Nucleic Acids Research, 1994, Vol. 22, No. 3 427–433

Ubiquitous expression of the 43- and 44-kDa forms of transcription factor USF in mammalian cells

Mario Sirito+, Qun Lin, Tapati Maity and Michèle Sawadogo*
Department of Molecular Genetics, The University of Texas, M.D.Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA

Received September 16, 1993; Revised and Accepted December 7, 1993 EMBL accession nos U01662 and U01663

ABSTRACT

USF is a helix-loop-helix transcription factor that, like Myc, recognizes the DNA binding motif CACGTG. Two different forms of USF, characterized by apparent molecular weights of 43,000 and 44,000, were originally identified in HeLa cells by biochemical analysis. Clones for the 43-kDa USF were first characterized, but only partial clones for the human 44-kDa USF (USF2, or FIP) have been reported. Here we describe a complete cDNA for the 44-kDa USF from murine cells. Analysis of this clone has revealed that the various USF family members are quite divergent in their N-terminal amino acid sequences, while a high degree of conservation characterizes their dimerization and DNA-binding domains. Interestingly, the 3′ noncoding region of the 44-kDa USF cDNAs displayed an unusual degree of conservation between human and mouse. In vitro transcription/translation experiments indicated a possible role for this region in translational regulation. Alternative splicing forms of the 44-kDa USF messages exist in both mouse and human. Examination of the tissue and cell-type distribution of USF by Northern blot and gel retardation assays revealed that while expression of both the 43- and 44-kDa USF species is ubiquitous, different ratios of USF homo- and heterodimers are found in different cells.

INTRODUCTION

The gene-specific transcription factor USF (also called MLTF or UEF) was first identified because of its involvement in transcription of the adenovirus major late gene (1–3). DNA sequences recognized by USF are characterized by the presence of a central CACGTG motif (1), which includes the canonical CANNTG recognition sequence of helix-loop-helix (HLH) transcription factors (4). Although the exact cellular function of USF remains unknown, it is noteworthy that most USF binding sites are also recognized by other families of transcription factors, such as Myc (5–7), MAX/Myn (8, 9), Mad/Mxi (10, 11) and TFE3/TFEB (12, 13). USF binding sites have been found in a variety of cellular and viral genes (14–25).

Characterization of USF-DNA complexes by electrophoretic mobility shift assays has revealed that crude as well as purified preparations of this transcription factor are apparently heterogeneous (2, 26). In agreement with this observation, two different polypeptides, with apparent molecular masses of 43- and 44-kDa, have been shown to be responsible for the USF activity present in HeLa cell nuclei (26, 27). These two different forms of human USF showed identical DNA-binding properties (27).

Isolation of cDNA clones to the 43-kDa form of human USF (USF1) has revealed that this transcription factor contains a leucine zipper motif immediately adjacent to the basic-HLH region responsible for dimerization and DNA-binding (28). Expression of this particular USF gene seems to be controlled by alternative splicing (28). Partial cDNA clones for a second human USF gene, designated USF2 or FIP (29, 30), as well as clones for USF family members from *Xenopus* (31) and sea urchin (32) were subsequently isolated. Comparison of these clones revealed a high degree of conservation in the C-terminal region of the various USF proteins. Interestingly however, the leucine zipper is absent in the sea urchin USF, while this same motif seems to play an important role in the dimerization and DNA-binding activities of all other USF family members (28, 29). Isolation of the FIP clone was made possible by the existence of a physical interaction between this protein and Fos (30).

Here, we report the characterization of a complete cDNA clone for the murine USF2 gene. Analysis of this clone confirmed the suspected relationship between USF2 (29) and the 44-kDa form of USF originally observed in HeLa cells (26). We show that USF2 expression is characterized by differential splicing patterns and alternative utilization of translation start sites and polyadenylation signals. Finally, gel retardation assays and Northern blotting analysis revealed that while USF is expressed ubiquitously, the relative abundance of the various USF proteins vary in different cell types.

* To whom correspondence should be addressed

+ Present address: Istituto di Biologia e Genetica, Università di Genova, Genova, Italy

MATERIALS AND METHODS

Isolation of probes for murine USF by RT-PCR

Two degenerate oligonucleotides (CARTTYTWYGTNATGATG and ATCCARTTRTTNATYTTRTC) were designed according to amino acid sequences most highly conserved in the various USF family members (i.e., USF-specific region and helix 1) (29). These sequences encompassed amino acids 159–164 and 213–219 of the human 43-kDa USF (28). A murine cDNA was prepared by extension of random primers with reverse transcriptase using poly(A)$^+$ RNA isolated from S194 cells (a B-cell line). PCR reactions were carried out with Taq DNA polymerase using 1 μM concentration of the degenerate oligonucleotides and two successive series of cycles (5 cycles at 92°C, 1 min, 55°C, 5 min, 72°C, 1 min, followed by 30 cycles at 92°C, 1 min, 45°C, 2 min, and 72°C, 1 min). The amplified 180-bp USF cDNA fragments were isolated, phosphorylated, end-filled with Klenow, and cloned into a pBKS$^-$ vector (Stratagene). Sequencing of several of the resulting clones revealed that a portion of the cDNAs for both the 43- and 44-kDa murine USF had been amplified. The insert from one of these subclones, corresponding to the 44-kDa USF, was isolated and used as a probe for library screening.

The amino acid sequences of the various USF genes were found to be almost entirely identical between human and mouse. This property was used to isolate a longer fragment of the murine 43-kDa USF cDNA for use in Northern analysis. This fragment was amplified by PCR using an oligonucleotide matching a portion of the region initially amplified and a degenerate oligonucleotide corresponding to amino acids 1–6 of human USF1 (28).

