Yonago Acta medica 2002;45:1–8

Genomic Organization of the Human *Arpp* Gene

Akiko Miyazaki*†, Yoshiyuki Tsukamoto*, Kenzo Sato*, Shigetsugu Ohgi† and Masatsugu Moriyama*

**Department of Molecular Biology, and †Second Department of Surgery, Tottori University Faculty of Medicine, Yonago 683-8503 Japan*

A novel ankyrin-repeated protein, *Arpp*, is specifically expressed in skeletal and cardiac muscles. *Arpp* protein is homologous, in its amino acid sequences (52.7% identity), to *Carp* protein which is proposed to be a putative genetic marker for cardiac hypertrophy. In this study, we isolated the human *Arpp* gene by screening a human genomic library and analyzed the genomic structure and its 5' flanking region. The *Arpp* gene was found to encompass a sequence of 11 kb and to consist of 9 exons. The translational initiation site and the stop codon were found to be located at exon 1 and exon 9, respectively. Each exon from 5 to 8 was found to encode 1 of the 4 ankyrin-like domains, respectively. The 2.7 kb upstream of exon 1 was sequenced. The TATA box was identified 29 bp upstream of the transcriptional start site, and multiple putative regulatory elements including the E box and upstream stimulating factor-1 were distributed within the proximal promoter regions. Since these elements were also found in the promoter region of the mouse *Arpp* gene, they may play an important role in the transcriptional regulation of both human and murine *Arpp* genes.

Key words: *Arpp*; *Carp*; E box; heart; skeletal muscle

We recently isolated a novel human gene, *Arpp*, encoding a 43 kDa protein that is characterized by the presence of 4 ankyrin-repeated domains in its central portion (Moriyama et al., 2001). *Arpp* has homologies in its amino acid sequence (52.7% identity), as well as in its size and structural feature, to cardiac ankyrin-repeated protein (*Carp*) which is highly expressed in the heart but rarely expressed in skeletal muscles (Zou et al., 1997). In contrast to *Carp*, human *Arpp* (h*Arpp*) transcripts were found to be expressed both in skeletal and cardiac muscles (Moriyama et al., 2001). Interestingly, expression of h*Arpp* is restricted to the ventricles, but it is very low or undetectable in the atria or large vessels (Moriyama et al., 2001). Furthermore, h*Arpp* expression in the heart is high in the adult but very low in the fetus (Moriyama et al., 2001), suggesting that the expression of h*Arpp* may be regulated developmentally or tissue-specifically. To elucidate the molecular mechanism for the tissue-specific and age-dependent expression of h*Arpp*, it is necessary to isolate and characterize the structure of the h*Arpp* gene and its promoter. In this study, we analyzed the structure of the h*Arpp* gene and its 5' flanking region.

Materials and Methods

Screening of the human genomic library

A human genomic library constructed from normal placenta (Clontech, Palo Alto, CA) was used for screening. Full-length h*Arpp* cDNA

Abbreviations: bp, base pair; *Carp*, cardiac ankyrin repeat protein; cDNA, complementary DNA; DDBJ, DNA Data Bank of Japan; h*Arpp*, human *Arpp*; kb, kilobase; m*Arpp*, mouse *Arpp*; mRNA, messenger RNA; MyHC, myosin heavy chain; PCR, polymerase chain reaction; pfu, plaque-forming unit; SDS, sodium dodecil sulfate; SSC, sodium chloride/sodium citrate; USF, upstream stimulating factor

(1169 bp) labeled with [^{32}P]-dCTP was used as a probe. For the 1st screening, 3.0×10^8 pfu were incubated with K802 host cells, followed by plating on NZY-agar plates. After incubation for several hours, nylon membranes (Hybond-N extra; Amersham Pharmacia BioTech, Tokyo, Japan) were put on the plate. The resulting replica membranes were hybridized overnight at 42˚C with the radio labeled probe. Membranes were washed twice for 30 min with a 0.2-fold amount of sodium chloride/sodium citrate (SSC) containing 0.1% sodium dodecil sulfate (SDS) at 65˚C, followed by autoradiography for 12 h at –80˚C with intensifying screens. The 2nd and 3rd screenings were performed in a similar way. Out of 21 positive clones obtained in the 1st screening, 5 clones survived after the 3rd screening. Phage DNA was extracted and purified using a Qiagen Lambda Kit (Qiagen, Tokyo) which was also used for further studies. The inserted DNA fragments were digested appropriately with restriction enzymes and subcloned into pBlueScript sk(–) phagemid vector (Stratagene, La Jolla, CA). The resulting plasmids were used for sequencing.

