Acta Medica et Biologica
Vol. 51, No.1, 13-24, 2003

The Genomic Organization, Alternative Splicing, and Promoter Assay of the Mouse *Ankhzn* Gene

Hiroshi MARUYAMA[1,2], Hideyuki KURIYAMA[3], Naoya ISHII[1,2], Kazuhisa ITO[3], Shoji ODANI[2] and Ryozo KUWANO[1]

[1]Department of Molecular Genetics, Genome Science Branch, Center for Bioresource-based Researches, Brain Research Institute, Niigata University, [2]Course of Biosphere Science, Graduate School of Science and Technology, Niigata University, and [3]Department of Internal Medicine (II), Niigata University School of Medicine, Niigata, Japan

Received December 27 2002; accepted January 20 2003

Summary. **We have previously established a GT3-12 mouse line[1] in which a novel gene was trapped by a promoterless gene trap (GT) vector. The gene was designated as *Ankhzn* because of the presence of 17 ankyrin-repeats and a zinc-finger domain FYVE[2]. In this study, we characterize the genomic organization of the mouse *Ankhzn* gene by analyzing two overlapping P1 (bacteriophage P1 cloning system) clones and a single bacterial artificial chromosome (BAC) clone. The *Ankhzn* gene spans more than 95 kb and comprises 25 exons, where the splice site conforms to the GT-AG rule except for the GC splice donor site instead of GT of intron 4. This GC splice donor site and several junctions are conserved between the mouse *Ankhzn* and the human *ANKHZN*. Furthermore, results of RT-PCR revealed that the 49 bp segment at the 3' end of exon 9 is alternatively spliced to generate two splice variants. One splice variant is a long form that has a calculated molecular mass of 130 kDa. The other is a short form that has a molecular mass of 50 kDa resulting from a frameshift leading to premature termination within exon 10. Results of RT-PCR analysis showed that the Ankhzn long form was expressed much more strongly than the short form in all tissues and developmental stages examined. The *Ankhzn* gene was identified as a single-copy gene by Southern blot analysis. Examination of the T31 Mouse Radiation Hybrid Database RH Chr 11 Public Map Data at the Jackson Laboratory showed its localization to mouse chromosome 11 which has high synteny to human chromosome 17. A promoter assay with a luciferase reporter gene revealed that an approximately 200 bp upstream region to the transcription start site is essential for the transcription of *Ankhzn*.**

Key words—alternative splicing, RT-PCR, intron, exon, FYVE finger.

INTRODUCTION

Ankhzn, which was first characterized as a trapped gene in GT3-12 ES (embryonic stem) cell line mice, was established by means of a gene trap strategy[1]. We previously reported the cDNA cloning of mouse *Ankhzn*[2] and its human homologue, *ANKHZN*[3]. The cDNA sequence of mouse *Ankhzn* revealed a unique domain structure comprised of an N-terminal coiled-coil domain, BTB/POZ domain, four and thirteen ankyrin repeats, and a FYVE finger domain at the extreme C-terminal end. As this is the first gene to be characterized that has both ankyrin repeats and a FYVE finger domain, we named this gene *Ankhzn* (ankyrin repeats hooked to a zinc finger motif[2]). Human *ANKHZN* was also characterized from fetal brain by means of PCR with reference to EST databases, and it shares the same domain organization with its mouse counterpart. Both mouse and human proteins were ubiquitously expressed in the tissues examined and at the developmental stages

Correspondence: Hiroshi Maruyama, Department of Molecular Genetics, Genome Science Branch, Center for Bioresource-based Researches, Brain Research Institute, Niigata University, Niigata 951-8510, Japan.

Abbreviations—Ankhzn, ankyrin repeat hooked to zinc finger motif, BAC, bacterial artificial chromosome; bp, base pair, ES, embryonic stem; GT, gene trap; ICR, Institute of Cancer Research; kb, kilo base; PCR, polymerase chain reaction, RT, reverse transcriptase; UTR, untranslated region.

Table 1. Primers used in this study

Primer name	Sequence	Location
GT3-12-41	5'-TGCAGCTTTGAATCAGCAGGAC-3'	intron 1
GT3-12-44	5'-GCCACTCATCTTCCATGAGC-3'	intron 1
exon 7	5'-CGCTCGAGACCTTCATCACGGAGGT-3'	GT vector
neo 3	5'-GCCTCGAGAGGCTATTCGGCTATGACT-3'	GT vector
GT3-12-3	5'-CGACTGTGTGTTCAGTCAGC-3'	exon 9
GT3-12-20	5'-CCTCATTTCCAGCCCCAGCTG-3'	exon 10
GT3-12-17	5'-GCAGGATGACGCTGATCATTG-3'	exon 25
GT3-12-18	5'-CGTGTTCTAACACACCACTGAG-3'	exon 25
GAPDH-UP	5'-GAAATCCCATCACCATCTTCCAGG-3'	GAPDH
GAPDH-DOWN	5'-CATGTGGGCCATGAGGTCCACCAC-3'	GAPDH

from embryonic day 10.5 to adulthood. Its subcellular localization to endosomal membrane[2] and presence of the FYVE finger motif imply that Ankhzn may have a function in vesicular transport or related processes, since FYVE finger proteins (Fab1, YOTB, Vac1 and EEA1.[4]) are involved in endocytic membrane traffic[5]. For example, early endosome antigen 1 (EEA1) specifically localizes to phosphatidylinositol 3-phosphate-enriched membranes through its FYVE domain and recruits regulatory Rab5 and syntaxins[6,7]. The Ankhzn protein appears to be essential for viability of the animal because we observed embryonic lethality of GT3-12 homozygotes (Ishii et al. unpublished observation). In order to understand the function and regulation of Ankhzn, we first attempted to determine the genomic structure of the *Ankhzn* gene that would afford information on the exact location of the GT vector integration, regulatory elements, and evolutionary relationship to other genes.

We here describe the genomic organization and promoter position of the mouse *Ankhzn* gene. Identification and structure of two alternative splice variants of 130 kDa and 50 kDa and their expression pattern in tissues and development stage are also described.

MATERIALS AND METHODS

Materials

Mouse *Ankhzn* genomic clones were obtained from P1 and BAC libraries purchased from the Genome Systems. Genomic DNA was prepared from GT3-12 cell line mice. ICR mice were purchased from Charles River Japan. Two P1 clones (18215 and 18216) and one BAC clone (20090) were isolated from mouse genomic libraries supplied by GENOME SYSTEMS. B103 and C6 were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum at 37℃ with 5% CO_2 and used for transfection and promoter activity assay.

PCR

The reaction mixture in a total volume of 10 μl contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 2.5 mM of dNTPs, 10 pmol each of forward and reverse primers, 0.25 U of Taq polymerase (TAKARA), and 20 pg of plasmid or 20 ng of genomic DNA. The primers are shown in Table 1. The PCR conditions were as follows: initial denaturation at 93℃ for 5 min, 30 cycles of denaturation at 93℃ for 1 min, annealing at a temperature optimized for individual sets of primers for 1 min, extension at 72℃ for 2 min, and final extension at 72℃ for 7 min.