Library screening and DNA sequencing

A commercial cDNA library from NIH3T3 cells in a lambda ZapII vector (Stratagene) was used to isolate murine USF2 clones. From a total of 300,000 plaques screened, two different cDNA clones, M2-2 and M2-5, were each isolated twice. DNA sequencing was performed using the dideoxy termination method (33) with double-stranded DNA templates and Sequenase (US Biochemical). The sequence of the entire M2-2 clone was determined on both strands using a combination of Exonuclease III deletions, subclones of various restriction fragments and specific oligonucleotide primers. Most of the M2-5 sequence information was determined for only one of the DNA strands.

In vitro transcription/translation and bacterial expression of USF cDNAs

USF1 and USF2 RNAs were produced by transcription of the human 43-kDa USF (28) or murine 44-kDa USF (clone M2-2) cDNAs with T7 RNA polymerase. These RNAs were purified and used to program rabbit reticulocyte lysates (Promega), either separately or in combination at various ratios. Translation reactions were carried out in the presence of [^{35}S]methionine for 2 h at 30°C. Control reactions contained either buffer or Brome Mosaic Virus (BMV) RNA. In vitro translated products were analyzed on 12% polyacrylamide gels.

For bacterial expression of the 44-kDa USF, the *Bam*HI fragment corresponding to nucleotides 93–1532 of M2-2 was subcloned into a pV2 vector (34). This process yielded clone pVM2, which allowed inducible expression in *E.coli* of the murine USF (amino acids 9–346) with a stretch of six histidines at its N-terminus. Expression of the recombinant VM2 protein was found to be relatively low, independent of the bacterial strain used, whether XL1-Blue or any of the Top strains (Stratagene). The bacterially produced protein was purified by chromatography on a column of immobilized Ni^{2+} as previously described (34), followed by an additional purification on an FPLC Mono Q column (Pharmacia) precisely as described previously for HeLa USF (26).

USF–DNA interaction studies

DNase I footprinting was carried out using bacterially produced human 43-kDa (34) and murine 44-kDa USF using a protocol essentially identical to the one originally described for human USF (26). The relative affinity of each protein for the USF site present in the adenovirus 2 major late promoter was determined using the standard USF filter binding method (26). Gel retardation assays performed as previously described using as a probe a 260-bp DNA fragment containing the adenovirus major late promoter (26,29).

Northern blot analysis

A blot containing poly(A)$^+$ RNA (2 μg per lane) isolated from various mouse tissues (Clontech) was probed repetitively to determine the relative abundance of messages hybridizing with the 43- and 44-kDa murine USF cDNAs and a human glyceraldehyde 3-phosphate dehydrogenase (GAPDH) probe. For low stringency screening, the blot was washed twice in 2×SSC at room temperature, followed by two washes at 42°C in this same solution. After exposure to film, high stringency washes were carried out in 0.1×SSC at 60°C. The GAPDH probe was obtained by RT-PCR from HeLa cell poly(A)$^+$ RNA and two oligonucleotides corresponding to positions 29–53 and 715–724 of human GAPDH (35).

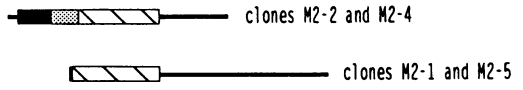
RESULTS

Various USF2 messages result from differential splicing and alternative poly(A) site usage

Isolation of clones to different USF family members revealed that several domains in the C-terminal region of these proteins were highly conserved (29). This observation was used to design degenerate PCR primers in two of the most conserved regions (the USF-specific domain and helix 1 of the HLH motif) that would allow amplification of a small region of the cDNA for all known USF genes. Using these primers and a cDNA made from poly(A)$^+$ RNA isolated from the S194 mouse cell line, we obtained small cDNA probes for both the murine USF1 and USF2 genes (see Materials and Methods). The USF2 probe was used in turn for screening an NIH3T3 cDNA library. This screen resulted in identification of four positive plaques (M2-1, M2-2, M2-4 and M2-5), which corresponded to two original clones, each represented twice.

The two types of clones represented two different forms of USF2 messages, as illustrated schematically in Fig. 1. Although clone M2-2 did not contain any poly(A) sequences, it terminated immediately downstream of a consensus polyadenylation signal located precisely at the same position as in several (poly(A)-containing) human USF2 cDNAs clones previously isolated (29). By contrast, the 3′ untranslated region of clone M2-5 was extended by another 743 nucleotides, terminating at a polyadenylate stretch just downstream of a second polyadenylation signal. This finding indicated a possible regulation of USF2 messages by poly(A) site selection. More significantly, clone

