide primer complementary to the 5′ end of CBP mRNA (see below). Synthesis of first- and second-strand cDNA was carried out with reverse transcriptase, *Escherichia coli* polymerase I, RNase H, and T4 DNA polymerase in the same reaction tube in the presence of RNasin. Methylation of DNA and covalent attachment of *Eco*RI linkers were performed essentially as described (26). Ligation of cohesive ends of cDNA with *Eco*RI- and phosphatase-treated λgt10 DNA (Promega Biotec, Madison, WI) was performed using the manufacturer's protocol. DNA was then packaged into bacteriophage particles, using packaging extract GP10-P (Stratagene, San Diego, CA), and plated with *E. coli* strain C600Hfl. The yield was 10^6 plaques per μg of mRNA; one CBP recombinant was obtained per 10^4 plaques. A 22-base oligodeoxynucleotide, complementary to the 5′ end of the CBP mRNA sequence (see below), was used for screening of the CBP cDNA-enriched λgt10 cDNA library.

Preparation of CBP-Derived Peptides. CBP was purified from human erythrocyte or rabbit reticulocyte lysate, reduced, carboxymethylated (13), lyophilized, and fragmented by (*i*) dissolving at 500 μg/ml in 50 mM ammonium bicarbonate, pH 8.3, and digesting with L-1-tosylamido-2-phenylethyl chloromethyl ketone-trypsin at an enzyme/substrate ratio of 1:50 (wt/wt) for 18 hr at 37°C, (*ii*) dissolving at 2

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: CBP, the 25-kDa mRNA cap-binding protein, also termed eukaryotic initiation factor 4E; nt, nucleotide residue.

*Present address: University of Chicago, Chicago, IL 60637.
†To whom reprint requests should be addressed.

Table 1. Amino acid sequence of human and rabbit peptides derived from CBP

Peptide name	Amino acid sequence	Position number	Retention time, min
HT2	E A V T H I G/S R	174	7
HT5	I V I G T G S S A D T A T	193	4.5
HT6	T W Q A N L	55	14.5
HT11	W L I T L N K	113	19
HT12	I A I W T/P T/M E E E N R	163	19.5
HT15	W A L W F F K	43	24.5
HT18	F D T V E D F W A L Y N H I Q L S S N L M P G	66	27
HT23	F X L E T L L A L I G E	129	34
RT1	D E I G P M W E D E K	96	2
RT5	I V I G Y	193	4.5
RT11	W L I T L E K	113	19
RT15	W A L W F F K	43	24.5
RT18	F D T V E D F W A L Y N H I Q L S S N L M P G X D Y	66	27
RV10	C E N A D A V T X I G R V Y X Q	170	14
RV20	T T P T P N P P P A E E E	8	25

Human (H) and rabbit (R) peptides that were similar or identical were labeled with the same numbers. Tryptic peptides (T) and V8-protease (V) peptides are also identified in the peptide name. Ambiguous amino acid assignments are indicated (using the single-letter code) by two residues being presented for one position or by X (for unknown amino acid); mismatches with sequences predicted from the cDNA sequence are underlined. Position numbers are as in Fig. 2. Retention times are for HPLC using C_3 column (13).

mg/ml in 50 mM sodium acetate, pH 4.0, and incubating with V8 protease at an enzyme/substrate ratio of 1:50 (wt/wt) for 4 hr at 37°C, or (*iii*) cleaving with CNBr (27). Peptides were fractionated by reverse-phase HPLC using a Beckman C_3 column (13) or with a Waters μBondapak C_{18} column (4.6 × 250 mm) under the same elution conditions.

Determination of Peptide Sequence. The amino acid compositions of purified peptides were determined after hydrolysis in 6 M HCl using a Beckman System 6300 high performance analyzer (13). Peptides were sequenced by automated Edman degradation using a model 470A protein sequencer (Applied Biosystems, Foster City, CA).

