Volume 16 Number 18 1988 Nucleic Acids Research

Identification of a novel TA-rich DNA binding protein that recognizes a TATA sequence within the brain creatine kinase promoter

Grace M.Hobson[1], Mark T.Mitchell, George R.Molloy[1], Mark L.Pearson and Pamela A.Benfield*

E.I. du Pont de Nemours and Company, Experimental Station, Wilmington, DE 19898 and [1]University of Delaware, School of Life and Health Sciences, Newark, DE 19713, USA

Received May 17, 1988; Revised and Accepted August 23, 1988

ABSTRACT

The rat brain creatine kinase gene possesses a structurally complex promoter with multiple potential regulatory elements. Two CCAAT sequences, a TATAAATA sequence and a TTAA sequence are found within the first one hundred base pairs. We present evidence that favors the allocation of the downstream TTAA sequence as the potential TATA box. We show that the CCAAT sequences and the upstream TATAAATA sequence are binding sites for potential regulatory factors and that sequences in this region are capable of regulating expression from the downstream TTAA sequence. We suggest that the protein that binds to the upstream TATAAATA sequence is not a classical TFIID factor but rather may serve to block the binding of TFIID and/or to promote transcription from the downstream start site. We have been able to define conditions *in vitro* under which binding to this upstream TATAAATA sequence does not occur. Under these conditions we are able to detect transcription from both potential TATA sequences, a situation which we have been unable to detect *in vivo*. Our experiments suggest the existence in HeLa and brain nuclei of a protein that recognizes the concensus TATAAATA sequence, that is distinct from TFIID, and that may function in part to deny access of TFIID to this potential promoter element.

INTRODUCTION

Creatine kinase (CK) is a metabolic enzyme involved in maintenance of high ATP levels at cellular sites where energy requirements are high (1,2). The brain isoform (CKB) has a broad tissue specificity but is expressed at highest levels in brain (3). Several reports have been presented that indicate expression of the *ckb* gene may be responsive to differentiation signals in multiple cell types (4), peptide (5) and steroid (6) hormone regulation, and the abnormal growth conditions that accompany transformation (7). A second cytoplasmic isoform, muscle creatine kinase (CKM) shows tissue specific expression. It is expressed at appreciable levels only in adult cardiac and skeletal muscle (8). The *ck* genes therefore, present an interesting system in which to study differential gene regulation. One isoform (CKM) is highly tissue specific whereas the second (CKB) is more reminiscent of a housekeeping function that nevertheless is differentially expressed in a variety of tissues and in response to a wide range of stimuli.

The cytoplasmic *ck* genes from rat have both been cloned (9). The structural organization of the two genes is identical. However, their 5'-flanking regions show little sequence homology except in regions immediately surrounding potential CCAAT and TATA sequences. The sequence of the *ckb* promoter is potentially complex. It contains multiple

sequence features that have been associated with gene regulatory elements and/or regulatory DNA binding proteins. In particular two potential CCAAT sequences and two potential TATA sequences are found in close juxtaposition, suggesting the presence of overlapping promoter elements. However, we have been able to detect only one set of transcription start points for brain mRNA (9).

Little is known about the mechanisms that allow differential expression for the *ck* genes. Recently it has been shown that tissue specific expression of the *ckm* gene is in part dependent on the function of a muscle specific enhancer element (10, Horlick and Benfield, manuscript in preparation). However, there is no information about the sequences within the *ckb* gene that are important for either expression or regulation of this important locus. In this report we extend our previous studies on this gene (11) and examine the sequences necessary for expression of the rat *ckb* gene *in vivo* and *in vitro*. The relationship of these sequences to the complex structure of the promoter of this gene, and to interaction of the promoter with potential regulatory DNA binding proteins is investigated.

MATERIALS AND METHODS

Gene transfer

Calcium phosphate mediated gene transfer was performed into mouse Ltk$^-$aprt$^-$ cells (a kind gift of Dr Nat Sternberg) according to the method of Graham and Van der Eb (12). Cells were plated at a density of 10^6 cells per 85 mm dish in DMEM + 10% fetal calf serum. DNA (20 μg/dish) was left in contact with the cells for 16h. The cells were then refed with fresh medium and 48 hours later RNA was harvested for analysis.

RNA isolation and analysis

RNA was isolated from transfected cells and from tissues by the procedure of Chirgwin et al. (13). HeLa poly-A$^+$ RNA was a kind gift of Dr Rick Scott. S1-nuclease and primer extension analysis were performed as previously described (9). Primer extension analysis of human *ckb* mRNA was carried out using the human *ckb* (14) oligodeoxynucleotide E50336-103 with the sequence 5'-TGGGCGCTCAGGTCGG-3'. as primer.

Preparation of nuclear extracts

Nuclei from the brains of adult male Sprague-Dawley rats were prepared by a modification of the method of Blobel and Potter (15). Four rat brains were homogenized in 50 ml of 0.35M sucrose in nuclei buffer (5 mM Tris-HCl pH 7.5, 25 mM KCl, 5 mM $MgCl_2$ 1 mM EGTA, 0.5 mM spermine, 0.15 mM spermidine, and 0.1 mM PMSF) in a tissue homogenizer with a motor driven pestle at a moderate speed for 10 sec. The homogenate was further homogenized in a glass Dounce homogenizer 10 times with a B-type pestle. 5 ml aliquots of the homogenate were then adjusted to 1.2 M sucrose by addition of 1.9 M sucrose in nuclei buffer and centrifuged at 12,000 rpm in Sorvall SS34 rotor for 30 min at 4° C. The pelleted nuclei were used for nuclear extract preparation (16). HeLa and Daudi nuclear extracts were prepared by a modification of the procedure of Parker and Topol (17) as described by Shapiro et al. (18).

In vitro transcription analysis

Transcription reactions (20 µl) contained 150 µM each NTP, 5 mM creatine phosphate, 20 units RNAsin (Promega Biotec), and 12 µl nuclear extract freshly diluted into nuclear dialysis buffer (20 mM Hepes pH 7.9, 20% v/v glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.2 mM EGTA, 2 mM DTT). Each reaction was adjusted to contain approximately 50 µg nuclear protein. DNA template and $MgCl_2$ concentrations for each reaction are indicated in the figures. In the run-off assays 10 µCi (α-^{32}P) UTP (NEN: 800 Ci/mmol) was included and the cold UTP reduced to 15 µM. Reactions were incubated for 45 min at 30° C and extracted as described by Manley et al. (19). For primer extension analysis RNA was first ethanol precipitated in the presence of 40 µg carrier tRNA and then reprecipitated out of 1 M LiCl to remove residual SDS.

Gel retardation analysis

DNA-protein interactions were analyzed by gel retardation analysis (20) as previously described (21). Probes were labeled at the 5' end with ^{32}P using polynucleotide kinase. Standard binding reactions were carried out in a final volume of 25 µl and contained 10 mM Tris-HCl pH 7.6, 50 mM NaCl, 1 mM DTT and 1 mM EDTA. Varying amounts of crude nuclear extracts were incubated with 0.1 - 0.5 ng labelled DNA probe (specific activity 10^8 cpm/pmol) in the presence of 0-4 µg sonicated *E. coli* DNA at room temperature for 30 min. Modifications to these conditions are indicated in the figures.

DNase I footprinting

DNase I footprinting was carried out by a modification of the procedure of Babiss et al. (22). Reactions (20 µl) contained 2 mM Hepes (pH7.4), 10 mM KCl, 1 mM $MgCl_2$, and 0.2 mM DTT. Nuclear extracts (5-30 µg protein) were preincubated for 15 min on ice with 0-10 µg sonicated *E. coli* DNA . Labelled fragment DNA was then added and the reactions incubated for a further 15 min at room temperature. DNase I freshly diluted in 20 mM Tris (pH 7.6), 50 mM NaCl, 1 mM DTT, 100 µg/ml BSA and 50% v/v glycerol, was then added to a final concentration of 10-100 µg/ml and digestion was carried out for 90 sec at room temperature. Reactions were stopped by the addition of 2 volumes 50 mM Na_2·EDTA, 0.2% SDS, 100 µg/ml tRNA, 100 µg/ml proteinase K, phenol chloroform extracted, ethanol precipitated and loaded onto 7% polyacrylamide, 8M urea sequencing gels. Marker sequencing reactions were performed according to Maxam and Gilbert (23) as modified by Chuvpilo and Kravchenko (24).

