Biochemical and Biophysical Research Communications **283,** 209–218 (2001)
doi:10.1006/bbrc.2001.4753, available online at http://www.idealibrary.com on IDEAL®

Conserved Structure and Promoter Sequence Similarity in the Mouse and Human Genes Encoding the Zinc Finger Factor BERF-1/BFCOL1/ZBP-89

Salvatore Feo,*,†,[1] Vincenzo Antona,* Giuseppe Cammarata,‡ Fatima Cavaleri,* Rosa Passantino,*,‡ Patrizia Rubino,‡ and Agata Giallongo‡

**Dipartimento di Biologia Cellulare e dello Sviluppo, Università di Palermo and †Centro di Oncobiologia Sperimentale, Viale delle Scienze, 90128 Palermo, Italy; and ‡Istituto di Biologia dello Sviluppo del Consiglio Nazionale delle Ricerche, Via U. La Malfa 113, 90146 Palermo, Italy*

Received March 20, 2001

We have characterized the genomic structure of the mouse Zfp148 gene encoding Beta-Enolase Repressor Factor-1 (BERF-1), a Kruppel-like zinc finger protein involved in the transcriptional regulation of several genes, which is also termed ZBP-89, BFCOL1. The cloned Zfp148 gene spans 110 kb of genomic DNA encompassing the 5′-end region, 9 exons, 8 introns, and the 3′-untranslated region. The promoter region displays the typical features of a housekeeping gene: a high G+C content and the absence of canonical TATA and CAAT boxes consistent with the multiple transcription initiation sites determined by primary extension analysis. Computer-assisted search in the human genome database allowed us to determine that the same genomic structure with identical intron–exon organization is conserved in the human homologue ZNF 148. Functional analysis of the 5′-flanking sequence of the mouse gene indicated that the region from nucleotide −205 to +144, relative to the major transcription start site, contains *cis*-regulatory elements that promote basal expression. Such sequences and the overall promoter architecture are highly conserved in the human gene. Furthermore, we show that the complex transcription pattern of the Zfp148 gene might be due to a combination of alternative splicing and differential polyadenylation sites utilization.

Key Words: **BERF-1; ZBP-89; gene structure; alternative splicing; promoter analysis; *cis*-elements; zinc-finger 148.**

Eukaryotic transcription factors containing Cys_2-His_2-type (Kruppel-like) zinc fingers are involved in various processes of cell growth and differentiation and many of these factors exhibit highly restricted patterns of expression in adult tissues and during development (1, 2). These proteins share a common structural organization with conserved zinc finger domains associated to activation and/or repression domains which may significantly differ in each factor. We have characterized a Kruppel-like zinc finger protein that binds in a sequence-specific manner to the G-rich box within the muscle-specific enhancer of the ß enolase gene (ENO3) (3). This factor, termed Beta Enolase Repressor Factor-1 (BERF-1), functions as a repressor of the ß enolase gene transcription (4). Independently other investigators reported the isolation of a mouse cDNA encoding an N-terminal shorter form of BERF-1, termed BFCOL1, which participates in the transcriptional control of the two mouse type I collagen genes (5). BFCOL1 has been found to interact with PTRF (Polymerase I and transcript-release factor) through its zinc-fingers and this interaction results in a more efficient binding to the GC-rich box within the mouse type-I collagen proximal promoter (6). Several other genes have been found to be regulated by the rat and the human homologues of BERF-1/BFCOL1 protein. The rat homologue, ZBP-89, has been identified as a repressor of the basal and EGF-induced transcription of the gastrin gene (7), and it also represses the bovine adrenodoxin gene (Adx) (8). The human homologue of BERF-1/BFCOL1/ZBP-89 has been isolated and found to act as a repressor of both the human ornithine decarboxylase (ODC) (9) and the vimentin gene expression (10) and as an activator of the stromelysin (matrix metalloproteinase-3) gene (11). A human isoform, termed hTβ, has been shown to activate transcription of the human T-cell receptor (TCR) Vβ8.1 gene counteracting the silencing effect of the mouse TCR α gene silencer (12) and it has been suggested that this factor may play an important role in T cell differentiation (13). More recently ZBP-89 has also been found to play

[1] To whom correspondence should be addressed. Fax: (+39-091) 6577-347. E-mail: feo@unipa.it.

0006-291X/01 $35.00

a role in the transcriptional regulation of the $p21^{waf1}$ gene, binding to a proximal SP1 element of the mouse promoter (14) and participating in the p300/Sp1 activation complex (15). Finally, ZBP-89 has been found to be involved in the replication of Epsten-Barr virus (EBV) genome from the lytic origin (16).

Despite the fact that many target genes have been identified for BERF-1/BFCOL1 activity, as well as for the human and rat homologous proteins, the genomic structures of the mouse gene, named Zfp148 (17), the human ZNF148 gene and the rat gene for ZBP-89 remain still unknown. We have previously mapped the Zfp148 gene to mouse chromosome 16 and identified in the mouse genome a pseudogene localized to chromosome 8 (17). The human ZNF148 has been mapped to human chromosome 3q21 (17, 18), more precisely between STS markers D3S1551 and D3S1765 (19).

To have a better understanding of the function and the expression of the gene encoding this transcription factor, we have isolated genomic clones for the Zfp148 genes. The genomic structure of both the mouse functional gene and the processed pseudogene was characterized and the promoter region controlling Zfp148 transcription was identified by functional analysis. Furthermore, we compared the structure of the mouse gene and its promoter sequence to those of the human ZNF148 gene we obtained by the alignment of several unordered contigs available in the HTGS section of the GenBank, showing that both sequence and structural features have been conserved through evolution. By sequence comparison of several cDNAs with the isolated genomic clones, we propose that the different transcripts generated by the mouse Zfp148 gene result by a combination of alternative splicing and utilization of different polyadenylation signals. This mRNA diversity is conserved in the human ZNF148 gene.

MATERIALS AND METHODS

Cloning and sequencing of cDNAs. A mouse 12 day post coitum (dpc) limb cDNA expression library in Lambda ZapII vector (Stratagene, La Jolla, CA) was screened with a ^{32}P-labeled BERF-1 cDNA (A22 in Fig. 1, Accession No. X98096), as described previously (20). Specific clones were identified by screening 1.5×10^6 plaques and three clones were completely sequenced on both strands using the Sequenase Version 2.0 or the Thermo Sequenase kit (US Biochemical, Cleveland, OH) according to the instructions of the manufacturer. The nucleotide sequences of the identified cDNAs have been submitted to GenBank with the following Accession Nos. AF316548–AF316550.

Isolation of genomic clones and Southern blot analysis. Genomic clones containing the Zfp148 and the Zfp148-ps1 genes were isolated from a mouse strain 129SvJ spleen cells genomic DNA library in Lambda Fix II vector (Stratagene, La Jolla, CA). Fourteen positive clones were identified by screening 2×10^6 phages using the labeled A22 BERF-1 cDNA as a probe. All the clones were probed with different regions of the cDNA and on the basis of the hybridization results seven clones were selected to be analyzed further. Pools of the RPCI-21 mouse female (strain 129/SvEvTac) spleen genomic DNA library in pPAC4 vector (21) were obtained from the Yac Screening Centre (DIBIT San Raffaele, Milan, Italy), and analyzed by PCR with exon-specific primers, as described previously (17). Additional screening of the PAC library was performed by hybridization with a 300-bp *Eco*RI-*Bam*HI fragment derived from A23 cDNA and containing the most 5′-UTR region of the mRNA. DNA fragments containing the exons were subcloned into the pBluescript vector (Stratagene, La Jolla, CA) and the intron/exon junction sequences were determined as described above. Intron lengths were determined by restriction mapping and Long-PCR analysis of the genomic lambda and PAC clones and verified using mouse genomic DNA. The 5′-flanking sequence of the Zfp148 gene was analyzed from a plasmid (p11-B1) containing a 6-kb *Spe*I–*Spe*I fragment isolated from the PAC clone. Southern blot analyses were carried out by restriction enzyme digestions of mouse genomic DNA followed by agarose gel separation and hybridization as described previously (22). The Zfp148 genomic sequences have been submitted to GenBank with Accession Nos. AF328790–AF328796.

