BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **233,** 258–264 (1997)
ARTICLE NO. RC976371

cDNA Cloning and Tissue-Specific Expression of a Novel Basic Helix–Loop–Helix/PAS Protein (BMAL1) and Identification of Alternatively Spliced Variants with Alternative Translation Initiation Site Usage[1]

Masaaki Ikeda[2] and Masahiko Nomura
Department of Physiology, Saitama Medical School, Moroyama, Saitama 350-04, Japan

Received February 13, 1997

Basic helix-loop-helix (bHLH)/PAS proteins, such as Sim, act as transcriptional factors, playing a critical role in the control of central nervous system (CNS) development. To isolate novel bHLH/PAS factors in the CNS an iterative search of a database for expressed sequence tags (ESTs) resulted in the location of several bHLH/PAS protein-like sequences. The rapid amplification of cDNA end (RACE) method was applied to isolate full-length cDNAs of these ESTs. Several 5′ and 3′ terminal sequences were isolated using primers derived from an EST from the human brain cDNA library. The predicted novel factor polypeptide had bHLH and PAS domains that were highly homologous with those of Ah receptor nuclear translocator (Arnt) and Arnt2. Combination of the isolated cDNA fragments revealed the existence of several alternatively spliced variants. The distribution of the novel bHLH/PAS factor message was analyzed by Northern blot hybridization. This detected only one transcript, which was 2.9 kb in size. Strong hybridization was found in the brain, skeletal muscle and heart. Expression of the novel bHLH/PAS factor, brain and muscle Arnt-like protein 1 (BMAL1), was different from that of Arnt and Arnt2, suggesting that BMAL1 has a different function in the CNS and muscle than Arnt and Arnt2.

Transcriptional regulatory factors with a basic amino acid region and a helix-loop-helix region (bHLH) structural motifs, such as *achaete-scute* complex and *daughterless* (1, 2), constitute a superfamily with several constituents that are capable of forming homo- and heterodimers through the HLH (3-5). These factors have been shown to play important and specific roles in the differentiation of specific cell types (6, 7). For example the Achaete protein confers upon ectodermal cells the ability to become neural precursors (8). The bHLH/PAS factors are transcription factors consisting of a bHLH structural motif followed by an approximately 300-amino-acid segment of sequence similarity, called the PAS domain, which is conserved between the *Drosophila* circadian rhythm regulatory protein Per (9, 10), Ah receptor nuclear translocator (Arnt) (11) and the *Drosophila* single-minded (dSim) (12, 13). The PAS domain, as well as the HLH domain, have been shown to be involved in forming hetero- and homodimers among bHLH/PAS transcription factors (14-17). The bHLH/PAS transcription factors have been implicated in various biological events: The AhR/Arnt system has been reported to mediate many biological effects such as epithelial dysplasia, tumor promotion, and immunosuppression (18). The Hif-α/Arnt system has been defined by its capability of mediating hypoxic induction of erythropoietin gene transcription (19, 20). Finally, Sim is necessary for the induction of midline cell fates in the embryonic central nervous system (CNS) (12, 13). To elucidate the function of bHLH/PAS transcription factors in the CNS it is necessary to identify full bHLH/PAS transcription factors expressed in the brain. An iterative search of expressed sequence tags (EST) in the Genbank™ and the TIGR Human cDNA Database (HCD) (21) resulted in the location of several bHLH/PAS protein-like sequences. We applied the rapid amplification of cDNA end (RACE) method to isolate a novel cDNA clone (BMAL1a) using primers derived from sequences of an EST that was expressed in the human brain.

[1] The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases under Accession Nos. D89722 (BMAL1a), AB000812 (BMAL1b), AB000813 (BMAL1c), AB000814 (BMAL1f), AB000815 (BMAL1e), and AB000816 (BMAL1d).

[2] To whom correspondence should be addressed. Fax: +81-492-95-5573. E-mail: mikeda@saitama-med.ac.jp.

Abbreviation: PCR, polymerase chain reaction.

0006-291X/97 $25.00

MATERIALS AND METHODS

Database search. The Genbank and the TIGR HCD (21) were searched for PAS domain sequences derived from mouse AhR, Arnt, Sim and Drosophila Per.