RESULTS

Comparative structure of the rat *ckb* promoter

The rat *ckb* promoter contains two potential TATA sequences and two potential CCAAT sequences (Fig. 1). However, S1-analysis shows that *ckb* mRNA isolated from rat brain contains one clustered set of transcription start points between positions +1 and +9. The location of these start points is inconsistent with the operation of the upstream TATAAATA sequence at -60 as a TATA box and suggests the TTAA sequence at -22 as a possible candidate for fulfilling the TATA function. The sequence TTAA has been identified in the immediate

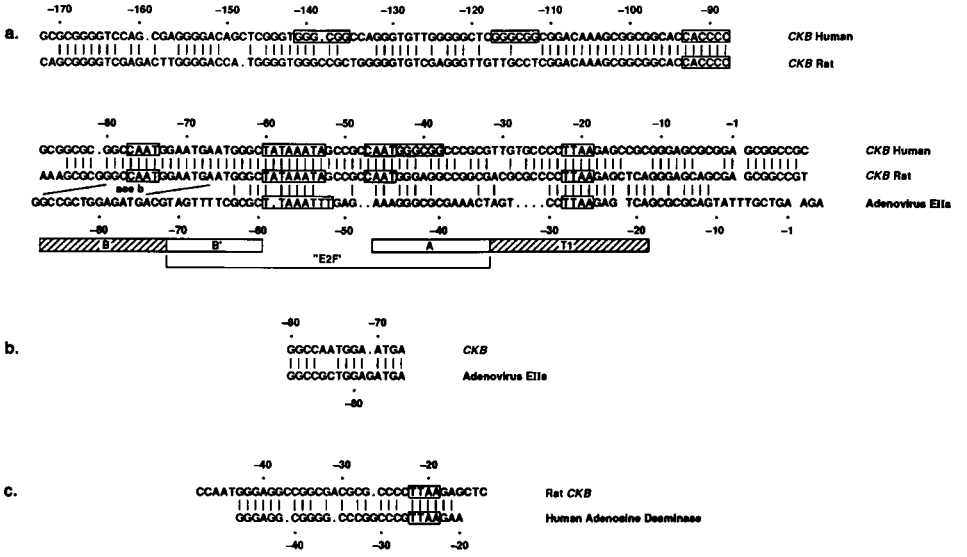

Figure 1. Sequences in the promoter regions of the rat and human *ckb* genes. TTAA, TATA, CCAAT , CACCC and potential Sp1 binding sites are enclosed in boxes. In (a) the rat and human sequences are compared with each other and with the promoter region of the adenovirus EIIaE gene. Beneath the adenovirus sequence regions implicated as important for E1a transactivation (27) are shown as boxes. Boxes are cross-hatched if protein binding has also been suggested based on footprinting experiments (28). The binding site for E2F (45) is also indicated by the lower bracket. In (b) the sequences for rat *ckb* between -67 and -80 and for adenovirus EIIaE between -74 and-88 are compared. In (c) the rat *ckb* promoter region between -13 and -48 is compared with the human adenosine deaminase promoter between -20 and -46.

promoter region of other genes e.g., adenosine deaminase (25) and the E1a responsive adenovirus EIIa promoter (26). The similarity of the *ckb* promoter to these two promoters extends beyond the TTAA sequence itself. This is observed in the 3' direction for EIIa (21 base pair match with two substitutions and one gap, Fig. 1a), and in the 5' direction for adenosine deaminase (28 base pair match with three substitutions and three gaps, Fig. 1c). Furthermore, like the *ckb* promoter the adenovirus EIIa promoter has a second upstream TATA element. In the adenovirus promoter each of these sequences functions as a promoter element although the downstream TTAA sequence forms part of the stronger promoter. Additional sequence conservation between the *ckb* and adenovirus EIIa promoters is observed in more upstream regions around the second TATA element and including regions B and A (Fig 1a). Regions B, B' and A along with a more upstream region, have been implicated in mediating the E1a induction of the EIIa promoter (27,28). However, unlike the rat *ckb* promoter the adenovirus EIIa promoter contains no potential CCAAT sequences.

Strong sequence conservation is observed between the human (29,30) and rat *ckb*

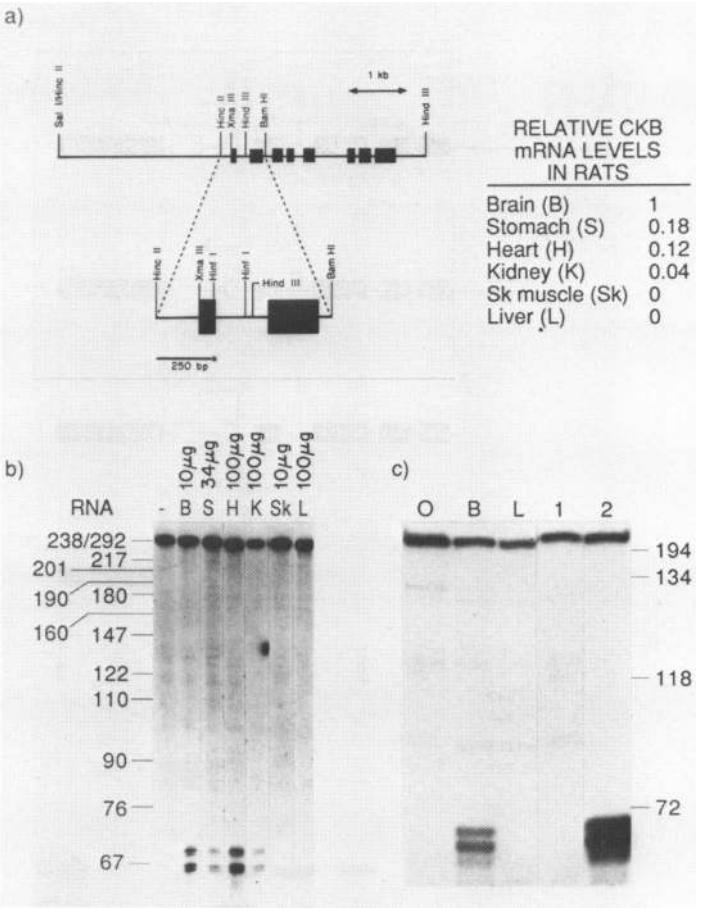

Figure 2. a.) S1-nuclease analysis was carried out using a 5'-α^{32}P end labelled probe derived from a *Hin*fI-*Hin*cI fragment of the rat *ckb* gene as indicated in (a). Relative levels of *ckb* RNA in different tissues were determined by scanning densitometry of the gel shown in (b). Each lane is derived from analysis of total RNA isolated from B, brain 10 µg; S, stomach, 34 µg; H, heart, 100 µg; K, kidney, 100 µg; Sk, skeletal muscle, 10 µg and L, liver, 100 µg. c) S1-analysis of 0, RNA control, and 10 µg total RNA derived from B, brain; L, liver; 1, mouse L cells and 2, mouse L cells transfected with plasmid PM97 that contains the rat *ckb* gene between the *Hin*cII site and the 3'-most *Hin*dIII site as shown in (a) cloned into the plasmid vector pUC8.

promoters (Fig. 1). The TTAA, TATAAATA, CCAAT and CACCC sequences are absolutely conserved, however, no potential Sp1 sites are found in the rat promoter.