Primer extention analysis. An oligonucleotide, 5′-TAGCCAGTGATCTGGCGGGATGGGCTTAT-3′, complementary to the 5′-untranslated sequence (bases 183 to 212) of the A22 cDNA was end-labeled with T4 polynucleotide kinase and used as primer with Poly(A)+ RNA from limbs of 13 dpc mouse embryos. RNA isolation and primer extention analysis were performed as described previously (23). The extended products were resolved by electrophoresis on a 6% polyacrilamide/7 M urea gel and detected by autoradiography. Plasmid M13mp18 was sequenced with the −40 primer and the sequencing reactions loaded on the same gel were used as size markers.

Plasmid constructs, transfection, and CAT assay. For construction of the Zfp148-CAT expression vectors, fragments of the p11-B1 plasmid, containing the exon 1 and the 5′-flanking region of the gene (see Fig. 2), were inserted into the polylinker of the promoterless plasmid pBLCAT-3 (24). pCAT5B contains an *Eco*RI-*Bss*HII genomic fragment spanning from nt −1020 to nt +144, relative to the major transcription initiation site, while pCAT5N contains a *Nar*I-*Bss*HII genomic fragment spanning from nt −205 to nt +144 (see Fig. 6). Both fragments were blunt-ended by using the Klenow fragment of DNA polymerase I and ligated into the filled *Bam*HI site of the pBLCAT-3 vector, according to standard procedures (25).

HeLa, COS7, and C2C12 myogenic cells were cultured, transfected with plasmid vectors, and assayed for CAT and β-galactosidase activities described as previously (4). β-galactosidase expression plasmid pON1 (36), was used to monitor transfection efficiency. All transfections were performed on multiple sets of cultures with at least two different DNA preparations for each plasmid.

Computer analysis of sequencing data. Several software tools were used to analyze the genomic and cDNA sequences. BLAST (http://www.ncbi.nlm.nih.gov/blast) was used to search databases and to align cDNA and genomic sequence to give a view of the actual splice sites. Genomic sequences containing ZNF148 related sequences were detected in the HTGS division of GenBank. These were downloaded and sequence contigs were aligned and analyzed using DNASIS analysis software (Hitachi). Regulatory sequence in the putative promoter regions were analyzed by applying PROMOTER SCAN, TESS, and TRANSFAC to make a comprehensive prediction and POLYAH was used for the prediction of potential polyadenylation sites (all above are from http://dot.imgen.bcm.tmc.edu:9331).

RESULTS AND DISCUSSION

Isolation and Sequence Analysis of BERF-1 cDNAs

Using a Southwestern screening approach we have previously isolated a full length c-DNA encoding BERF-1 from an embryonic muscle expression library (4). Further screening of the same library, performed

by hybridization with the previously isolated A22 cDNA, brought to the isolation of several clones subsequently analyzed by restriction mapping and hybridization assays. Three cDNA clones resulted to differ either in length or restriction map and were further analyzed by nucleotide sequencing. The obtained sequences were compared and aligned to the nucleotide sequence of A22 cDNA (Fig. 1A). Clone A23 contains an identical coding region but it has a longer 5′-UTR and a 3060-bp-long 3′-UTR that is very rich in A+T nucleotides (more then 70%) and contains six putative polyadenylation sites (pA1–pA6 in Fig. 1A). Clone A26 contains a truncated open reading frame (ORF), it lacks most of the amino-terminal coding region but has a 1311-bp-long 3′-UTR suggesting that it has been generated by polyadenylation at the pA5 site. Clone A24 also has a truncated ORF, it differs in the length of the 3′-UTR which is consistent with the utilization of the pA1 polyadenylation site and, interestingly, it lacks of an internal 749-bp coding sequence corresponding to the carboxyl-terminal portion of BERF-1 (Fig. 1A). As a result of this deletion the A24 cDNA theoretically specifies a protein that lacks a 249-aminoacids segment as compared to BERF-1 but has 20-extra aminoacids at the carboxyl-terminus due to an out-of-frame inclusion of a 58-nt sequence (data not shown). These results indicated that the A24 cDNA might represent an alternative-spliced variant form of mRNA with the potential of producing a shorter BERF-1 isoform and the heterogeneity that we and others have previously observed in the size of BERF-1 transcripts (4, 5), might be due to both a combination of alternative splicing and the usage of different polyadenylation sites. This hypothesis was supported by BLAST search of ESTs that identified several mouse sequences ending at the predicted sites and a human clone (AI318282) whose sequence organization is indicative of alternative splicing.

Cloning of the Mouse Zfp148 Gene

In order to investigate the structure of the Zfp148 gene and its potential to generate different mRNAs, recombinant lambda phages from a mouse strain 129SvJ genomic library were screened using the BERF-1 cDNA as a probe. Several hybridizing clones were isolated and characterized by restriction mapping and Southern blot analysis with different portions of the A22 cDNA. Two overlapping clones contained a single *Bam*HI fragment that hybridized to cDNA probes derived from both 5′ and 3′ regions of the mRNA. Further sequence analysis indicated that these clones contain a Zfp148 pseudogene we termed Zfp148-ps1 (17). A comparison of the obtained sequence (Accession No. AJ001165) with the nucleotide sequence of the A22 cDNA confirmed that Zfp148-ps1 does not contain any intron and revealed the presence of multiple insertions and deletions that generate frameshifts and termination codons suggesting that the gene does not encode a functional BERF-1-like protein (Fig. 1C).

The other isolated clones (λ128, λ2.3, λ3.1, λ77, and λ49 in Fig. 2) showed distinct *Bam*HI hybridizing bands with the different probes and did not give a positive signal with the most 5′ cDNA fragment. Additional restriction map and hybridization analyses of these clones indicated that they did not completely overlap each other suggesting the presence in the gene of very large introns. Therefore, in order to isolate recombinant clones encompassing the 5′ end of the Zfp148 gene and to establish the genomic distances among the different exons, a mouse genomic PAC library (21) was screened with a 300-bp *Eco*RI-*Bam*HI fragment spanning the 5′-end of the A23 cDNA. Several positive clones were identified and by Southern hybridization, long-PCR analysis and nucleotide sequencing, one clone (RPCI-21-359-P11 in Fig. 3) was established to contain the entire gene. A partial restriction map of the Zfp148 locus is shown in Fig. 2. Southern blot analysis was performed with probes derived from two regions of the A23 cDNA, corresponding respectively to exons 4 and 5 (probe A in Fig. 1A) and to the 5′ portion of the 3′-UTR (probe B in Fig. 1A and Fig. 2), and with a genomic fragment derived from intron 7 (probe C in Fig. 2). Mouse genomic DNA cleaved with a variety of enzymes gave simple patterns, with all the detected bands accounting for the gene or the pseudogene, indicating that both Zfp148 and Zfp148-ps1 genes are present in a single copy and there are not other closely related genes in the mouse genome (Fig. 1B).