Determination of exon/intron boundaries

The exon/intron boundary and the size of each exon and intron were determined by polymerase chain reaction (PCR). Primer pairs used were designed within the introns adjacent to the exon/intron boundaries. Consequently, the resulting DNA fragments contained the whole length of each exon and both its exon/intron and intron/exon boundaries. Primer pairs used to amplify the boundaries were as follows: F1 (5'-CAGTGAGCTCATGGCCAAAGG-3') and R1 (5'-GTAGCA GACAGGATCTCCCT-3'), F2 (5'-TTGGGGGAGAGGCTGCCTGT-3') and R2 (5'-GCCATGGCTACCTGGCTGGT-3'), F3 (5'-GGTCTCCCAAGCCCCCCCAAA-3') and R4 (5'-TGACCCAGGAGCCGAACGGA-3'), F5 (5'-GCTTGGGTGGGAGAT GGGCT-3') and R5 (5'-GTTGGCACAGGTGGCAGG GA-3'), F6 (5'-ACCCCCGAATACTTTCTT CC-3') and R6 (5'-CAGCCAGGGGAATAT CTGTT-3'), F7 (5'-CCCGCCTTCCCAGGG GTACC-3') and R8 (5'-AGGACAGCACCT CCACTGTA-3'), F9 (5'-TGAAAGCCTTCA GGACAGTT-3') and R9 (5'-TCCCTCCTC CCCAAACCAGC-3'), respectively. Ten nanograms of the cloned phage DNA were used as a template for the PCR. PCR conditions were 94˚C for 2 min, followed by 35 cycles of 94˚C for 1 min, 55˚C for 45 s and 72˚C for 1 min and finally incubated at 72˚C for 5 min. Subsequently, the PCR products were subcloned into a pGEM-T Easy Vector (Promega, Madison, WI), followed by sequencing.

Determination of the transcriptional start site

The cap site hunting method was used to determine the transcriptional start site of the h*Arpp* gene. The Cap Site cDNA constructed from the human skeletal muscle was purchased from Nippon Gene (Toyama, Japan), and used as a template for the following PCR. The procedure to generate cap site cDNA was described in detail previously (Moriyama et al., 2001). The primer pair used for the 1st round of PCR was a sense primer complementary to r-oligo, 1RDT (5'-GATGCTAGCTGCGAGTCAAGTC-3') and a reverse primer complementary to h*Arpp* mRNA, Rev1 (5'-GTGCTTCTCATCCTCCAG CAC-3'). The primer pair for the nested PCR was a sense primer complementary to r-oligo, 2RDT (5'-CGAGTCAAGTCGACGAAGTGC-3') and a reverse primer complementary to h*Arpp* messenger RNA (mRNA), Rev2 (5'-GCAAGTCCATGGGCAGCTTC-3'). PCR was performed according to the manufacturer's instructions.

Results and Discussion

*Genomic structure of the h*Arpp *gene*

We 1st screened the human genomic library (Clontech, Palo Alto, CA) using h*Arpp* cDNA as a probe and isolated 21 independent positive clones from a total of 3.0×10^8 pfu. Out of the 21 clones, 9 clones were subjected to a 2nd screening and then 5 out of the 9 clones were