Location of the transcription start points for rat ckb mRNA in multiple tissues

S1-nuclease analysis (Fig. 2), was used to determine the transcription start points in a range of tissues in which the *ckb* gene is expressed at varying levels. Relative RNA levels (Fig.

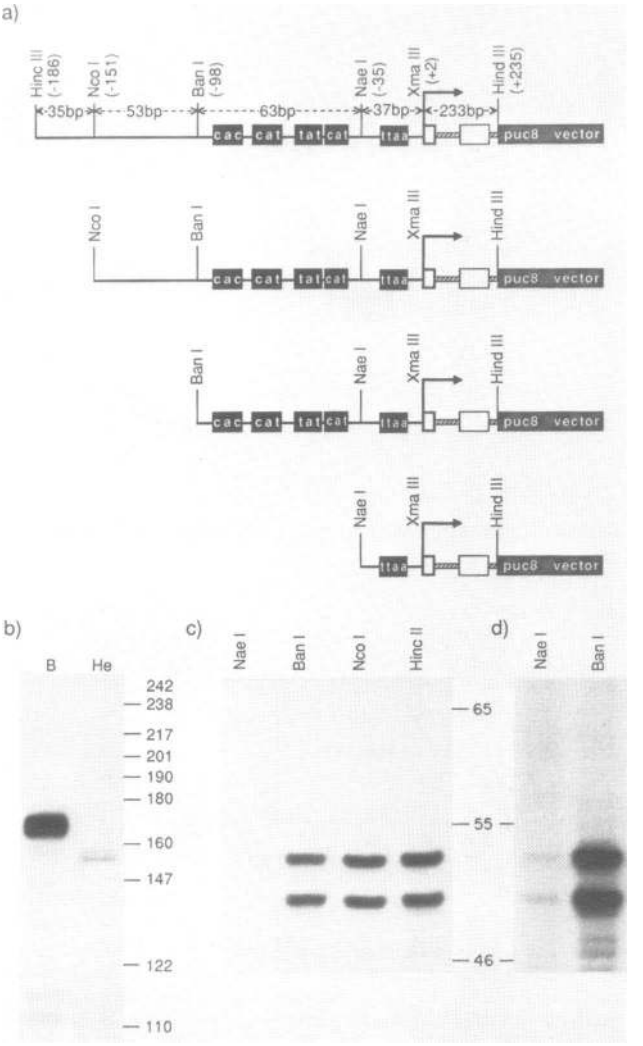

Figure 3. *In vitro* transcription analysis using Daudi cell nuclear extracts was performed with a series of linearized *ckb* templates truncated at various distances from the transcription start point as indicated in (a). Potential regulatory sequences CACCC, CCAAT, TATA and TTAA are in highlighted boxes. The broken arrow represents the transcript start point and the open boxes the first two exons of the *ckb* gene. c) and d), transcription reactions contained 150 ng template DNA and products were analyzed by primer extension using an oligodeoxynucleotide E50336-119 as primer with the sequence 5'-CAGGAACGGAGACTGC-3' that represents bases +43 to +59 in the first exon of the rat *ckb* gene. Panel (d) shows a 4X exposure of the first two lanes in panel c). Panel b) shows the results of primer extension analysis of rat brain RNA (2 µg) and HeLa poly-A+ RNA (1 µg) using oligodeoxynucleotide E35561-49 as primer. This oligodeoxynucleotide originates in the second exon of the *ckb* gene (9).

2a) were calculated from the data in Fig. 2b by correcting for the differences in the amounts of total RNA analyzed. The same transcription start points were observed in all tissues even when the *ckb* gene was expressed at much lower levels than in brain, e.g. in kidney (Fig. 2b), and primary rat myoblasts (data not shown).

The cloned rat *ckb* gene that contains only 186 base pairs of promoter DNA can be expressed upon transfection into mouse L cells, although expression from the endogenous mouse gene is undetectable (manuscript in preparation). RNA transcribed from the transfected rat gene shows the same set of transcription start points as rat brain *ckb* mRNA (Fig. 2c). These experiments suggest that only one *ckb* TATA sequence is functional in the sources examined. Based on its distance from the determined transcription start points, and its similarity to the adenovirus EIIa downstream TATA sequence, this is probably the nonconcensus downstream TATA sequence, TTAA, at position -22.

Transcription from the ckb promoter *in vitro*

Primer extension analysis (Fig. 3b), shows that HeLa cells express their endogenous *ckb* gene. Rat brain RNA produces extension products of 164 and 167 base pairs (Fig. 3a), consistent with initiation at positions +7 and +10 in the *ckb* sequence (9). HeLa poly-A+ RNA produces a similar doublet but reduced in size (Fig.3b). The published human *ckb* sequence of Daouk et al., (29) predicts a shorter transcript than that from rat by 10 base pairs. This results from a difference in size of the 5' untranslated region that arises principally from variation in the number of GCC repeats that immediately precede the translation initiation codon ATG.

The 186 base pair rat *ckb* promoter is recognized in HeLa and Daudi transcription extracts and transcription occurs from both supercoiled and linear templates. An RNA product is generated (Figs. 3c and d) that initiates at the same start site as *in vivo* rat brain mRNA (see Fig. 9b). Linearized *ckb* templates cut in the 5'-flanking region at the *Nco*I site (-151) or at the *Ban*I site (-98) are similarly transcribed with little loss in efficiency. However, a dramatic drop in transcription efficiency (approx. 50-fold) is observed with a template cut at the *Nae*I site to leave only the 37 base pairs of 5'-flanking sequence (Fig 3c and d), that include the downstream TTAA sequence. Thus sequence between -98 and -37 has a dramatic effect on the transcriptional efficiency of the *ckb* promoter (probably off the downstream TTAA sequence) *in vitro*, however the transcript start points remain the same.

Recognition of the *ckb* promoter region by sequence specific binding activities

Gel retardation assays were performed to determine if sequences within the rat *ckb* promoter could be recognized by sequence specific DNA binding proteins (Fig. 4) that might be candidates for potential regulatory molecules (31). A series of overlapping DNA test fragments were used as substrates as indicated in Fig. 4. With rat brain nuclear extracts an identical pattern of retarded bands was observed with any fragment that contains the region between the *Ban*I site (-98) and the *Nae*I site (-37). This region contains a TATAAATA sequence, two CCAAT sequences and a CACCC sequence that has been implicated in β-globin (32) and tryptophan

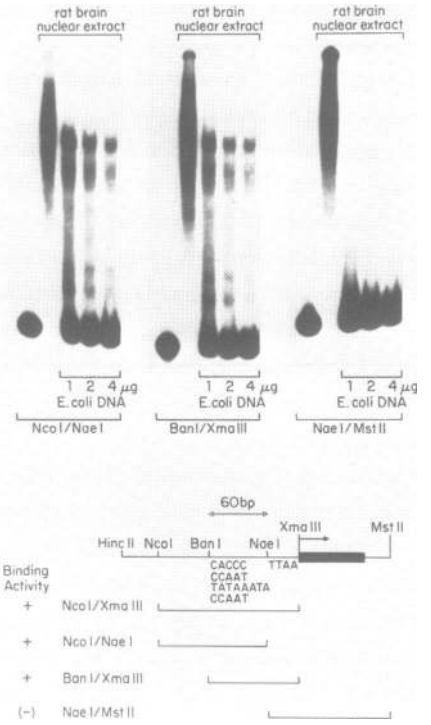

Figure 4. Gel retardation analysis was performed using rat brain nuclear extract and a series of fragments derived from the rat *ckb* promoter as shown at the bottom of the figure. For each gel retardation series the first lane contains the fragment alone and the second lane the fragment plus nuclear extract in the absence of competitor DNA. Subsequent lanes show a titration of fragment plus nuclear extract and increasing amounts of *E. coli* competitor DNA as indicated. The solid black box denotes the first exon of the *ckb* gene. The bent arrow designates the start point of transcription.

oxidase (33) regulation. It is the same region that was shown to be important for *in vitro* transcription of the *ckb* promoter (Fig. 3).

