Nucleic Acids Research, Vol. 19, No. 17 4631–4637

Cloning and structural analysis of cDNA and the gene for mouse transcription factor UBF

Koji Hisatake, Tetsuji Nishimura, Yasushi Maeda, Ken-ichi Hanada, Chao-Zhong Song and Masami Muramatsu*
Department of Biochemistry, Faculty of Medicine, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Received June 10, 1991; Revised and Accepted August 6, 1991 EMBL accession no. X60831

ABSTRACT

The gene and protein structure of the mouse UBF (mUBF), a transcription factor for mouse ribosomal RNA gene, have been determined by cDNA and genomic clones. The unique mUBF gene consists of 21 exons spanning over 13 kb. Two mRNAs coding for mUBF1 and mUBF2 having 765 a.a. and 728 a.a., respectively, are produced by an alternative splicing of exon 8. It specifies 37 amino acids constituting a part of the regions homologous to high mobility group proteins (HMG box 2). A human UBF (hUBF) cDNA obtained by polymerase chain reaction also indicates the presence of two kinds of mRNAs, the shorter form lacking the same regoin as mUBF2. Comparison of the cDNAs from hUBF and mUBF revealed an unusual conservation of nucleotide sequence in the 3'-terminal non-coding region. We examined the relative amounts of expression of mUBF1 and mUBF2. The eight tissues studied contained both molecular species, although mUBF2 was the predominant form of UBF. The mRNA of mUBF1 was expressed one half of the mUBF2 in quiescent mouse fibroblasts but reached the same amount in growing state.

INTRODUCTION

Transcription of ribosomal RNA gene (rDNA) requires multiple protein factors beside RNA polymerase I (1–6). Among these factors, proteins that bind to the promoter region of rDNA have been characterized in a number of laboratories (3,7–18). One of these proteins, referred to as TFID (5–9) or SL-1 (10–12), recognizes the CPE (core promoter element) (1,19–21) and is required for species-specific transcription in vitro. UBF is another factor that stimulates transcription of rDNA 10–100 fold by binding to UCE (upstream control element) (5,12–14,22,23) that is located −100 ~ −160 bp upstream of the transcription initiation site. UBF, together with TFID/SL1, shows broad footprints at the promoter region of rDNA (8,11,15). Thus, UBF plays an important role in the programming of initiation complex on the rDNA promoter. However, little is known about the molecular mechanisms through which UBF exerts its potentiating function on rDNA transcription. To make clear the molecular mechanisms, UBF was purified by using conventional and sequence-specific DNA affinity chromatography from *Xenopus laevis* (24), rat (25), mouse (26) and human (27). No differences were found between hUBF and mUBF in terms of reactivity to antibody, DNA binding activity, pattern of footprinting and activation of transcription (27). In the case between hUBF and *Xenopus laevis* UBF (xUBF), whereas their DNA recognition properities are very similar, they could not be interchanged functionally (24). The comparison in four verbetrates has revealed UBFs are well conserved. The purified UBFs consist of two polypeptides of molecular weight 97 kd and 94 kd in human, rat and mouse (13,15) and in the case of *Xenopus laevis*, 85 kd and 82 kd (26). These two UBFs may originate from two distinctive genes. Alternatively, they may be produced by posttranscriptional or posttranslational modifications. Recently, Jantzen et al. have cloned and reported the complete nucleotide sequence of one form of hUBF cDNA (27). When this clone was expressed in mammalian cells using a recombinant vaccinia expression system, only 97 kd polypeptide seen in SDS polyacrylamide gel electropforesis was produced. However, the origin of 94 kd form remained unclear.

In order to address the origin of the two polypeptides and also for the purpose of utilizing mUBF in in vitro transcription systems, we have cloned cDNAs and the chromosomal gene for mUBF and examined their sequence and organization. We show that mUBF contains two molecular forms, designated mUBF1 and mUBF2. The smaller molecule, mUBF2, lacks 37 amino acids consisting of a part of HMG box 2 that is homologous to the nuclear high mobility group (HMG) protein 1 and 2 in mUBF1. Using PCR, we have also identified a shorter form of hUBF mRNA that lacks the same region. The deleted region in mUBF2 is found to form exon 8 (111 bp) itself. Because the other part of these mRNAs have identical nucleotide sequence and mUBF gene appears to exisit only once in mouse genome according to Southern blot analysis, we conclude that these two mRNAs are formed by alternative splicing and both mUBFs are the translation products of the two different mRNAs.

* To whom correspondence should be addressed

While this paper was in preparation, two reports appeared that described the cloning and characterization of the cDNA of rat UBFs (rUBFs) (28) and an xUBFs (29). Their results on the cDNA are in good agreement with ours described here.

MATERIALS AND METHODS

Cells and RNA preparation

Mouse FM3A cells, a mouse mammary tumor cell line, and HeLa cells were cultured in ES medium containing 5% fetal calf serum and harvested at the density of 3×10^5 cells/ml. Mouse NIH3T3 cells were cultured in α-medium containing 10% fetal calf serum. A quiescent state was achieved by the additional 2-day culture after confluence and the shift-up culture was carried out by changing the growth medium to a fresh one. RNA was extracted as described previously (30). Poly $(A)^+$ RNA was purified by an oligo(dT)-cellulose (Pharmacia type 7) column according to the standard procedure (31).