DNA sequencing

DNA fragments were ligated into pCR2.1 (Invitrogen) or pBluescript II SK (-) (Stratagene) and sequenced using the ALFexpressTM AutoCycleTM Sequencing Kit (Pharmacia Biotech) on ALFred DNA Sequencers (Pharmacia Biotech). Analysis of DNA sequences was carried out with GENETYX-MAC computer software (SOFTWARE DEVELOPMENT) and the BLAST search program on the web site (http://www.ncbi.nlm.nih.gov/BLAST/).

Analysis of genomic organization

PCR was performed with primers GT3-12-41 and GT3-12-44 (Table 1, annealed at 65℃) for screening the P1 clone. Both primers were designed near the

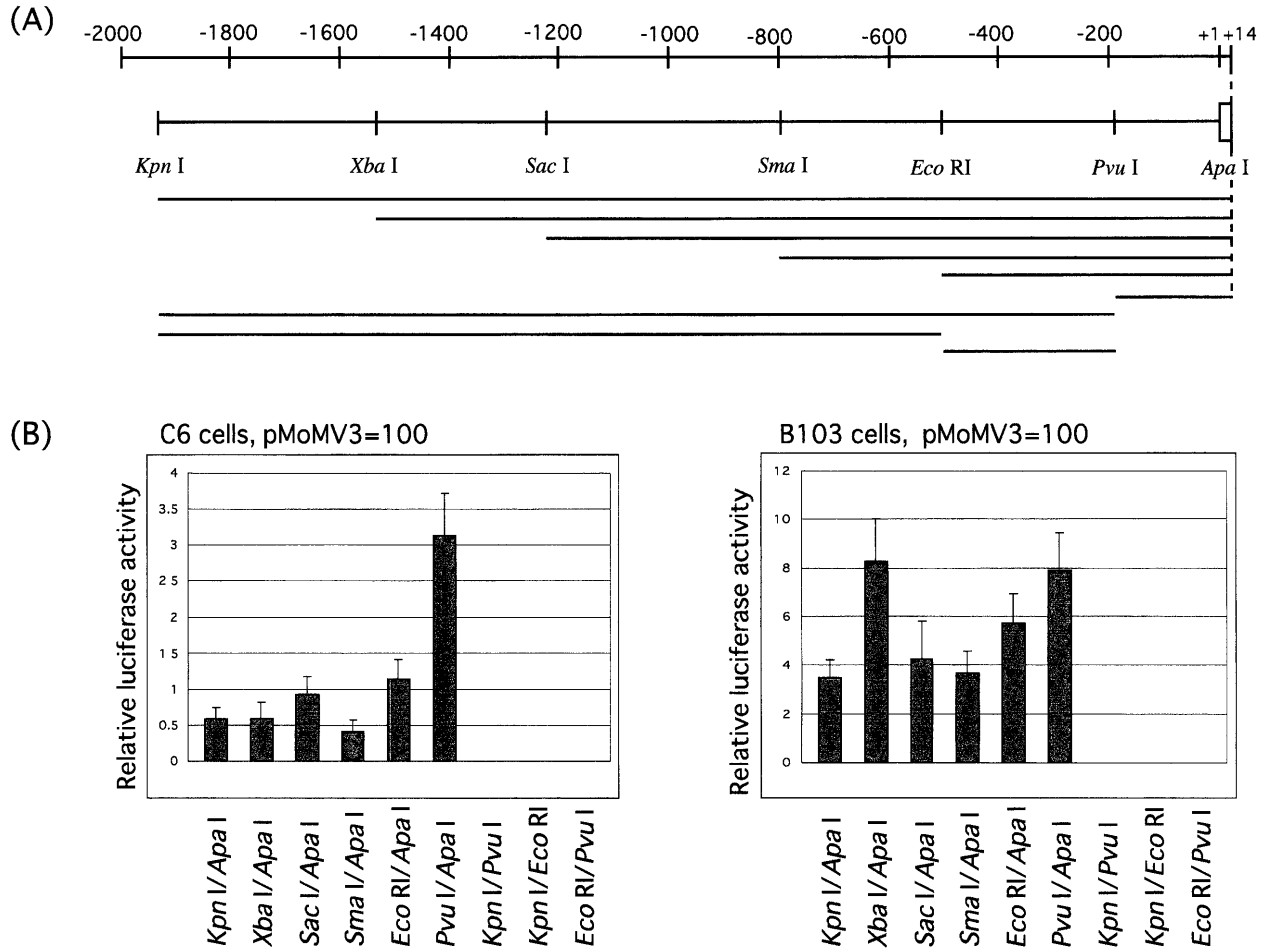

Fig. 1. Promoter activity assay. (**A**) Restriction map of the 5' flanking region of the Ankhzn gene. Promoter fragments were subcloned from P1 clone 18215 by the restriction enzymes indicated and ligated into the promoterless luciferase reporter vector pGV-B. (**B**) Luciferase assay. Luciferase assays were performed in six wells for each vector with a control LacZ plasimid vector. Transfections were performed twice in different experiments. No activities were observed for the construct from -197 to +14 in both B103 and C6 cells.

GT vector inserted region as follows. The genomic DNA of GT3-12 heterozygotes was digested with *Pst* I, and then ligated itself. Inverse PCR was carried out using these self-ligated fragments as templates and GT vector specific primers with exon7 and neo 3 (Table 1, annealed at 65℃). For screening of the BAC library, PCR was performed with primers GT3-12-17 and GT3-12-18 (Table1, annealed at 60℃). Both P1 and BAC plasmids were isolated according to the manufacturer's protocol of Genome Systems, and then subcloned into pBluescript II SK (-) (Stratagene). These plasmids were used to determine the nucleotide sequence and make a restriction map with *Bam* HI, *Eco* RI and *Pst* I.

RT-PCR

Total RNA was isolated with ISOGEN according to the manufacturer's instruction (Wako Chemical, Japan). To eliminate any possible contamination of genomic DNA in the RNA preparation, 10 μg of total RNA was treated with 2 units of DNase I (amplification grade, LIFE TECHNOLOGIES) for 15 min at 20℃, and DNase was inactivated by incubation for 10 min at 65℃. First-strand cDNAs were synthesized from 5 μg of the RNA template and 74 ng of random nucleotide hexamer with 200 units of Super Scripts II reverse transcriptase (LIFE TECHNOLOGIES) at 45℃ for 90 min. The cDNAs were dissolved in 20 μl water, and 1 μl aliquots were used