To determine which sequences are recognized by the site binding activities a series of oligodeoxynucleotide competitors was designed to separate sequences that are believed to be potential factor binding sites based on data from other laboratories i.e., the CCAAT, TATA and CACCC sequences (Fig. 5C). Only oligodeoxynucleotide 262/263 that contains the upstream TATAAATA sequence was able to compete the binding activity in brain extracts. HeLa cell extracts also contain binding activities that recognize the *ckb* promoter. However, with HeLa cell extracts no one oligodeoxynucleotide is able to compete for the binding activity. Three major retarded bands a, b and c are observed (Fig. 5b). Oligodeoxynucleotide 262/263, which contains the TATAAATA sequence, competes bands a and b while band c is increased in intensity. No change in binding pattern is observed with any other oligodeoxynucleotide

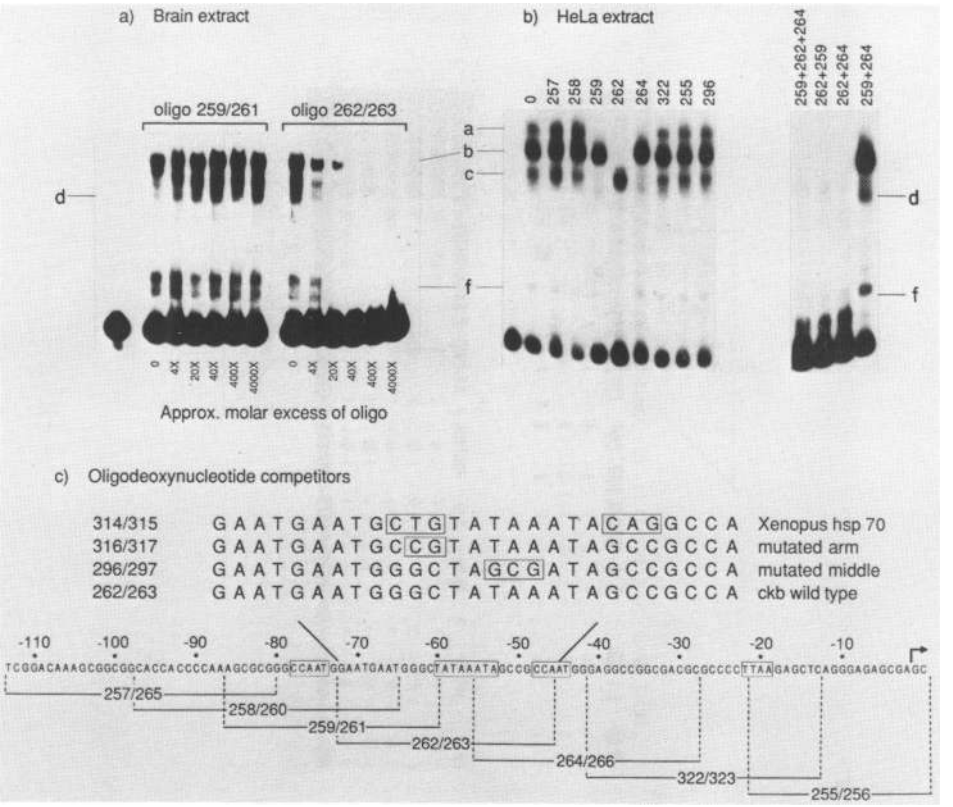

Figure 5. Gel retardation analysis of binding to the *NaeI/NcoI* fragment derived from the rat *ckb* promoter with brain extract (a) or HeLa extract (b). Binding is competed with double stranded oligodeoxynucleotides derived from the rat *ckb* promoter region as indicated in (c). In (a) competition was performed with increasing concentrations of the oligodeoxynucleotides indicated in the presence of 1 μg sonicated *E. coli* competitor DNA. In (b) retarded bands generated using HeLa extract (11 μg) were competed with 6 μg sonicated *E. coli* competitor DNA and 1000X excess of the oligodeoxynucleotides indicated. Each double-stranded oligodeoxynucleotide is indicated by the first number of its single stranded components. In (c) potential regulatory sequences are enclosed in boxes. In addition mutated regions in each of the mutated versions of 262/263 are also enclosed in boxes.

except 259/261 and 264/266. These contain the CCAAT sequences that flank the TATAAATA sequence and compete away bands a and c but leave band b. However, a mixture of oligodeoxynucleotides 262/263, 259/261 and 264/266 provides effective competition of all bands. Similar competition can be achieved with oligodeoxynucleotide 262/263 in combination with either of the CCAAT sequence containing oligodeoxynucleotides 259/261 or 264/266. Conversely, a mixture of the two CCAAT-containing oligodeoxynucleotides 259/261 and 264/266 does not result in complete competition. Band b remains along with enhanced band d. The simplest interpretation of these data is that HeLa nuclear extracts contain binding

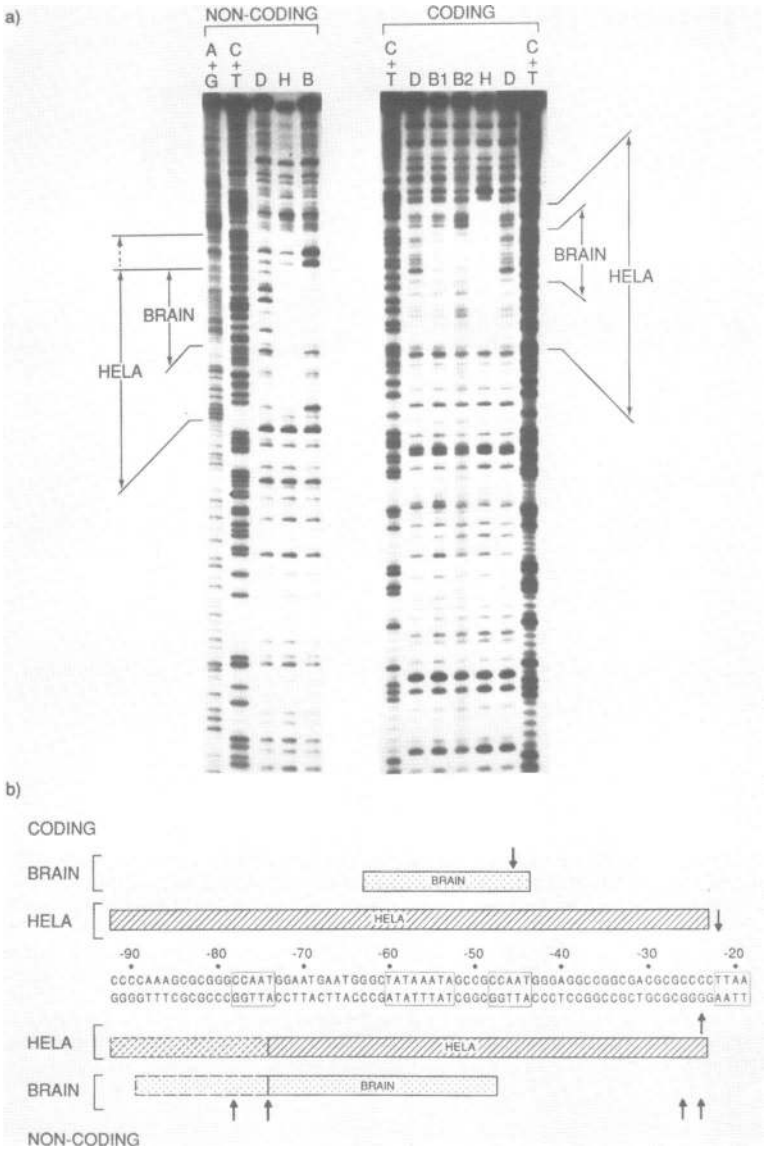

Figure 6. a) DNase I footprinting analysis of the rat *ckb* promoter region. Footprinting was carried out using a 243 base pair *Hin*fI-*Pst*I or *Hin*fI-*Hin*dIII fragment end labeled at the *Hin*dIII site (coding strand) or at the *Hin*fI site (non-coding strand) using polynucleotide kinase. The *Hin*fI site lies at +57 in the *ckb* gene and the *Pst*I and *Hin*dIII sites are in the polylinker beyond the *Hin*cII site at -186 in the gene. A+G or C+T Maxam and Gilbert sequencing lanes are used as markers. D indicates DNase I digestion alone, H digestion in the presence of 25 µg HeLa extract and 10 µg sonicated *E. coli* DNA. B represents digestion in the presence of 6 µg brain extract and 3 µg sonicated *E. coli* DNA. Results of footprinting are summarized in (b). Arrows indicate the location of hypersensitive sites. CCAAT, TATA and TTAA sequences are enclosed in boxes.