Oligodeoxynucleotide Probes and Primers. Two oligodeoxynucleotides were designed using peptide sequence data. One of these [14 nucleotides (nt)] contained mixtures of nucleotides at selected locations to allow for codon degeneracy (see Table 2). The other (45 nt) had a unique sequence, based on the codon utilization frequency calculations of Lathe (28). Other oligonucleotides of 22 and 23 nt were complementary to cDNA sequences 112–133 and 218–240, respectively (see Figs. 1 and 2). Synthesis was performed using a model 380A DNA synthesizer (Applied Biosystems). Each oligonucleotide was purified by polyacrylamide gel electrophoresis in the presence of urea (26). Purified oligonucleotides were 5′-^{32}P-labeled to specific activities of 1–2 × 10^9 cpm/μg (29).

Oligonucleotide Screening of Phage Libraries. Phage-infected bacteria were blotted on 137-mm Colony/Plaque Screen (New England Nuclear) or Magna Nylon 66 (Micron Separations, Honeoye, NY) filters and amplified *in situ* (26). After denaturation of the DNA in NaOH, the filters were air dried and irradiated with a Fotodyne UV 440 DNA transilluminator for 3 min and then washed in 1× SSC (0.15 M NaCl/0.015 M sodium citrate, pH 7.0) containing 0.1% $NaDodSO_4$ for 4–18 hr at 65°C. Filters were prehybridized for 4 hr at 60°C, 58°C, 46°C, and 37°C for the 22-, 23-, 45-, and 14-nt probes, respectively, in 6× SSC, 0.6% $NaDodSO_4$, 0.05% sodium pyrophosphate, 20 mM Hepes·NaOH (pH 7.0), *E. coli* tRNA at 100 μg/ml, and 1× Denhardt's solution (30). Hybridization with labeled oligonucleotides (2 × 10^6 cpm/ml of hybridization solution) was carried out for 18 hr under the same conditions. Filters were washed after hybridization in 6× SSC containing 0.1% $NaDodSO_4$ and 0.05% sodium pyrophosphate (1 ml/cm^2), twice at room temperature for 15 min, and once at the temperature of hybridization for 2–3 hr.

Nucleotide Sequence Analysis. DNA was extracted from recombinant phage (31), treated with *Eco*RI endonuclease, inserted into M13mp19 replicative form DNA in both orientations, and prepared for sequencing (32, 33). The nucleotide sequences of the two subclones were determined in their entirety by the chain-termination method (34) as modified by Biggin *et al.* (35). Portions of the single-stranded DNA inserts of the M13 subclones were deleted to permit sequencing of internal regions (36).

RNA Gel Blot Analysis. The sizes of specific mRNA species were estimated by RNA gel blot analyses (37). Ten-microgram RNA samples were separated on 4% polyacrylamide gels containing 8 M urea and electrophoretically transferred to Magna Nylon 66 filters (Micron Separations). Hybridization was performed in 50% (vol/vol) formamide following the manufacturer's recommendations. Radioactive probes were prepared by the method of Feinberg and Vogelstein (38).

RESULTS

Selected peptides from rabbit reticulocyte and human erythrocyte CBP were subjected to amino acid analysis and Edman degradation. Amino acid sequences were determined for eight human and seven rabbit CBP-derived peptides, providing a total of 87 residues from human and 85 from rabbit (Table 1). The sequence of peptide HT18 was used to synthesize two oligodeoxynucleotide probes (Table 2). The first was 45 nt and consisted of a unique, "best probability" sequence based on the calculations of Lathe (28), with one exception. We examined GenBank* for the frequency of occurrence of coding sequences for the tripeptide Trp-Ala-Leu. A higher frequency was found for UGGGCUCUG than the UGGGCCCUG sequence proposed by Lathe (28), so the former sequence was chosen for the construction of the probe. The second probe, based on peptide HT18, was one of 14 nt, having 8-fold degeneracy.

*National Institutes of Health (1985) *Genetic Sequence Data Bank: GenBank* (Research Systems Div., Bolt, Beranek, and Newman, Inc., Boston), Tape Release 38.0.