Oligonucleotides. The following oligonucleotides were chemically synthesized and used in the experiments described below: R594-U1, 5′-ACTTCTAGGCACATCGTGTT-3′; R594-D2, 5′-TATTCTACTTCCTTGGTCC-3′; R594-FLANK-EX1U1, 5′-ACCGCAAACGGAAAGGCAGC-3′; R594-FLANK-EX1D2, 5′-CCCGACGCCGCTTTTCAATC-3′; R594-339U1, 5′-GCTGGATCTGGGGTGTAAG-3′; FITC-6081D, 5′-CACTGGAAGGAATGTCTGAGTCCCT-3′; FITC-7063C, 5′-GCCCAAAGAGGACCCACCCCACTGT-3′; FITC-7066C, 5′-AGGCTTAGTTCCACTTTGTCTGAAG-3′; FITC-417-5-273D, 5′-TTTCTTTGAGCAGGTAGAGGGGAAG-3′; FITC-417-6-485U, 5′-CACGACGTTCTTTCTTCTGTAGGA-3′; FITC-R594-496, 5′-CAATCCATACACAGAAGCAAACTA-3′.

Cloning and sequencing of cDNA for human BMAL1. To clone full-length novel PAS protein cDNAs, the Marathon cDNA amplification kit (Clontech) was used according to the manufacturer's instructions. Briefly, the 5′ and 3′ cDNA ends were amplified by polymerase chain reaction (PCR) using possible PAS protein cDNA specific primers (R594-U1, R594-D2), the Marathon cDNA adaptor sequence specific primer AP1, and human brain Marathon-Ready cDNA (Clontech). The 5′ and 3′ RACE products were cloned using the Invitrogen TA cloning kit. After determination of the 5′ and 3′ sequences of the RACE products, full-length cDNAs were amplified by PCR using 5′ and 3′ cDNA terminal-specific primer pairs.

The nucleotide sequence was determined by the fluorescence dye-labeling cycle sequencing system (Amersham) using fluorocein isothiocyanate-labeled primers and an automated DNA sequencer (DSQ1000, Shimadzu). Three independent clones each for 5′ RACE, 3′ RACE, and five independent clones for full-length cDNAs were sequenced to check for PCR errors in the RACE and PCR products. The *ex* Taq polymerase used was purchased from Takara.

Northern blot analysis. Northern blot analysis of the human multiple tissue and adult brain multiple tissue RNA blot (Clontech) was performed according to the instructions of the manufacturer using an 856-base pairs (bp) fragment from the 3′ end of the BMAL1 cDNA probe.

Identification of alternative splicing in BMAL1 using PCR. To investigate the existence of several forms of BMAL1 mRNAs, reverse transcriptase (RT)-PCR was carried out, the first-strand cDNA templates being transcribed with random primers from poly(A)+ RNAs derived from human whole brain (Clontech). The cDNA was amplified using R594-FLANK-EX1U1 (U2) and R594-339U1 (U1) as forward primers, and R594-FLANK-EX1D2 (D1) as reverse primer. The authenticity of these PCR products was confirmed by subcloning of the products into the plasmid pCRII vector (Invitrogen), followed by cycle sequencing. The Avian Myeloblastosis Virus (AMV) RT was purchased from Seikagaku.

RESULTS

An iterative EST database search revealed the existence of a number of PAS factors. In this study, we isolated full-length cDNA coding for a novel PAS factor using the RACE method with primers derived from R59448, which was one of the EST candidates for the PAS factors. Three different-sized fragments from the 5′ RACE and two different fragments from the 3′ RACE were isolated (Fig. 1B). Sequencing analysis revealed that all of these clones contained sequences derived from R59448 and were different in size and in exon-intron boundaries. To obtain full-length cDNAs and examine the exon-intron boundaries, PCR was performed using 5′ and 3′ end-specific primers and the resulting fragments were subcloned into the plasmid pCRII vector. Several clones were obtained, and five randomly selected clones were sequenced. Sequencing analysis revealed that three isoforms existed (Fig. 1B). Clone NC005 was the longest among the isoforms, and was named BMAL1a (Fig. 1,2). The nucleotide sequence around the estimated initiation codon, GCTATGA, matched the consensus sequence reported by Kozak (22) in five out of seven of the nucleotides. The