A



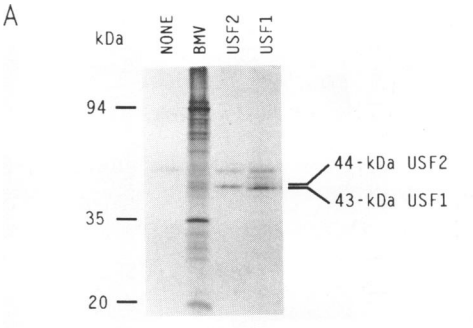
B

```
CACTTTCCGT TCAAGGTCAA GACGTGCCAG AGAAGCATTG TCCCCTCCCC CCCGCCCCCC   60
CCATGGACAT GCTGGACCCG GGTCTGGATC CCGCTTCCTC GGCCACCGCT GCTGCCGCCG  120
CCAGCCACGA CAAGGGACCC GAGGCCGAGG AGGGCGTTGA GCTGCAGGAA GGCGGGGACG  180
GCCCGGGGGC TGAGGAGCAG ACGGCGGTGG CCATCGCCAG TGTCCAGCAG GCGGCGTTCG  240
GCGACCACAA TATCCAGTAC CAGTTCCGTA CAGAGAGTAA TGGAGGCCAG GTGACATACC  300
                                                                    28
GCGTAGTCCA GGTGACTGAT GGTCAGCTGG ACGGCCAGGG CGACGCGGCC GGCGCCGTCA  360
GCGTCGTGTC CACCGCTGCC TTCGCGGGGG GGCAGCAGGC TGTGACCCAG GTGGGTGTGG  420
ACGGGGCAGC CCAGCGCCCC GGCCCTGCCG CTGCCTCTGT ACCCACAGGT CCCGCAGCAC  480
CCTTCCCGCT GGCTGTAATT CAAAATCCAT TCAGCAATGG TGGCAGCCCG GCGGCCGAAG  540
                                                                    77
CTGTCAGCGG GGAGGCACGA TTTGCTTATT TTCCAGCATC CAGTGTGGGC GATACCACAG  600
                                                                   137
CTGTGTCTGT TCAAACTACA GACCAGAGCC TACAGGCCGG AGGCCAGTTC TATGTCATGA  660
                                                                   197
TGACACCCCA GGATGTGCTT CAGACAGGCA CACAGAGGAC AATTGCACCC CGGACACACC  720
                    -A                                             257
CCTATTCTCC GAAAATTGAT GGAACCAGAA CTCCTCGAGA TGAGAGAAGA AGAGCTCAGC  780
                                                                   317
ACAATGAAGT GGAACGGAGA AGGAGGGACA AGATCAACAA CTGGATCGTC CAGCTTTCCA  840
                                                                   377
AAATCATTCC AGATTGTCAT GCAGACAACA GCAAGACAGG AGCAAGTAAA GGGGGCATCC  900
                                                                   437
TGTCCAAGGC TTGCGATTAC ATCCGGGAGC TACGCCAGAC CAACCAGCGC ATGCAGGAGA  960
                                                                   497
CCTTCAAGGA GGCAGAGCGG CTGCAGATGG ACAATGAGCT CCTCAGGCAG CAGATCGAGG 1020
                                                                   557
AGCTGAAGAA TGAGAATGCC CTGCTTCGAG CCCAGCTACA GCAGCACAAC CTGGAAATGG 1080
                                                                   617
TGGGCGAGAG TACCCGGCAG TGATGCCCGC CGGCTCACCC TTCAGCTGCT GTCGGCCTCT 1140
                                                                   677
GCTGCCCCTC CCCCAGCCCT TAGCACAGAG AGGGAGATGC CCCTCCCCAA GCTGCGTTTT 1200
                                                                   737
TTTATAGTAG ATTTTTAACA AAAACAAGGA GACATAATGC ATTTCTGTTG ATACGTTGCC 1260
                                                                   797
CACTACCCTC CTCCATGCGA AAACGATGCC CACCTGCCCA TCTGTCCATC AGCCTCCCTT 1320
                                                                   857
CTCGGTCCTC ACCAGCCCAT CACTTATAGT GGGCTCATCA GGAGGAGCAG GGGAGGAGGA 1380
                                                                   917
CAGAGGCCCC GCTGCCCGCT GCCTCCTTGG TCTCTAGAGG TATTGAGACA GGGTGCTCAT 1440
                                                                   977
GGGAAGGAGG GAGCTTTGGG GATGCGGGCC TTCCCTGGGG CTTGGACCTT TGCAGGGAGG 1500
                                                                  1037
CCATGCCCGC CACCCCTCTT GTTCTGGATC CCTGCTCCCC TCTGGGTGTG TGGGTGTGTG 1560
                                                                  1097
TGTACGTTTG GGTGTGTGTG TATGCGCGTG TGTGCGTGTG TGTGTTTTAA TTTTCTTTAT 1620
                                                                  1157
GGAAAATGGA CAAAAAATAG AGAGAGAGGT ATTTAACTGC AATAAACTGG CCAGTGTGGC 1680
                                                                  1217
CCCGCCT                                                           1687
-------TGT CTGTCTGTGT GTCCATGTTG GGCATGGGGT TGGGTAGGGT TCCTAGGGTC 1277
TCCCTGATAT CTAAGCTACA GTTATACTGT AGCTGTGTGA CAATGGCTGG GGGCATTAAC 1337
TGAGGCCCTG GGGTGCTGAT TAGGGGTGTG GGAGCTGAGG TCAAGGACAT ACCCCTGCTG 1397
TTAGAGGGGA GGCTGGAGTA TCCCTGGGCT CTGGGGCTGA AGCCCTCCCT CTGTACCAGC 1457
CAATGATGAG GGCTGGCCTT GAGCCCCCCA CCCCCAAGAG AGGGCAGATG GCCAGAAAGC 1517
CAGGCTTGGC CCGGCAGTCT GTCGCCTCAA ATTGTAGCTC TGGCGCCTGT GCTGTGACCC 1577
TTGTGGCCCT TGGCTGATGA TGAAACCTGT CCTGAGAAGC CACAGTGATG GTTGGCGGGG 1637
TCGGGGAATA CCATCATGTC CCCAAGAGAT TGGCTCACCA TTCCCAGGAT GGAGGCCTGC 1697
GCGTCAGGTT CCCTTTGCTA CACTTTGTAC CTTGCCTCAA TCTAAGTACT CCTGTTGGCC 1757
CCTGGTGTCT TGAGAACCCT GTCTCGAGCT GGAGAGATGG CTCATCAGTT AAGAGCACCT 1817
GCTGCTCTTC CATAGGTCCT GAGTTCAATT CCCAGCAACC ACATGATGGC TCACAAACCT 1877
CTGTAATGGG ATCCGATGCC CTCTTTAGTG TGTCTGAAGA CATCGACCGT GTACTCATAT 1937
AAATTAAGTA AATAAAACTT AAAAAAAAAA A                                 1968
```