Genomic Organization of the Mouse Zfp148 Gene

By matching the sequences of the isolated genomic clones with the known cDNA sequences, the complete structure of the mouse Zfp148 gene was established (Fig. 2). The entire gene covers a region of about 110 kb of genomic DNA and it is composed of 9 exons and 8 introns. The size of the exons ranges from 82 bp (exons 2 and 7) to 5275 bp (exon 9) and all the donor and acceptor splice sites conform to the GT-AG rule (Table 1). Exons 1, 2, 3, and part of exon 4 encode the 5′-UTR sequence present in the A23 cDNA (Fig. 1A). The coding region starts at base 17 of exon 4, spans exon 5 to 8 and extends up to the first 1599 bp of exon 9. Exon 9 is the largest among the exons, and most of it corresponds to the long 3′-UTR region of the A23 cDNA. The size of the introns was estimated by comparing the size of the intron–containing restriction fragments and by long-PCR of both the PAC clone and mouse genomic DNAs using exon-specific primers (data not shown). Intron 7 is the largest, spanning about 24 kb, the smallest is intron 8, which is 280 bp in length (Table 1).

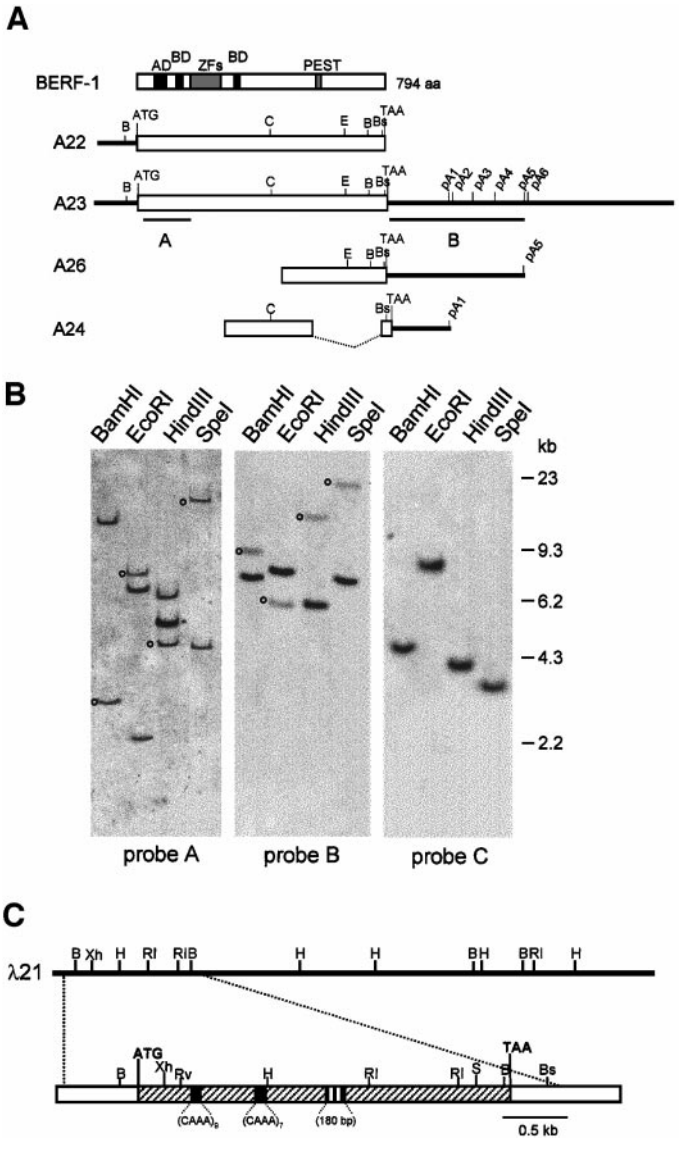

FIG. 1. Organization of the Zfp148 gene products, Southern blot analysis, and structure of the isolated mouse pseudogene. (A) Partial restriction maps of the isolated cDNAs. Black lines represent the UTR regions and the coding regions are represented as open boxes. Positions of the translation start site (ATG), the stop codon (TAA), the putative polyadenylation sites (pA1–pA6) and relevant restriction sites are indicated. B, *Bam*HI; C, *Cla*I; E, *Eco*RI; Bs, *Bst*XI. For comparison a schematic representation of the BERF-1 protein is shown at the top. The acidic (AD), basic (BD), zinc finger (ZF), and PEST domains are labeled. (B) Southern blot analysis of mouse genomic DNA. Mouse genomic DNA (10 μg) from the fibroblast cell line A31 was digested with the indicated restriction enzymes and analyzed by Southern blot hybridization as described under Materials and Methods. The position of the probe with respect to the cDNA and genomic clone is indicated in Figs. 1A and 2. Open circles adjacent to some hybridizing bands indicate the genomic fragments corresponding to the Zfp148 processed pseudogene (Zfp148-ps1). Molecular size (kb) markers from *Hin*dIII digested lambda DNA are shown at left. (C) Organization of the Zfp148-ps1 processed pseudogene. Schematic representation and partial restriction map of the λ21 clone. The 2.7 kb region containing the processed pseudogene is enlarged, positions of the inserted $(CAAA)_n$ repeats and a 180-bp insertion, identified by sequence alignment with the A22 cDNA, are shown. B, *Bam*HI; R, *Eco*RV; S, SpeI.

The exon/intron architecture of the Zfp148 gene suggests a correlation with the protein structure predicted from amino acid sequence and functional studies, showing a domain-specific organization of almost all the exons (Fig. 2). An N-terminal domain, containing a highly acidic region reported to be involved in the binding to p300 (15) is encoded by exon 4 (aa 1–111). A highly basic domain and most of the sequence responsible for transcriptional repression (4) are encoded by exon 5 (aa 112–153). Two of the four zinc-fingers are encoded by exon 6 (aa 134–194) and exon 7 (aa 195–222), respectively. Both exons 6 and 7 are interrupted on either side by introns in codon position 1 (see Table 1), suggesting that Zfp148 may generate, by alternative splicing, related proteins with distinct DNA binding specificities like Drosophila tramtrack and CF2 genes (26, 27) and the human Evi-1 gene (28). The third zinc-finger and the cysteins of the fourth one are encoded by exon 8 (aa 223–262). Exon 9 encodes the second half of the fourth zinc-finger and the C-terminus of the protein (aa 263–794) which includes a basic domain (aa 313–335) and a PEST domain (aa 569–596), identified in a number of eukaryotic proteins with a short half-life (32). In the carboxyl-terminal portion of the protein (aa 444–794) has been also identified a strong transcriptional activation domain (4, 5), while the zinc-finger domain have been found to mediate interaction with PTRF (6).