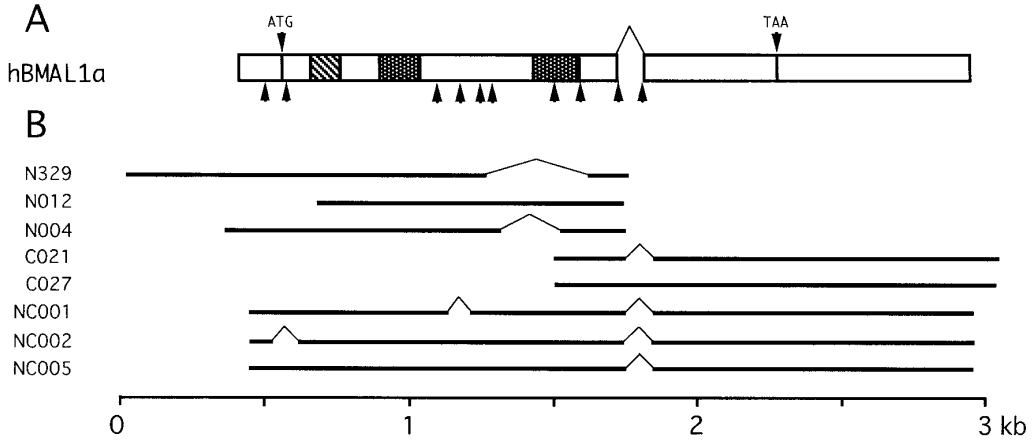

FIG. 1. Brain and muscle Arnt-like protein 1 (BMAL1) cDNAs. A) A schematic diagram of the structure of human BMAL1a (hBMAL1a) is represented by a box. The basic helix-loop-helix domain is represented by a hatched box. The PAS-A and -B domains are represented by shaded boxes. The identified exon-intron boundaries are indicated by arrows below the boxes. The start codon (ATG) and stop codon (TAA) are indicated by arrows above the boxes. B) Multiple alternative splice variants of BMAL1 cDNA isolated by the rapid amplification of cDNA end (RACE) method or polymerase chain reaction (PCR).