Construction of hUBF fragment by PCR

A pair of oligonucleotides, 5′-CCGAATTCATGAACGGAG-AAGCCGACTGCCCC-3′ and 5′-CCGTCGACCTGCTTCTT-GTTGATGTTCAGCAT-3′, which correspond to the hUBF mRNA sequences from +1 to +24 and from +1134 to +1159, respectively (27), were synthesized using an Applied Biosystem 381A DNA Synthesizer. The primers contained the EcoRI or SalI recognition sequence at 5′-end. First strand cDNA was made from HeLa cell poly $(A)^+$ RNA using cDNA synthesis kit (Boehringer Mannheim) according to manufacturer's protocol. Polymerase chain reaction (PCR) was done using cDNA as template in the standard condition (30) for 30 cycles at 94°C for 1 min, 55°C for 2 min and 72°C for 3 min. The PCR products were extracted and cloned between EcoRI and SalI sites of pBluescript SK^+ (Stratagene).

Genomic Southern blot and Northern blot

Mouse liver DNA and HeLa DNA were digested to completion by each restriction enzyme, and electrophoresed on a 0.8% agarose gel. DNA was transferred to a nitrocellulose filter (Schleicher & Schuell) as described previously (30). Hybridization was performed according to the standard procedure (30). Washing was done at room temperature in 2×SSC, and then at 60°C in 0.5×SSC. The filter was dried and autoradiographed.

Construction and screening of a cDNA library

Poly $(A)^+$ RNA was used for cDNA synthesis. The cDNA library was constructed with FM3A cDNA according to the standard procedure (30). PCR products were used to screen a mouse brain cDNA library (kindly provided by Dr. H.Hamada) as described in section 3 except that the final washing was done at 60°C in 1×SSC. Another FM3A cDNA library was screened with a N-terminal EcoRI-SacI fragment derived from λUBFb-2.

Screening of a mouse genomic library

A mouse liver genomic library was constructed using λFixII (Strategene) and screened with a fragment derived from the cDNA clone λUBFb-2. Hybridization was carried out as described above except that final washing was done at 65°C in 0.1×SSC.

Sequencing

Sequencing was done using Sequenase kit version 2.0 (USB) with electrophoresis in 6% polyacrylamide−8M urea gel.

Expression of mUBF1 and mUBF2

Total RNA was extracted from mouse MH134 cells and mouse NIH3T3 cells at various times after shift-up culture as described above. Total RNA from different organs was extracted by guanidinium/cesium chloride method as previously described

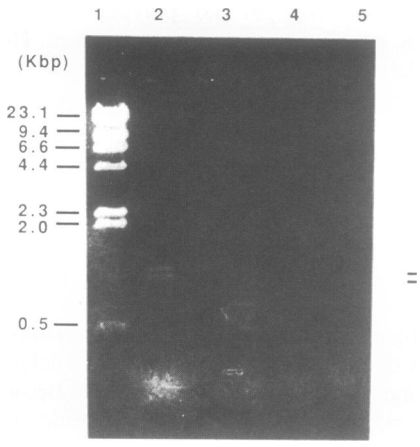

Figure 1. PCR products of hUBF. Oligo(dT) primed first-strand cDNA from HeLa cells was amplified by PCR using the primer pair (+1−+24) /(+1134−+1159) as described in Materials and Methods. PCR products were analyzed by agarose gel electrophoresis. Lane 1 shows marker λ DNA digested with HindIII. Two PCR products were synthesized using a pair of primer (lane 2), but not by one primer alone lane 3 (primer: +1−+24), lane 4 (primer: +1134−+1159), lane 5 (no primer).

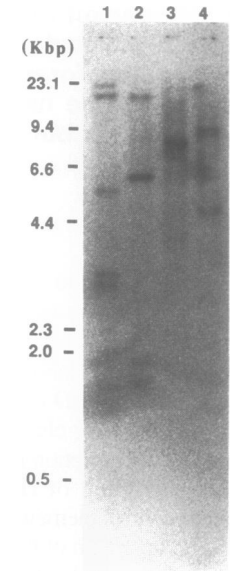

Figure 2. Genomic Southern blot of human and mouse DNA probed with hUBF probe. Ten μg of HeLa (lane 1 and lane 2) or mouse liver DNA (lane 3 and lane 4) was digested with EcoRI (lanes 1 and 3) and HindIII (lanes 2 and 4) and subjected to electrophoresis on a 0.8% agarose gel. After blotting to a nitrocellulose filter, the filter was hybridized with hUBF DNAs labeled by random primed method, washed in 1×SSC at 60°C and autoradiographed.

(30). The mUBF1 fragment from BstNI (+746) to SacI (+931) was inserted into pBluescript to make a template plasmid pUBF7BS which was linealized with PvuII and used for preparing ^{32}P-labeled probe by TransProbe T kit (Pharmacia). Total RNA (5 μg), ^{32}P-labeled probe and carrier tRNA (5 μg) were mixed and dissolved in hybridization buffer (80% formamide, 0.4 M NaCl, 0.05 M PIPES (pH6.4), 1 mM EDTA), heated at 80°C for 10 min, hybridized at 50°C overnight and then digested with RNaseT1 (2 μg/ml) plus RNaseA (40 μg/ml) in 0.3 M NaCl, 10 mM Tris-HCl (pH7.5), 5 mM EDTA. Samples were analyzed by 4% polyacrylamide−8 M urea gel electrophoresis.

RESULTS

Cloning and primary structure of mUBF cDNA

In order to clone the cDNA for mUBF, we first cloned N-terminal half of the hUBF by PCR using the published sequence of hUBF cDNA clone (27). A pair of oligonucleotides were synthesized as primers that corresponded to the hUBF mRNA sequences from +1 to +24 and from +1134 to +1159. PCR was performed on the first strand cDNA prepared from HeLa cell poly (A)$^+$ RNA. Two specific bands appeared (Figure 1, lane 2), one of which was slightly shorter than the expected length (1184 bp).

