Biochemical and Biophysical Research Communications **267,** 317–320 (2000)
doi:10.1006/bbrc.1999.1942, available online at http://www.idealibrary.com on IDEAL®

Sequence, Organization, and Expression of the Human FEM1B Gene

Tereza Ventura-Holman* and Joseph F. Maher†,[1]
*Department of Medicine, Division of Hematology, University of Mississippi Medical Center, and
†G. V. (Sonny) Montgomery Veterans Affairs Medical Center, Jackson, Mississippi 39216

Received November 17, 1999

The FEM-1 protein of *Caenorhabditis elegans* functions within the nematode sex-determination pathway. Two mouse homologs, encoded by the *Fem1a* and *Fem1b* genes, have been reported. We report here the characterization of a novel human gene, designated FEM1B, that is highly homologous to the mouse *Fem1b* gene. FEM1B encodes a protein, designated FEM1*β*, that shows >99% amino acid identity to the corresponding mouse Fem1b protein, including 100% amino acid identity in the N-terminal ANK repeat domain. FEM1*β* represents the first characterized human member of the FEM-1 protein family. The human and mouse genes show conservation of coding sequence and its intron/exon organization, flanking untranslated and genomic sequences, and expression pattern in adult tissues. These findings suggest that there may be evolutionary conservation of regulation and function between the mouse and human FEM1B genes.

The *fem-1* gene of *Caenorhabditis elegans* is one of three genes necessary for all aspects of male development in the nematode (1, 2). The product of this gene, FEM-1, is an ANK repeat protein that functions within the nematode sex-determination pathway (3, 4). Along with the products of the *fem-2* and *fem-3* genes, FEM-1 negatively regulates the *tra-1* gene, which encodes a sequence-specific DNA-binding transcription factor that is necessary for female development (5, 6).

Two mouse genes have recently been characterized, designated *Fem1a* and *Fem1b,* that encode proteins with 36 and 30% amino acid identity, respectively, with *C. elegans* FEM-1 (7). We have recently identified a gene, designated FEM1B (8), that appears to be a homolog of the mouse *Fem1b* gene. We report here the cloning and sequencing of the human FEM1B gene and cDNA, its intron/exon organization, expression patterns, and comparison of these with the mouse *Fem1b* gene.

GenBank Accession Number: AF204883.
[1] To whom correspondence should be addressed at VAMC–Research Service (151), 1500 E. Woodrow Wilson Blvd., Jackson, MS 39216. Fax: (601) 364-1390. E-mail: jmaher@teclink.net.

MATERIALS AND METHODS

Cloning of the FEM1B gene. The dbEST database (http://www.ncbi.nlm.nih.gov/dbEST) at NCBI was searched for human sequences with homology to the mouse *Fem1b.* Human testis clone #781955 was obtained through the IMAGE Consortium (9). This clone was sequenced and its 600 bp insert used as a probe to identify the human FEM1B gene in a bacterial artificial chromosome (BAC) library (Genome Systems, Inc.). The BAC clone was characterized by restriction analysis and Southern blotting, and a 7.5-kb *Eco*RI fragment was subcloned into pBluescriptKS(+) (Stratagene). A full-length mouse *Fem1b* cDNA (7) was also used as a probe and identified a 6.5-kb *Sph*I fragment not detected with the human probe. This fragment was subcloned into pGEM-5Zf(+) (Promega).

Cloning of the FEM1B cDNA. Five microliters of human testis Marathon-Ready cDNA (Clontech) was used in a PCR with primers derived from the 5′ end and 3′ end of the putative open reading frame of FEM1B, respectively: m/hF-B(5′*Bam*HI) (5′-CTCGGGATCCCC-ACCATGGAGGGCCTGGCTGGCTATG-3′); and m/hF-B(3′*Eco*RI) (5′-CCGGAATTCATTAATGAAATCCAACAAACTCTTCAAG-3′). PCRs were performed in a 50-μl reaction volume containing PCR buffer, 20 ng of DNA, 200 μM of each dNTP, 0.4 μM of each primer, and 2.5 units of *Taq* DNA Polymerase (Boehringer-Mannheim). PCR conditions were an initial denaturation step at 94°C for 6 min, followed by 30 cycles of 94°C for 1 min, 65°C for 1 min, and 72°C for 2 min. After an additional elongation step at 72°C for 7 min, the reactions were terminated by incubation at 4°C. The product obtained in this reaction was diluted 1:500 and 1 μl of the dilution was subjected to a second round of PCR using the same primers and under the same PCR conditions. The product of the second round of PCR was digested with *Bam*HI and *Eco*RI, gel purified and subcloned into *Bam*HI/*Eco*RI-digested pBluescript KS plasmid.