Table 2. Oligonucleotide probes and primers

45-residue probe*	
Amino acid sequence	F D T V E D F W A L Y N H I Q
Synthesized sequence	3' AAACTGTGTCACCTCCTGAAGACCCGAGACATGTTGGTGTAGGTC 5'
Observed sequence†	5' TTTGATACTGTTGAAGACTTTTGGGCTCTGTACAACCATATCCAG 3'
Residue number	196 * * * * * △ * 240
14-residue probe†	
Amino acid sequence	E D F W A
Synthesized sequence	3' CTYCTRAARACCCG 5'
Observed sequence	5' GAAGACTTTTGGGC 3'
Residue number	208 221

*The 45-residue probe sequence was based on the amino acid sequence of peptides HT18 and RT18 (Table 1). With one exception, △, the codon choices for the 45-base oligonucleotide were those most frequently observed in mammalian genes (28). Mismatches between predicted and observed sequence are indicated (*).
†Mixture of all eight possible oligonucleotides complementary to the mRNA encoding a portion of peptides HT18 and RT18.

Initial screening of a λgt11 human fibroblast cDNA library with each of the probes independently resulted in a large number of signals. However, when plaques selected on the basis of hybridization with the 45-nt probe were rescreened using the 14-nt degenerate probe, a single positive clone was obtained. The DNA sequence of the 503-base-pair (bp) insert (Fig. 1, insert A) suggested that, although clearly related to CBP mRNA, it represented neither the 5′ nor 3′ terminus. Consequently, a second human cDNA library, prepared from lymphocytes, was screened using a 23-nt probe derived from the sequence of insert A. In 10^6 plaques, only one positive clone was found, having an insert of 1816 bp (Fig. 1, insert B). Sequence analysis indicated that the coding region for the N terminus of CBP was still not represented. Therefore, a third library, specifically enriched in CBP cDNA, was constructed by priming cDNA synthesis with the 23-nt probe. Out of 13 positive clones that were analyzed, 11 contained inserts of ≈250 bp, beginning near position −20 (Fig. 1, inserts C–M).

The composite sequence of CBP mRNA is shown in Fig. 2. The various inserts overlapped over regions totaling ≈700 bp, and the sequences obtained from these regions were identical. Position 1 is assigned to an AUG codon found after 18 nt of presumably untranslated leader, and a UAA termination codon is found at position 652. A 3′-untranslated region of 1226 nt follows this.

Several lines of evidence argue that the entire coding region of CBP mRNA is represented in Fig. 2 and that the AUG codon at position 1 is the initiation codon. First, the molecular weight of CBP has variously been estimated to be 24,000 (2), 26,000 (19, 39), and 28,000 (13). That of the polypeptide encoded by the sequence shown is 25,117. Second, all 15 of the rabbit or human peptides whose amino acid sequences were determined are coded for by the nucleotide sequence (amino acid sequence locations are given in Table 1). Third, the putative initiation codon is flanked by adenosine and guanosine at positions −3 and +4, respectively, which are the most conserved elements of the consensus sequence for eukaryotic initiation sites (40). Fourth, as noted above, nearly all of the phage inserts obtained by cDNA synthesis primed by the 23-base oligonucleotide, located at positions 218–240, were ≈250 bp, suggesting the end of the mRNA had been reached.

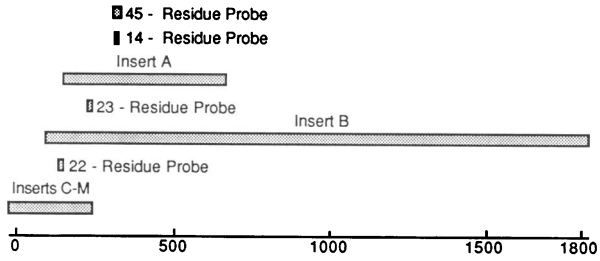

FIG. 1. Cloning strategy for CBP mRNA. The insert portions of the recombinant λ phage (inserts A–M) used to obtain the nucleotide sequence of CBP mRNA are shown in relation to the entire sequence. Above each insert are the synthetic oligodeoxynucleotide probes used to select it.