We have previously shown that the Zfp148 gene is ubiquitously expressed in mouse tissues and three major forms of mRNA corresponding to 7.6, 4.1, and 3.4 kb are present at different levels in different tissues (4). Computer-assisted nucleotide sequence analysis predicted seven putative polyadenylation signals within the long 3′-untranslated region of the Zfp148 gene. Six of these putative sites are present in the longest of the cloned cDNAs (Fig. 1A, sites pA1-pA6 in A23). The nucleotide positions of pA1/pA2 and pA5/pA6 correspond to the 3′-end of the isolated A24 and A26 cDNA, respectively, and are consistent with the size of the 3.4 and 4.1 kb transcripts. The absence of canonical polyadenylation sites at the 3′-end of the A22 and A23 cDNAs (Fig. 1A) and the presence of long stretches of adenine (14n and 15n, respectively) in the corresponding genomic sequences, suggests that both cDNAs were generated by false oligo-dT priming at these A-rich sites and the A23 cDNA may represent a truncated form of the 7.6-kb transcript. This long transcript is probably generated by polyadenylation at one of the two sites (or both) present in the gene 240 and 790 bp downstream of the sequence corresponding to the 3′-end of the A23 cDNA, as further supported by the presence in the mouse ESTs database of five sequences (Accession Nos.: AW227275, AW986940, AA511015, AA047174, BE573816) with 100% similarity to this region (data not shown). All these data further support that transcription of the Zfp148 gene gives rise to at

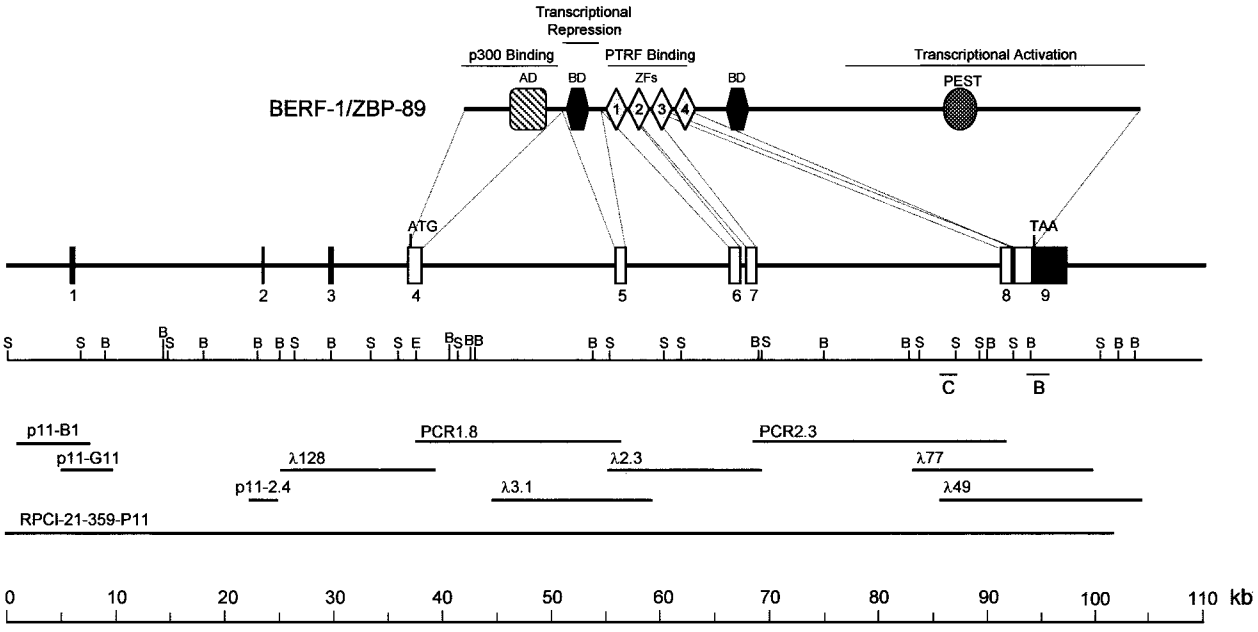

FIG. 2. Genomic organization of the mouse Zfp148 gene. A schematic representation of the different domains of BERF-1/ZBP-89 protein is shown at the top. Broken lines connect the starting and the ending of each coding exon with the corresponding position in the protein. Coding exons are shown as open rectangles, introns as thick lines, noncoding exons as black rectangles. Positions of the translation initiation codon (ATG) and stop codon (TAA) are indicated. A partial restriction map, the relative position and extention of the isolated lambda clones (λ), the PAC clone containing the Zfp148 gene (RPCI), the PCR-amplified fragments (PCR), and the recombinant subclones generated (p11) are shown below. B, *Bam*HI; S, *Spe*I.

least three major forms of mRNA due to differential usage of polyadenylation signals.

Genomic Organization of the Human ZNF148 Gene

Like the mouse Zfp148 gene the human homologue ZNF148 shows an ubiquitous expression pattern and three major transcripts corresponding to 8.6, 7.6, and 4.2 kb have been detected by Northern blot analysis (7, 12). This observation is not in accordance with the length of the isolated cDNAs (Accession Nos. AF039019, L04282, and AJ236885), indicating that none of the identified clones corresponds to a full-length cDNA and suggesting that the ZNF148 gene, like the mouse homologue, may generate a transcript with a long 3′-UTR. Using the mouse A23 cDNA sequence to search the human dbEST, a large number (>90) of ZNF148-positive ESTs were identified, however only one cDNA contig was detected indicating that all the ZNF148-related human ESTs are transcribed from the same gene. The compiled human sequence, derived from the alignment of the ESTs and the cDNAs nucleotide sequences, extends for 5.900 bp with a 391-bp-long 5′-UTR, a 2381-bp coding region and 3128-bp 3′-UTR and it shows 83% similarity with the mouse A23 cDNA (data not shown). The human cDNA sequence was used to search the current release of the HTGS division of the GenBank. Four distinct unfinished (phase 1) ZNF148-related sequences were identified, two sequences (Accession Nos. AC026618 and AC067998) were derived from BAC clones assigned to chromosome 3 (clones RP11-775J23 and RP11-627D21, respectively), one sequence (AC19289) was derived from a BAC clone assigned to chromosome 4 (clone RP11-203I19) and one sequence (AC032043) was derived from a clone assigned to chromosome 11 (PR11-775I17). The unordered contigs of the four clones were aligned and a continuous sequence spanning more than 200 kb was obtained, indicating that all the analyzed clones derive from the same region of human chromosome 3 between the STS D3S2370 and D3S1551 and encompass the ZNF148 gene (Fig. 3). By comparing the assembled BAC derived sequence with the human cDNA sequence, we identified all exon sequences and the 5′- and 3′-flanking sequences of the gene obtaining a precise localization of each individual ZNF148 exon. The ZNF148 gene spans a region of about 130 kb, it is composed of 9 exons and its genomic organization is quite similar to that of the homologous mouse gene. Like the mouse gene the 5′-UTR of ZNF148 is contained in three exons with the translation initiation site located in exon 4 and the termination site in exon 9. Exon–intron boundaries of the human ZNF148 gene follow the typical GT-AG rule and map at the same nucleotide positions as in the mouse Zfp148 homologue, therefore the size of each exon is conserved between the mouse and human genes (Table 1). Sequence of the analogous exons is very similar (93–100% homology) and, as shown in Table 1, even the sequences around the exon–intron junctions are highly