Figure 1. Isolation of different cDNA clones to murine USF2. (**A**) Schematic representation of the two types of USF2 cDNA clones isolated from an NIH3T3 cell library using as a probe a small PCR fragment amplified from the USF-specific and helix 1 regions of this protein (see Materials and Methods). The 5′ and 3′ untranslated regions are represented as lines, while the coding regions of the cDNAs are depicted as rectangular boxes filled in various patterns to illustrate the regions that are present in both clones. (**B**) Nucleotide sequence of clones M2-2 and M2-5. The sequences of clones M2-2 (top) and M2-5 (bottom), as determined by the dideoxy method, are shown with the identical regions aligned. Dashes represent nucleotides in clone M2-5 that are identical to those of clone M2-2. Underlined are the putative translation initiation sites for the full-length and mini-USF2 proteins, their translation termination codon and the two polyadenylation signals.

M2-2 contained an intervening sequence of 201 bp (residues 291−491) that was absent in clone M2-5. Characterization of genomic clones to the murine USF2 gene confirmed that this intervening sequence corresponded to a discrete exon (Q.Lin and M.Sawadogo, unpublished result). From these observations, we concluded that the production of USF2 messages was controlled by alternative splicing mechanisms. It is clear that these same two alternatively spliced forms of USF2 mRNAs are also present in human cells. Indeed, several clones similar to M2-5 were previously isolated from two different human cDNA libraries (29). By contrast, the human cDNA clone designated FIP, although incomplete, contained at its 5′ end sequences encoding amino acids present in the intervening sequence of clone M2-2 (30). Note that the number of base pairs present in the alternative exon is a multiple of 3. Consequently, the two different splicing patterns maintain the same open reading frame.

A



B

Figure 2. Comparison between the 43- and 44-kDa USF proteins. (**A**) SDS-PAGE analysis of the in vitro translation products made using T7 RNA polymerase transcripts of the mouse USF2 (clone M2−2, lane 3) and human USF1 (lane 4) cDNAs. Apparent molecular weights of 43,000 and 44,000 for the two USF species could be determined using molecular weight markers prepared by in vitro translation of BMV RNA (lane 2). (**B**) The deduced amino acid sequence of the mouse 44-kDa USF (top) was aligned with that of the human 43-kDa protein (bottom). Blocks of identical amino acids are boxed.

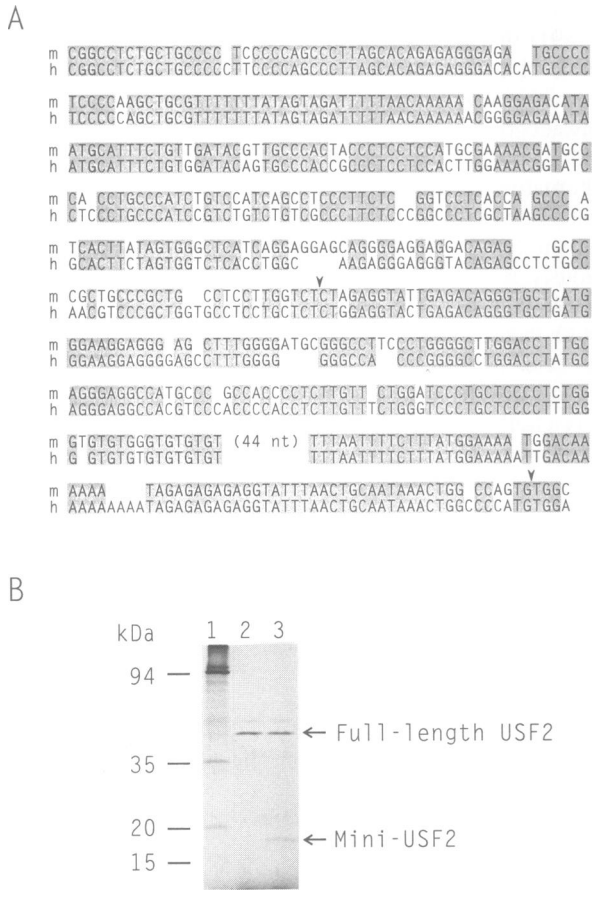


Figure 3. Sequence conservation in the 3′ untranslated region of the USF2 messages. (**A**) The 3′ untranslated region of the mouse USF2 cDNA (m, residues 1133−1675) was compared with that of human USF2 (h, residues 742−1260 of clone hUSF2-C, ref. 29). Gaps have been introduced in the sequences to maximize homologies. Identical bases are outlined in gray. The 44 nucleotides in the murine cDNA that are absent in the human clone represent an expansion of the preceding GT repeat region. The two arrows indicate cleavage sites for restriction endonucleases *Xba*I and *Bgl*I in the mouse sequence. (**B**) SDS-PAGE analysis of in vitro translation products obtained using mouse USF2 RNA containing only the 3′ untranslated sequences located upstream of the *Xba*I site at position 1413 (lane 2) or the entire 3′ region down to the *Bgl*I site at position 1676 (lane 3). Migration of the full-length 44-kDa USF protein, as well as that of the 17.5-kDa mini-USF2, is indicated at right.