Another argument that Fig. 2 represents the complete coding sequence comes from amino acid composition data. The composition of CBP is in reasonable agreement with that of the predicted polypeptide (Table 3). More precise information was obtained from CNBr fragments. Treatment of CBP with CNBr yielded three peptides, consistent with there being two internal and one initiator methionine. One of the three peptides was resistant to Edman degradation, suggesting a blocked N terminus. The same behavior was noted for intact CBP. Comparision of the amino acid compositions of the blocked CNBr peptide and the equivalent peptide predicted from the cDNA sequence gives good agreement (Table 3). If amino acid residues potentially encoded by the −18 to −1 sequence are added to the predicted composition of the CNBr peptide, the agreement is considerably worse.

Hybridization of insert B to two human mRNA preparations, one from placenta and one from HeLa cells, revealed multiple bands ranging from 1925 to 2250 nt (Fig. 3). In experiments not shown, a second probe, containing only sequences corresponding to the CBP mRNA coding region, was hybridized to RNA samples denatured with CH_3HgOH prior to electrophoresis. This gave identical patterns to that shown. Also, no hybridization was detected with RNA samples that did not bind to oligo(dT)-cellulose.

DISCUSSION

The cDNA sequence presented in Fig. 2 is sufficient to encode a 25-kDa polypeptide, which is in good agreement with the molecular weight, amino acid composition (Table 3), and selected amino acid sequences (Table 1) of the CBP isolated from reticulocytes and erythrocytes. Minor differences between the predicted human translation product and some rabbit peptides are expected from compositional differences between rabbit and human CBP (13). Five differences in amino acid sequence between that predicted from the cDNA and that experimentally determined from human peptides were observed. The three peptides in which these discrepancies occurred were obtained in the lowest yields. It is thus likely that they represent errors in the interpretation of peptide sequence data. The amino acid sequence of the peptide that was used for probe construction (peptide HT18) agreed perfectly with the cDNA sequence.