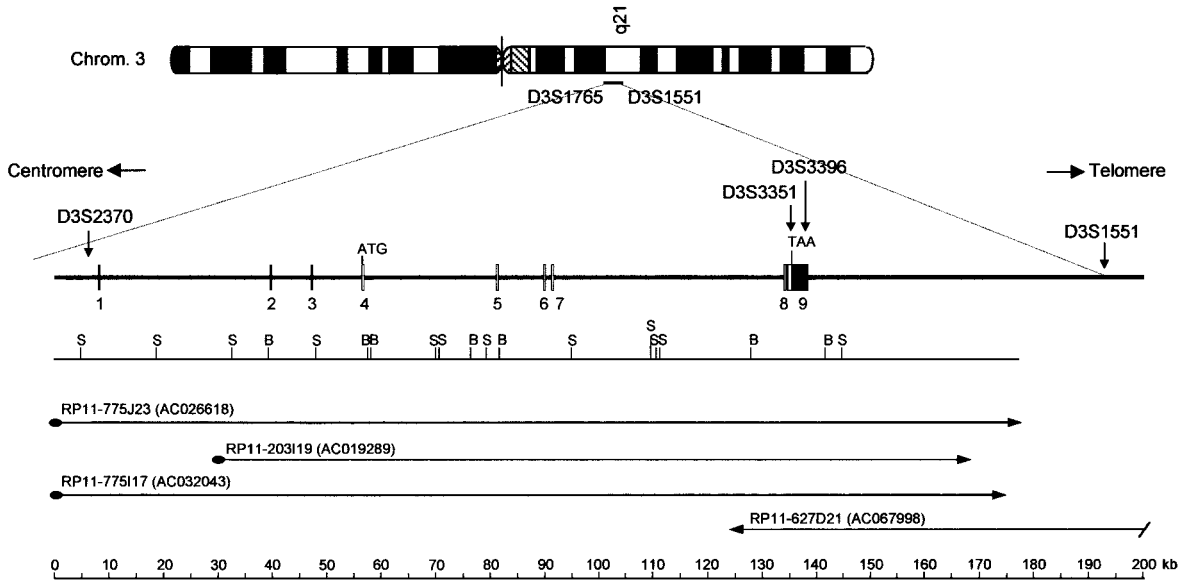

FIG. 3. Genomic organization and chromosomal localization of the human ZNF148 gene. A schematic representation of human chromosome 3 where the position of the ZNF148 gene relative to STS markers and its orientation respect to the centromere and telomere is shown at the top. Coding exons are shown as open rectangles, introns as thick lines, noncoding exons as black rectangles. Positions of the translation initiation codon (ATG) and stop codon (TAA) are indicated. A partial restriction map, relative position and extention of the BAC clones with corresponding clone name and GenBank Accession No. (in brackets) are shown below. Sequence ends of each clone are indicated by filled circles (Sp6 end) or black arrowheads (T7 end). The position of STS markers, localized by sequence analysis, is indicated by vertical arrows. B, *Bam*HI; S, *Spe*I.

conserved. Additional searches identified the positions of other STS loci relative to the exons of the gene, D3S2370 is located at about 2.5 kb from the 5′-end of exon 1, D3S3351 and D3S3396 are both located in exon 9 within the 3′-UTR and D3S1551 is located about 50 kb downstream of the 3′-end of exon 9 (Fig. 3), therefore the ZNF148 gene, which localizes centromere–5′→3′–telomere, physically links D3S2370 to D3S1551. Recently the Psoriasis Susceptibility 5 locus (PSORS5) has been mapped at marker D3S1551 (29), making ZNF148 a positional candidate gene for this disease.

Mapping of the Transcription Initiation Sites and Characterization of the 5′-End Region of the Mouse Zfp148 Gene

The transcription initiation site(s) of the mouse Zfp148 gene were defined by primer extention analysis performed with an end-labeled synthetic oligonucleotide complementary to bases 183 to 212 of the 5′-UTR present in the A22 cDNA, which corresponds to a region of exon 2. In the presence of mouse limb poly(A)-rich RNA isolated from 13 dpc embryos, one major extended product of 228 bp and several minor bands ranging from 122 to 354 bp were observed (Fig. 4A, lane 1). None of these bands was detectable in the control lane containing unrelated RNA (Fig. 4A, lane 2), even after longer exposure (data not shown). The exact positions of the major and minor start sites with respect to the sequence of the Zfp148 5′-end region are indicated in Fig. 5. The major start and at least three of the closest sites are consistent with the transcription start sites predicted from the isolated cDNAs (see Fig. 1A) and the mouse ESTs found in the database (AI158501, AA967471, and BE847435). Our attempts to confirm these data with oligonucleotides complementary to sequences in exon 1 and by S1 protection analysis have been unsuccessful, most likely because of the presence of quite stable secondary structures in the 5′-UTR of the Zfp148 mRNA, due to the high G+C content (see Fig. 4B).

To characterize the 5′-end region of the Zfp148 gene, a 2.1 kb *Eco*RI-*Spe*I fragment, derived from the genomic subclones p11-B1 (Fig. 2), which contains the first untranslated exon and 5′- and 3′-flanking regions was sequenced. The entire sequence was analyzed for GC content, CpG and GpC dinucleotides density, and recognition sites for methylation-sensitive restriction enzymes (Fig. 4B). The overall GC content of the 2133 bp fragment is 59%, however the 480 bp immediately upstream of the major transcription initiation site and in the 370 bp immediately downstream of the exon 1 are particularly G+C-rich (63 and 79%, respectively). The ratio of observed/expected CpG is 0.75 for the entire 2133-bp fragment and 0.76 for the 480 bp upstream of the major transcription initiation site, both of which are greater than 0.6, defining a CpG island (30). A restriction map analysis for several methylation-sensitive enzymes is shown in Fig. 5B, the region appeared loaded of these sites indicating that the sequence surrounding the first

TABLE 1

Exon–Intron Junctions of Mouse Zfp148 and Human ZNF148 Genes

	Exon size	Splice donor site	Intron size	Splice acceptor site
	Exon 1		Intron 1	
Zfp148	(53–302)	*ccagccccgcg***gt**aaacgct	(17.5)	cttttac**ag***gatagaagag*
ZNF148	(150)	g*cccgcccgcg***gt**aaagcgc	(32.6)	cttttgt**ag***gatagaggga*
	Exon 2		Intron 2	
Zfp148	(82)	*attttatagag***gt**aagtcat	(6.2)	ctcctgt**ag***gcattggatg*
ZNF148	(82)	*attttatagag***gt**aagtcat	(7.7)	ctcctgt**ag***gcattggatg*
	Exon 3		Intron 3	
Zfp148	(135)	t*tcagcctcag***gt**aactttat	(7.6)	tctttct**ag***gcagttatac*
ZNF148	(135)	*atttaatggag***gt**aattcag	(9.6)	gttttct**ag***gcagttatac*
	Exon 4		Intron 4	
Zfp148	(344)	TCAATGTCCCT**gt**aagtact	(18.8)	tttaaac**ag**ATAAGTGTAA
ZNF148	(344)	TTAATGTCCCT**gt**aagtaat	(25.0)	tttaaac**ag**ATAAGCGTAA
		AsnValPro		IleSerVal
		111		112
	Exon 5		Intron 5	
Zfp148	(125)	CTCCTGCAAAA**gt**aagagac	(10.2)	ttaaact**ag**ATCCTTACAA
ZNF148	(125)	CTCCCGCAAAA**gt**aagacaa	(8.85)	ttgaact**ag**ATCCTTACAA
		ProAlaLys		IleLeuThr
		153		154
	Exon 6		Intron 6	
Zfp148	(123)	CATTCATACAG**gt**attgaat	(1.12)	tattttaagGCGAAAAACC
ZNF148	(123)	CATTCATACAG**gt**attcttg	(1.31)	tattttaagGTGAAAAACC
		IleHisThrG		lyGluLys
		194		196
	Exon 7		Intron 7	
Zfp148	(82)	GATCCATACCG**gt**gagtgct	(23.8)	atgtactagGTGAAAAACC
ZNF148	(82)	GATTCATACTG**gt**gagtgtt	(43.4)	atgtgctagGTGAAAAACC
		IleHisThrG		lyGluLys
		222		224
	Exon 8		Intron 8	
Zfp148	(118)	ACTGCTTACAG**gt**aagtca	(0.28)	tttcaacagTATTTTTCCA
ZNF148	(118)	ACTGTTTACAG**gt**aagagag	(0.27)	attcaacagTATTTTTCCA
		CysLeuGln		TyrPheSer
		262		263
	Exon 9			
Zfp148	(5275)	*gctatattctgcaaataaaaagtgtgtgttcctttttttccacctaaactttggt*		
ZNF148	(4955)	*gctatattctgcaaataaaaaccgtgttt*ccttttttcacttaaactttggtagg		