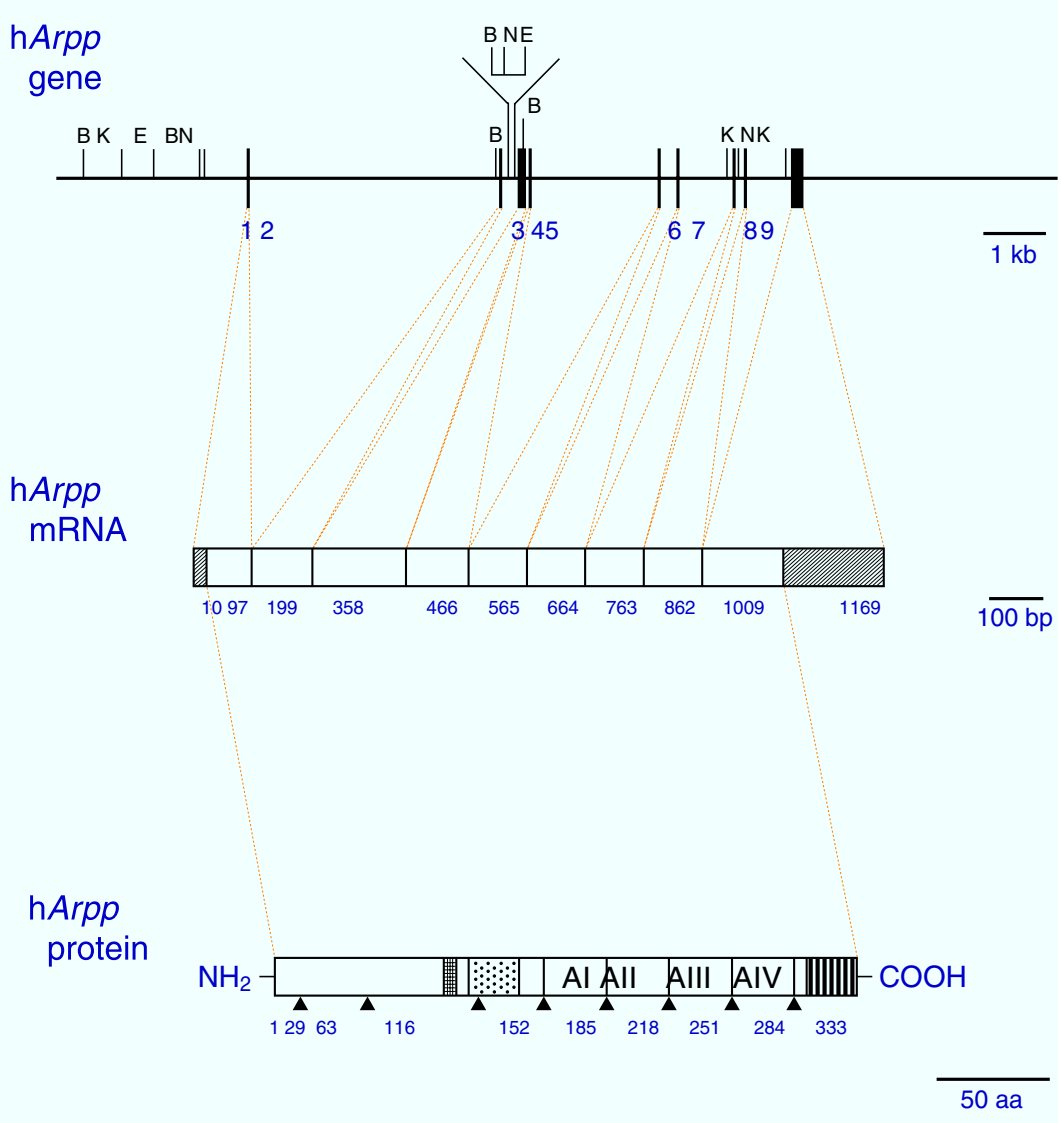

Fig. 1. Structure of the h*Arpp* gene, its transcripts and the protein product. λ-*arpp*1 covers the entire h*Arpp* gene. Exons (1 to 9) of the h*Arpp* gene are indicated by closed boxes with restriction sites; *Bam*HI (B), *Eco*RI (E), *Kpn*I (K) and *Nhe*I (N), respectively. In the schematic representation of the h*Arpp* mRNA, each exon is indicated by an open box and the 5' and 3' untranslated regions are indicated by hatched boxes. In the schematic representation of h*Arpp* protein, the 4 ankyrin repeat domains (AI to AIV), lysine-rich region, PEST sequence and proline-rich region are indicated by open, checked, dotted and striped boxes, respectively. bp, base pair; h*Arpp*, human *Arpp*; kb, kilobase; mRNA, messenger RNA.

subjected to the 3rd screening. The inserts of these clones were analyzed by restriction mapping. As shown in Fig. 1, a clone designated as λ-*arpp*1 carrying a 13 kb insert was further characterized by hybridization with a complementary DNA (cDNA) sequence. By the analysis of 1.1 kb of the cDNA sequence using the DNA Data Bank of Japan (DDBJ) data base, we