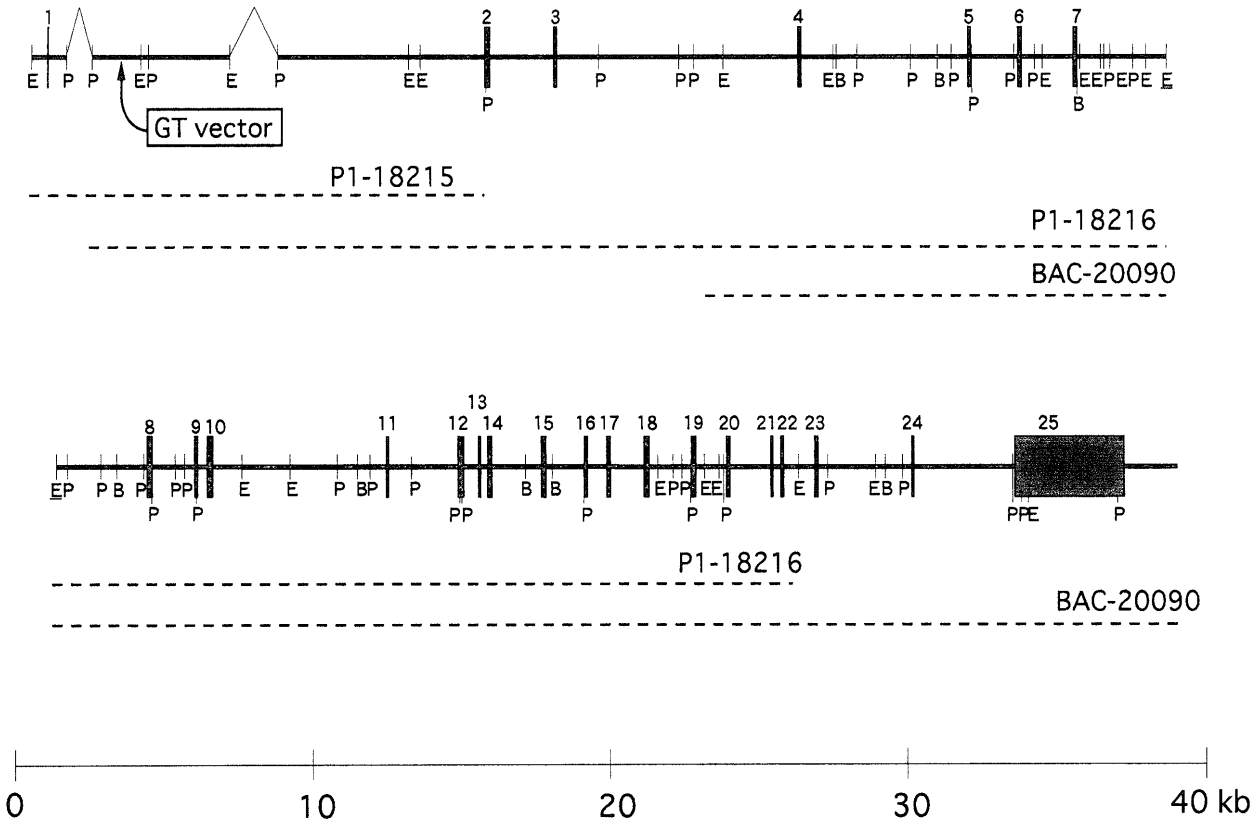

Fig. 2. Genomic organization and restriction map of the mouse *Ankhzn* gene. The *Ankhzn* comprises 25 exons indicated as boxes and 24 introns as a solid line. The restriction map is drawn with *Bam* HI (B), *Eco* RI (E) and *Pst* I (P) restriction sites. E indicates the same *Eco* RI site. The GT vector was inserted into the first intron. Dotted lines show an overlap of genomic clones derived from P1 (18215, 18216) and BAC (20090) libraries.

for PCR. PCR was carried out with primers GT3-12-3 and GT3-20 (Table 1, annealed at 58℃) and a dNTP mixture containing [^{32}P]-labeled dCTP. The cDNAs syntheses were confirmed with a primer set for glyceraldehyde 3-phosphate dehydrogenase (GAPDH-UP and GAPDH-DOWN. Table 1, annealed at 60℃). Amplified fragments were separated by 7.5% polyacrylamide gel electrophoresis. The gels were dried for 30 min at 80℃, and then exposed to X-ray film for 3 h. Comparisons of each splice variant expression pattern were analyzed using a computer software, Quantity One (pdi). Southern blot analysis was performed to confirm that detected bands were actually derived from *Ankhzn*. A 321-bp PCR fragment, amplified from full length *Ankhzn* cDNA with GT3-12-3 and GT3-12-20 primers (Table 1, annealed at 58℃) and labeled with [^{32}P], was used as a probe.

Southern blot analysis

Ten μg of genomic DNA of GT3-12 wild-type mice was digested either with *Bgl* II, *Eco* RI, *Pst* I, *Sca* I, or *Xba* I (TOYOBO), separated by 0.8% agarose gel electrophoresis, transferred to nitrocellulose membrane, and then hybridized with a specific [^{32}P]-labeled 718 bp PCR fragment amplified from the 3' UTR with primers GT3-12-17 and GT3-12-18 (Table 1, annealed at 65℃) at 42℃ in a 40% formamide solution. The filter was washed for 30 min in 0.5×SSC, 0.1% SDS at room temperature and exposed to X-ray film for 3 days.

Isolation of Ankhzn 5' flanking region

To isolate the 5' flanking region of exon 1, the P1 genomic clone 18215 was digested with *Kpn* I and ligated into the pBluescript II SK (-). These plasmids were transfected to the HB101 competent cells (Ta-

Table 2. Exon and intron bounderies of the *Ankhzn* gene

Splice acceptor site				Splice donor site	
Intron	Exon	Numbers of exons	Length bp	Exon	Intron
		1	31	GCGGAAG	**gt**aggtt
cttgc**ag**	AGGAGGT	2	193	AGTACAG	**gt**gagct
ttgtc**ag**	TGATTTGA	3	119	TTGTCAG	**gt**aagtc
taact**ag**	ATGCGA	4	136	GGGAGAG	**gc**caagt
tttac**ag**	ATGTGAG	5	124	TCACTGG	**gt**aagta
tcctt**ag**	GATGACT	6	150	TTCTCAG	**gt**attgt
tgtgc**ag**	CTACCTG	7	166	CAAAGAG	**gt**aagga
ttttaa**g**	GGGATCT	8	205	AGGGCAG	**gt**gagtt
gcctc**ag**	GACTCCT	9 (short)	118	TGGGATT	**gt**gagtg
		9 (long)	69	GTAAACA	**gt**gcgta
atttc**ag**	ATTAGAT	10	200	ATGACAG	**gt**agaac
ctcgt**ag**	GAAATTG	11	98	CAAATGG	**gt**aagtt
cattc**ag**	GGAGAAA	12	229	CAGAAAG	**gt**aggta
tattt**ag**	CCAATGC	13	99	TGGACTG	**gt**aagtc
ttcca**ag**	GCATGCA	14	154	ATGTCAG	**gt**aggac
ttcat**ag**	GACTCAG	15	169	CACTCTC	**gt**gagtc
tctgc**ag**	GTCAGAC	16	119	TTCGCAG	**gt**cttaa
acttc**ag**	TGGCTGT	17	157	TGCACAG	**gt**atata
ttcct**ag**	GATGCAG	18	201	TGAGCAG	**gt**cagtg
gctgc**ag**	GTAGATA	19	177	AAATCTG	**gt**aagcc
tgaac**ag**	CTTCTTG	20	148	AATAATG	**gt**aactc
tcccc**ag**	CTCTTCA	21	91	ATCTGAG	**gt**aagtg
tcatc**ag**	AGGCCAG	22	125	AACACAG	**gt**atgta
gttct**ag**	TGCTGCT	23	147	TTATTAG	**gt**aagtg
gttac**ag**	ATATGCT	24	91	ATCACTG	**gt**aaggc
tccac**ag**	TCGTCAC	25	3700		

Sequences are displayed at the 3' and 5' splice junctions of the *Ankhzn* gene. Donor splice site and acceptor splice site are shown in bold. Introns are indicated as lower letters and exons are capital letters.