activities that are capable of recognizing the TATAAATA sequence and the two flanking CCAAT containing sequences. Band b is derived from the TATAAATA binding activity and band c from CCAAT-related binding activities. Band a could be derived from a combination of TATAAATA and CCAAT activities. The origin of bands d and f is unknown but they may represent proteolysis products of the TATAAATA binding activity and appear prominent in Fig. 5a. In addition the oligodeoxynucleotide competition results imply that CCAAT and TATAAATA binding activities can bind independently, as CCAAT binding activity remains after TATAAATA binding has been competed away and vice versa.

Localization of potential protein binding sites by DNAaseI footprinting

DNAase I footprinting experiments (Fig. 6) show that both brain and HeLa extracts generate footprints on both strands of the *ckb* promoter. The brain footprints span a region of approximately 30 base pairs centered on the TATAAATA sequence. The HeLa footprint includes this region but extends over a considerable area of flanking sequence that extends from the CACCC sequence at -92 to the TTAA sequence at -22. Although the brain footprint is less extensive, hypersensitive sites are observed distant from the protected region, on the non-coding strand at positions -24 and -26.

The TATAAATA-containing oligodeoxynucleotide 262/263 is able to compete for the brain footprint (Fig. 7b). Other oligodeoxynucleotides derived from the wild-type sequence are all without effect. In addition, oligodeoxynucleotide 296/297 where the TATAAATA sequence has been replaced by TAGCGATA (Fig. 5) is unable to compete for binding as judged by the gel retardation assay (data not shown) and is also unable to compete in the footprinting reaction.

The *ckb* TATAAATA sequence is contained in a region of dyad symmetry (Fig. 5). The *Xenopus* heat shock 70 promoter (34) has a functional TATA box that is also a perfect concensus TATA with surrounding dyad symmetry (Fig. 5). A mutant oligodeoxynucleotide 314/315 is identical to 262/263 except that it replaces the three bases that flank the TATAAATA sequence with those derived from the heat shock promoter. This mutant oligodeoxynucleotide is a poor competitor for the brain footprinting activity suggesting that in addition to the TATAAATA sequence, flanking sequences are important for binding activity. To test whether the flanking dyad symmetry was important for binding, oligodeoxynucleotide 316/317 was created (Fig. 5C). This contains an inversion mutation that destroys the symmetry flanking the TATAAATA sequence, and is able to compete for the brain binding activity at an efficiency similar to that of the wild type sequence. These experiments suggest that the brain footprinting activity observed on the *ckb* promoter is related to the activity assayed by gel retardation. This activity is sensitive to the TATAAATA sequence and sequences flanking this region but the presence of the full dyad symmetry does not appear to be essential for binding.

Similar competition experiments were performed with the HeLa derived footprinting activity. (Fig 7). Competition was observed with oligodeoxynucleotides that contained the TATAAATA sequence (262/263) or the two flanking CCAAT sequences (259/261 and 264/266).

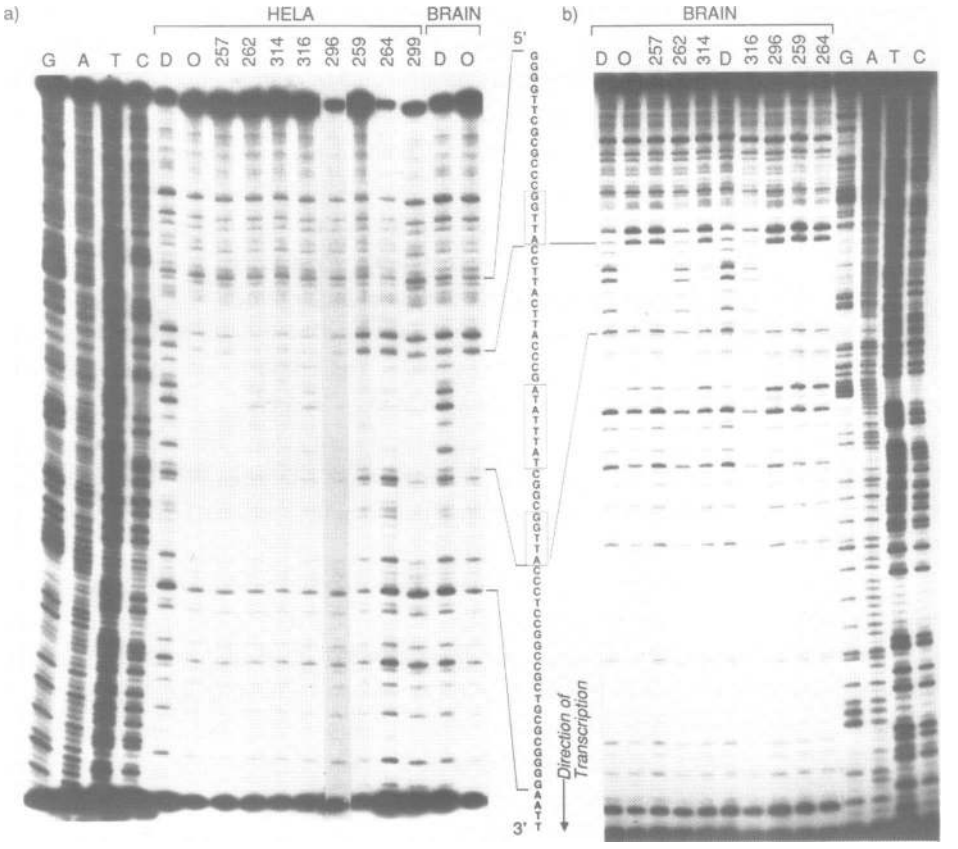

Figure 7. DNase I footprinting on the non-coding strand indicated in Fig. 6, in the presence of oligodeoxynucleotide competitors (Fig. 5). a) Oligodeoxynucleotide competition on HeLa extract derived footprints; b) Similar competition on brain extract derived footprints. GATC indicates Maxam and Gilbert sequencing ladders. Footprinting was performed using 25 μg HeLa extract and 6 μg brain extract as indicated in the figure. D indicates DNase alone; O extract but no competing oligodeoxynucleotide. In all other lanes competing oligodeoxynucleotides are present as indicated by the numbers that indicate the first member of each complementary annealed pair.

Competition with oligodeoxynucleotide 262/263 was weak. Incomplete restoration of bands in the TATAAATA segment of the sequence was observed. The same relative ability to compete was observed with mutant oligodeoxynucleotides as was observed for the brain footprint. More dramatic competition was observed with either of the CCAAT containing oligodeoxynucleotides 259/261 and 264/266 which generated a pattern indistinguishable from that obtained with brain nuclear extract. An oligodeoxynucleotide that contains the potential CCAAT sequence from the *ckm* promoter (299/300 sequence 5'-GGTGCCTGGGCAATGAGCTGAAAGCTC-3') was a much less effective competitor.