Table 1. DNA sequences at the exon/intron boundaries of the h*Arpp* gene

Exon No.	Exon Size (bp)	3' Splice site (acceptor)	5' Splice site (donor)	Intron (bp)
1	97	ctgtggcctgcAGAGGCGGTTATGGA	GAGGAGGAGAATGAG**gt**gcgagcaggggtg	4925
2	102	tctttggatcacc**ag**CAACTCCGAGGAGAC	CTGCAGAAGGTGAAG**gt**aagcctgggagga	338
3	159	ggccatcccgcgc**ag**GGCCAAGAGCGCGTG	GAGCCCGAGGAGATC**gt**aagggtcctgggg	99
4	108	cgtccacatctgc**ag**ACTGGCCCTGTGGAT	GACACGTGCGACCAG**gt**gatgctcctagcc	2249
5	99	cctctctggggac**ag**TTCCGTCGGACAGCA	GACTTCCAGGATCGG**gt**gagtgagagggca	362
6	99	ctcatttctttct**ag**CTGGACTGCACAGCC	AATGTGAGGGATAAG**gt**gaggcaaaaacac	900
7	99	attcctcccaccc**ag**CTGCTGAGCACCCCG	AATGCCAGAGACAGG**gt**gagtgctagcctg	190
8	99	ccacactgactct**ag**GAAGGGGATACTGCC	ATGACCAAGAACCTG**gt**aagctcattccct	873
9	307	tcccgccccctcc**ag**GCAGGAAAGACCCCG	AAAGCTGTTTTTGCTaattgcgatgttcat	

h*Arpp*, human *Arpp*.

found that the cDNA sequence was colinear with the human genomic sequences published in the DDBJ data base (accession number: AL 359388). Based on these data, we determined the precise exon/intron boundaries by PCR using the λ-*arpp*1 DNA as a template. Primers used for the PCR are shown in Materials and Methods. Following subcloning of the PCR products, sequence analysis revealed that the h*Arpp* gene was split into 9 exons distributed over 11 kb (Fig. 1). The nucleotide sequences of the exon/intron boundaries are shown in Table 1. Concerning the structure of the exon/intron junctions, all of them except for exon 9 were consistent with the AG/GT rule (Table 1). Exon 1 is short (97 bp), although it contains a 5' untranslated region and a translational initiation codon. On the other hand, intron 1 is relatively long (4925 bp). Exons 2 to 5 are relatively short, ranging in size from 99 to 159 bp. Each exon from 5 to 8 encodes 1 of the 4 ankyrin-like domains, respectively. Exon 9 (307 bp) contains a stop codon and a 3' untranslated region of mRNA. It also contains the polyadenylation signal AATAAA 136 bp downstream of the stop codon. The sizes of the introns varied from 99 bp to 4925 bp (Table 1). Although the *Carp* gene contains a couple of destabilizing motifs (ATTTA) in the 3' untranslated region (Shaw and Kamen, 1986), these were absent in the 3' untranslated region of the h*Arpp* gene. Further studies will be required to determine whether the presence or absence of this sequence may be involved in the stability of h*Arpp* and *Carp* mRNA.

We recently reported that h*Arpp* is homologous to *Carp* in its amino acid sequences (52.7% identity) and is specifically expressed in the heart and skeletal muscles. Since *Carp* has been proposed to be a putative genetic marker for cardiac hypertrophy, we can easily hypothesize that h*Arpp* may also be involved in the hypertrophy pathway of skeletal muscles as well as cardiac muscles. Interestingly, we found the CAGA motif (CAGAGC) at position +386 to +398. The CAGA motif has been reported to be 1 of the target motifs for transforming growth factor β signaling, which is known to play an important role in cardiac hypertrophy (Kanai et al., 2001).

In addition, we recently isolated a mouse *Arpp* (m*Arpp*) gene and found that the struc-