KaRa) and screened by a [^{32}P]-labeled oligonucleotide probe GT3-12-1 that contained exon 1 of *Ankhzn*.

Promoter activity assay

The 5' flanking regions of *Ankhzn* were digested with several restriction enzymes (see Fig. 1. A), and inserted into the pGV-B (Toyo Ink) promoterless luciferase vector. C6 and B103 cells were transfected with the LIPOFECTIN Reagent (LIFE TECHNOLOGIES) in 6-well plates (COSTER 3506) in 2×10^5 cells per well with 3 μg luciferase vector and 1 μg control vector, and pMoSneoIN lacZA containing *LacZ* promoted by Molony murine leukemina virus long-terminal report promoter. The cells were cultured for 48 h. After lysis of the transfected cells, the activity of luciferase was measured on a luminometer according to the manufacturer's instruction. β-Galactosidase activity was measured by a spectrophotometer.

RESULTS

Genomic organization of Ankhzn

The entire *Ankhzn* gene overlapped in these three clones, two P1 clones (18215 and 18216) and one BAC

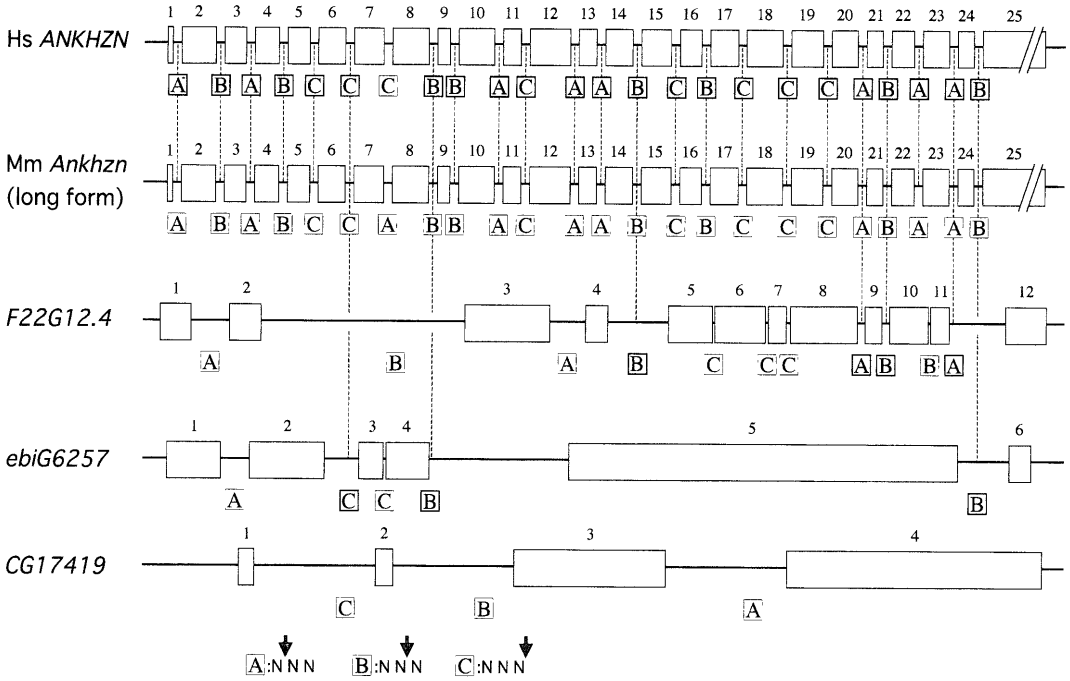

Fig. 3. Genomic organization of the mouse *Ankhzn* and other related genes. The mouse *Ankhzn* gene structure is compared with *F22G. 12.4* from *C. elegans*, *ebiG6257* from *A. gambiae*, *CG17419* from *D. melanogaster*, and the human *ANKHZN* gene. Genomic organization is shown schematically with boxes (exons) and solid lines (introns). Alignment of intron-exon junctions is determined by a search of amino acid homology. Intron and exon junctions conserved among another species against the mouse are indicated by dotted lines. [A], A codon across two exons is divided between the first and second codon position by an intron-exon junction; [B], A codon is divided between the second and third codon positon; [C], A codon is not divided by a junction.

clone (20090). The genomic organization of the *Ankhzn* gene of 25 exons was identified by comparison between sequences of previously reported cDNA[2] and subcloned genomic DNA. The length of the *Ankhzn* gene was estimated to be at least 95 kb by mapping with restriction enzymes, *Bam* HI, *Eco* RI and *Pst* I. Furthermore, a database search identified the *Ankhzn* gene in a mouse cosmid clone, 95,661 bp in length, (accession number AL808023) which confirmed the expected size of the gene. A schematic drawing of the genomic organization of the murine *Ankhzn* gene is given in Fig. 2, and sequences of intron-exon junctions are listed in Table 2. The site of the integration of the GT vector was identified within the first intron by comparison of the genomic DNA sequence of GT3-12 heterozygotes with that of the *Ankhzn* gene. The accession numbers of the Gen Bank were given for each exon from AB098135 to AB098157. Splice sites conform to the GT-AG rule, except for the splice donor site of intron 4, which is GC instead of GT (Table 2). Exon and intron junctions and the GC splice donor site of intron 4 is highly

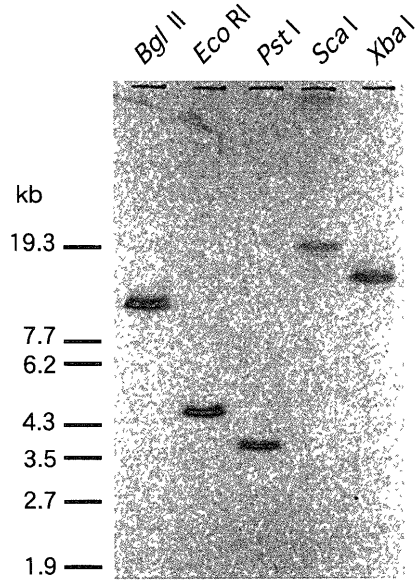

Fig. 4. Southern blot analysis. Ten μg of GT3-12 wild type mouse tail DNA was digested by either *Bgl* II, *Eco* RI, *Pst* I, *Sca* I or *Xba* I, separated by 0.8% agarose gel electrophoresis, and then hybridized with a [^{32}P]-labeled 718 bp PCR fragment from the 3' UTR.