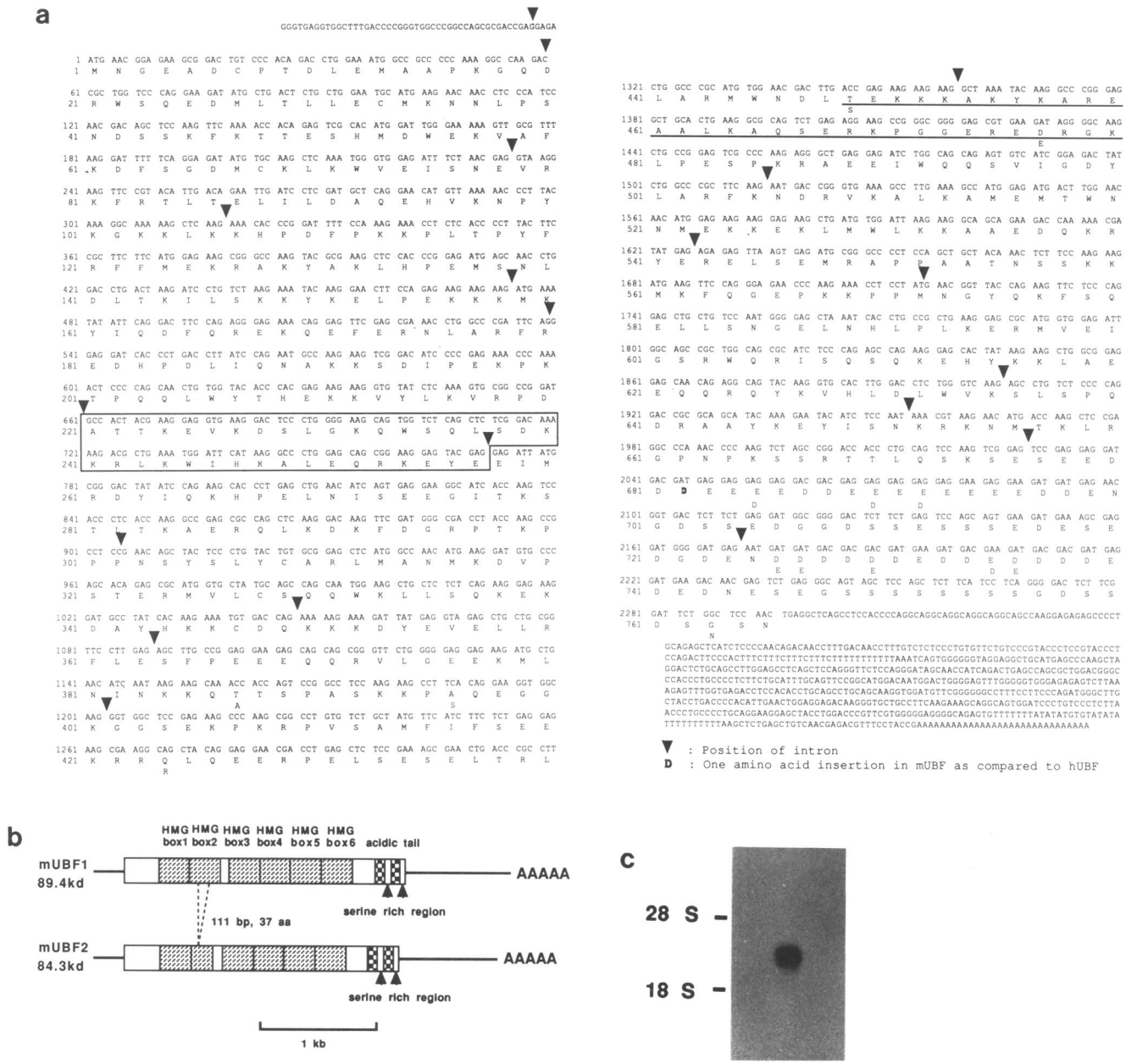

Figure 3. Structure of cDNA of mUBF. a) Nucleotide sequence of the cDNA coding for mUBF1 and the predicted amino acid sequence. DNA sequence from λC7, λC9, UBFb-2 were combined. hUBF amino acids that are different from mUBF are shown in the third line. Arrow heads indicate the position of introns. Outlined D indicates the one asparagine insertion in mouse UBF. Boxed region is the exon 8 that is deleted in mUBF2. Possible nuclear localization sequence is underlined. b) Two alternatively spliced forms of mUBF. Overall structure of mUBF1 and mUBF2 is shown. The position of a 37 a.a. deletion of HMG box 2 in UBF1 is indicated in the Figure. c) Northern blot analysis of mUBF transcripts. Two μg of poly(A)$^+$ RNA from mouse FM3A cells was electrophoresed on a 1.0% agarose/formaldehyde gel and transferred to a nitrocellulose filter. The insert fragment of λC9 was labeled by random primed method. Hybridization and washing was done as described in Materials and Methods. Positions of 28S and 18S rRNA are indicate.