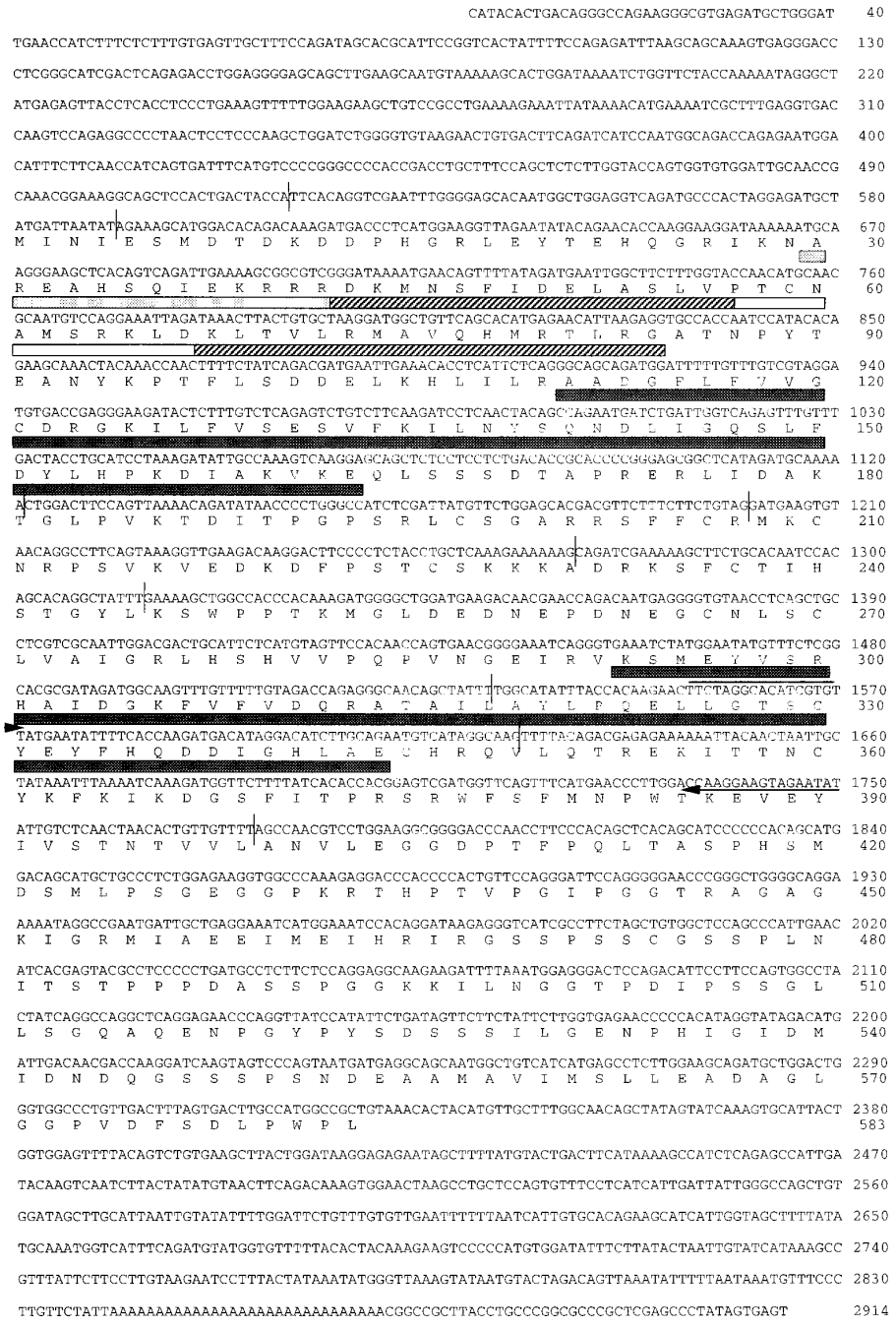

FIG. 2. Nucleotide and amino acid sequences of BMAL1a cDNA. Basic region (▤), helix 1 (▨), loop (□), helix 2 (▨), and two of the repeated sequences (▩) in the PAS domain are shown by the bars under the amino acid sequences. The identified exon-intron boundaries are shown by bars. The oligonucleotides used for RACE are indicated by arrows with a single line.

BMAL1a sequence of 1749 bp corresponded to an open reading frame (ORF) of 583 amino acids and revealed a remarkable similarity with human Arnt (11) and mouse Arnt2 (23) (Fig. 3A). The calculated molecular weight of BMAL1a was 64 kd. The overall sequence identity of BMAL1a with human Arnt was 29%. The BMAL1a had bHLH and PAS domains that were highly homologous with those of Arnt and Arnt2 (Fig. 3A). However, no significant similarity was found in the other regions between BMAL1a and other known bHLH/PAS proteins. Clone NC002 (BMAL1b) did not exhibit the 71-bp exon which contained the start codon of BMAL1a (Fig. 1B, 5). The start codon of the clone without the 71-bp exon was located 140 bp upstream