Note. The exons coding sequences are shown in uppercase letters and exons untranslated sequences are shown in italic lowercase letters; introns sequences have been depicted in lowercase letters. The invariant 5′-gt · · · ag-3′ sequence are shown in boldface. Size of exons and introns is indicated in parenthesis in base pairs and kilobases, respectively. Amino acids are numbered as reported previously (4). Putative 3′ mRNA processing signals are underlined.

exon meets all the criteria to be designated as a CpG island (30). A sequence comparison with the corresponding sequences of the human ZNF148 gene indicated the existence of a high degree of similarity in the first exon (93%), as well as in the region containing the putative promoter (85%) and a large portion of the first intron (73–85%) (Fig. 4B). A 1200-bp sequence containing exon 1 and its 5′-flanking region was also analyzed for the existence of potential binding sites for known transcription factors. The sites with a consensus sequence of at least six bases and with a match ratio higher than 0.9 are shown in Fig. 5. The majority of the binding sites are in a 540-bp region (nt −374 to nt +166) that is highly conserved between mouse and human genes. Furthermore, the position and the relative distance of eight specific binding sites, Sp1, TCF-1, Oct1, NF-kB, MyoD, Ets-1, LyF-1, and BERF-1, are also conserved (Fig. 5), suggesting the actual involvement of these transcription factors in the transcriptional control of both genes.

Identification of the Zfp-148 Gene Promoter Region

To test whether the 5′-flanking region of the mouse Zfp-148 gene contained *cis*-acting elements capable to

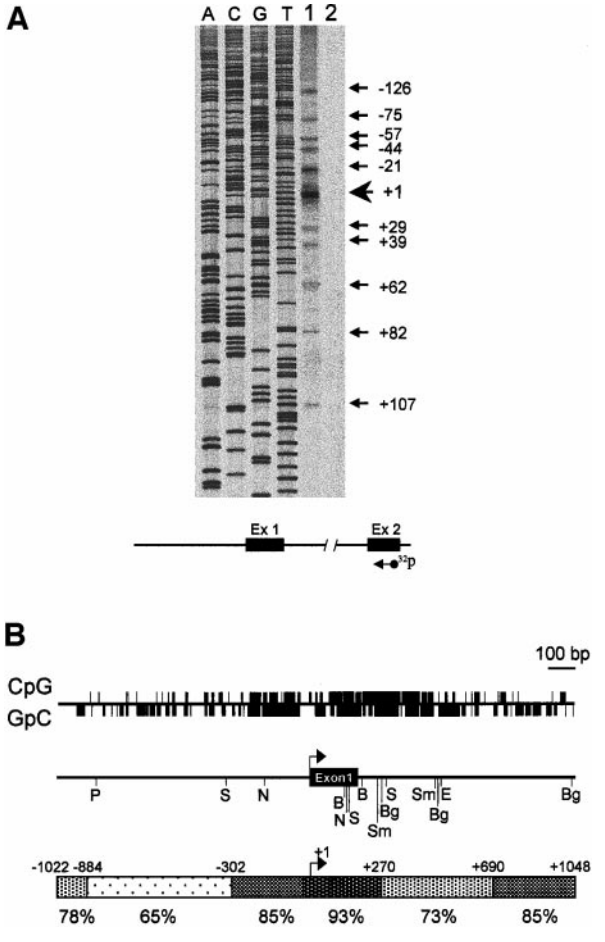

FIG. 4. (A) Identification of the transcription start site of the mouse Zfp148 gene. A 29-nucleotide oligomer, complementary to bases +183 to +212 in exon 2, was labeled at its 5′ end and used for primer extension in the presence of 5 μg of poly(A)-rich RNA from 13 dpc-mouse embryos (lane 1) or 10 μg of yeast tRNA as a control (lane 2). The lenghts of the extended fragments, indicated by arrows on the right side, were estimated by the sequence reactions generated using the −40 primer and M13mp18 DNA (lanes A, C, G, and T). A big arrow indicates the major extented product (labeled +1), minor products are indicated by small arrows and numbered from +107 to −126 relative to the major start (see also Fig. 6). (B) Analysis of the sequences surrounding the first untranslated exon of the mouse Zfp148 gene. Distribution of the CpG and GpC dinucleotides, positions of the restriction sites for a series of rare CG-cutters and a schematic representation of nucleotide sequence similarity in the 5′ end, first exon and first intron of the human and mouse genes, are shown. Nucleotide positions are numbered relative to the major transcription start site (arrow). Regions of the mouse gene that show different percentages of similarity with the homologous regions of the human gene are indicated (shaded boxes). B, *Bss*HI; Bg, *Bgl*I; E, *Eag*I; N, *Nar*I, S, *Sac*II; Sm, *Sma*I; P, *Pst*I.

regulate transcription of the gene acting as a functional promoter, two chimeric ZNF148-CAT constructs, one including 5′-upstream sequences from nt −1020 relative to the major transcription start site, the other a 5′-deletion up to nt −205, both ending at nt +144 were used in transient transfection assays. Both CAT constructs were active in proliferating and differentiated C2C12 cells as well as in non muscle cells (Fig. 6B), indicating that a basal promoter is present within the 205-bp sequence upstream of the major start site. In all cases deletion from nt −1020 to nt 205 resulted in a decrease of the CAT activity indicating the presence of at least one positive regulatory element in the deleted region. It is interesting to note that the deleted construct, CAT5N, is significantly less active in myoblasts as compared to differentiated myotubes or COS7 cells (40% residual activity vs 80%) suggesting that myoblast-specific regulatory elements might be located between nt −1020 and nt 205. Additional experiments will be required to determine whether