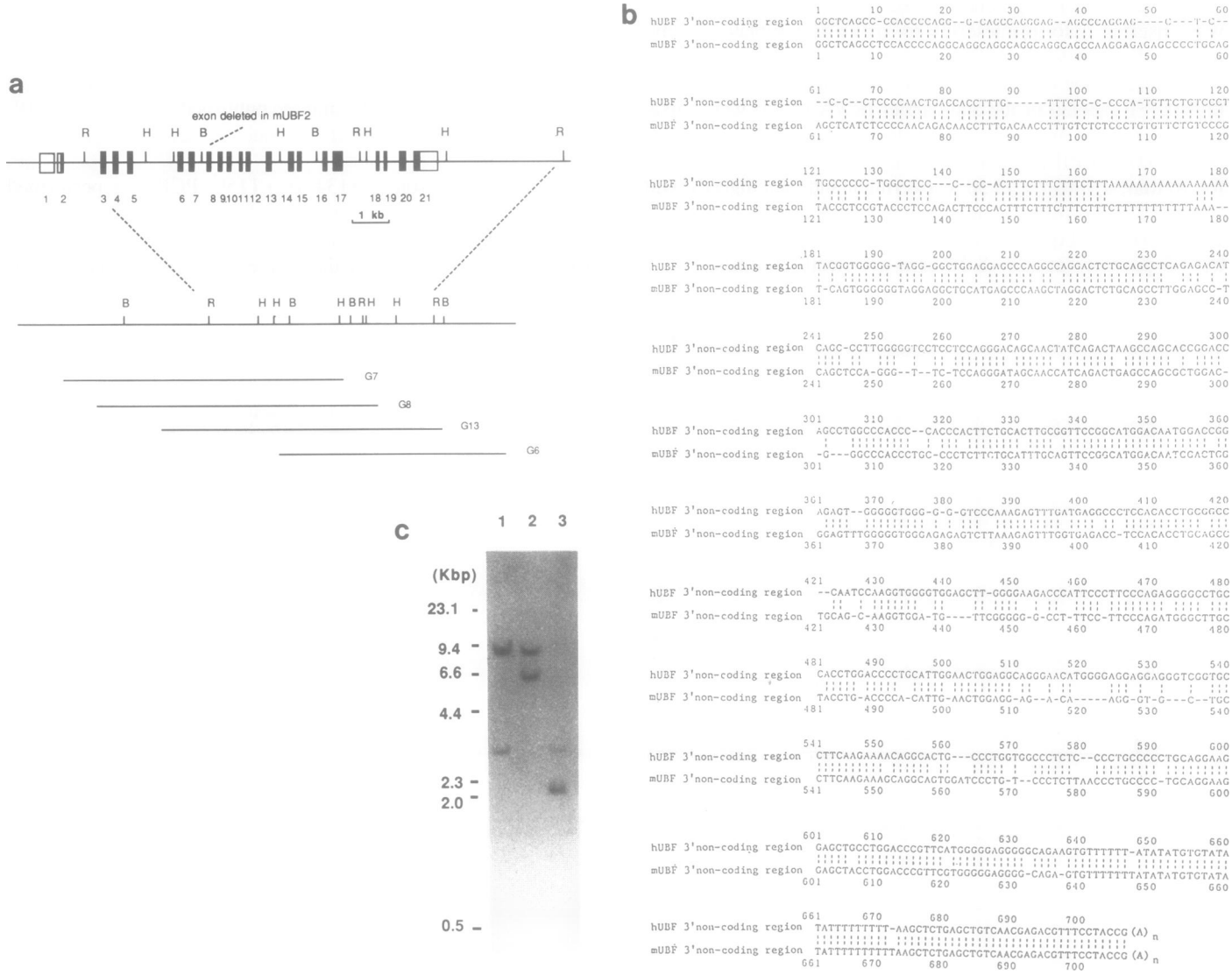

Figure 4. Structure of mUBF gene. a) Exon-intron structure and restriction map of the mUBF gene region covering about 20 kb. It was constructed by aligning four overlapping phage clones G7, G8, G13 and G6. Restriction sites of EcoRI, HindIII, and BamHI are indicated by R, H and B. Positions of exons are indicated by boxes numbered 1−21. Filled-in boxes indicate coding region. Open boxes in exon 1,2 and 21 indicate 5′ and 3′ non-coding regions. Exon 8 that is deleted in mUBF2 is indicated in the Figure. b) Comparison of 3′ non-coding region between human and mouse UBF cDNAs. Nucleotide sequences of the 3′ non-coding region of human and mouse UBF are compared. Gaps are inserted to allow maximum matching. Nucleotide sequences right after the A of TGA stop codon are shown. Poly (A) tract begins immediately after the last G. c) Mouse genomic Southern blot. Ten μg of mouse liver DNA was digested with BamHI (lane 1), EcoRI (lane 2) and HindIII (lane 3), electrophoresed on a 0.8% agarose gel, and then transferred to a nitrocellulose filter. The insert fragment of λC9 was labeled by random primed method and used for hybridization. Washing was done under stringent conditions (see Materials and Methods) and the filter was autoradiographed.

No specific band was observed with either one (Figure 1, lane 3 and 4) or no primer (Figure 1, lane 5). These two PCR products were subcloned and sequenced. As both of the clones contained the expected sequences, the inserted DNA fragments were used as a probe for screening a mouse cDNA library. Genomic Southern blots of human and mouse DNA with these probes revealed several bands with EcoRI or HindIII digestion, respectively (Figure 2). These results indicate that both human and mouse UBF are encoded by one or a few genes. We have screened 5×10^5 phage plaques of a mouse brain cDNA library and two clones were obtained. The longer 1.8 kb clone λUBFb-2 was sequenced and found to contain a part of the coding region for mUBF. This clone lacks 111 bp that code for 37 amino acids compairing the published sequence of hUBF. We have also sequenced the two kinds of hUBF PCR products (Figure 1, lane 2) and found that the shorter hUBF clone lacks the same 111 bp region. We designate the larger and the smaller mUBF proteins as mUBF1 and mUBF2, respectively. We could not obtain the mUBF1 clone from a mouse brain cDNA library. However, by screening 5×10^5 unamplified phage plaques of the mouse FM3A cDNA library, we obtained 10 clones and found that two clones, λc7 and λc9, had a complete open reading frame for UBF1. λc9 was sequenced for both strands (Figure 3). As already shown for UBFs from other species (25,28,29), mUBF has six HMG boxes covering entire molecule and two acidic domains with alternate serine rich sequences in the C-terminal region (Figure 3b). There is a highly basic domain in the HMG box 4 which may represent a nuclear localization sequence (Maeda et. al., unpublished).