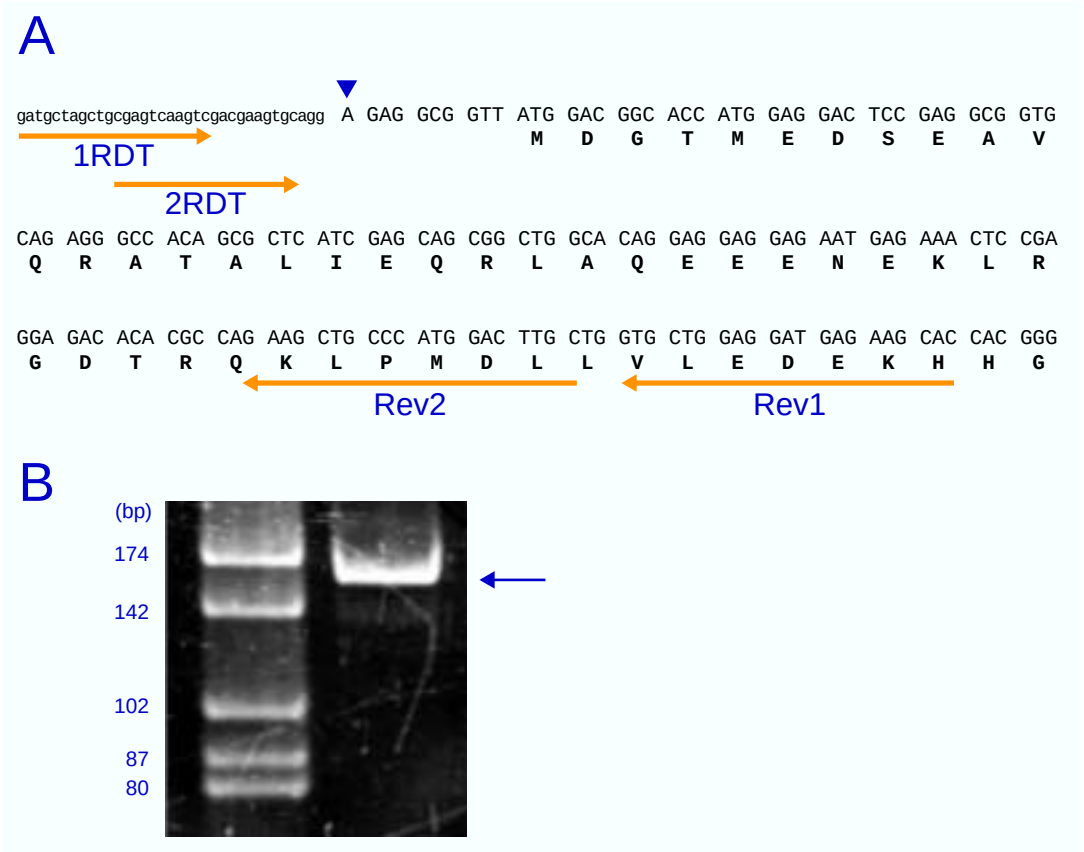

Fig. 2. Determination of the transcriptional start site by the cap site hunting method. **A:** Positions of the primers (1RDT and Rev1, 2RDT and Rev2) are shown by arrows, respectively. The sequence of the synthetic r-oligo ligated to the 5' end of the mRNA is shown in small letters. The putative transcriptional start site is indicated by the vertical arrowhead. **B:** Electrophoresis of the nested-PCR products. Following the 1st and 2nd nested PCRs, the resulting PCR products were electrophoresed. A single band at a size of 150 bp is shown to be detected (arrow). bp, base pair; mRNA, messenger RNA; PCR, polymerase chain reaction.

tural feature of the h*Arpp* gene is well conserved in the m*Arpp* gene.* Comparison of the nucleotide sequences corresponding to the ankyrin repeated domain revealed that h*Arpp* is 88.9 % identical to m*Arpp*, suggesting that these regions are likely to represent a functional domain and play an important role both in humans and mice.

Determination of the transcriptional start site

The transcriptional start site of the h*Arpp* gene was determined by the cap site hunting method as described in Materials and Methods. Briefly, the primary PCR and the secondary nested PCR were performed using 1RDT and 2RDT as the sense primers complementary to r-oligo and Rev1 and Rev2 as the h*Arpp* gene-specific reverse primers (Fig. 2A). The resulting PCR products at a size of 150 bp were subcloned (Fig. 2B), followed by sequencing. We sequenced 6 clones and found that all of them terminated at the same nucleotide at their 5' ends of the h*Arpp* cDNA, clearly showing that it is the major transcriptional start site of the h*Arpp* gene.

* Tsukamoto Y. Unpublished observation.