DNA sequencing. Sequencing was done in our laboratory by the dideoxy-chain termination method (10) using Sequenase version 2.0 (U.S. Biochemical Corp.), or at the Molecular Biology Core Facilities of the Dana-Farber Cancer Institute. All sequences were confirmed in both DNA strands. Primers used in sequencing were vector primers (T3, T7, M13 universal and reverse), as well as gene-specific primers: hF-B#2F (5′-CATTGTGGAGCCACAGCATTG-3′); hF-B#2R (5′-CAATGCTGTGGCTCCACAATG-3′); hF-B#3F (5′-CAG-GAACACCCACAAGGATC-3′); hF-B#3R (5′-GATCCTTGTGGGT-GTTCCTG-3′); hF-B#4F (5′-CATTAGCCTAGTTGAAGCCG-3′); hF-B#4R (5′-CGGCTTCAACTAGGCTAATG-3′); hB-in1.1R (5′-

0006-291X/00 $35.00

AGGGCCCGTGAATTAACGAGTC-3′); hB-in2.1F (5′-CATATGTTAG-GCTGTGGCCTC-3′); hB-5′U#1F (5′-CTGTGAGGGCCCAGGTTTA-AAG-3′); hB-3′G#1R (5′-AGGGTAGTGGAAAGAAAAGCAG-3′).

Sequence alignments. Amino acid and nucleic acid alignments and identity/similarity scores were done using ALIGN Plus, Version 2.0 (Scientific and Educational Software, Inc.).

Northern blot analysis. Human Multiple Tissue Northern (MTN) Blots containing 2 μg of poly(A)$^+$ RNA per lane were purchased from Clontech: Human Blot (lot #63344), Blot III (lot #8070150), Blot IV (lot #54426), Brain II (lot #8070634), and Brain IV (lot #8080221). Human Normal Blot III (lot #7901034), containing 20 μg of total RNA, was purchased from Invitrogen. The blots were hybridized to radiolabeled probes in ExpressHyb buffer (Clontech) and washed under high stringency conditions (0.2× SSC/0.1% SDS, 50°C).

RESULTS AND DISCUSSION

Cloning and sequence analysis of the FEM1B gene and cDNA. A search of the dbEST database at NCBI for human sequences with homology to the mouse *Fem1b* gene resulted in the isolation of a human testis clone containing a 600-bp cDNA fragment (EST #781955). This EST clone did not represent a full-length cDNA, but rather appeared truncated at the 5′ end. An alignment of this human EST showed 95% nucleotide sequence identity to the 3′ portion of the mouse *Fem1b* cDNA; this homology extended past the stop codon and included the 3′ UTR. To identify a genomic clone containing this gene, the partial human testis EST was used as a probe to screen a human BAC genomic library. In the BAC clone obtained in this screen, the EST probe identified a 7.5-kb *Eco*RI fragment and a 5.5-kb *Sph*I fragment. When a full-length mouse *Fem1b* cDNA was used as a hybridization probe, it identified an additional *Eco*RI fragment of approximately 12 kb, as well as a second *Sph*I fragment of 6.5 kb. These results suggested that the coding region homologous to the 3′ exon of the mouse *Fem1b* gene is located in the 5.5-kb *Sph*I fragment, while the 6.5-kb *Sph*I fragment contains the coding region homologous to the 5′ end of the mouse *Fem1b* gene. The 7.5-kb *Eco*RI fragment and both SphI fragments were subcloned and sequenced. Primers derived from the 5′ end and 3′ end of the putative open reading frame (ORF) were used for PCR from a human testis cDNA library, yielding a single product of 1.8 kb. This PCR product was sub-cloned and sequenced to verify continuity of the coding region in the mRNA and confirm intron/exon boundaries (see below).