We analyzed poly $(A)^+$ RNA of mouse FM3A cells by

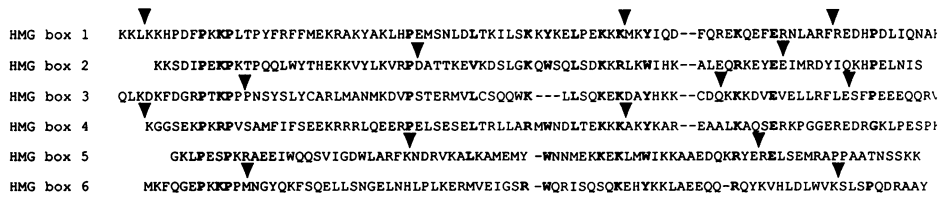

Figure 5. Alignment of six HMG boxes to indicate relative positions of introns. Six HMG boxes are aligned to allow maximum matching. Positions of introns are indicated by arrow heads above amino acid sequences. Arrows directly above amino acids R in HMG box 1 and P in HMG box 3 indicate that these introns fall between 2nd and 3rd codons of the respective amino acids. For exact positions of introns, see Figure 3a. Bold amino acids indicate conserved amino acid residues among six HMG boxes.

Northern blots with the λC9 insert DNA as probe and detected only one band of 3.2 kb. It was previously reported (25) that bands of 3.2 kb and additional weak 4.5 kb were detectable using human UBF cDNA as a probe in mouse and human cells. However, from the small size difference we suggest that our 3.2 kb band contains mRNAs for both mUBF1 and mUBF2, and that 4.5 kb species may represent a partially processed form of the precursor RNA.

Cloning and structural analysis of mUBF gene

To address the question of whether these two transcripts represented by different cDNAs are derived from two distinct genes or they are derived from a single gene but alternatively spliced, we cloned mouse chromosomal UBF gene from 1×10^6 plaques of mouse genomic library. We obtained 10 positive clones and examined the exon-intron organization by sequencing λG13 clone using synthetic oligonucleotides as primer. Mouse UBF gene extends over 13 kb of genome and contains 21 exons (Figure 4a). Organization of exons and introns and exact position of introns are shown in Figure 3. The 111 bp region lacking in the mUBF2 cDNA was found to be exon 8 itself.

We mapped the 3′ end of mUBF mRNA by comparing cDNA and genomic clones (Figure 4b). Curiously, a poly (A) addition signal was not found in 650 bp of 3′ non-coding region in spite of the presence of a poly (A) stretch. Also, comparison of nucleotide sequence between mouse and human has disclosed unusual conservation of 120 bp of DNA just upstream of poly (A) addition site; 95% identical as compared to 91% in the UBF coding region. It is possible that, in addition to the signal for polyadenylation, this region could have some other important functions such as regulation of mRNA stability and translation.

Although we expected that six HMG boxes were similarly divided into distinct exons, this was found not to be the case (Figure 5). Only two pairs of introns fall in the same positions in HMG box 1 and 4, which can be most readily explained by duplication of a common ancestor. Other discordant introns may be the result of de novo insertion of introns after six HMG boxes have evolved from a common ancestral HMG box. On the other hand, two repeats of highly acidic amino acids in so-called acidic tail are separately encoded in two distinct exons together with each serine rich cluster, probably reflecting a duplication event that produced these two repeats from a single primordial sequence seen in HMG proteins (33,34).

Mouse genomic DNA was digested with BamHl, EcoRl or Hindlll and Southern blot was carried out using the λC9 insert as a probe (Figure 4c). The bands detected are two 8.5 kb and 2.8 kb with BamHl, 8.5 kb and 6 kb with EcoRl and 10 kb, 2.8 kb, 2.2 kb and 2.1 kb with Hindlll. These fragments are all present in our genomic clones which overlap each other. This result proves that mUBF has a single gene in mouse genome and mUBF1 and mUBF2 transcripts are indeed the alternatively spliced forms originating from this unique gene.

Expression of the two forms of mUBF

We first examined the relative amount of expression of mUBF1 and mUBF2. Total RNA was extracted from various tissues and examined for mRNA by RNase protection assay. As shown in Figure 6a (also see Figure 6b), two specific bands of 189 bp and 165 bp were protected, each representing UBF1 and UBF2 mRNA. Table 1 shows that all the tissues examined contained both molucular species, although the relative amounts were different significantly one another. mUBF2 was the more predominant form of UBF in most tissues. Table 1 also shows that mUBF2 was expressed about twice higher than mUBF1 in stationary phase but that both were expressed at almost the same level in the growing phase in both mouse fibriblast NIH3T3 cells and mouse ascites hepatoma MH134 cells. Next, RNA was extracted from mouse NIH3T3 cells at various times after a shift-up culture and examined (Figure 6b). The mRNA of mUBF2 was expressed nearly twice as much as that of the mUBF1 in quiescent cells as well as at 1 and 2 hr after medium change. Although both mRNA species tend to increase after nutritional shift-up, the ratio of increase of mUBF1 mRNA exceeded that of mUBF2 mRNA and both concentrations became the same at 4 hr, when the expession level reached a maximum, several times that of quiecent cells (Figure 6b, lane7; Figure 6c).