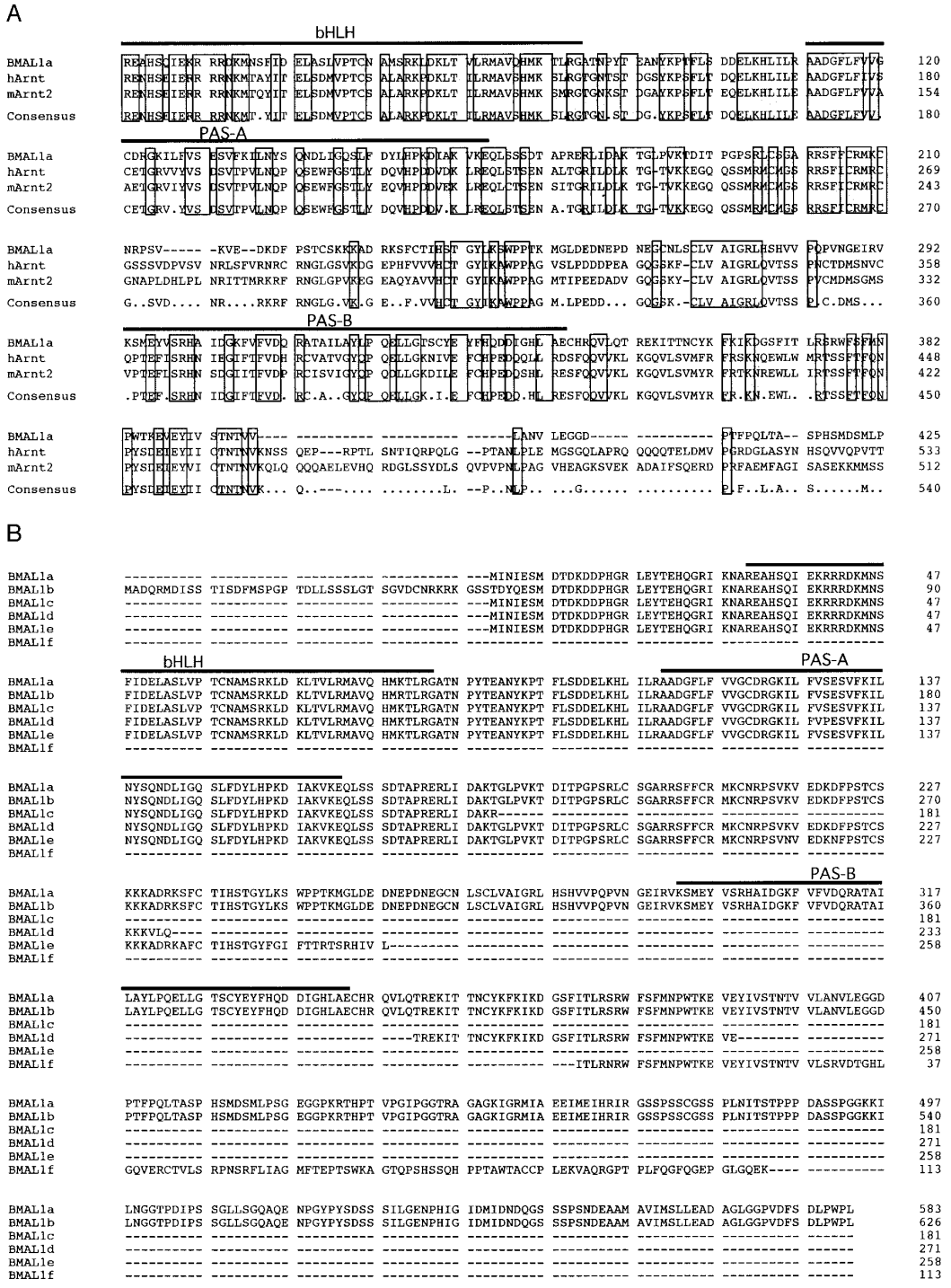

FIG. 3. Predicted amino acid sequences of BMAL1. A) Predicted amino acid sequence of BMAL1a and sequence alignment with human Arnt (hArnt) and mouse Arnt2 (mArnt2) sequences (11, 23) in the bHLH motif and PAS domain. B) Sequence alignment of the deduced amino acid sequences for BMAL1 cDNA isoforms.

of the exon-exon junction (Fig. 2). The other isoform clone, NC001 (BMAL1c), exhibited no 79-bp exon downstream of the PAS-A domain coding sequence (Fig. 1B). The full-length cDNA BMAL1b and BMAL1c contained open reading frames encoding 626 and 181 amino acids, and their calculated molecular weights were 69 and 21 kd, respectively (Fig. 1B, 3B). The predicted amino acid sequence encoded by BMAL1c exhibited no PAS-B do-

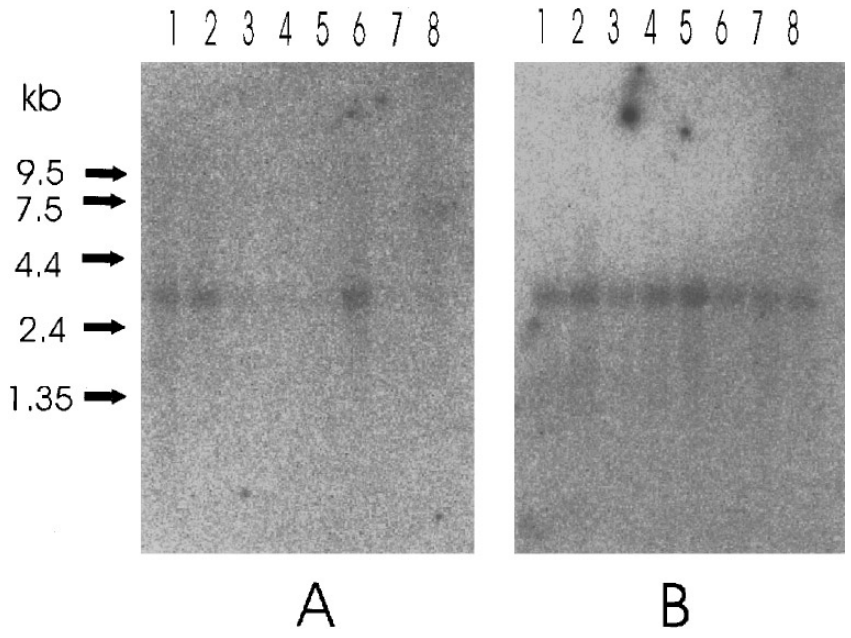

FIG. 4. Expression pattern of BMAL1. A) Human adult multiple tissue RNA blot: 1. heart; 2. brain; 3. placenta; 4. lung; 5. liver; 6. skeletal muscle; 7. kidney; 8. pancreas. B) Human adult brain multiple tissue RNA blot: 1. amygdala; 2. caudate nucleus; 3. corpus callosum; 4. hippocampus; 5. whole brain; 6. substantia nigra; 7. subthalamic nucleus; 8. thalamus. The positions of RNA size markers in kb are shown on the left. The signal migrated at 2.9 kb on each RNA blot.

main (Fig. 1B, 3B). The C′ terminal region sequence of N329 (BMAL1d) and N004 (BMAL1e) have not yet been identified, however, the predicted proteins encoded by N329 and N004 contain no PAS-B domain (Fig. 1B, 3B). Three isoforms of the proteins that lack a PAS-B domain, encoded by BMAL1c, BMAL1d and BMAL1e, may have different properties from the BMAL1a protein.