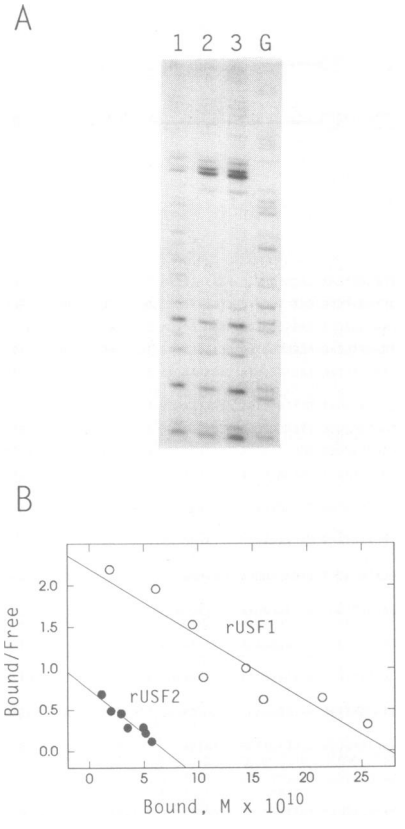


Figure 4. Interaction of the 43- and 44-kDa USF homodimers with the major late promoter binding site. (**A**) Interaction of the human 43-kDa USF (lane 2) and the mouse 44-kDa USF (lane 3) with their specific binding site in the adenovirus 2 major late promoter was analyzed by DNase I footprinting using recombinant proteins produced in *E.coli* (see Materials and Methods). Lane 1 is a control DNase I cleavage reaction performed in the absence of USF, while lane G shows a G-specific chemical sequencing reaction of the same DNA fragment. (**B**) The affinity of the recombinant 44-kDa USF (rUSF2, ○−○) and 43-kDa USF (rUSF1, ○−○) for the specific major late promoter binding site was determined using a filter binding assay (26). The results are shown as Scatchard plots with lines best-fitted by a linear regression method.

In all likelihood, these various forms of USF2 messages account for the multiplicity of RNAs observed for this gene in Northern blot analysis (29, see also Fig. 5). Messenger RNAs of the M2-2 type, although apparently less abundant, are clearly those encoding the full-length 44-kDa form of USF. Indeed, translation of RNA made by in vitro transcription of clone M2-2 yielded a major protein product migrating at precisely 44-kDa, just slightly slower than the 43-kDa translation product of the USF1 cDNA (Fig. 2A). Since a complete cDNA of the M2-5 type has not yet been isolated, it remains to be determined whether these messages actually encode a functional product or whether their function is simply to regulate the synthesis of the 44-kDa USF protein.

Sequence conservation in USF

Comparison of the deduced amino acid sequences of USF2 from human (29) and mouse revealed almost complete identity. Only two differences were present, one located in the loop region of the HLH motif and one very close to the C-terminal end (namely, the histidine 66 and serine 343 in murine USF2 were respectively replaced by an asparagine and a glycine in the human protein). Alignment of the primary sequence of the 44-kDa USF with that of the 43-kDa protein (Fig. 2B) revealed that domain conservation between various USF family members was limited to the C-terminal region encompassing the dimerization and DNA-binding domains. In the N-terminal region, only very small blocks of identical amino acids could be found.

Interestingly, a strong conservation of the entire 3′ untranslated region could be demonstrated by comparing the human and mouse USF2 genes. As shown in Fig. 3, more than 78% of the nucleotides were identical in the two cDNAs over a region larger than 540 bp. Such a degree of conservation between evolutionarily distant species of vertebrates has been previously observed in several instances, often in genes that are widely expressed and that encode DNA-binding proteins or cytoskeletal components (36). These untranslated sequences are thought to play an important role in either transcriptional or post-transcriptional control mechanisms. In the case of the USF2 gene, a possible role in translational regulation was suggested by the

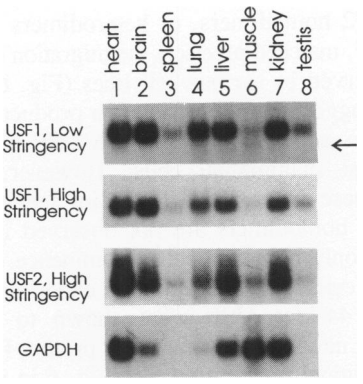


Figure 5. The USF1 and USF2 genes are both ubiquitously expressed. The distribution of USF messages in various mouse tissues was determined by probing a Northern blot of poly(A)$^+$ RNA with labelled DNA fragments isolated from the murine USF1 or USF2 cDNAs. In the USF1 case, the blot was initially washed under low stringency conditions, exposed, and washed again under high stringency conditions (see Materials and Methods). The arrow indicates the migration of a muscle-specific mRNA that cross-hybridizes with USF1 at low stringency. As a control for the amount of RNA present in each lane, the same blot was finally hybridized with a glyceraldehyde 3-phosphate dehydrogenase (GAPDH) probe.