Initial attempts at cloning CBP cDNA were made with short, degenerate oligonucleotide probes and with an affinity-purified anti-CBP antibody (41). These efforts were unsuccessful, presumably because CBP mRNA is very rare. The number of false-positive signals was substantially reduced by

```
                                    5                                      15                                      25
                    M   A   T   V   E   P   E   T   T   P   T   P   N   P   P   T   T   E   E   E   K   T   E   S   N
CGATCAGATCGATCTAAG ATG GCG ACT GTC GAA CCG GAA ACC ACC CCT ACT CCT AAT CCC CCG ACT ACA GAA GAG GAG AAA ACG GAA TCT AAT
      -10       -1 1                                   30                                      60
                                    35                                      45                                      55
 Q   E   V   A   N   P   E   H   Y   I   K   H   P   L   Q   N   R   W   A   L   W   F   F   K   N   D   K   S   K   T
CAG GAG GTT GCT AAC CCA GAA CAC TAT ATT AAA CAT CCC CTA CAG AAC AGA TGG GCA CTC TGG TTT TTT AAA AAT GAT AAA AGC AAA ACT
                90                                      120                                     150
                                    65                                      75                                      85
 W   Q   A   N   L   R   L   I   S   K   F   D   T   V   E   D   F   W   A   L   Y   N   H   I   Q   L   S   S   N   L
TGG CAA GCA AAC CTG CGG CTG ATC TCC AAG TTT GAT ACT GTT GAA GAC TTT TGG GCT CTG TAC AAC CAT ATC CAG TTG TCT AGT AAT TTA
                180                                     210                                     240
                                    95                                      105                                     115
 M   P   G   C   D   Y   S   L   F   K   D   G   I   E   P   M   W   E   D   E   K   N   N   R   G   G   R   W   L   I
ATG CCT GGC TGT GAC TAC TCA CTT TTT AAG GAT GGT ATT GAG CCT ATG TGG GAA GAT GAG AAA AAC AAC CGG GGA GGA CGA TGG CTA ATT
                270                                     300                                     330
                                    125                                     135                                     145
 T   L   N   K   Q   Q   R   R   S   D   L   D   R   F   W   L   E   T   L   L   C   L   I   G   E   S   F   D   D   Y
ACA TTG AAC AAA CAG CAG AGA CGA AGT GAC CTC GAT CGC TTT TGG CTA GAG ACA CTT CTG TGC CTT ATT GGA GAA TCT TTT GAT GAC TAC
                360                                     390                                     420
                                    155                                     165                                     175
 S   D   D   V   C   G   A   V   V   N   V   R   A   K   G   D   K   I   A   I   W   T   T   E   C   E   N   R   E   A
AGT GAT GAT GTA TGT GGC GCT GTT GTT AAT GTT AGA GCT AAA GGT GAT AAG ATA GCA ATA TGG ACT ACT GAA TGT GAA AAC AGA GAA GCT
                450                                     480                                     510
                                    185                                     195                                     205
 V   T   H   I   G   R   V   Y   K   E   R   L   G   F   P   P   K   I   V   I   G   Y   Q   S   H   A   D   T   A   T
GTT ACA CAT ATA GGG AGG GTA TAC AAG GAA AGG TTA GGC TTT CCT CCA AAG ATA GTG ATT GGT TAT CAG TCC CAC GCA GAC ACA GCT ACT
                540                                     570                                     600
                                    215
 K   S   G   S   T   T   K   N   R   F   V   V
AAG AGC GGC TCC ACC ACT AAA AAT AGG TTT GTT GTT TAA GAAGACACCTTCTGAGTATTCTCATAGGAGACTGCGTCAAGCAATCGAGATTTGGGAGCTGAACCA
                630                                      660                            690                            720
AAGCCTCTTCAAAAAGCAGAGTGGACTGCATTTAAATTTGATTTCCATCTTAATGTTACTCAGAGTATAAGAGAAGTCTCATTCGCCTTTGTCTTGTACTTCTGTGTTCATTTTTTTTTT
                           750                           780                           810                           840
TTTTTTTTGGCTAGAGTTTCCACTATCCCAATCAAAGAATTACAGTACACATCCCCAGAATCCATAAATGTGTTCCTGGCCCACTCTGTAATAGTTCAGTAGAATTACCATTAATTACAT
                           870                           900                           930                           960
ACAGATTTTACCTATCCACAATAGTCAGAAAACAACTTGGCATTTCTATACTTTACAGGAAAAAAAATTCTGTTGTTCCATTTTATGCAGAAGCATATTTTGCTGGTTTGAAAGATTATG
                           990                          1020                          1050                          1080
ATGCATACAGTTTTCTAGCAATTTTCTTTGTTTCTTTTTACAGCATTGTCTTTGCTGTACTCTTGCTGATGGCTGCTAGATTTTAATTTATTTGTTTCCCTACTTGATAATATTAGTGAT
                          1110                          1140                          1170                          1200
TCTGATTTCAGTTTTCATTTGTTTTGCTTAAATTTTTTTTTTTTTTTTCCTCATGTAACATTGGTGAAGGATCCAGGAATATGACACAAAGGTGGAATAAACATTAATTTTGTGCATTC
                          1230                          1260                          1290                          1320
TTTGGTAATTTTTTTTGTTTTTTGTAACTACAAAGCTTTGCTACAAATTTATGCATTTCATTCAAATCAGTGATCTATGTTTGTGTGATTTCCTAAACATAATTGTGGATTATAAAAAAT
                          1350                          1380                          1410                          1440
GTAACATCATAATTACATTCCTAACTAGAATTAGTATGTCTGTTTTTGTATCTTTATGCTGTATTTTAACACTTTGTATTACTTAGGTTATTTTGCTTTGGTTAAAAATGGCTCAAGTAG
                          1470                          1500                          1530                          1560
AAAAGCAGTCCCATTCATATTAAGACAGTGTACAAAACTGTAAATAAAATGTGTACAGTGAATTGTCTTTTAGACAACTAGATTTGTCCTTTTATTTCTCCATCTTTATAGAAGGAATTT
                          1590                          1620                          1650                          1680
GTACTTCTTATTGCAGGCAAGTCTCTATATTATGTCCTCTTTTGTGGTGTCTTCCATGTGAACAGCATAAGTTTGGAGCACTAGTTTGATTATTATGTTTATTACAATTTTTAATAAATT
                          1710                          1740                          1770                          1800
GAATAGGTAGTATCATATATATGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
                          1830                          1860
```

FIG. 2. Nucleotide sequence of cDNA corresponding to CBP mRNA, determined from recombinant λ phage. The predicted amino acid sequence of CBP is shown above the cDNA sequence. Nucleotides are numbered below the sequence and amino acids, above. Potential polyadenylylation signals are underlined.

the use of the unique-sequence, 45-base oligonucleotide. The one deviation from the rules of Lathe (28), in which an adenosine residue was used in place of the recommended guanosine complementary to position 222, resulted in a perfectly base-paired stretch of 17 residues (Table 2). This undoubtedly increased the stability of the hybrid, since the longest base-paired stretch would have otherwise been only 11.