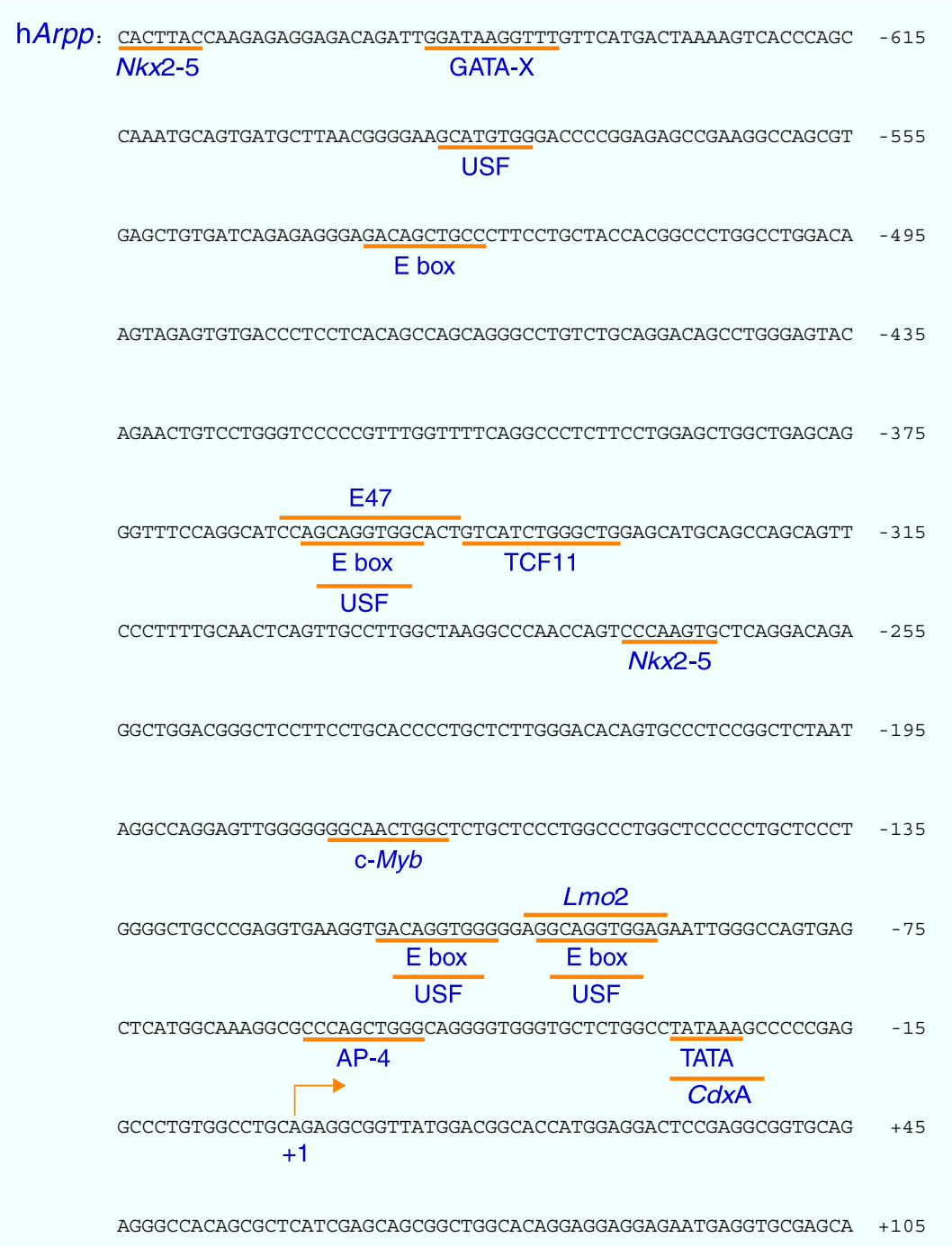

Fig. 3. Sequence alignment of the 5' flanking region of the h*Arpp* gene. The putative binding sites for transcription factors are underlined. The transcriptional start site is shown by the arrow. AP-4, activator protein 4; GATA-X, GATA binding site; h*Arpp*, human *Arpp*; TCF11, transcription factor 11; USF, upstream stimulating factor.