DISCUSSION

mUBF is highly homologous to hUBF

We have cloned two kinds of cDNAs and a gene for mUBF and determined their primary structures. Sequence comparison between mUBF and hUBF has shown identities of 98% in amino acid sequence and 91% in DNA sequence. In mUBF, there is one aspartic acid insertion in position 683 of hUBF. C-terminal region of UBF has two highly acidic repeats consisting almost exclusively of aspartic and glutamic acids. The nucleotide sequence of these repeats are $(GAX)_n$, forming repeats of simple DNA sequences that are prone to change their number. The changes in amino acids are most prominent in this region and there are 8 reciprocal changes between Asp and Glu. Since these two amino acids are acidic, the only constraint of this region may be its acidic negative charge and β-turn structure and not its primary amino acid sequence. Among other amino acid changes, Thr to Ala, Gln to Arg and Thr to Ser, two are conservative in nature.

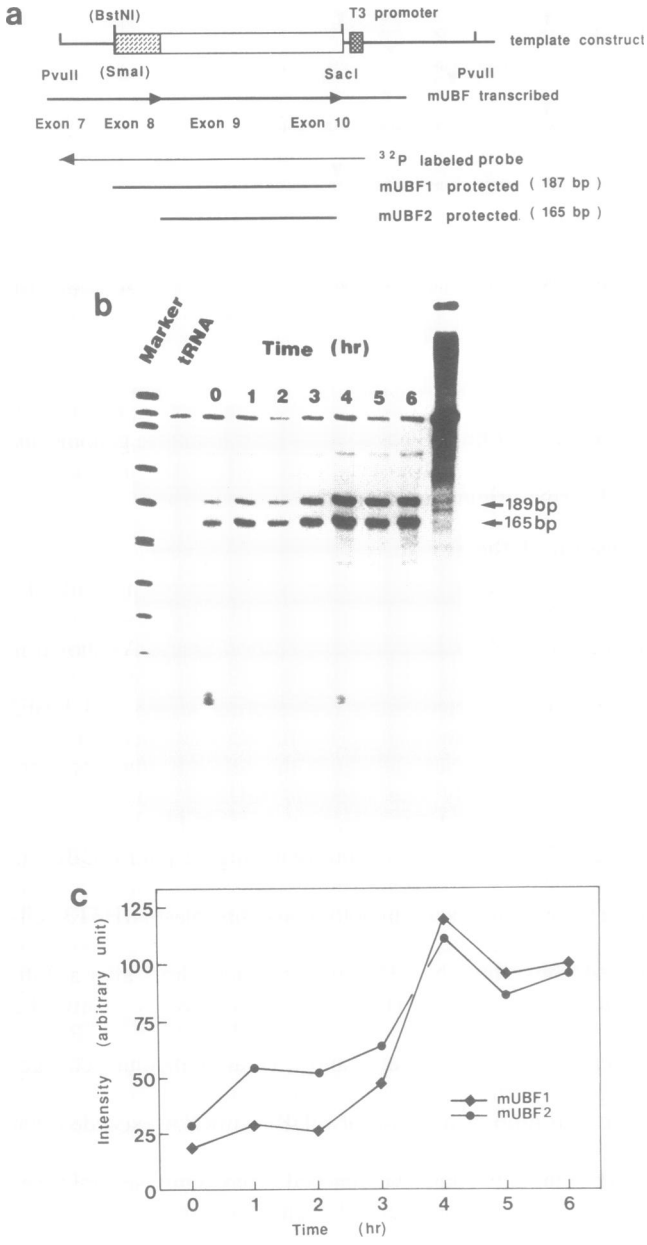

Figure 6. Comparison of the relative amounts of expression of mUBF1 and mUBF2. a) The map of the construct for preparing probe. b) RNase protection assay. Total RNA were prepared from NIH3T3 cells at various time after the shift-up. lane 1; marker lane 2; tRNA lane 3–9; total RNA from the cells incubated for 0,1,2,3,4,5 and 6 hr after medium change lane 10; ^{32}P-labeled probe only. The longer fragment (189 bp) represents the content of mUBF1 mRNA, the shorter fragment (165 bp) represents the content of mUBF2 mRNA. c) The time course of the relative amounts of expression of mUBF1 and mUBF2. For lane 5 (2 hr), the intensity is normalized by the non-specific bands and tRNA control.

A poly (A) addition signal, AATAAA, was not found in 650 bp of 3′ non-coding region of mUBF gene. The lack of AATAAA or its derivatives in polyadenylated mRNA is reminiscent of yeast mRNAs in which three kinds of consensus sequences for transcriptional termination and poly (A) addition signal are proposed (35–37). The consensus sequence of Henikoff et al., TTTTTATA is found about 60 bp upstream of poly (A) addition site of mouse and human UBF genes. The lack of AATAAA and conservation of this sequence between human and mouse

Table 1. The expression of mUBF1 and mUBF2 in various of tissues and in the stationary and growing phase.

		mUBF1/mUBF2			mUBF1/mUBF2
Mouse C3H/He	Lung	0.20	NIH3T3 cells	Stationary phase	0.60
	Liver	0.27		Growing phase	1.06
	Heart	0.56			
	Muscle	0.73			
	Kidney	0.73	MH134 cells	Stationary phase	0.59
	Stomach	0.83		Growing phase	1.06
	Brain	0.86			
	Pancreas	0.91			

The tissues were used from mouse C3H/He. NIH3T3 cells were cultured in α-medium containing 10% fetal calf serum for 2 days after confluence (stationary phase) and changed the growth medium, incubated for 4 hr (growing phase). MIH134 cells were grown in peritoneal and then incubated in α-medium containing 0.5% fetal calf serum (stationary phase) or 10% (growing phase). The intensity of the autoradiographs was determined by densitometer.

make it reasonable to assume that it may be the poly (A) signal for UBF gene.