The sequencing data obtained revealed that the ORF of the FEM1B gene comprises 1,884 nucleotides and exhibits 90% nucleotide sequence identity to the mouse *Fem1b* coding region. FEM1B encodes a protein, designated FEM1β, which has greater than 99% amino acid identity to the corresponding mouse Fem1b protein, with only six amino acid differences over a total length of 627 amino acids (Fig. 1). There is 100% identity of the N-terminal 269 amino acids of both proteins, which includes the domain of seven sequential ANK

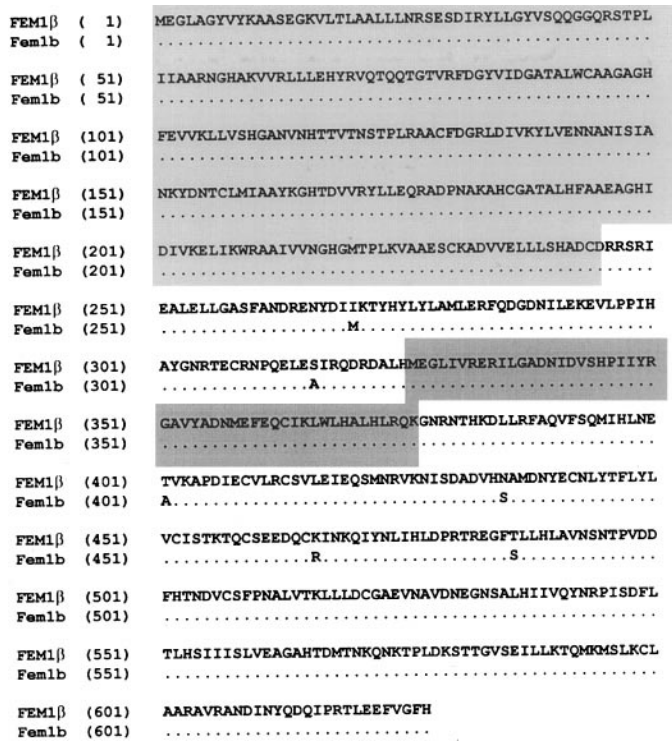

FIG. 1. Amino acid sequence alignment of the encoded human FEM1β and mouse Fem1b proteins. Single-letter amino acid code is used, and only amino acid differences in the mouse protein are shown, with dots (. . .) representing amino acid identities. The light-gray shaded box at the N-terminus of the proteins highlights the ANK-repeat domain, consisting of seven sequential ANK repeats. The medium-gray shaded box in the central region of the proteins highlights a 51-amino-acid sequence with homology to Kinesin light chains.

repeats (11). A high degree of conservation of ANK repeats has been observed in the mouse Fem1a and Fem1b proteins relative to each other and to the *C. elegans* FEM-1 protein (7). The central region of the mouse Fem1a and Fem1b proteins contains a 51-amino-acid sequence that has been noted to be highly conserved between them and the *C. elegans* FEM-1 protein, and to have homology to Kinesin light chains (7, 12). This amino acid sequence is 100% conserved between mouse Fem1b and human FEM1β. There is also 100% identity of the C-terminal 140 amino acids of mouse Fem1b and human FEM1β. This includes a 10-amino-acid C-terminus motif which we had noted to be conserved between mouse Fem1a, Fem1b, and *C. elegans* FEM-1, and which has the sequence P xx L xx F xxx H [where x = variable amino acid] (7).

As is the case with mouse *Fem1b,* the FEM1B coding sequence is divided into two exons. The intron interrupts the coding sequence at identical sites in the two species (Fig. 2). The intron sequence adjacent to the splice-donor site is very conserved, with the first 15 nucleotides of the intron being identical in the mouse and human sequences. The overall nucleotide sequence

FIG. 2. Conservation of intron/exon boundaries between human FEM1B and mouse *Fem1b.* (A) Sequence alignment of the proximal intronic sequence at the splice donor site (indicated by the arrowhead) shows identity in the first 15 nucleotides. The nucleotide sequence homology then decreases to 74% over the first 150 nucleotides of the intron. (B) The 36 nucleotides immediately upstream of the acceptor splice site (indicated by the arrowhead) share 66% nucleotide sequence homology between human and mouse.