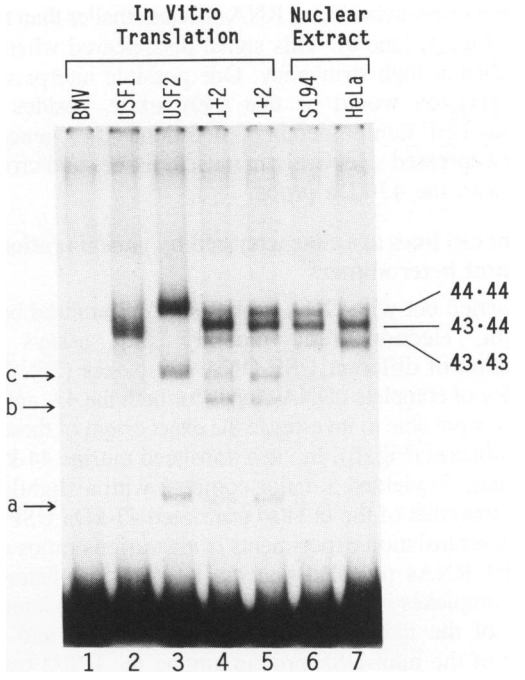


Figure 6. Various complexes of USF homo- and heterodimers are present in different cell types. An electrophoretic mobility shift assay was used to determine the origin of the various USF complexes present in nuclear extracts prepared from either mouse B cells (S194 cell line) or HeLa cells. Controls were provided by in vitro transcription/translation products of the mouse 44-kDa USF (USF2) and human 43-kDa USF (USF1) translated separately (lanes 2 and 3, respectively) or cotranslated at different ratios of the two RNAs (lanes 4 and 5). The migration of protein-DNA complexes containing the two different full-length USF homodimers, as well as those containing the 43−44 heterodimers, is indicated at right. The minor complexes indicated at left can be attributed to homodimers of the mini-USF (complex a) and heterodimers of the mini-USF with the 43-kDa (complex b) or 44-kDa full-length USF (complex c).

results shown in Fig. 3B. As was previously noticed for the human USF2 gene (29), in vitro translation of murine USF2 RNA often yielded not only the expected full-length 44-kDa translation product, but also a smaller protein with a molecular weight of 17,500 (Fig. 3B, lane 3). The size of this 'mini-USF' protein corresponded precisely to that expected for a translation start site located at one of the two consecutive methionines (positions 199 and 200) in the USF-specific domain, immediately upstream of the basic region. Consequently, this mini-USF contained the intact dimerization and DNA-binding domains of USF2 and was observed to bind DNA in gel retardation assays as both homodimers and heterodimers with the full-length USF2 protein (see example in Fig. 6). Interestingly, the ratio of in vitro translated mini- to full-length USF2 decreased considerably when the reticulocyte lysate was programmed with a truncated USF2 RNA lacking half of the 3′ untranslated region (Fig. 3B, lane 2). This result indicates a critical role for the USF2 3′ untranslated region in controlling, at least in vitro, the selection of translation start sites. Whether this control involves particular nucleotide sequences in this region, or particular RNA structures that are stabilized by these sequences, remains to be investigated.

The 43- and 44-kDa USFs have very similar DNA-binding properties

To compare the DNA-binding properties of the 43- and 44-kDa forms of USF, we expressed both proteins in bacteria as polyhistidine fusions (34). As shown in Fig. 4A, DNase I footprinting of the two recombinant proteins on their specific binding site in the adenovirus 2 major late promoter failed to reveal any difference in their specific contacts with the DNA. The observed DNase I cleavage interference pattern was furthermore identical to that previously obtained with USF purified from HeLa cells (1). To determine whether there could be a difference in the affinity of the two proteins for the DNA, quantitative analysis was performed using the standard filter binding assay originally developed for human USF (26). Under these particular conditions, an apparent dissociation constant of 1.24×10^{-9} M was determined for the recombinant 43-kDa protein, while a value of 9.3×10^{-10} M was obtained for the 44-kDa recombinant USF (Fig. 4B). These figures were very close to the value reported for purified human USF (7.5×10^{-10} M, 26). From these experiments, we concluded that the different USF family members had very similar DNA-binding properties. In addition, these DNA-binding properties were probably not modified by post-translational modifications, since these would not be present in the bacterially-produced proteins.

USF1 and USF2 messages are both ubiquitously expressed

USF expression has been observed in a variety of tissues and cell lines using transcription assays and gel retardation analyses. However, the exact distribution of USF messages had not been investigated previously. For this analysis, we used a blot containing poly(A)$^+$ RNA from various mouse tissues (Fig. 5). A control for the amount of RNA loaded in each lane was provided by hybridization with a cDNA probe to GAPDH. The patterns of USF1 and USF2 transcripts, as revealed by high stringency hybridization with the corresponding mouse cDNA probes, were almost identical to those of GAPDH, indicating that these two USF genes were expressed at a similar level in most tissues. The only exception was skeletal muscles, where both the USF1 and USF2 messages seemed somewhat less abundant. Interestingly, hybridization of the USF1 cDNA probe at low stringency revealed the existence, exclusively in skeletal muscle

tissue, of a cross-hybridizing RNA species smaller than the USF1 mRNA (Fig. 5, lane 6). This signal disappeared when the blot was washed at high stringency. One possible interpretation for this observation would be that there exist, besides the two ubiquitous USF family members, additional USF genes, one of which is expressed selectively in muscle tissues and cross-reacts weakly with the 43-kDa probe.

Different cell lines are characterized by various ratios of USF homo- and heterodimers

When carried out with DNA probes several hundred base pairs in length, electrophoretic mobility shift assays allowed visualization of different USF-DNA complexes (26). With the availability of complete cDNA clones for both the 43- and 44-kDa USFs, we were able to investigate the exact origin of these various USF complexes (Fig. 6). In vitro translated murine 44-kDa USF (USF2, lane 3) yielded a major complex with a slightly slower mobility than that of the in vitro translated 43-kDa USF (USF1, lane 2). Cotranslation experiments using various ratios of USF2 and USF1 RNAs revealed that the USF1-USF2 heterodimers formed complexes with an electrophoretic mobility intermediate to those of the individual homodimers (lanes 4 and 5). The presence of the mini-USF protein among the USF2 translation products was also visualized in these experiments, resulting in additional complexes attributable to mini-USF homodimers (complex a) and heterodimers formed between this protein and either the 43-kDa (complex b) or 44-kDa (complex c) full-length USF species (lanes 3–5).