UBF is encoded by two alternatively-spliced mRNAs

Our novel finding is that the two mUBFs are encoded by one gene and their mRNAs are produced by an alternatine splicing of exon 8. mUBF1 and mUBF2 most probably correspond to the previously identified two polypeptides of 97 kd and 94 kd (11) for the following reasons. 1) Differences in calculated molecular weight between mUBF1 and mUBF2 is 5.1 kd which is close to 3 kd difference between two purified polypeptides. 2) A shorter form of mRNA lacking the same sequence was also detected by PCR in hUBF (Figure 1, lane 2). 3) What we now call hUBF1 cloned by Jantzen et al. (27), when expressed in HeLa cells, produced a 97 kd polypeptide. 4) Partial V8 proteolysis suggested that 97 kd and 94 kd forms of hUBF are structurally very similar (27). Purifications of UBFs from other laboratories have shown the presence of two polypeptides in hUBF (27), mUBF (26), rUBF (25) and xUBF (24). In human,rat and mouse, there are 97 kd and 94 kd polypeptides and in *Xenopus laevis*, there are 85 kd and 82 kd polypeptides. It is noteworthy that although overall lengths are variable between species, the difference in size is constantly 3 kd. These facts may suggest that the existence of two forms of UBF is conserved and that these two forms (UBF1 and UBF2) are produced by a similar mechanism. In the case of *Xenopus laevis*, xUBFs are a little smaller in size than mammalian counterparts. In a recent paper, Bachvarov and Moss indicate that xUBFs delete a part of HMG box 3 of mammalian UBF as seen by cDNA sequence (29).

Functional significance of the two forms of UBF

Functional implication of the existence of two forms of UBF is still unknown. Recent cloning of transcription factors such as Ubx (38), USF (39), BTF3 (40), FosB (41), CREB (42) and CTF/NF-1 (43) has revealed several examples of presence of alternatively spliced forms. A general transcription factor BTF exists as two alternatively spliced forms, BTF3a and BTF3b, the latter lacking the first 44 amino acids of the former. While both BTF3a and BTF3b can interact with RNA polymerase II, only BTF3a is transcriptionally active (40). ΔFosB is a truncated form of FosB that lacks the C-terminal amino acids of FosB and has a similar ability to form a heterodimer with Jun and bind with AP-1 site. However, this ΔFosB/Jun complex is defective in transcriptional activation and repression (41). Two forms of CREB proteins, CREB and ΔCREB, exist that are probably

produced from alternatively spliced mRNAs (42); ΔCREB lacks a region which interacts cooperatively with kinase A phosphorylation motif, and has 10-fold less activity in transcriptional activation. In the present work, we have shown that relative amounts of mUBF1 and mUBF2 mRNAs are different among different tissues and the ratio of these mRNAs changes after nutritional shift-up. These facts suggest a possiblity that UBF1 and UBF2 are functionally distinct, although we do not have any evidence for that at present. Although UBF2 lacks large part of HMG box 2, difference in DNA binding is not very likely for the following reasons. First, both 97 kd and 94 kd polypeptides, that probably correspond to UBF1 and UBF2, respectively can be efficiently recovered by purification using DNA affinity column. Second, in one report only HMG box 1 was shown to be necessary and enough for DNA binding (27). This point, however, has to be reexmamined according to our preliminary data (to be published). Another possibility is the difference in transcriptional activity. Transcriptional activity of UBF1 expressed by recombinant vaccinia virus is shown to be equivalent to a mixture of UBF1 and UBF2 purified from cells (27). But this does not eliminate the possibility that UBF2 alone might have a different transcriptional activity. Studies using separate preparations of UBF1 and UBF2 are necessary to determine the functional significance of the two forms of UBF proteins.

Secondary structure prediction and evolutionary implications

Mouse UBF is a highly charged protein. Hydrohpathy profile determined by the method of Hopp and Wood confirms this hydrophilic nature of this protein. Two prominent hydrophilic regions in the C-terminus correspond to acidic tail. Secondary structure analysis predicts that the each C-terminal half of 6 HMG boxes have a strong tendency to form α-helix and acidic tail to form β-turn. We compared Edmundson wheels of α-helix forming region of 6 HMG boxes (data not shown). These helices have in general hydrophilic amino acids (especially rich in lysine and arginine) on one side and hydrophobic amino acids on the other forming an amphipathic helix. This amphipathy is more prominent in HMG box 1, 2 and 6 than in HMG box 3, 4 and 5. We speculate that the C-terminal half of each HMG box, which is rich in basic amino acids, contacts directly DNA via hydrophilic and basic region of the amphipathic helix.

Recent literature has documented a number of HMG box-containing proteins. UBF is unique in that it has six HMG boxes. HMG proteins are non-specific DNA binding proteins that preferentially bind to AT rich and single-stranded DNA, but the UBF, which has a similar DNA binding domain, is a sequence specific DNA binding protein, albeit the sequence selectivity is not so stringent. It is interesting to know the structural basis for specific DNA binding activity in UBF HMG boxes and how they are developed from the primordial HMG box with non-specific DNA binding activity seen in HMG 1 and 2 proteins.