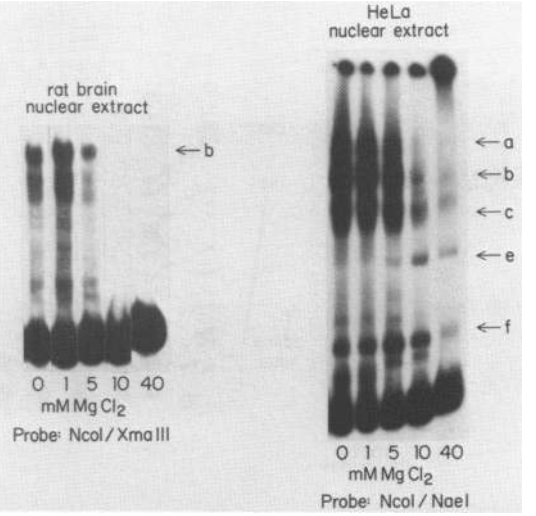

Figure 8. Gel retardation analysis showing sensitivity of binding of brain extract (left) and HeLa extract (right) to increasing concentrations of magnesium chloride. Probes are as indicated, see Fig. 4. Rat brain reactions contained approximately 1 µg nuclear extract, 0.1 ng probe DNA and 2 µg sheared *E. coli* competitor DNA. HeLa reactions contained 11 µg nuclear extract, 0.1 ng probe DNA and 6 µg sheared *E. coli* competitor DNA.

Competition was observed for the upstream CCAAT binding, but much weaker competition was observed around the downstream CCAAT sequence. These results are consistent with those obtained using the gel retardation assay. They suggest HeLa cells contain binding activities that recognize the TATAAATA sequence and the two flanking CCAAT-containing sequences in the rat *ckb* promoter. The TATAAATA and CCAAT binding activities can bind independently.

Sensitivity of binding activities and transcription to divalent cations

The binding activities from brain and HeLa extracts are sensitive to magnesium. (Fig. 8). The brain binding activity shows optimal binding in 1mM $MgCl_2$. Binding decreases with increasing magnesium concentration until at 10mM $MgCl_2$ it is completely abolished. A similar observation is made using HeLa extracts but a new band, band e, becomes prominent.

In contrast to transcription from the adenovirus major late promoter, transcription from the *ckb* promoter was highly sensitive to magnesium (Fig. 9a). Two transcripts were generated from the *ckb* promoter. One was consistent with transcription initiation at the *in vivo* rat brain start site. The second appeared to be initiated from a point 30-40 base pairs upstream. This is the approximate location of the expected transcript start point if the upstream TATAAATA sequence could function as a classical TATA box. A point mutation in this TATA sequence (TAGAAATA) destroys transcription from this upstream start point (manuscription in preparation). At low magnesium concentrations (< 6mM) the downstream

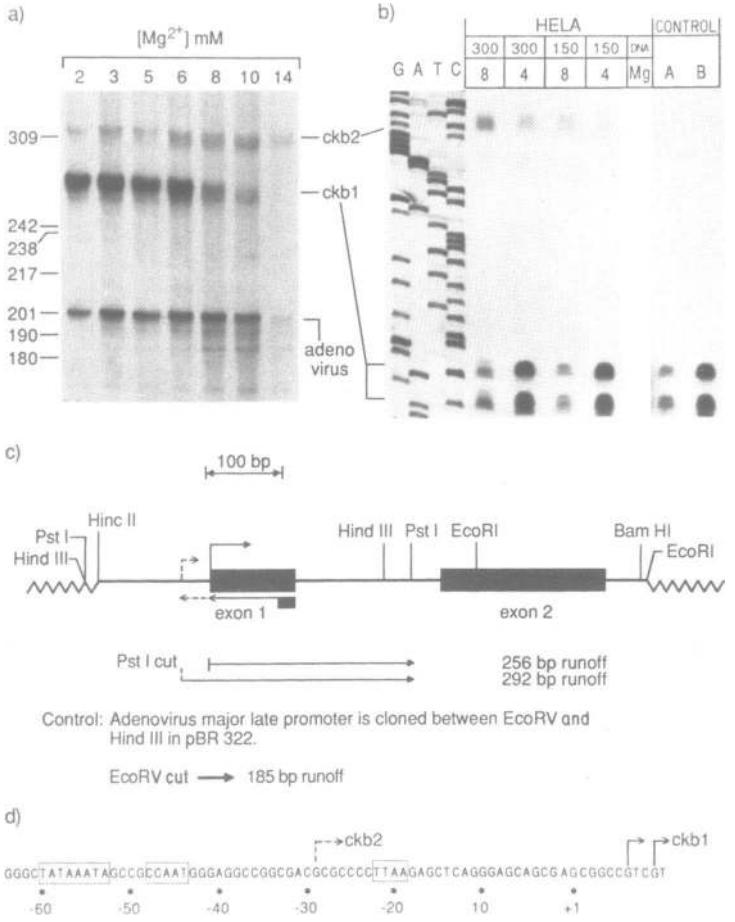

Figure 9. Sensitivity of the *in vitro* transcription of the *ckb* promoter using HeLa extracts to magnesium chloride. a). Run off transcripts generated using linearized templates as indicated in (c). The adenovirus control is derived using pø4 (49) cut at the *Eco*RV site and represents run-off transcription driven by the adenovirus major late promoter. The marker lane contained pBR322 DNA digested with *Hpa*II. These DNA markers run 10% faster than equivalently sized single-stranded RNA. This accounts for the apparent anomaly in size of the run-off RNA transcription products. In (b) products are generated from equivalent *ckb* supercoiled templates and analyzed by primer extension using oligodeoxynucleotide E50336-119. GATC sequencing marker lanes were generated by dideoxy sequencing (50) off the template with the same primer. HeLa transcription products were analyzed at different DNA template concentrations per reaction (20 µl) and magnesium concentrations as indicated. 14 µg (A) rat heart and 2 µg (B) rat brain RNA were used as control templates. The deduced transcript start points for ckb1 and ckb 2 are indicated in (d). CCAAT, TATA and TTAA sequences are enclosed in boxes.

start point was much stronger than the upstream start point. As the magnesium concentration is raised to 10 mM, transcription from the downstream start point decreases and that from the upstream start point remains unchanged. At 10 mM $MgCl_2$ the two transcripts have

approximately equal intensity. The location of these start points has been confirmed by primer extension analysis of transcripts from supercoiled templates Fig. 9 b. Again the relative use of the two start sites is dependent on magnesium concentration. At low DNA concentration (150 ng/reaction) and low magnesium (4 mM) the downstream start point is preferred over the upstream start point by about ten fold. As the Mg^{2+} is raised to 8 mM transcription off the downstream start point drops to a level similar to that off the upstream start point. Similar results are obtained at higher DNA concentrations (300 ng reaction). Therefore, conditions that may eliminate binding to the upstream TATAAATA sequence i.e. high Mg^{2+}, competing DNA concentrations (data not shown), increase transcription from the upstream start point and decrease transcription from the downstream start point.

DISCUSSION

The *ckb* promoter contains several sequence features that have been associated with functional activity in other promoters, a nonconcensus TATA sequence (-22) and an upstream concensus TATA sequence (-59) flanked by two CCAAT sequences (-47, and -76). Further upstream (-97) is a CACCC-like sequence. These features are highly conserved in the human *ckb* promoter although this promoter contains in addition three potential Sp1 binding sites that are not conserved in rat. Our data strongly suggest that in the tissues we have examined, the nonconcensus TATA sequence region at -22 is used preferentially to direct the start point of transcription despite the presence of a perfect concensus TATA sequence nearby. To our knowledge the *ckb* gene provides the first example where a nonconcensus TATA sequence may be used in preference to an alternative more canonical TATA sequence. We cannot rule out the possibility that the upstream TATA sequence may be used under conditions that we have not tested and we have shown that in the context of a HeLa *in vitro* transcription extract both TATA sequences can be functional under appropriate conditions of salt and template concentration.