The mRNA structure itself, as deduced from cDNA sequences and RNA gel blot hybridization, contains several interesting features. The 3′-noncoding region (1226 nt) is nearly twice as long as the coding region. It contains three oligo(U) sequences, 18 nt beginning at position 831, 16 nt at position 1234, and 15 nt at position 1329, interrupted by two guanosine residues, thus falling into the class of oligo(U)-containing mRNAs observed by Edmonds and collaborators (42). Several other specific oligo(U)-containing mRNAs have been reported, including human β-actin (43), human MYC (44), and eukaryotic initiation factor 4A (45). The 3′-noncoding region also contains three AAUAAA sequences at positions 1297, 1602, and 1793. The latter is located 26 nt from the start of the poly(A) segment and thus probably serves as the polyadenylylation signal for the mRNA. By RNA gel blot analysis there are at least two species of CBP mRNA; the predominant one is ≈1900 nt (Fig. 3). The mRNA for eukaryotic initiation factor 4A was also found to exist in two forms, due to alternate polyadenylylation sites (45). Especially intriguing in the case of CBP mRNA is the minor band of ≈2500 nt. In the construction of the specific oligonucleotide-primed library, 11 of the recombinant phage inserts were of ≈250 bp (Fig. 1, inserts C–M). One, however, was ≈800 bp. Additional studies will be required to determine whether these multiple forms of CBP mRNA represent alternate start sites for transcription, as in the case of mouse α-amylase (46) and dihydrofolate reductase (47) mRNAs.

In examining the amino acid sequence of CBP, it would be of particular interest to identify the site of phosphorylation and the cap-binding site. Preliminary evidence suggests serine-53 is phosphorylated. This exists in the context Asp-Xaa-Ser, which resembles somewhat the consensus sequence recognized by casein kinase I, Glu-Xaa-Ser (48). Regarding the cap-binding site, various GTP-binding proteins have been found to contain the following three consensus sequences for nucleotide binding (49): Gly-Xaa-Xaa-Xaa-Xaa-Gly-Lys (phosphoryl), Asp-Xaa-Xaa-Gly (phosphoryl),

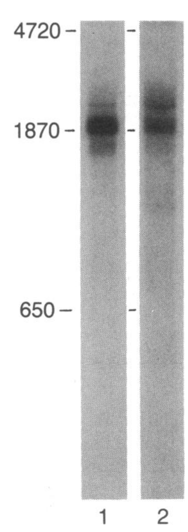

FIG. 3. Characterization of human CBP mRNA. HeLa cell (blot 1) or human placental (blot 2) mRNA fractions were subjected to blot hybridization using insert B as probe (see Fig. 1). Molecular size markers (in nt) correspond to globin mRNA, 18S and 28S rRNA.

Table 3. Amino acid composition of intact CBP and N-terminal CNBr-cleaved peptide

	Amino acid composition, residue per mol			
	CBP		N-terminal CNBr peptide	
Amino acid	Analyzed	Predicted	Analyzed	Predicted
Ala	10.2	11	5.4	5
Arg	12.0	12	2.7	2
Asx	25.5	28	10.4	11
Cys	5.3	4	0.0	0
Gly	13.2	11	0.0	0
Glx+Hsl*	24.7	25	13.9	14
His	5.2	5	3.4	3
Ile	10.5	11	2.6	3
Leu	19.9	16	4.9	7
Lys	13.7	15	7.5	6
Met	3.2	3	0.0†	0
Phe	8.4	9	4.4	4
Pro	15.1	11	7.0	7
Ser	12.1	12	4.4	5
Thr	16.6	18	8.0	9
Trp	6.6	8	3.2	4
Tyr	5.8	6	2.3	2
Val	12.0	12	3.5	3

Molecular weights of 25,117 for CBP and 10,044 for the N-terminal peptide were assumed for calculation purposes. Amino acid composition (average of six determinations) was performed as described (13). The predicted amino acid composition was from the cDNA sequence (Fig. 2). A single amino acid analysis was made for the N-terminal CNBr peptide, and no corrections were made for degradation of serine, threonine, and tryptophan or for slow release of isoleucine and valine.