REFERENCES

1. Grummt, I. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 727–731.
2. Mishima, Y., Yamamoto, Y., Kominami, R. and Muramatsu, M. (1981) *Nucleic. Acid Res.*, **9**, 6773–6785.
3. Mishima, Y., Finacsek, I., Kominami, R. and Muramatsu, M. (1982) *Nucleic Acids Res.*, **10**, 6659–6670.
4. Schnapp, A., Pfleider, C., Rosenbauer, H. and Grummt, H. (1990) *EMBO J.*, **9**, 2857–2863.
5. Tower, J., Culotta J. C. and Sollner-Webb, B. (1986) *Mol. Cell. Biol.*, **6**, 3451–3462.
6. Kato, H., Nagamine, M., Kominami, R. and Muramatsu, M., (1986) *Mol. Cell. Biol .*, **6**, 3481–3427.
7. Nagamine, M., Kishimoto, T., Aono, J., Kato, H., Kominami, R. and Muramatsu, M. (1987) *Mol. Cell. Biol.*, **7**, 1486–1495.
8. Tanaka, N., Kato, H., Ishikawa, Y., Hisatake, K., Tashiro, K., Kominami, R. and Muramatsu, M. (1990) *J. Biol. Chem.*, **256**, 13836–13842.
9. Cavanaugh, A. H. and Thompson, E. A. (1985) *Nucleic Acids Res.*, **13**, 3357–3369.
10. Learned, R. M., Cordes, S. and Tjian, R.(1985) *Mol. Cell. Biol.*, **5**, 1358–1369.
11. Bell, S. P., Jantzen, H-M. and Tjian, R. (1990) *Genes Dev.*, **4**, 943–954.
12. Learned, R. M., Learned, T. K., and Haltiner, M. M. and Tjian, R. (1986) *Cell*, **45**, 847–857.
13. Cassidy, B., Haglund, R. and Rothblum, L. (1987) *Biocim. Biophys. Acta*, **909**, 133–144.
14. Smith, S. D., Oriahi, E., Lowe, D., Yang-Yen, H-F., O'Mahony, D., Rose, K., Chen, K. and Rothblum, L. I. (1990) *Mol. Cell. Biol.*, **10**, 3105–3166.
15. Bell, S. P., Leraned, R. M., Jantzen, H-M. and Tjian, R. (1988) *Science*, **241**, 1192–1197.
16. Clos, J., Buttgereit, D. and Grummt, I. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 604–608.
17. Schnapp, A., Clos, J., Haedelt, W., Schreck, R., Cvekl, A. and Grummt, I. (1990) *Nucleic Acids Res.*, **18**, 1385–1393.
18. Bateman, E. and Paul, M. R. (1988) *Nucleic Acids Res.*, **8**, 1940–1946.
19. Yamamoto, O., Takakusa, N., Mishima, Y., Kominami, R. and Muramatsu, M (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 299–303.
20. Kishimoto, T., Nagamine, M., Sasaki, T., Takakusa, N., Miwa, T., Kominami, R. and Muramatsu. M. (1985) *Nucleic Acids Res.*, **13**, 3515–3532.
21. Miller, K., Tower, J. and Sollner-Webb, B. (1985) *Mol. Cell. Biol.*, **5**, 554–562.
22. Haltiner, M., Smale, S. and Tjian, R. (1986) *Mol. Cell. Biol.*, **6**, 227–235.
23. Smale, S. and Tjian, R. (1985) *Mol. Cell. Biol.*, **5**, 352–362.
24. Pikaard, C. S., MacStay, B., Schultz, M. C., Bell, S. P. and Reeder, R. (1990) *Genes Dev.*, **3**, 1179–1788.
25. Pickaard, C., Smith, S. D., Reeder, R. and Rothblum, L. I.(1990) *Mol. Cell. Biol.*, **10**, 3810–3812.
26. Bell, S., Pikaard, C., Reeder, R. and Tjian, R. (1989) *Cell*, **59**, 489–497.
27. Jantzen, H-M., Admon, A., Bell, A. P. and Tjian, R. (1990) *Nature*, **344**, 830–836.
28. O'Mahony, D. J. and Rothblum, L. I. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3180–3184.
29. Bachvarov, D and Moss, T. (1991) *Nuceic Acids Res.*, **19**, 2331–2335.
30. Sambrook, J., Fritsch, E. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
31. Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1408–1412.
32. Gubler, U. and Hoffman, B. S. (1983) *Gene*, **25**, 263–269.
33. Tsuda, K., Kikuchi, M., Mori, K., Waga, S. and Yoshida, M. (1988) *Biochemistry*, **27**, 6159–6163.
34. Shirakawa, H., Suda, K. and Yoshida, M. (1990) *Biochemistry*, **29**, 4419–4423.
35. Zeret, K. S. and Sheeman, F. (1982) *Cell*, **28**, 563–573.
36. Bennetzen, E. and Paule, M. R. (1986) *Cell*, **47**, 445–450.
37. Henikoff, S., KElly, J. D. and Coohen, E. H. (1983) *Cell*, **33**, 607–614.
38. Kornfeld, K., Saint, R. B., Beachy, P. A., Harte, P. J., Peattie, D. A. and Honess, D.(1989) *Genes Dev.*, **3**, 243–258.
39. Gregor, P. D., Sawadogo, M. and Roeder, R. G. (1990) *Genes Dev.*, **4**, 1730–1740.
40. Zheng, X. M., Black, D., Chambon, P. and Egly, J. M. (1990) *Nature*, **344**, 556–559.
41. Nakabeppu, Y. and Nathans, D. (1991) *Cell*, **64**, 751–759.
42. Yamamoto, K. K., Gonzalez, G. A., Menzel, P., Rivier, J. and Montminy, M. R. (1990) *Cell*, **60**, 611–617.
43. Santro, C., Mermod, N., Andrew, P. C. and Tjian, R.(1988) *Nature*, **334**, 218–224.