*Sequences of the 5' flanking region of the h*Arpp *gene*

Next, to isolate the 5' flanking region of the h*Arpp* gene, a *Bam*HI/*Bam*HI fragment (–2682 to –775) at a size of 1907 bp and a *Bam*HI/*Bam*HI fragment (–775 to +4771) at a size of 5546 bp were subcloned into pBluescript sk(–) vector. Sequencing of the isolated clones revealed that the TATA box and multiple consensus sites for transcription factors were identified in the 5' flanking region of the h*Arpp* gene (Fig. 3). The TATA box was found at 29 bp upstream of the transcriptional start site (position –29). Several putative regulatory elements including the E box, upstream stimulating factor (USF), c-*Myb*, sex-determining region Y gene product, GATA and *Nkx*2-5 site were identified. Four E box consensus motifs (CNNTG) were identified at –100 to –90, –113 to –103, –359 to –350 and –535 to –524, respectively. The former 3 sites were well conserved in the 5'-flanking region of the murine *Arpp* gene, but the latter did not appear in mice.* It is well known that promoters of muscle specific genes often contain multiple E boxes, and that expressions of these genes are actually regulated by transcription factors that target the E box. Accordingly, it is possible that *Arpp* may also be regulated by transcription factors that target the E box.

Skeletal muscles are known to be classified based on the expression patterns of myosin heavy chain (MyHC) isoforms. Furthermore, the expression of the MyHC isoforms is shown to be regulated under the control of the cis-regulatory elements of the MyHC promoter. We recently found that *Arpp* is preferentially expressed in type 1 skeletal muscle fibers.† This expression pattern was observed not only in humans but also in mice,* suggesting that cis-elements responsible for the specification of the type 1 muscle fiber may exist within the promoter sequences conserved between humans and mice. In addition to the E box, 4 consensus binding sites for USF-1 (–14 to –5, –99 to –92, –112 to –104 and –357 to –351) were also found in the upstream of the h*Arpp* gene. Furthermore, all of the USF sites were completely conserved between humans and mice. It has been reported that the ventricular myosin light chain-2v promoter contains multiple USF sites and is actually transactivated by USF-1 (Harvey and Rosenthal, 1999), leading us to speculate that the USF sites may be involved in the transcriptional regulation of the h*Arpp* gene in the human heart.

Since recent reports have stressed that *Carp* expression is activated during cardiac hypertrophy and that *Carp* is a genetic marker for cardiac hypertrophy (Kuo et al., 1999; Aihara et al., 2000), we can easily hypothesize that *Arpp* may also participate in muscle hypertrophy of the heart and skeletal muscles. Further studies will be required to address this possibility.

Acknowledgments: The authors are grateful to Professor H. Ito of the First Department of Pathology and Professor J. Hasegawa of the Department of Clinical Pharmacology, Tottori University Faculty of Medicine, for their helpful discussion.

* Tsukamoto Y. Unpublished observation.

† Ishiguro Naoko (Dept. of Molecular Biology, Tottori University Faculty of Medicine). Unpublished observation.

References

1 Aihara Y, Kurabayashi M, Saito Y, Ohyama Y, Tanaka T, Takeda S, et al. Cardiac ankyrin repeat protein is a novel marker of cardiac hypertrophy: role of M-CAT element within the promoter. Hypertension 2000;36:48–53.

2 Kanai H, Tanaka T, Aihara Y, Takeda S, Kawabata M, Miyazono K, et al. Transforming growth factor-β/Smads signaling induces transcription of the cell type-restricted ankyrin repeat protein CARP gene through CAGA motif in vascular smooth muscle cells. Circ Res 2001;88:30–36.

3 Kuo HC, Chen J, Ruiz-Lozano P, Zou Y, Nemer M, Chien KR. Control of segmental expression of the cardiac-restricted ankyrin repeat protein gene by distinct regulatory pathways in murine cardiogenesis. Development 1999;126:4223–4234.

4 Moriyama M, Tsukamoto Y, Fujiwara M, Kondo G, Nakada C, Baba T, et al. Identification of a novel human ankyrin-repeated protein homol-

ogous to CARP. Biochem Biophys Res Com 2001;285:715–723.
5 Harvey RP, Rosenthal N, eds. Heart development. San Diego: Academic Press; 1999.
6 Shaw G, Kamen R. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 1986;46:659–667.
7 Zou Y, Evans S, Chen J, Kuo H-C, Harvey RP, Chien KR. CARP, a cardiac ankyrin repeat protein, is downstream in the *Nkx*2-5 homeobox gene pathway. Development 1997;124:793–804.

Received October 12, 2001; accepted October 23, 2001

Corresponding author: Dr. Masatsugu Moriyama