Following Northern blot analysis of human multiple tissues (Fig. 4A), BMAL1 mRNA was detected only in the brain, skeletal muscle and heart. The expression pattern of BMAL1 mRNA was different from that of Arnt (24) and Arnt2 (23). The length of the BMAL1 transcript(s) was estimated to be about 2.9 kb, which is almost same as that of the isolated cDNA. To examine BMAL1 expression in the brain, Northern blots from several CNS regions (Fig. 4B) were hybridized with a BMAL1 cDNA. BMAL1 mRNAs were expressed in all of the regions that were tested, and particularly high levels of expression were observed in the hippocampus, caudate-putamen and amygdala. Several alternatively spliced isoforms were isolated from human whole brain cDNA. Of these isoforms, the majority were 2.9-kb transcripts.

Using RT-PCR to selectively amplify the 5′ untranslated region (5′ UTR) and the translation initiation region of the BMAL1 mRNA, multiple bands were generated from human brain mRNA (Fig. 5). As shown in Fig. 5, products of the expected sizes of 221 bp and 150 bp, representing BMAL1a and BMAL1b, respectively, were generated (B; lane 2). The authenticity of these PCR products was confirmed by subcloning of the products into the plasmid pCRII vector. Two partial cDNA clones, named clone NN011 and NN013, were isolated. The transcript of the BMAL1b form was the most abundantly expressed in brain tissue.

DISCUSSION

In the study presented here, we isolated a novel human bHLH/PAS factor (BMAL1) and several of its alternatively spliced isoforms from a library of adaptor-ligated double-stranded (ds) cDNAs of whole brain. We determined that its sequence is similar to, but distinct from that of Arnt and Arnt2. The most characteristic region of sequence homology found in BMAL1 and bHLH/PAS proteins has been termed the PAS domain, since it is highly conserved in Per, Arnt, AhR and dSim (13). Although a large number of bHLH proteins have been described to date, the presence of the PAS domain is rare and has been observed in only eight eukaryotic proteins: Per (9,10); Trh (25, 26); AhR (27, 28); Arnt (11); Arnt2 (23); Sim1 (29); Sim2 (30-33), (dSim) (10), and Hif- α (20, 34). The sequences of the PAS domain of this family contain two copies of an approximately 50-amino-acid repeat, referred to as the PAS-A and PAS-B repeats. The BMAL1a and BMAL1b sequences contain these two repeats, whereas BMAL1c, BMAL1d