Complexes with electrophoretic mobilities identical to those of the USF2-USF2 and USF1-USF1 homodimers, as well as complexes corresponding to the USF1-USF2 heterodimers, were also observed with nuclear extracts from tissue culture cells. In extracts from the mouse B cell line S194, the relative abundance of these three complexes indicated a nearly equimolar abundance of the two USF proteins (lane 6). An identical pattern was observed using extracts from Namalwa cells, which is also a B cell line but of human origin (data not shown). By contrast, the complex formed by the 44–44 homodimers was essentially absent in extracts from HeLa cells (lane 7), indicating that this protein was less abundant in these cells than was the 43-kDa species. Longer exposure of the gel revealed that faster migrating complexes attributable to the presence of mini-USF species could also be observed in the two types of nuclear extract, although their abundance was less than in the in vitro translation reactions. Taken together, these results indicated that the relative expression of the various forms of USF varied in different cell types. Whether this variation results from transcriptional or posttranscriptional control mechanisms remains to be determined.

DISCUSSION

We had previously reported immunological evidence that the human USF2 gene encodes the 44-kDa USF protein originally identified in HeLa cells (26, 29). With the isolation of a complete cDNA clone for the corresponding gene in mouse, we were able to obtain further proof of the identity between USF2 and the 44-kDa form of USF. First, in vitro translation reactions confirmed that the full-length USF protein encoded by clone M2-2 migrated in SDS gels with an apparent molecular weight of 44,000 (Fig. 2A), even though the calculated molecular weight of this 346-amino-acid protein was only 36,952. Second, the electrophoretic migration of the protein-DNA complexes containing USF2 homodimers, or heterodimers of USF2 with the 43-kDa USF, matched precisely the migration of natural USF complexes observed in various cell lines (Fig. 6).

Our results suggest that the translation products of the USF1 and USF2 genes account for all of the USF binding activity present in most mammalian cells. However, the relative abundance of these different USF proteins seems to vary. For instance, USF2 homodimers are not observed in HeLa cells. Clearly, this is only because of the prominence of the 43-kDa species in these cells, since renaturation experiments using gel-purified HeLa 44-kDa USF were shown to result in the appearance of a new, slower migrating protein-DNA complex absent in the original preparation (see Fig. 6 in ref. 26). With our present knowledge, it is obvious that this new species was due to reconstitution of a HeLa 44–44 USF complex. However, the observation could not be understood at the time because it was then thought that USF interacted with the DNA as a monomer (37).

Noticeably, the fact that translation products from the USF1 and USF2 genes are the predominant USF species in many cells does not preclude the existence of other USF family members, especially if these are expressed in a tissue-specific manner. In other families of USF-related HLH proteins such as Myc, some family members are ubiquitously expressed (e.g., c-Myc), while others display more restricted patterns of expression (e.g., N-Myc) (reviewed in 38). The observed cross-hybridization of the murine 43-kDa USF cDNA with a muscle-specific RNA may well reflect the existence of a muscle-specific USF family member (Fig. 5).

Strong conservation of non-coding sequences during vertebrate evolution has been previously observed for a number of genes. Since these conserved regions are much longer than would be necessary for specifying the binding of a regulatory protein, speculations as to their function have included a possible role in mRNA export, localization, translation or degradation (reviewed in 36). Here, we report evidence that the conserved 3′ untranslated RNA sequences in the USF2 gene may be involved in translational control. Indeed, the frequency at which translation of USF2 messages initiated at the first methionine or at an internal site was dependent upon the presence of these 3′ untranslated sequences. Production from a single RNA of both a full-length transcription factor and an N-terminally truncated species starting at the second in-phase ATG was previously observed in the case of the liver-enriched transcription factor LAP (39). In this case, the smaller protein, designated LIP, was also fully capable of specific DNA-binding, but lacked the transcription activation domain and therefore functioned as a dominant negative mutant (39). Since the transcription activation domain of USF is similarly located in the N-terminal portion of the protein (40), heterodimer formation between the mini-USF and full-length USF species could well be a mechanism by which the activity of this family of transcription factors is regulated.

Expression of the USF1 and USF2 genes is ubiquitous. However, various patterns of messages exist in both human and mouse, resulting from differential splicing as well as alternative poly(A) site usage. It is possible that some of these messages do not encode functional protein products but instead serve to regulate the abundance of USF under various physiological states of the cells. Also, it is intriguing that the relative abundance of USF1 and USF2 proteins varies among different cell types. Our studies so far have not revealed a difference in the DNA-binding properties of the two proteins. Furthermore, their transcription

activation potential, as determined by transient transfection assays, appears also quite similar (X.Luo and M.Sawadogo, unpublished observation). However, these studies represent artificial and greatly simplified situations. In the normal context, the two forms of USF may well be involved in regulating different sets of genes. Indeed, their extreme divergence in N-terminal amino acid sequences could specify differential interactions with other families of transcription factors. Such selective contacts between transcription factors are clearly important for stabilizing the different multiprotein complexes present at the promoters of the various genes.