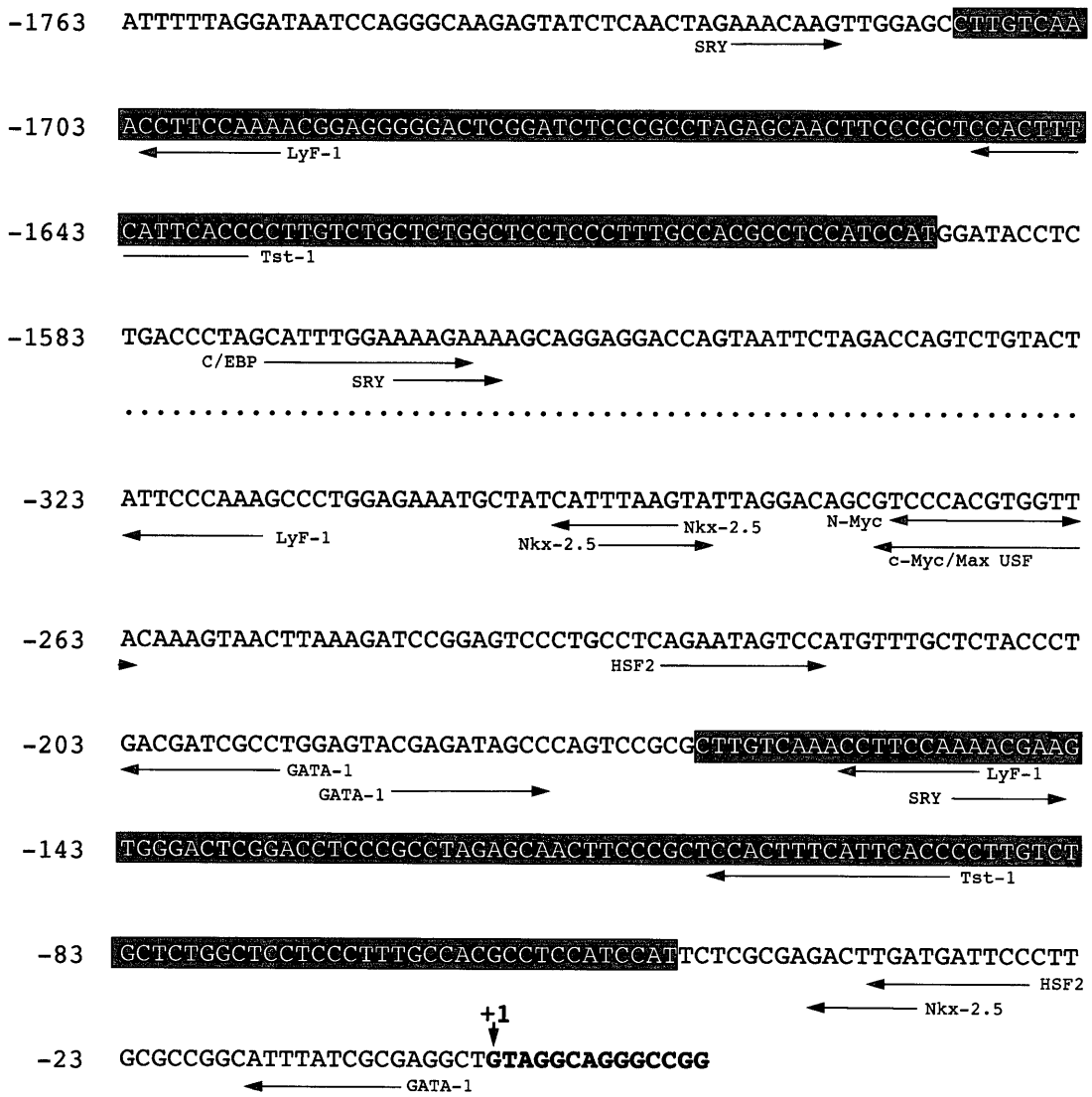

Fig. 5. Nucleotide sequence of the 5' flanking region of the Ankhzn gene and potential transcription factor-recognition sequences. A TF search program was used to predict the transcription factor binding sites. Areas highlighted in black from -1711 to -1597 and from -167 to -49 are highly homologous regions. The transcription start site is indicated by +1, and exon 1 is in bold.

conserved between the mouse *Ankhzn* and the human *ANKHZN* (Fig. 3).

Southern blot analysis

To investigate the mode of *Ankhzn* gene expression, we first examined whether *Ankhzn* is present as a single-or multiple-copy gene. Total mouse genomic DNA was digested with a series of restriction enzymes, *Bgl* II, *Eco* RI, *Pst* I, *Sca* I and *Xba* I, and analyzed with 3' UTR of the cDNA as a probe. Fig. 4 shows that only one hybridization band with each digested fragment was detected by Southern blotting, indicating that there is a single copy of the gene in the mouse genome.

Promoter activity assay

The 5' upstream region of the *Ankhzn* gene was analyzed to determine the essential promoter region. A 1957-bp long fragment (-1943 to +14) 5' upstream to the putative *Ankhzn* gene transcription start site