The role of CCAAT sequences in gene regulation is not completely understood and they are not found as promoter elements in all genes. However, functional CCAAT sequences are usually located approximately 50 base pairs upstream of the TATA box although this distance can be variable (35). The *ckb* promoter has the appearance of two tandemly arranged promoters with closely juxtaposed TATA and CCAAT sequences. This is of interest in view of the recent finding that the chicken *ckb* gene generates two distinct message products (and ultimately two proteins) from the same gene by alternate splicing (Perriard J.-C. personal communication). Although the promoter structure for this gene is currently unknown, this observation raises the possibility that the ancestral gene may have had two promoters that were differentially regulated.

A 61 base pair sequence between -98 and -37 is important for efficient *in vitro* transcription from a minimal *ckb* promoter. This sequence contains both the CCAAT sequences and the upstream TATA sequence and is recognized by sequence specific binding

activities found in nuclear extracts from both HeLa cells and brain. One of these activities recognizes the upstream concensus TATA motif and can be detected by both gel retardation assays and DNase I footprinting using crude nuclear extracts. Several laboratories have identified a factor that recognizes the functional TATA box element [TFIID (36), BTFI (37)] but this factor is usually undetectable by the gel retardation assay using standard crude nuclear extracts. Furthermore, *in vitro*, conditions that favor binding of this *ckb* TATAAATA (TA-rich) recognition protein (TARP), suppress transcription driven from the upstream TATA motif. Binding of this protein is not competed by the TTAA sequence region that provides the principal TATA box function for the *ckb* gene. Therefore, we favor a model in which binding to the upstream TATA element by TARP enhances transcription from the downstream TTAA box. In the absence of TARP the upstream TATAAATA sequence might in addition compete with the downstream promoter elements for the transcriptional machinery thus providing an additional level of control.

In addition to an activity that recognizes the upstream TATA motif, HeLa nuclear extracts contain activities that recognize the two CCAAT-containing flanking sequences. Several proteins have been identified that interact with the CCAAT sequence motif e.g. CTF/NF1 (38), CBP (39) and CTY (40). Thus it appears that there is a family of proteins that recognize sequences that contain the CCAAT core motif but which have differential affinities for different CCAAT containing sequences (41). These proteins activate transcription and/or replication and in some instances have been found to be the prime mediators of regulatory responses e.g., the cell type specificity of the heat shock response for *Xenopus hsp*70 (42), and the response of the collagen gene to transforming growth factor β (43). As oligodeoxynucleotides directed against each of the two CCAAT containing regions both compete away the CCAAT region binding activity, and the CCAAT motif is the sequence feature that they share in common we conclude that this binding is a reflection of a CCAAT binding activity. This conclusion is further supported by the observation that an oligodeoxynucleotide containing the *ckm* CAAT sequence also has the ability to compete partially, and that mutations in these CCAAT-containing sequences affect the transcriptional efficiency from the *ckb* promoter (Hobson et al., manuscript in preparation). We are however, unable to explain the lack of ability of oligodeoxynucleotide 258/260 that also contain the CCAAT motif to compete in either the gel retardation assay or the footprinting assay. Two possibilities are that 1) there is a technical problem with this oligodeoxynucleotide or 2) that sequences downstream of the CCAAT motif are important for binding and these are not present in this synthetic sequence.

Nuclear extracts from rat brain appear to be deficient in the binding activity that recognizes the CCAAT domains, but still contain the TARP. We have observed that nuclear extracts from rat brain that, unlike those analyzed here, are transcriptionally active, do contain additional activities that recognize the CCAAT domains (Mitchell et al., manuscript in

preparation). Therefore, we believe that the lack of this CCAAT recognition activity is not a fundamental difference between brain nuclei and HeLa nuclei, but is rather an artifact of the two different nuclear preparation schemes. Nevertheless, these particular brain extracts do allow us to examine binding of the TARP in the absence of CCAAT binding.

DNase I footprinting data (Figs. 6 and 7) suggests that the combined TATA and CCAAT recognition activities in HeLa cells lead to the protection of a stretch of DNA that extends from the TTAA box at -22 to the CACCC motif at -94. Several potential arrangements of factors on the *ckb* promoter are possible and we cannot determine at present which combinations may be functionally important. Oligodeoxynucleotide competition experiments coupled with the gel retardation assay and DNase I footprinting suggest that the TATA and CCAAT recognition activities can bind independently. The presence of a gel retarded band a that competes with both CCAAT and TATA oligodeoxynucleotides suggests that both sequences can be bound simultaneously. The competition of the upstream CCAAT footprint in the absence of competition of the downstream footprint (Fig. 7a, oligodeoxynucleotide 299) suggests that downstream CCAAT-binding can occur in absence of binding to the upstream CCAAT. Whether all three activities can bind simultaneously is unclear from this data.

The region protected from DNase I extends all the way to the TTAA sequence at -22. This implies that proteins binding in this region probably come close enough to this downstream sequence to interact with a potential TFIID-like TATA transcription factor. This would allow for a mechanism of activation where binding of these proteins stabilizes TFIID binding by direct protein interaction without the need for DNA bending.

The sequence TTAA is rarely used as a TATA box. Two examples where this sequence may perform this function are the adenovirus EIIa promoter and the human adenosine deaminase promoter. Each of these promoters shows some additional sequence homology with the *ckb* promoter. In the adenovirus EIIa promoter this homology extends beyond the immediate TTAA sequence to include a region 15 base pairs downstream. Of particular interest is the observation that the adenovirus EIIa promoter also contains tandem TATA elements both of which are functional although the downstream TTAA provides the principal TATA function. Deletion of the upstream TATA increases transcription from the downstream TTAA (26) and is therefore, functionally similar in its effect to the binding of TARP by the upstream TATA element for *ckb*. The *ckb* and adenovirus EIIa promoters show significant homology upstream of the TTAA in regions including the second TATA element and regions corresponding to boxes A and B as defined by Kedinger and co-workers (Fig. 1) (27,28). Among the differences between the two promoters is the lack of CCAAT or potential Sp1 binding sites for the adenovirus promoter, and the observation that the upstream TATA box in adenovirus is also non-concensus. The EIIa promoter is an E1a activated promoter and the sequence requirements for its activation appear to be complex. Sequences between -29 and -81 have been implicated in E1a transactivation (45) and in particular sequences defined by boxes B and A,

(Fig. 1) although these sequences alone are not sufficient to convey E1a inducibility to a heterologous promoter. Furthermore, binding (E2F) has been identified to the region between -33 and -71 (45), and to boxes B and A (28). No binding has been reported that recognizes the upstream TATA element. By contrast protein binding to the *ckb* promoter in HeLa cells appears to target the CCAAT and TATA regions, a feature that is not conserved in the adenovirus promoter.

Whether the similarity between the adenovirus EIIa promoter and the *ckb* promoter is of any functional significance remains to be determined. For example does the similarity reflect common evolutionary origins or common mechanisms of regulation? With respect to the latter, two observations are of interest. Firstly,*ckb* has been reported to respond to a wide variety of hormonal stimuli (5,6). Furthermore, Colberg-Poley and Santomenna (manuscript submitted) have recently shown that *ckb* message levels show a slight but consistent increase upon infection of human foreskin fibroblasts with cytomegalovirus. Thus expression of this cellular gene may respond to conditions of viral infection. One theory that has been proposed for the mechanism of action of viral activators e.g., E1a, pseudorabies virus IE proteins, is that they may serve to stabilize the rate-limiting interaction of TFIID with its cognate promoter (46). In fact for some genes the principal mediator of E1a activation appears to be a specific TATA box (47). It has already been observed that not all TATA boxes are equivalent in their ability to respond to particular regulatory signals (48). It will be of interest to determine if the structure of the cellular *ckb* promoter reflects the presence of common regulatory mechanisms shared with viral promoters.