ACKNOWLEDGEMENTS

We wish to thank D.K.Hawley for her gift of Namalwa nuclear extract, S.Walker for technical assistance, D.R.Pratt and M.A.Tainsky for the GAPDH oligonucleotides and M.Szentirmay and M.W.Van Dyke for critical reading of the manuscript. This work was supported by Grants GM38212 from the National Institutes of Health and G1195 from the Robert A.Welch Foundation.

REFERENCES

1. Sawadogo,M. and Roeder,R.G. (1985) Cell, **43**, 165–175.
2. Carthew,R.W., Chodosh,L.A. and Sharp,P.A. (1985) Cell, **43**, 439–448.
3. Miyamoto,N.G., Moncollin,V., Egly,J.M. and Chambon,P. (1985) EMBO J., **4**, 3563–3570.
4. Murre,C., Schonleber-McCaw,P. and Baltimore,D. (1989) Cell, **56**, 777–783.
5. Blackwell,T.K., Kretzner,L., Blackwood,E.M., Eisenman,R.N. and Weintraub,H. (1990) Science, **250**, 1149–1151.
6. Prendergast,G.C. and Ziff,E.B. (1991) Science, 251, 186–189.
7. Halazonetis,T.D. and Kandil,A.N. (1991) Proc. Natl. Acad. Sci. USA, **88**, 6162–6166.
8. Blackwood,E.M. and Eisenman,R.N. (1991) Science, **251**, 1211–1217.
9. Prendergast,G.C., Lawe,D. and Ziff,E.B. (1991) Cell, **65**, 395–407.
10. Ayer,D.E., Kretzner,L. and Eisenman,R.N. (1993) Cell, **72**, 211–222.
11. Zervos,A.S., Gyuris,J. and Brent,R. (1993) Cell, **72**, 223–232.
12. Beckmann,H., Su,L.K. and Kadesch,T. (1990) Genes Dev., **4**, 167–179.
13. Carr,S.C. and Sharp,P.A. (1990) Mol. Cell. Biol., **10**, 4384–4388.
14. Carthew,R.W., Chodosh,L.A. and Sharp,P.A. (1987) Genes Dev., **1**, 973–980.
15. Chodosh,L.A., Carthew,R.W., Morgan,J.G., Crabtree,G.R. and Sharp,P.A. (1987) Science, **238**, 684–688.
16. Lemaigre,F.P., Courtois,S.J., LaFontaine,D.A. and Rousseau,G.G. (1989) Eur. J. Biochem., **181**, 555–561.
17. Chang,L.S., Shi,Y. and Shenk,T. (1989) J. Virol., **63**, 3479–3488.
18. Scotto,K.W., Kaulen,H. and Roeder,R.G. (1989) Genes Dev., **3**, 651–662.
19. Sato,M., Ishizawa,S., Yoshida,T. and Shibahara,S. (1990) Eur. J. Biochem., **188**, 231–237.
20. Greaves,R.F. and OHare,P. (1991) J. Virol., **65**, 6705–6713.
21. Potter,J.J., Cheneval,D., Dang,C.V., Resar,L.M., Mezey,E. and Yang,V.W. (1991) J. Biol. Chem., **266**, 15457–15463.
22. Zwartkruis,F., Hoeijmakers,T., Deschamps,J. and Meijlink,F. (1991) Mech. Dev., **33**, 179–190.
23. Bungert,J., Kober,J., Düring,F. and Seifart,K.H. (1992) J. Mol. Biol., **223**, 885–898.
24. Riccio,A., Pedone,P.V., Lund,L.R., Olesen,T., Olsen,H.S. and Andreasen,P.A. (1992) Mol. Cell. Biol., **12**, 1846–1855.
25. Reisman,D. and Rotter,V. (1993) Nucleic Acids Res., **21**, 345–350.
26. Sawadogo,M., Van Dyke,M.W., Gregor,P.D. and Roeder,R.G. (1988) J. Biol. Chem., **263**, 11985–11993.
27. Sawadogo,M. (1988) J. Biol. Chem., **263**, 11994–12001.
28. Gregor,P.D., Sawadogo,M. and Roeder,R.G. (1990) Genes Dev., **4**, 1730–1740.
29. Sirito,M., Walker,S., Lin,Q., Kozlowski,M.T., Klein,W.H. and Sawadogo,M. (1992) Gene Expression, **2**, 231–240.
30. Blanar,M.A. and Rutter,W.J. (1992) Science, 256, 1014–1018.
31. Kaulen,H., Pognonec,P., Gregor,P.D. and Roeder,R.G. (1991) Mol. Cell. Biol., **11**, 412–424.
32. Kozlowski,M.T., Gan,L., Venuti,J.M., Sawadogo,M. and Klein,W.H. (1991) Dev. Biol., **148**, 625–630.
33. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA, **74**, 5463–5467.
34. Van Dyke,M.W., Sirito,M. and Sawadogo,M. (1992) Gene, **111**, 99–104.
35. Tokunaga,K., Nakamura,Y., Sakata,K., Fujimori,K., Ohkubo,M., Sawada,K. and Sakiyama,S. (1987) Cancer Res., **47**, 5616–5619.
36. Duret,L., Dorkeld,F. and Gautier,C. (1993) Nucleic Acids Res., **21**, 2315–2322.
37. Lennard,A.C. and Egly,J.M. (1987) EMBO J., **6**, 3027–3034.
38. Zimmerman,K. and Alt,F.W. (1990) Crit. Rev. Oncog., **2**, 75–95.
39. Descombes,P. and Schibler,U. (1991) Cell, **67**, 569–779.
40. Kirschbaum,B.J., Pognonec,P. and Roeder,R.G. (1992) Mol. Cell. Biol., **12**, 5094–6101.