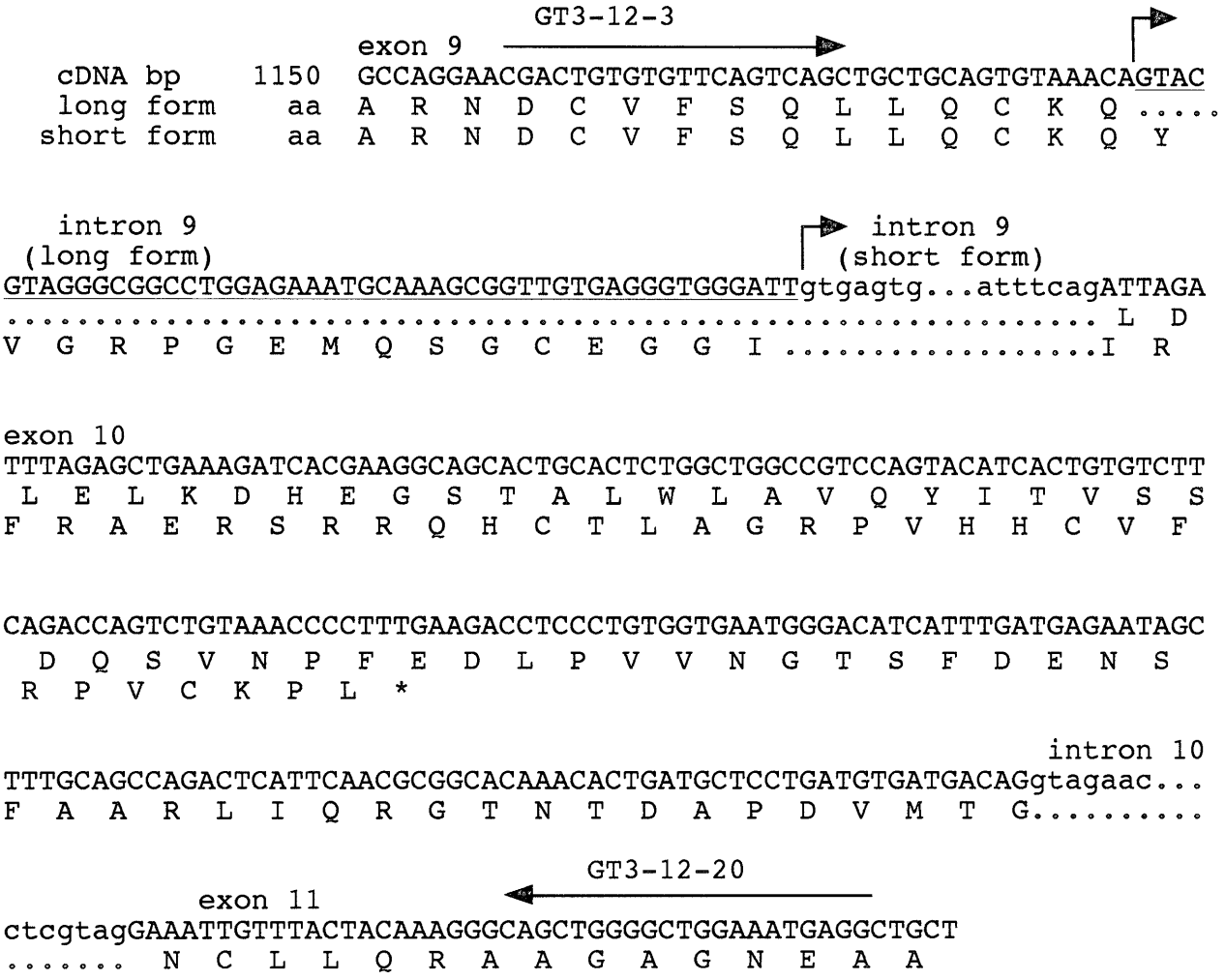

Fig. 6. Nucleotide sequence from exon 9 to 10. Forty-nine bp at the 3' end of exon 9 are alternatively spliced and affect the open reading frame, leading to the formation of long and short protein products.

(+1) was generated by *Apa* I digestion of a *Kpn* I-fragment containing a 5' upstream region and exon 1 of the gene (Fig. 1. A). The following fragments were obtained by digestion of the fragment at various restriction sites and subcloned into the luciferase vector; *Kpn* I / *Apa* I (-1943 to +14), *Xba* I / *Apa* I (-1540 to +14), *Sac* I / *Apa* I (-1236 to +14), *Sma* I / *Apa* I (-807 to +14), *Eco* RI / *Apa* I (-515 to +14), *Pvu* I / *Apa* I (-197 to +14), *Kpn* I / *Eco* RI (-1943 to -512), *Kpn* I / *Pvu* I (-1943 to -200) and *Eco* RI / *Pvu* I (-515 to -200). A β-galactosidase expressing vector was co-transfected to allow corrections for transfection efficiency. There were significant differences in the relative expression level among the constructs and between the two cell lines, but no luciferase activities were observed for those constructs lacking the region of -197 to +14 in both cell lines (Fig. 1. B). These results suggest that this region plays an essential role in the transcription of the murine *Ankhzn* gene. The sequence of the 5' upstream region was analyzed for potential transcription factor-binding sites by the TFSearch program in the Genome Net Database Service, and results are shown in Fig. 5.

Alternative splicing of Ankhzn

During the analysis of the genomic structure of *Ankhzn*, we found some discrepancies between the sequence of cDNA (AB011370) and that of genomic DNA. Detailed examination of the sequence data

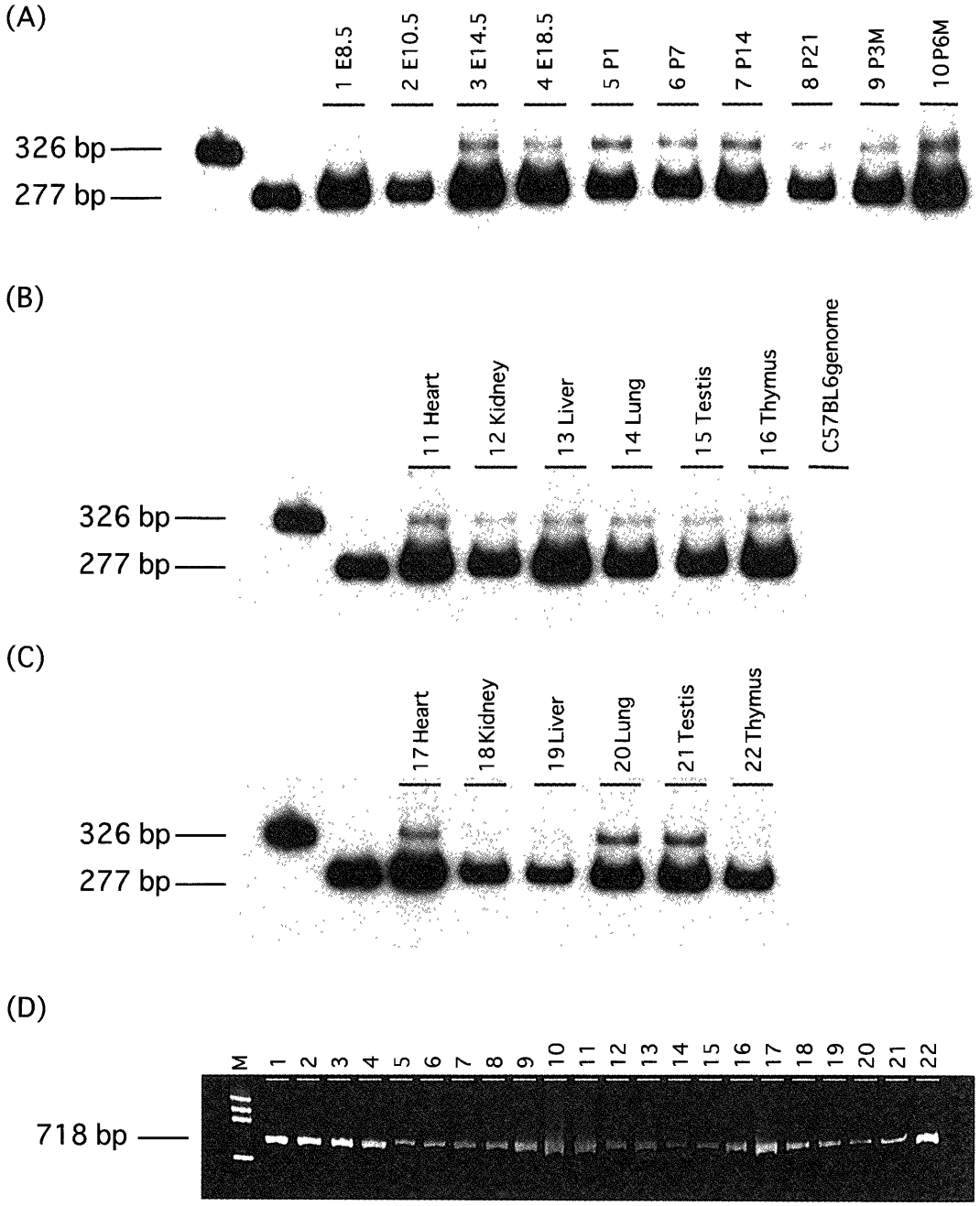

Fig. 7. Results of RT-PCR analysis in ICR mice. (**A**), (**B**) and (**C**) each experiment was performed with Ankhzn specific primer GT3-12-3 and GT3-12-20. The long form is more strongly expressed than the short form. (**A**) Comparison of developmental changes of expression levels between the long form and short form. The whole body and cerebrum were used for RNA extraction from embryos and postnatals, respectively. Cerebrum, heart, kidney, testis, and thymus each of P (postnatal day) 1(**B**) and P 21(**C**). (**D**) Basal expression level of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) in each cDNA pool was examined for normalization among the preparations. Numbers correspond to (**A**), (**B**) and (**C**). Marker used phi x174 *Hae* III digest.