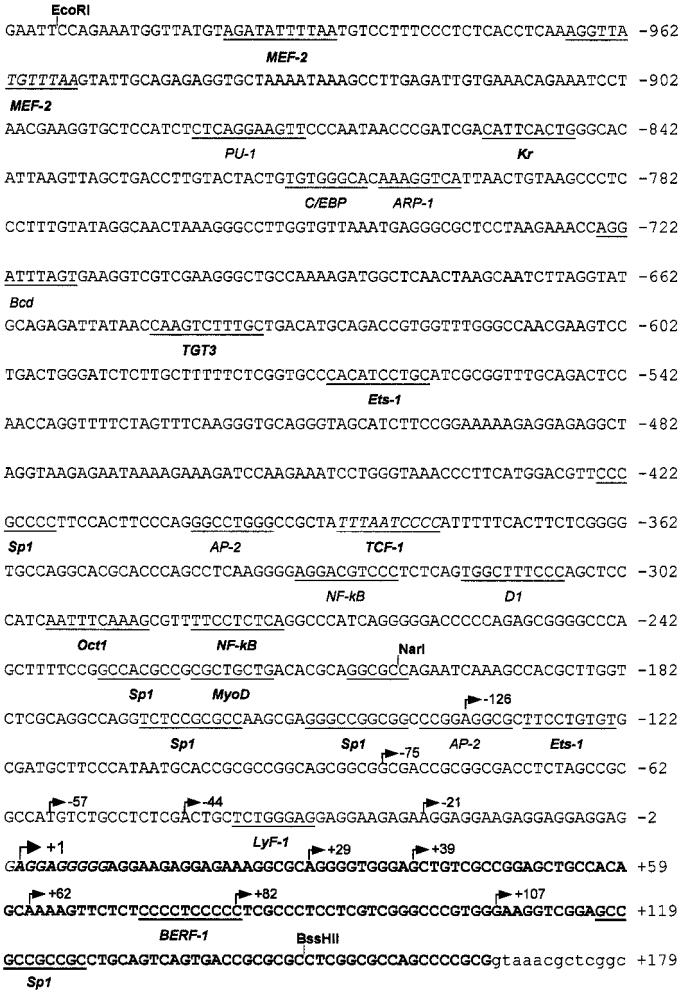

FIG. 5. Nucleotide sequence of the 5′ end of the mouse Zfp148 gene. The nucleotide sequence is numbered with the major transcription start site designated as +1 (big arrow), nucleotides 5′ to this position are negatively numbered, nucleotides downstream are indicated with positive numbers, sequence of the first untranslated exon is printed in boldface letters and nucleotides within the first intron are shown in lower case letters. The minor multiple start sites of transcription demonstrated by primer extention are indicated by small horizontal arrowheads and numbered relatively to the major site. Putative regulatory elements are underlined and specified, elements conserved in the human ZNF148 promoter sequence are in bold. Positions of restriction sites relevant for the preparation of Zfp148-CAT constructs are also indicated.

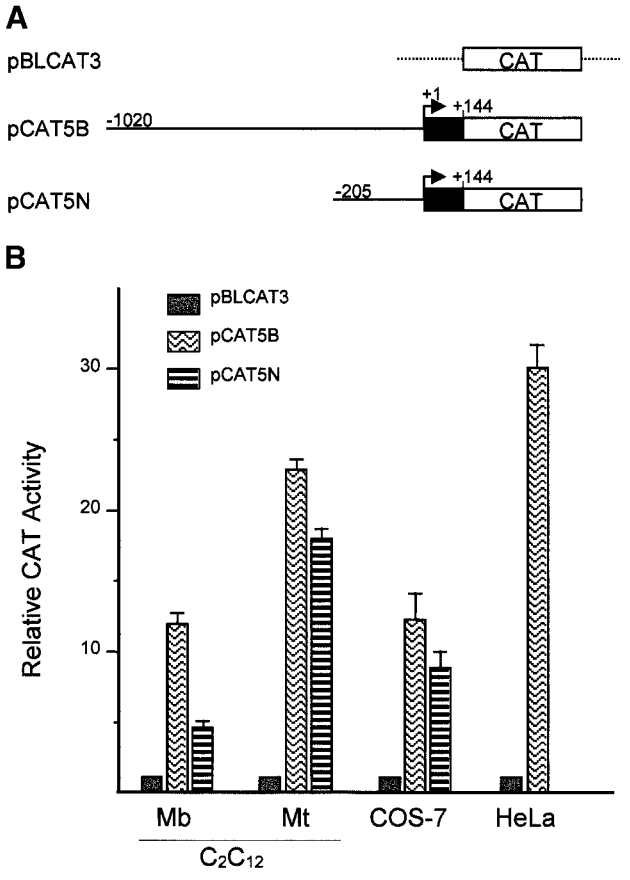

FIG. 6. Identification of transcriptional regulatory elements in the 5′-flanking region of the mouse Zfp148 gene. (A) Schematic representation of the promoterless plasmid pBLCAT3 and Zfp48-CAT constructs used in transient transfection assays. The nucleotide positions relative to the major transcription start site (+1) are indicated, the first untranslated box is indicated (black box). (B) Activity of the different Zfp148-CAT constructs in C2C12 myoblasts (Mb) and myotubes (Mt), COS7 and HeLa cells. Relative CAT activities, corrected for differences in transfection efficiencies, are expressed in arbitrary units relative to the basal level produced by the control plasmid pBLCAT-3 (assigned a value of 1, which correspond to 0.8 to 1% conversion of (^{14}C) chloramphenicol). The data are average +/− standard deviations of three to five independent experiments.

sequences responsible for transcriptional regulation of the Zfp-148 gene in different cell types are present in a more distal portion of the 5′-flanking region or within the first intron. Although, the Zfp148 gene is ubiquitously transcribed it is differentially expressed in a cell type-specific manner during mouse skeletal muscle development (4) and T cells differentiation (13), suggesting that, besides of common regulatory mechanisms ubiquitously used, distinct mechanisms that control expression of the gene relative to cell lineage and developmental stage should exist.

CONCLUSIONS

The present study has characterized the genomic structures of the mouse Zfp148 gene and the human homologue ZNF148, that encode a zinc-finger transcription factor known as BERF-1/ZBP89/BFCOL1/hTβ. The intron/exon organization and nucleotide sequence are highly conserved, both mouse and human genes consisting of 9 exons, the coding ones are organized in a domain-specific manner, each encoding a distinct functional domain. Comparison of the murine 5′-flanking region with the corresponding human 5′-end sequence revealed a significant degree of similarity. In both genes sequences surrounding the first untranslated exon are imbibed in a CpG island and contain several *cis*-acting elements whose position and relative distance are also well conserved, indicating a functional role for these regions. Preliminary transfection studies indicate that sequences up to 1.0 kb upstream of the major transcription start site do not significantly contribute to a cell type-specific transcription of the mouse Zfp148 gene, therefore, *cis*-acting elements located further upstream or in the introns or other transcriptional or posttranscriptional mechanisms might be involved in the regulation of expression. The data reported indicate that several mRNAs with different 3′-UTR are generated by the usage of distinct polyadenylation signals. Large blocks with high sequence similarity (>90%) are also present in the 3′-UTR of the mouse Zfp148 and the human ZFP148 genes. These regions contain several ATTTA motifs that have been found to be involved in controlling the stability of mRNA molecules coding for highly inducible proteins such *c*-fos, *c*-myc, GM-CSF, and IIFN-β (31). The presence of these sequences in the mRNA of the mouse and human transcripts associated with the utilization of several alternative polyadenylation sites, support the hypothesis of the existence of cell type-specific posttranscriptional regulatory mechanisms. We have also isolated a BERF-1 cDNA lacking an internal 749-bp-long coding sequence corresponding to the carboxyl-terminal portion of the protein. Sequence comparison with the isolated Zfp148 genomic clones indicated that this variant form of mRNA might be generated by a splice event occurring within exon 9. In addition a human EST was found to correspond to a message theoretically resulting by alternative splicing of the human primary transcript. This message has the potential to encode a protein that, like in the mouse, lacks most of the carboxyl-terminal region. This preliminary evidence suggests that alternatively spliced transcripts of the Zfp148 may produce protein isoforms that may have distinct functions.

ACKNOWLEDGMENTS

This work was supported by Telethon-Italia (Project No. 943 to A.G.) and by M.U.R.S.T. and Associazione Italiana per la Ricerca sul Cancro (AIRC) grants (to S.F). We are indebted to the Fondazione Telethon for the excellent work in administering the Telethon grant and to the Yac Screening Centre (DIBIT-San Raffaele, Milan, Italy), for providing us with the mouse RPCI-21 genomic library. R.P. was

supported by fellowships under the Telethon and M.U.R.S.T. grants and F.C. was under a Ph.D. program supported by the FSE (Fondo Sociale Europeo).