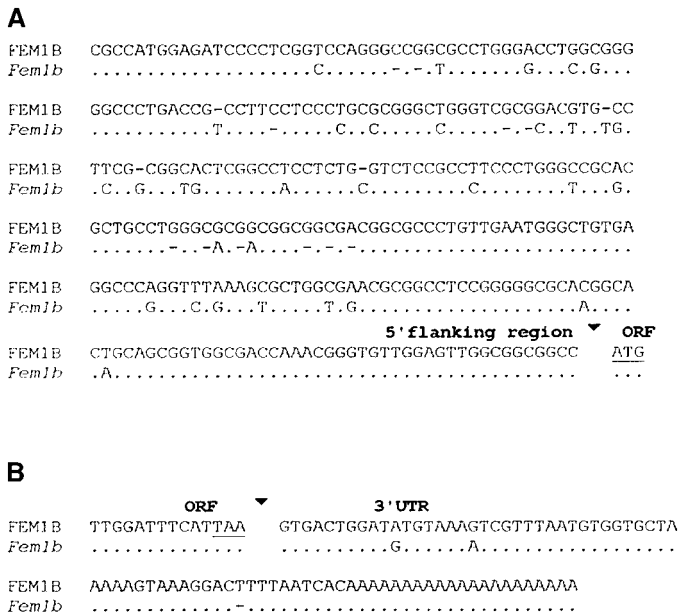

FIG. 3. Sequence conservation between mouse and human Fem1b genes at the 5′ flanking region and 3′ untranslated region (UTR). (A) The 295 nucleotides immediately upstream of the predicted initiation codon (underlined) exhibit 87% nucleotide sequence homology. An arrowhead separates the 5′ flanking region from the start of the open reading frame (ORF). (B) The 62 nucleotides of the 3′ UTR of human FEM1B exhibit only three nucleotide differences from the mouse Fem1b 3′ UTR. The termination codon is underlined. An arrow separates the end of the ORF from the 3′ UTR.

homology in the first 150 nucleotides of the intron is 74% (Fig. 2A). The sequence adjacent to the intron splice-acceptor site is not as well conserved—a comparison of 36 nucleotides immediately upstream of the acceptor site reveals 66% nucleotide sequence homology (Fig. 2B).

The sequence conservation between the two species extends to sequences flanking the ORF. Genomic DNA sequence 295 nucleotides immediately upstream of the translation-initiation codon is 87% conserved (Fig. 3A). The 5′ untranslated region (UTR), transcription start site(s) and promoter(s) of FEM1B remain to be elucidated. As determined for mouse *Fem1b,* the sequence surrounding the initiator Methionine residue (CGGCCAUGG) is a good match with the Kozak translational start site consensus sequence (CC[A/G]CCAUGG) (7, 13). The 3′ UTR of FEM1B comprises 59 nucleotides, with only three nucleotide differences between it and the mouse *Fem1b* 3′ UTR sequence (found in four of six mouse testis cDNA clones previously studied—an additional four nucleotides in the 3′ UTR preceding the poly(A) tract were observed in two of six mouse cDNA clones—see Ref. 7). The sequence AGUAAA is conserved and found 14 nucleotides upstream of the poly(A) tract in FEM1B mRNA (Fig. 3B). This sequence, which deviates by one nucleotide from the consensus hexanucleotide polyadenylation signal AAUAAA, may be functional in directing 3′ end cleavage and polyadenylation (14, 15).

Tissue-specific expression of the FEM1B gene. Northern blots of poly(A)$^+$ mRNA from adult human tissues reveal abundant expression in testis, with a predominant message in this tissue of apparent size 1.6 kb, as well as a fainter message of 2.8 kb (Figs. 4B and 4C). In addition, low levels of two transcripts of 7.5 and 5.5 kb were observed in most other tissues analyzed (Figs. 4A and 4B). It is not known at present if these large messages represent alternatively spliced or processed FEM1B transcripts. Although gender differences in expression in the somatic tissues of adults have not yet been examined, FEM1B is much more highly expressed in testis than in ovary (Fig. 4C).

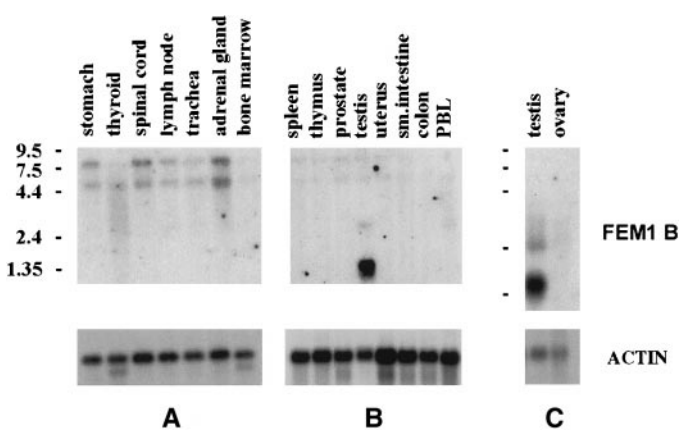

FIG. 4. Pattern of expression of FEM1B in multiple human adult tissues determined by Northern blot analysis. Each lane contains 2 μg of poly(A)$^+$ mRNA. Individual tissues are noted over each of the lanes. RNA size markers in kilobases are indicated on the left margin. Actin probe blot for the control of loading is shown below each blot.