Table 3. Quantitative analysis of expression pattern of long and short forms of Ankhzn splice variants

	1	2	3	4	5	6	7	8	9	10	11
Percentage	0.56	0.57	4.62	2.71	8.94	6.38	8.08	2.83	4.82	10.93	3.23
	12	13	14	15	16	17	18	19	20	21	22
Percentage	0.83	1.31	1.28	1.20	3.28	6.62	0.89	1.59	11.00	11.94	1.01

Intensity of bands in Fig. 7 was digitized and quantified by the software Quantity One. The results (%) are expressed in ratios of short and long form. Numbers correspond to those in Fig. 7A, B and C.

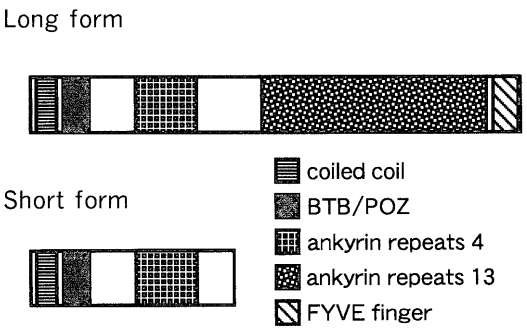

Fig. 8. Domain structure of Ankhzn. The long form encodes 1169 amino acids and its molecular mass is 130 kDa, the domain structure appears as a Coiled-coil, BTB/POZ, 4 and 13 ankyrin repeats and FYVE finger, respectively. The short form encodes 439 amino acids and its molecular mass is 50 kDa, the domain structure shows a Coiled-coil, BTB/POZ, 4 ankyrin repeats, respectively.

suggested the presence of two classes of cDNA clones, which had been erroneously interpreted and combined into a single cDNA sequence with a few sequencing errors. These cDNA clones turned out to be derived from two splice variants by an alternative splicing of the gene. Sequence comparison between *Ankhzn* genomic clones and *Ankhzn* cDNA showed that 3' end 49 bp of exon 9 is alternatively spliced to generate two splice variants, designated as long and short forms. In the long form, this 49 bp segment is not incorporated into mRNA and the resulting cDNA is 7,098-nucleotides long encoding a 1,169-amino acid protein having a molecular mass of 130 kDa. The revised cDNA sequence is slightly shorter than the previously reported one of 7,143 nucleotides encoding 1,184-amino acid residues. When the 49 bp segment was incorporated, the mRNA of 7,147 bp short form encodes a 50-kDa protein of only 439 amino acids because of premature termination by a frame-shift within exon 10 (Fig. 6). The revised cDNA sequences had been deposited in Gen Bank under the accession number of AB011370 for the short form and AB098329 for the long form.

To examine the expression pattern of these alternatively spliced transcripts, we performed RT-PCR experiments on RNA preparations extracted from several mouse tissues and from different developmental stages. First strand cDNAs were synthesized from total RNA isolated from the cerebrum, heart, kidney, liver, lung, testis and thymus from ICR mice, and amplified with GT3-12-3 and GT3-12-20 primers specific for the splicing region. The RT-PCR products were analyzed by Southern blotting using a fragment spanning the splice region as a probe. The expected two bands of 326 bp (short form splicing) and 277 bp (long form splicing) were detected, which means that these bands were actually derived from the splicing region (data not shown). To quantify amounts of the PCR products, we performed RT-PCR experiments with [^{32}P]-labeled dCTP followed by measurement of their intensity of radio-labeled bands. Both long and short forms were expressed ubiquitously, with the long form band being always much more intense than that of the short form in all the tissues examined and in all developmental stages from embryonic day 8.5 to adulthood (Fig. 7). The results were quantified by the data analysis software Quantity One (pdi) and summarized in Table 3.

DISCUSSION

Genomic organization of Ankhzn

Mouse *Ankhzn* comprises 25 exons and 24 introns, and it shares a nearly identical genomic structure with that of human *ANKHZN*[3]. Other than the human counterpart, a database search showed homologous genes in nematodes and insects. Structures of these genes were compared with that of *Ankhzn F22G.12.4* gene of *Caenorhabditis elegans,* which consists of 12 exons and 11 introns (Fig. 3). Four out of 11 intron-exon junctions were identical

with those of mouse *Ankhzn*. An insect homologue, *ebiG6257,* from *Anopheles gambiae,* comprises of 6 exons and 5 introns, and 3 of 5 junctions were conserved in the mouse gene. The other insect homologue from *Drosophila melanogaster* (CG17419) shows even more sequence similarity (49.6%) to mouse *Ankhzn* than the mosquito gene, but its 3 exon-intron boundaries do not match the mouse gene. This appears rather curious as these two insects, a mosquito and a fly, both belong to Diptera and are far more closely related to each other than to the mouse.

Southern blot analysis

Southern blot analysis of the restriction enzyme digests of mouse genomic DNA strongly suggests a single copy number of *Ankhzn* gene. Human *ANKHZN* gene has been mapped to chromosome 17p13 determined by Radiation Hybrid and FISH[3]. This region has high synteny to mouse chromosome 11[8]. The T31 Mouse Radiation Hybrid Database RH Chr 11 Public Map Data in the Jackson Laboratory (http://www.jax.org/resources/documents/cmdata/rhmap/11data.html) locates *Ankhzn* gene on chromosome 11. Furthermore, results of a BLAST search found the *Ankhzn* genome sequences within two genomic clones of RP23-32B6 (accession number AL808023) and RP23-327G1 (accession number AL663082) from the RPCI-23 Mouse PAC Library, both of which had been derived from chromosome 11.

Promoter activity assay

We have identified an essential promoter region within 211 bp upstream from the transcriptional start site. Examination of the 5' upstream sequence revealed several clusters of potential transcription factor binding sites, but we could not find a TATA box within this region. The lack of TATA boxes in the promoter regions is generally recognized as an indication for house-keeping genes[9,10]. *Ankhzn* may be a house-keeping gene judging by the results of Northern and Western blot analyses[2]. One feature in this region is that a 119-bp sequence (-49 to -167) is exactly repeated far upstream (-1593 to -1711), although its significance is not clear.

Alternative splicing of Ankhzn and protein function

In the present study we found a splice variant coding for a much smaller protein (short form). The domain structure of the long form comprises a coiled-coil, BTB/POZ, 17 ankyrin repeats and a FYVE finger, and the short form comprises a coiled-coil, BTB/POZ, 4 ankyrin repeats, respectively (Fig. 8). Results of an RT-PCR analysis indicated that the expression of the long form was always dominant over that of the short form in every tissue and developmental stage. FYVE domains serve to target proteins containing them to particular membrane compartments by binding to phosphatidylinositol 3-phosphate with high specificity, which is found mainly on the endosomal membrane. Proteins containing a FYVE finger domain are distributed among eukaryotes including yeast, Vps27p[11], Vac1p[12], Fab1p[13], Pep7p[14], and vertebrates, EEA1[15], Hrs[16] and SARA[17]. These proteins are involved in vesicular transport, signal transduction and cytoskeletal regulation, and, therefore, FYVE domains are crucial for these vital processes in the cell. Lack of the FYVE domain in the Ankhzn short form suggests a functional divergence between the two splice variants. At least 27 human FYVE finger proteins have been identified in the genome and classified into 14 goups[5], of which three proteins, Vac1p, Fab1p and Vps27p, were also found in yeast (*Saccharomyces cerevisiae*). A search in the complete yeast genome for an *Ankhzn* homologue failed to detect it, and Ankhzn proteins might have some roles specific for multicellular organisms.