REFERENCES

1. Turner, J., and Crossley, M. (1999) Mammalian Kruppel-like transcription factors: More than just a pretty finger. *Trends Biochem. Sci.* **24,** 236–240.
2. Suske, G. (1999) The Sp-family of transcription factors. *Gene* **238,** 291–300.
3. Feo, S., Antona, V., Barbieri, G., Passantino, R., Cali, L., and Giallongo, A. (1995) Transcription of the human beta enolase gene (ENO-3) is regulated by an intronic muscle-specific enhancer that binds myocyte-specific enhancer factor 2 proteins and ubiquitous G-rich-box binding factors. *Mol. Cell Biol.* **15,** 5991–6002.
4. Passantino, R., Antona, V., Barbieri, G., Rubino, P., Melchionna, R., Cossu, G., Feo, S., and Giallongo, A. (1998) Negative regulation of beta enolase gene transcription in embryonic muscle is dependent upon a zinc finger factor that binds to the G-rich box within the muscle-specific enhancer. *J. Biol. Chem.* **273,** 484–494.
5. Hasegawa, T., Takeuchi, A., Miyaishi, O., Isobe, K. I., and De, C. B. (1997) Cloning and characterization of a transcription factor that binds to the proximal promoters of the two mouse type I collagen genes. *J. Biol. Chem.* **272,** 4915–4923.
6. Hasegawa, T., Takeuchi, A., Miyaishi, O., Xiao, H., Mao, J., and Isobe, K. (2000) PTRF (polymerase I and transcript-release factor) is tissue-specific and interacts with the BFCOL1 (binding factor of a type-I collagen promoter) zinc-finger transcription factor which binds to the two mouse type-I collagen gene promoters. *Biochem. J.* **347** Pt 1**,** 55–59.
7. Merchant, J. L. *et al.* (1996) ZBP-89, a Kruppel-like zinc finger protein, inhibits epidermal growth factor induction of the gastrin promoter. *Mol. Cell Biol.* **16,** 6644–6653.
8. Cheng, P. Y., Kagawa, N., Takahashi, Y., and Waterman, M. R. (2000) Three zinc finger nuclear proteins, Sp1, Sp3, and a ZBP-89 homologue, bind to the cyclic adenosine monophosphate-responsive sequence of the bovine adrenodoxin gene and regulate transcription. *Biochemistry* **39,** 4347–4357.
9. Law, G. L., Itoh, H., Law, D. J., Mize, G. J., Merchant, J. L., and Morris, D. R. (1998) Transcription factor ZBP-89 regulates the activity of the ornithine decarboxylase promoter. *J. Biol. Chem.* **273,** 19955–19964.
10. Wieczorek, E., Lin, Z., Perkins, E. B., Law, D. J., Merchant, J. L., and Zehner, Z. E. (2000) The zinc finger repressor, ZBP-89, binds to the silencer element of the human vimentin gene and complexes with the transcriptional activator, Sp1. *J. Biol. Chem.* **275,** 12879–12888.
11. Ye, S., Whatling, C., Watkins, H., and Henney, A. (1999) Human stromelysin gene promoter activity is modulated by transcription factor ZBP-89. *FEBS Lett.* **450,** 268–272.
12. Wang, Y., Kobori, J. A., and Hood, L. (1993) The ht beta gene encodes a novel CACCC box-binding protein that regulates T-cell receptor gene expression. *Mol. Cell Biol.* **13,** 5691–5701.
13. Reizis, B., and Leder, P. (1999) Expression of the mouse pre-T cell receptor alpha gene is controlled by an upstream region containing a transcriptional enhancer. *J. Exp. Med.* **189,** 1669–1678.
14. Hasegawa, T., Xiao, H., and Isobe, K. (1999) Cloning of a GADD34-like gene that interacts with the zinc-finger transcription factor which binds to the p21(WAF) promoter. *Biochem. Biophys. Res. Commun.* **256,** 249–254.
15. Bai, L., and Merchant, J. L. (2000) Transcription factor ZBP-89 cooperates with histone acetyltransferase p300 during butyrate activation of p21waf1 transcription in human cells [In Process Citation]. *J. Biol. Chem.* **275,** 30725–30733.
16. Baumann, M., Feederle, R., Kremmer, E., and Hammerschmidt, W. (1999) Cellular transcription factors recruit viral replication proteins to activate the Epstein-Barr virus origin of lytic DNA replication, oriLyt [published erratum appears in *EMBO J.* (2000) Jan. 17;**19**(2), 315]. *EMBO J.* **18,** 6095–6105.
17. Antona, V., Cammarata, G., De Gregorio, L., Dragani, T. A., Giallongo, A., and Feo, S. (1998) The gene encoding the transcriptional repressor BERF-1 maps to a region of conserved synteny on mouse chromosome 16 and human chromosome 3 and a related pseudogene maps to mouse chromosome 8. *Cytogenet. Cell Genet.* **83,** 90–92.
18. Tommerup, N., and Vissing, H. (1995) Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics* **27,** 259–264.
19. Law, D. J., Tarle, S. A., and Merchant, J. L. (1998) The human ZBP-89 homolog, located at chromosome 3q21, represses gastrin gene expression. *Mamm. Genome* **9,** 165–167.
20. Oliva, D., Barba, G., Barbieri, G., Giallongo, A., and Feo, S. (1989) Cloning, expression, and sequence homologies of cDNA for human gamma enolase. *Gene* **79,** 355–360.
21. Osoegawa, K., Tateno, M., Woon, P. Y., Frengen, E., Mammoser, A. G., Catanese, J. J., Hayashizaki, Y., and de Jong, P. J. (2000) Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* **10,** 116–128.
22. Giallongo, A., Oliva, D., Cali, L., Barba, G., Barbieri, G., and Feo, S. (1990) Structure of the human gene for alpha-enolase. *Eur. J. Biochem.* **190,** 567–573.
23. Oliva, D., Cali, L., Feo, S., and Giallongo, A. (1991) Complete structure of the human gene encoding neuron-specific enolase. *Genomics* **10,** 157–165.
24. Luckow, B., and Schutz, G. (1987) CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. *Nucleic Acids Res.* **15,** 5490.
25. Ausubel, F. M., Brent, R., Kingston, R. F., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1999) John Wiley, New York.
26. Read, D., and Manley, J. L. (1992) Alternatively spliced transcripts of the Drosophila tramtrack gene encode zinc finger proteins with distinct DNA binding specificities. *EMBO J.* **11,** 1035–1044.
27. Hsu, T., Gogos, J. A., Kirsh, S. A., and Kafatos, F. C. (1992) Multiple zinc finger forms resulting from developmentally regulated alternative splicing of a transcription factor gene. *Science* **257,** 1946–1950.
28. Bordereaux, D., Fichelson, S., Tambourin, P., and Gisselbrecht, S. (1990) Alternative splicing of the Evi-1 zinc finger gene generates mRNAs which differ by the number of zinc finger motifs. *Oncogene* **5,** 925–927.
29. Enlund, F. *et al.* (1999) Psoriasis susceptibility locus in chromosome region 3q21 identified in patients from southwest Sweden. *Eur. J. Hum. Genet.* **7,** 783–790.
30. Gardiner-Garden, M., and Frommer, M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.* **196,** 261–282.
31. Zubiaga, A. M., Belasco, J. G., and Greenberg, M. E. (1995) The nonamer UUAUUUAUU is the key AU-rich sequence motif that mediates mRNA degradation. *Mol. Cell Biol.* **15,** 2219–2230.