*CNBr degrades methionine to homoserine lactone (Hsl), which comigrates with glutamic acid or glutamine during amino acid analysis.

†Degraded by CNBr method.

and Asn-Lys-Xaa-Asp (guanine). The first sequence is not found in CBP, but the second is found at amino acid position 148. Sequences similar to the third, Xaa-Lys-Xaa-Asp, occur at positions 48, 65, and 159. However, it should be noted that CBP is a m^7GTP-binding protein, and that this nucleotide has a charge distribution considerably different from that of GTP. Perhaps a more valid comparison is with the influenza polymerase PB2, which is also a cap-binding protein (50). Striking homologies are found between the sequence Lys-Xaa-Xaa-Xaa-Gln-Xaa-Baa-Trp-Ala-Leu (where Baa is a basic residue) at position 360 of the polymerase and position 36 of CBP, and the sequence Asp-Gly-Haa-Glu-Pro (where Haa is a hydrophobic residue) at positions 234 of the polymerase and 96 of CBP. Regarding the former sequence, it is interesting that the cap structure has been shown to have a high affinity for tryptophan (51).

We are indebted to Dr. Ernest Beutler, Scripps Clinic, La Jolla, CA, for providing the fibroblast λgt11 library. Special thanks are due to Dr. Jonathan Whittaker, University of Chicago, Chicago, IL, for providing the λgt10 library. We also wish to thank Dr. Janet Morgan, Dr. Mary Sue Coleman, Dr. Lela C. Riley, Dr. Thomas C. Vanaman, and Dr. Leena S. Hiremath for helpful advice and assistance. This work was supported by Grant GM20818 from the National Institute of General Medical Sciences.