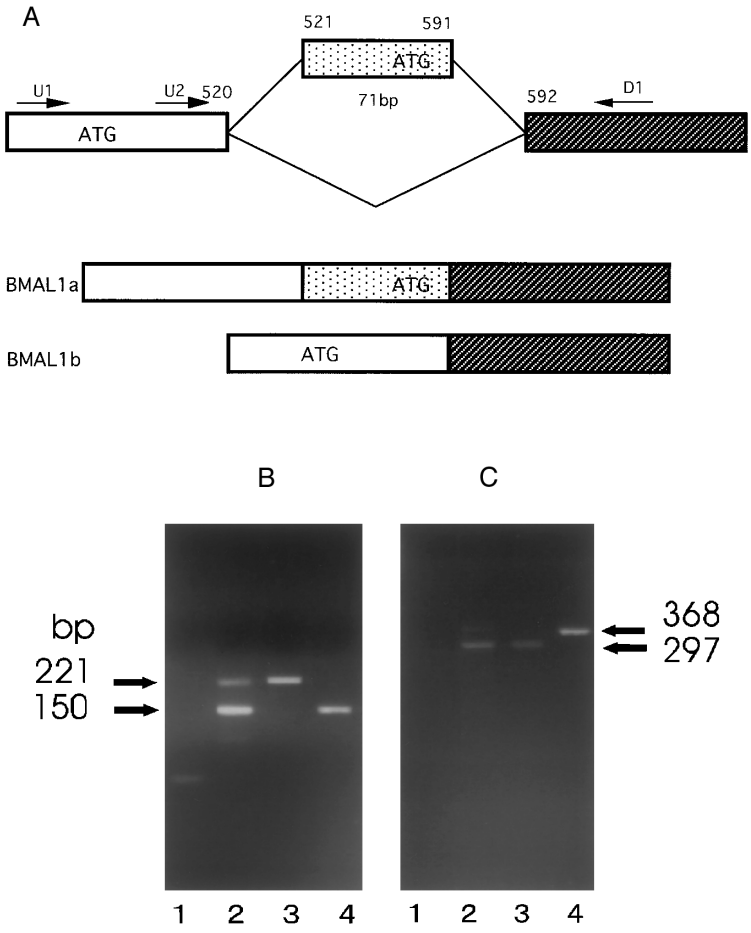

FIG. 5. Reverse transcriptase (RT)-PCR analysis of human BMAL1 isoform mRNA expression in the human brain. The mRNA that was isolated from human brain was subjected to RT-PCR. Two pairs of primers were used to amplify the translation initiation region of BMAL1 cDNA. A) Position of primers which induced amplification of the translation initiation region of BMAL1 cDNA. B) RT-PCR analysis of BMAL1 using the U2 and D1 primers. *Lane* 1, negative control using human brain mRNA not treated with reverse transcriptase, showing absence of any PCR products; *lane* 2, PCR amplification of human brain cDNA, showing the presence of a major 150-bp products, as well as the 221-bp product corresponding to the BMAL1a cDNA; *lane* 3, positive control using clone NC005 plasmid, showing that a 221-bp PCR product is obtained; *lane* 4, positive control (150-bp fragment, amplified from NC002 plasmid). C) RT-PCR analysis of BMAL1 using the U1 and D1 primers. Assay was carried out in absence (*lane* 1) or presence (*lane* 2) of RT. *Lane* 3, positive control using clone NN011 plasmid; *lane* 4, positive control using clone N329.

and BMAL1e have no PAS-B domain (Fig. 1, 3B). The PAS domains appear to be important for protein-protein interactions (17). The deletion of either the A or B segments of the PAS region of Arnt slightly reduces dimerization with AhR (35), suggesting alternative properties of those BMAL1 proteins that lack a PAS-B domain in the interactions with their partners.

BMAL1 cDNAs differ in their 5′ untranslated region and usage of the AUG initiation codon, suggesting that they are generated by alternative splicing of the primary transcript (Fig. 5). The alternative exons predict two isoforms with different initiation codons and lengths. The existence of two isoforms with different initiation codons may be important for protein processing or function.

It is reported that AhR and Arnt form a functional heterodimer in liver cells. The intrinsic partner of the BMAL1 protein is not known. The bHLH/PAS transcription factors are classified into two groups. Group 1, the Arnt group, includes Arnt, Arnt2, BMAL1a and Per. Group 2, the Ahr group, consists of AhR, dSim, Sim1, Sim2, Trh and Hif-α (23). It has been reported that the members of group 1 form homodimers with themselves, as well as heterodimers with group 2 molecules. BMAL1a may be a partner of AhR group proteins such as AhR or Sim2.

BMAL1 is expressed dominantly in the adult brain, skeletal muscle and cardiac muscle. It has been reported that AhR (24) and Sim2 (31) mRNAs are not detectable in adult rat and mouse brains, respectively.

It is not clear what the natural partners of BMAL1 are in the brain. One possible partner molecule of BMAL1 in skeletal muscle is Sim2, since the mRNA of both of these proteins is expressed in that tissue. Investigations toward the elucidation of the interaction between the BMAL1 and bHLH/PAS transcription factors, and finding a natural partner molecule for BMAL1 are in progress.