The second gene that contains a promoter TTAA sequence is the human adenosine deaminase gene. Again, homology extends beyond immediate TTAA sequence this time to include sequences 23 base pairs upstream in a region different from that shared with the adenovirus promoter. The human adenosine deaminase gene (25) also contains a series of potential Sp1 binding sites that precede the TTAA sequence and is in this respect reminiscent of the human *ckb* promoter. These potential Sp1 binding sites are not conserved in the rat and we have so far no evidence for Sp1 binding to the rat *ckb* promoter.

In conclusion we have examined the nature of protein binding to a small segment of the rat *ckb* promoter that appears to be essential for its ability to support transcription *in vitro*. HeLa cells and brain contain binding activities that recognize a TATA motif and two CCAAT containing segments. The *ckb* promoter demonstrates homology with other genes that appear to use the sequence TTAA as a functional TATA box. We do not know whether the TTAA sequence in the *ckb* promoter serves as a classical TATA box or whether it can interact with TFIID. However, our data suggest that the upstream TATAAATA sequence does not provide the principal TATA function in the sources we have examined. We are currently performing mutagenesis experiments aimed at a more detailed understanding of the sequences required for expression and regulation of this interesting gene, and examining the interaction of regulatory sequences with *trans*-activating factors.

ACKNOWLEDGEMENTS

We would like to thank Penni Lockhart for typing the manuscript. We are grateful to Dr. Rick Scott and Dr Anamaris Colberg-Poley for helpful comments and critical reading of the manuscript.

*To whom correspondence should be addressed

REFERENCES

1. Watts, D.C. (1973) In Boyer, P.D. (Ed.), TheEnzymes, Vol. 8, Academic Press, NY. pp.383-455.
2. Walliman, T., Walzthöny G., Wegmann, D., Moser, H., Eppenberger, H.M., and Barrantes, F.J. (1985) J. Cell Biol. 100, 1063-1072.
3. Urdal, P., Urdal, K., and Strømme, J.H. (1983), Clin. Chem. 29, 310-313.
4. Loike, J.D., Kozler, V.F., and Silverstein, S.C. (1984). J. Exp. Med. 159, 746-757.
5. Sõmjen, D., Kaye, A.M. and Binderman, I. (1985) Biochem J. 225, 591-596.
6. Reiss, N.A. and Kaye, A.M. (1981) J. Biol. Chem. 256, 5741-5749.
7. Carney, D.N., Gazdor, A.F., Bepler, G., Guccion, T. G., Marangos, P. J., Moody, T. A., Zweig, M.H. and Minna, T.D. (1985) Cancer Research 45, 2913-2923.
8. Dawson, D.M. and Fine, I.H. (1967) Arch. Neurol. 16, 175-180.
9. Benfield, P. A., Graf, D., Korolkoff, P. N., Hobson, G., and Pearson, M.L. (1988) Gene 63, 227-243.
10. Jaynes, J.B., Johnson, J.E., Buskin, J.N., Gartside, C.L. and Hauschka, S.D. (1988) Mol. Cell. Biol. 8, 62-70.
11. Hobson, G., Mitchell, M., Korolkoff, P.N., Molloy, G. and Benfield, P.A., (1987) in New Insights into the regulation of transcription , 6th Annual Stonybrook Symposium, Stonybrook, NY p.4.
12. Graham, F.L. and van der Eb, A.J. (1973) Virology 52, 456-467.
13. Chirgwin, J.M., Pryzbyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979). Biochemistry, 18, 5294-5299.
14. Villarreal-Levy, G., Ma, T.S., Kerner, S.A., Roberts, R. and Perryman, M.B. (1987) Biochem. Biophys. Res. Comm. 144, 1116-1127.
15. Blobel, G. and Potter, V.R., (1966), Science 154, 1662-1665.
16. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983), Nucl. Acids Res. 11, 1475-1489.
17. Parker, C.S. and J. Topol. (1984) Cell 36, 357-369.
18. Shapiro, D.J., Sharp, P.A., Wahl, W.W. and Keller, M.J. (1988) DNA 7, 47-55.
19. Manley, J.L., Fire, A., Cano, A., Sharp, P.A., and Gefter, M.L. (1980). Proc. Natl. Acad. Sci. USA 77, 3855-3859.
20. Fried, M. and Crothers D.M. (1981) Nucleic Acids Res. 9, 6505-6525.
21. Singh, H., Sen, R., Baltimore, D. and Sharp, P.A. (1986) Nature 319, 154-158.
22. Babiss, L.E., Herbst, R.S., Bennett, A.L. and Darnell, J.E., Jr. (1987) Genes and Development 1,256-267.
23. Maxam, A. M. and Gilbert, W. (1980), Methods Enzymol. 65, 499-560.
24. Chuvpilo, S.A. and Kravchenko V.V. (1984) FEBS Lett. 179, 34-36.
25. Valerio, D., Duyvesteyn, M.G.C., Dekker, B.M.M., Weeds, G., Berkvens, Th.M., Van der Voorn, L., Van Ormondt, H. and Van der Eb, A.J. (1985), EMBO J. 4, 437-443.
26. Zajchowski, D.A., Boeuf, H. and Kédinger, C. (1985) EMBO J. 4, 1293-1300.
27. Zajchowski, D.A., Boeuf, H. and Kédinger, C. (1987) Gene 58, 243-256.
28. Boeuf, H., Zajchowski, D.A., Tamura, T., Hauss, C. and Kedinger, C. (1987). Nucleic Acids Res. 15, 509-527.
29. Daouk, G.H., Kaddurah-Daouk, R., Putney, S., Kingston, R. and Schimmel (1988) J. Biol. Chem. 263, 2442-2446.
30. Mariman, E.C.M., Broers, C.A.M., Claesen, C.A.A., Tesser, G.I. and Wieringa, B. (1987) Genomics 1, 126-137.
31. Dynan, W.S. and Tjian, R. (1985) Nature 316, 774-778.
32. Myers, R. M., Tilly, K. and Maniatis, T. (1986) Science 232, 613-618.
33. Schüle, R., Muller, M., Otsuka-Murakami and Renkawitz, R. (1988) Nature 332, 87-90.
34. Bienz, M. and Pelham, H.R.B. (1986) Cell 45, 753-760.
35. Breathnach, R. and Chambón, P. (1981) Ann. Rev. Biochem. 50, 349-383.
36. Sawadogo, M. and Roeder, R.G. (1985) Proc. Natl. Acad. Sci. USA 82, 4394-4398.
37. Zheng, X., Moncollin, V., Egly, J. and Chambon, P. (1987) Cell 50, 361-368.

38. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. and Tjian, R. (1987) Cell 48, 79-89.
39. Graves, B.J., Johnson, P.F. and McKnight, S.L. (1986) Cell 44, 565-576.
40. Hooft van Huijsduijnen, R.A.M., Bollekens, J., Dorn, A., Benoist, C. and Mathis, D. (1987) Nucleic Acids Res. 15, 7265-7282.
41. Raymondjean, M., Cereghini, S. and Yaniv, M. (1988) Proc. Natl. Acad. Sci. USA 85, 757-761.
42. Bienz, M. (1986) Cell 46, 1037-1042.
43. Rossi, P., Karsenty, G., Roberts, A.B., Roche, N.S., Sporn, M.B. and de Crombrugghe, B. (1988) Cell 52, 405-414.
44. SivaRaman, L. and Thimmappaya, B. (1987) Proc. Natl. Acad. Sci. USA 84, 6112-6116.
45. Reichel, R., Kovesdi, I. and Nevins, J.R. (1988) Proc. Natl. Acad. Sci. 85, 387-390.
46. Abmayr, S.M., Feidman, L.D., and Roeder, R.G. (1985) Cell 48, 821-829.
47. Simon, M.C., Fisch, T.M., Benecker, B.J., Nevins, J.R. and Heinz, N. (1988) Cell 52, 723-729.
48. Horna, F.L., Glorioso, J.C. and Levine, M. (1988) Genes and Development 2, 40-53.
49. Hu, S.L. and Manley, J.K. (1981) Proc. Natl. Acad. Sci. USA 78, 820-824.
50. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.