As described above, a database search found a few *Ankhzn* homologues of similar domain organizations in C. *elegans,* A. *gambiae* and D. *melanogaster.* We examined the alignment of these homologous proteins to determine whether the domain organization is also evolutionally conserved among them. The amino acid sequence similarity (%) of the Ankhzn homologues to the mouse protein was: F22G12.4 from *C. elegans,* 35.0%; ebiP6257 from A. *gambiae,* 46.4%; CG17419 from *D. melanogaster,* 49.6% and human ANKHZN, 91.0%. All of them have BTB/POZ, ankyrin repeats, and FYVE finger domains except for CG17419, which is devoid of a BTB/POZ domain. The BTB/POZ domain is known as a homodimerization domain present at the N-terminus of proteins containing multiple copies of zinc-fingers, many of which are transcriptional regulators[18,19]. Tandem repeats of ankyrin modules usually function as interacting domains between proteins, and the identification of proteins that interact with Ankhzn protein should be a clue to the function of the protein[20,21].

The strong conservation of the domain structure, genomic organization, and chromosomal localization between mouse *Ankhzn* and human *ANKHZN* suggests that the biological functions of these proteins are essentially identical. Embryonic lethality of the homozygotes of *Ankhzn*-targeted mice indicates criti-

cal roles for the protein. The information on the genomic structure of the murine *Ankhzn* obtained here will allow us to develop a conditional gene targeting strategy, which can bypass embryonic lethality and will enable us to reveal the function of the Ankhzn protein and, hopefully, its relationship to human diseases.

Acknowledgments. This work was supported in part by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to R. K).

REFERENCES

1) Shirai M, Miyashita A, Ishii N, Itoh Y, Satokata I, Watanabe YG, Kuwano R: A gene trap strategy for identifying the gene expressed in the embryonic nervous system. *Zoolog Sci* **13:** 277-283, 1996.
2) Ito K, Ishii N, Miyashita A, Tominaga K, Kuriyama H, Maruyama H, Shirai M, Naito M, Arakawa M, Kuwano R: Molecular cloning of a novel 130-kDa cytoplasmic protein, Ankhzn, containing Ankyrin repeats hooked to a zinc finger motif. *Biochem Biophys Res Commun* **257:** 206-213, 1999.
3) Kuriyama H, Asakawa S, Minoshima S, Maruyama H, Ishii N, Ito K, Gejyo F, Arakawa M, Shimizu N, Kuwano R: Characterization and chromosomal mapping of a novel human gene, ANKHZN. *Gene* **253:** 151-160, 2000.
4) Stenmark H, Aasland R, Toh BH, D'Arrigo A: Endosomal localization of the autoantigen EEA1 is mediated by a zinc-binding FYVE finger *J Biol Chem* **271:** 24048-24054, 1996.
5) Stenmark H, Aasland R, Driscoll PC: The phosphatidylinositol 3-phosphate-binding FYVE finger. *FEBS Let* **513:** 77-84, 2002.
6) Lawe DC, Patki V, Heller-Harrison R, Lambright D, Corvera S: The FYVE domain of early endosome antigen 1 is required for both phosphatidylinositol 3-phosphate and Rab5 binding. Critical role of this dual interaction for endosomal localization. *J Biol Chem* **275:** 3699-3705, 2000.
7) Mills IG, Urbe S, Clague MJ: Relationships between EEA1 binding partners and their role in endosome fusion. *J Cell Sci* **114:** 1959-1965, 2001.
8) O'Brien SJ, Womack JE, Lyons LA, Moore KJ, Jenkins NA, Copeland NG: Anchored reference loci for comparative genome mapping in mammals. *Nat Genet* **3:** 103-112, 1993.
9) Azizkhan JC, Jensen DE, Pierce AJ, Wada M. Transcription from TATA-less promoters: dihydrofolate reductase as a model. *Crit Rev Eukaryot Gene Expr* **3:** 229-254, 1993.
10) Haun RS, Moss J, Vaughan M: Characterization of the human ADP-ribosylation factor 3 promoter. Transcriptional regulation of a TATA-less promoter. *J Biol Chem* **268:** 8793-8800, 1993.
11) Piper RC, Cooper AA, Yang H, Stevens TH: VPS27 controls vacuolar and endocytic traffic through a prevacuolar compartment in *Saccharomyces cerevisiae*. *J Cell Biol* **131:** 603-617, 1995.
12) Weisman LS, Wickner W: Molecular characterization of VAC1, a gene required for vacuole inheritance and vacuole protein sorting. *J Biol Chem* **267:** 618-623, 1992.
13) Yamamoto A, DeWald DB, Boronenkov IV, Anderson RA, Emr SD, Koshland D: Novel PI (4)P 5-kinase homologue, Fab1p, essential for normal vacuole function and morphology in yeast. *Mol Biol Cell* **6:** 525-539, 1995
14) Webb GC, Zhang J, Garlow SJ, Wesp A, Riezman H, Jones EW: Pep7p provides a novel protein that function in vesicle-mediated tranport between the yeast Golgi and endosome. *Mol Biol Cell* **8:** 871-895, 1997.
15) Mu FT, Callaghan JM, Steele-Mortimer O, Stenmark H, Parton RG, Campbell PL, McCluskey J, Yeo JP, Tock EP, Toh BH: EEA1, an early endosome- associated protein. EEA1 is a conversed helical peripheral membrane protein flanked by cysteine "fingers" and contains a calmodulin-binding IQ motif. *J Biol Chem* **270:** 13503-13511, 1995.
16) Komada M, Kitamura N: Growth factor-induced tyrosine phosphorylation of Hrs, a novel 115-kilodalton protein with a structurally conserved putative zinc finger domain. *Mol Cell Biol* **15:** 6213-6221, 1995.
17) Tsukazaki T, Chiang TA, Davison AF, Attisano L, Wrana JL: SARA, a FYVE domain protein that recruits Smad2 to the TGF beta receptor. *Cell* **95:** 779-791, 1998.
18) Bardwell VJ, Treisman R: The POZ domain: a conserved protein-protein interaction motif. *Genes Dev* **8:** 1664-1677, 1994.
19) Chen W, Zollman S, Couderc JL, Laski FA: The BTB domain of bric a brac mediates dimerization in vitro. *Mol Cell Biol* **15:** 3424-3429, 1997.
20) Sedgwick SG, Taylor IA, Adam AC, Spanos A, Howell S, Morgan BA, Treiber MK, Kanuga N, Banks GR, Foord R, Smerdon SJ: Structural and functional architecture of the yeast cell-cycle transcription factor swi6. *J Mol Biol* **281:** 763-775, 1998.
21) Jacobs MD, Harrison SC: Structure of an IkappaBalpha/NF-kappaB complex. *Cell* **95:** 749-758, 1998.