REFERENCES

1. Villares, R., and Cabrera, C. V. (1987) *Cell* **50,** 415–424.
2. Alonso, M. C., and Cabrera, C. V. (1988) *EMBO J.* **7,** 2585–2591.
3. Murre, C., McCaw, P. S., and Baltimore, D. (1989) *Cell* **56,** 777–783.
4. Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H., and Baltimore, D. (1989) *Cell* **58,** 537–544.
5. Cabrera, C. V., and Alonso, M. C. (1991) *EMBO J.* **10,** 2965–2973.
6. Tapscott, S. J., Davis, R. L., Thayer, M. J., Cheng, P. F., Weintraub, H., and Lassar, A. B. (1988) *Science* **242,** 405–411.
7. Skeath, J. B., and Carroll, S. B. (1994) *FASEB J.* **8,** 714–721.
8. Campuzano, S., and Modolell, J. (1992) *Trends Genet.* **8,** 202–207.
9. Jackson, F. R., Bargiello, T. A., Yun, S.-H., and Young, M. W. (1986) *Nature* **320,** 185–188.
10. Reddy, P., Jacquier, A. C., Abovich, N., Petersen, G., and Rosbash, M. (1986) *Cell* **46,** 53–61.
11. Hoffman, E. C., Reyes, H., Chu, F. F., Sander, F., Conley, L. H., Brooks, B. A., and Hankinson, O. (1991) *Science* **252,** 954–958.
12. Crews, S. T., Thomas, J. B., and Goodman, C. S. (1988) *Cell* **52,** 143–151.
13. Nambu, J. R., Lewis, J. O., Wharton, K. A., and Crews, S. T. (1991) *Cell* **67,** 1157–1167.
14. Reyes, H., Reisz-Porszasz, S., and Hankinson, O. (1992) *Science* **256,** 1193–1195.
15. Matsushita, N., Sogawa, K., Ema, M., Yoshida, A., and Fujii-Kuriyama, Y. (1993) *J. Biol. Chem.* **268,** 21002–21006.
16. Huang, Z. J., Edery, I., and Rosbash, M. (1993) *Nature* **364,** 259–262.
17. Lindebro, M. C., Poellinger, L., and Whitelaw, M. L. (1995) *EMBO J.* **14,** 3528–3539.
18. Swanson, H. I., and Bradfield, C. A. (1993) *Pharmacogenetics* **3,** 213–230.
19. Semenza, G. L., and Wang, G. L. (1992) *Mol. Cell. Biol.* **12,** 5447–5454.
20. Wang, G. L., Jiang, B. H., Rue, E. A., and Semenza, G. L. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 5510–5514.
21. Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult, C. J., Lee, N. H., Kirkness, E. F., Weinstock, K. G., Gocayne, J. D., White, O. *et al* (1995) *Nature* **377,** 3–174.
22. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
23. Hirose, K., Morita, M., Ema, M., Mimura, J., Hamada, H., Fujii, H., Saijo, Y., Gotoh, O., Sogawa, K., and Fujii-Kuriyama, Y. (1996) *Mol. Cell. Biol.* **16,** 1706–1713.
24. Carver, L. A., Hogenesch, J. B., and Bradfield, C. A. (1994) *Nucleic Acids Res.* **22,** 3038–3044.
25. Wilk, R., Weizman, I., and Shilo, B. Z. (1996) *Genes Dev.* **10,** 93–102.
26. Isaac, D. D., and Andrew, D. J. (1996) *Genes Dev.* **10,** 103–117.
27. Burbach, K. M., Poland, A., and Bradfield, C. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8185–8189.
28. Ema, M., Sogawa, K., Watanabe, N., Chujoh, Y., Matsushita, N., Gotoh, Y., Funae, Y., and Fujii-Kuriyama, Y. (1992) *Biochem. Biophys. Res. Com.* **184,** 246–253.
29. Ema, M., Morita, M., Ikawa, S., Tanaka, M., Matsuda, Y., Gotoh, O., Saijoh, Y., Fujii, H., Hamada, H., Kikuchi, Y., and Fujii-Kuriyama, Y. (1996) *Mol. Cell. Biol.* **16,** 5865–5875.
30. Dahmane, N., Charron, G., Lopes, C., Yaspo, M. L., Maunoury, C., Decorte, L., Sinet, P. M., Bloch, B., and Delabar, J. M. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 9191–9195.
31. Ema, M., Suzuki, M., Morita, M., Hirose, K., Sogawa, K., Matsuda, Y., Gotoh, O., Saijoh, Y., Fujii, H., Hamada, H., and Fujii-Kuriyama, Y. (1996) *Biochem. Biophys. Res. Com.* **218,** 588–594.
32. Yamaki, A., Noda, S., Kudoh, J., Shindoh, N., Maeda, H., Minoshima, S., Kawasaki, K., Shimizu, Y., and Shimizu, N. (1996) *Genomics* **35,** 136–143.
33. Moffett, P., Dayo, M., Reece, M., McCormick, M. K., and Pelletier, J. (1996) *Genomics* **35,** 144–155.
34. Wenger, R. H., Rolfs, A., Marti, H. H., Guenet, J. L., and Gassmann, M. (1996) *Biochem. Biophys. Res. Com.* **223** (1), 54–59.
35. Reisz-Porszasz, S., Probst, M. R., Fukunaga, B. N., and Hankinson, O. (1994) *Mol. Cell. Biol.* **14,** 6075–6086.
36. Sogawa, K., Nakano, R., Kobayashi, A., Kikuchi, Y., Ohe, N., Matsushita, N., and Fujii-Kuriyama, Y. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 1936–1940.