1. Banerjee, A. K. (1980) *Microbiol. Rev.* **44**, 175–205.
2. Sonenberg, N., Edery, I., Darveau, A., Humbelin, M., Trachsel, H., Hershey, J. W. B. & Lee, K. A. W. (1983) in *Protein Synthesis*, eds. Abraham, A. K., Eikhom, T. S. & Pryme, I. F. (Humana, Clifton, NJ), pp. 23–43.
3. Rhoads, R. E. (1985) in *Progress in Molecular and Subcellular Biology*, eds. Hahn, F. E., Kopecko, D. J. & Muller, W. E. G. (Springer, Berlin), Vol. 9, pp. 104–155.
4. Shatkin, A. J. (1985) *Cell* **40**, 223–224.
5. Edery, I., Humbelin, M., Darveau, A., Lee, K. A. W., Milburn, S., Hershey, J. W. B., Trachsel, H. & Sonenberg, N. (1983) *J. Biol. Chem.* **258**, 11398–11403.
6. Tahara, S. M., Morgan, M. A. & Shatkin, A. J. (1981) *J. Biol. Chem.* **256**, 7691–7694.
7. Grifo, J. A., Tahara, S. M., Morgan, M. A., Shatkin, A. J. & Merrick, W. C. (1983) *J. Biol. Chem.* **258**, 5804–5810.
8. Lax, S., Fritz, W., Browning, K. & Ravel, J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 330–333.
9. Hiremath, L. S., Webb, N. R. & Rhoads, R. E. (1985) *J. Biol. Chem.* **260**, 7843–7849.
10. Duncan, R. & Hershey, J. W. B. (1985) *J. Biol. Chem.* **260**, 5486–5492.
11. Ray, B. K., Brendler, T. G., Adya, S., Daniels-McQueen, S., Miller, J. K., Hershey, J. W. B., Grifo, J. A., Merrick, W. C. & Thach, R. E. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 663–667.
12. Panniers, R., Stewart, E. B., Merrick, W. C. & Henshaw, E. C. (1985) *J. Biol. Chem.* **260**, 9648–9653.
13. Rychlik, W., Gardner, P. R., Vanaman, T. C. & Rhoads, R. E. (1986) *J. Biol. Chem.* **261**, 71–75.
14. Buckley, B. & Ehrenfeld, E. (1986) *Virology* **152**, 497–501.
15. Duncan, R. F., Milburn, S. C. & Hershey, J. W. B. (1986) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **45**, 1769 (abstr.).
16. Sonenberg, N., Guertin, D., Cleveland, D. & Trachsel, H. (1981) *Cell* **27**, 563–572.
17. Zumbe, A., Stahli, C. & Trachsel, H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2927–2931.
18. Sonenberg, N. & Trachsel, H. (1982) *Curr. Top. Cell. Regul.* **21**, 65–88.
19. Trachsel, H., Zumbe, A., Stahli, C., Humbelin, M. & Sonenberg, N. (1982) *Biochem. Soc. Symp.* **47**, 145–161.
20. Patzelt, E., Blaas, D. & Kuechler, E. (1983) *Nucleic Acids Res.* **11**, 5821–5835.
21. Webb, N. R., Chari, R. V. J., De Pillis, G., Kozarich, J. W. & Rhoads, R. E. (1984) *Biochemistry* **23**, 177–181.
22. Sorge, J., West, C., Westwood, B. & Beutler, E. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7289–7293.
23. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
24. Le Meur, M., Glanville, N., Mandel, J. L., Gerlinger, P., Palmiter, R. & Chambon, P. (1981) *Cell* **23**, 561–571.
25. Desrosiers, R., Friderici, K. & Rottman, F. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 3971–3975.
26. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
27. Gross, E. (1967) *Methods Enzymol.* **11**, 238–255.
28. Lathe, R. (1985) *J. Mol. Biol.* **183**, 1–12.
29. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
30. Denhardt, O. (1966) *Biochem. Biophys. Res. Commun.* **23**, 641–646.
31. Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L. & Treiber, G. (1970) *Virology* **40**, 734–744.
32. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101–106.
33. Messing, J. (1983) *Methods Enzymol.* **101**, 20–89.
34. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
35. Biggin, M. F., Gibson, T. J. & Horig, G. G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
36. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13**, 31–40.
37. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
38. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
39. Hansen, J. & Ehrenfeld, E. (1981) *J. Virol.* **38**, 438–445.
40. Kozak, M. (1984) *Nature (London)* **308**, 241–246.
41. Hiremath, L. S. & Rhoads, R. E. (1986) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **45**, 1660 (abstr.).
42. Wood, W. M., Wallace, J. C. & Edmonds, M. C. (1985) *Biochemistry* **24**, 3686–3693.
43. Ponte, P., Ng, S. Y., Engel, J., Gunning, P. & Kedes, L. (1984) *Nucleic Acids Res.* **12**, 1687–1696.
44. Colby, W., Chen, E., Smith, D. & Lennson, A. (1983) *Nature (London)* **301**, 722–725.
45. Nielsen, P. J., McMaster, G. K. & Trachsel, H. (1985) *Nucleic Acids Res.* **13**, 6867–6880.
46. Young, R. A., Hagenbuchle, O. & Schibler, U. (1981) *Cell* **23**, 451–458.
47. Setzer, D. R., McGrogan, M., Nunberg, J. H. & Schimke, R. T. (1980) *Cell* **22**, 361–370.
48. Hathaway, G. M. & Traugh, Y. A. (1982) *Curr. Top. Cell. Regul.* **21**, 101–127.
49. Merrick, W. C., Abramson, R. D., Anthony, D. D., Dever, T. E. & Caliendo, A. M. (1987) in *Translational Regulation of Gene Expression*, ed. Ilan, J. (Plenum, New York), in press.
50. Braam-Markson, J., Jaudon, C. & Krug, R. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4326–4330.
51. Ishida, T., Katsuta, M., Inoue, M., Yamagata, Y. & Tomita, K. (1983) *Biochem. Biophys. Res. Commun.* **115**, 849–854.