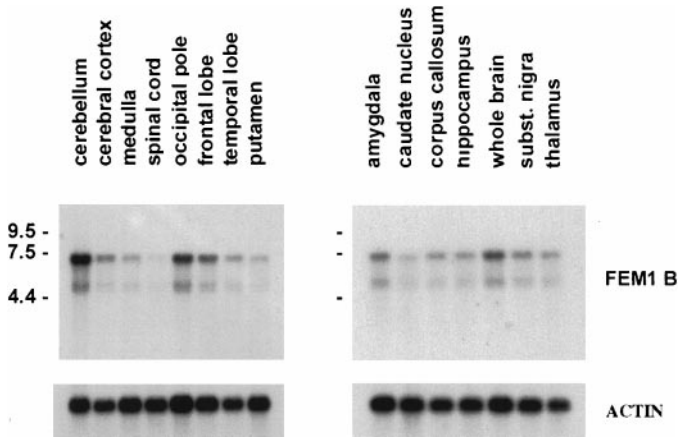

FIG. 5. Pattern of expression of FEM1B within different regions of human adult brain. Each lane contains 2 μg of poly(A)$^+$ mRNA. Individual brain regions are noted over each of the lanes. RNA size markers in kilobases are indicated on the left margin. Actin probe blot for the control of loading is shown below each blot.

Within the brain, FEM1B is expressed in all regions examined, although transcripts were especially abundant in cerebellum (Fig. 5). The high level expression of FEM1B in discrete adult human tissues such as testis and cerebellum suggests that it, like its mouse *Fem1b* counterpart, may have more tissue-specific functions compared to the more global role of *fem-1* in sex-determination in *C. elegans* (1, 4, 7).

In summary, we have characterized the full coding sequence of the human FEM1B gene, its intron/exon organization, flanking genomic sequences, and expression pattern in adult tissues. Comparison with the mouse *Fem1b* gene demonstrates that there is a high degree of conservation—of amino acid sequence, coding and flanking sequences, intron/exon organization, and adult expression pattern—between the two species. This suggests that the human and mouse genes may share evolutionarily conserved modes of regulation and physiologic function. The nematode *C. elegans* is at present the only organism that has provided functional information on the *fem-1* gene family. The studies reported here suggest that functional information obtained on the human FEM1B gene will give insight into the mouse *Fem1b* gene, and vice versa, allowing studies in both systems to contribute to a general understanding of the function of this novel gene family in mammals.

ACKNOWLEDGMENTS

This work was supported by a University of Mississippi Medical Center–Biomedical Research Grant (J.F.M.) and a Career Development Award from the Department of Veterans Affairs (J.F.M.).

REFERENCES

1. Doniach, T., and Hodgkin, J. (1984) *Dev. Biol.* **106,** 223–235.
2. Hodgkin, J. (1987) *Annu. Rev. Genet.* **21,** 133–154.
3. Spence, A. M., Coulson, A., and Hodgkin, J. (1990) *Cell* **60,** 981–990.
4. Gaudet, J., VanderElst, I., and Spence, A. M. (1996) *Mol. Biol. Cell* **7,** 1107–1121.
5. Zarkower, D., and Hodgkin, J. (1992) *Cell* **70,** 237–249.
6. Zarkower, D., and Hodgkin, J. (1993) *Nucleic Acid Res.* **21,** 3691–3698.
7. Ventura-Holman, T., Seldin, M. F., Li, W., and Maher, J. F. (1998) *Genomics* **54,** 221–230.
8. Maher, J. F., and Ventura-Holman, T. (1998) *Am. J. Hum. Genet.* **63**(S), A186.
9. Lennon, G. G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996) *Genomics* **33,** 151–152.
10. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
11. Bork, P. (1993) *Prot. Struct. Funct. Genet.* **17,** 363–374.
12. Cyr, J. L., Pfister, K. K., Bloom, G. S., Slaughter, C. A., and Brady, S. T. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 10114–10118.
13. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
14. Sheets, M. D., Ogg, S. C., and Wickens, M. P. (1990) *Nucleic Acids Res.* **18,** 5799–5805.
15. Wahle, E., and Keller, W. (1992) *Annu. Rev. Biochem.* **61,** 419–440.