Eur. J. Biochem. **262**, 755–764 (1999) © FEBS 1999

Complete nucleotide sequence, origin of isoform and functional characterization of the mouse hepsin gene

Shunsuke Kawamura, Sumiko Kurachi, Yoshihiro Deyashiki and Kotoku Kurachi

Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, USA

Hepsin, a type-II membrane-associated serine protease, has been implicated in cell growth and developement as well as possible initiation of blood coagulation. Here, we report on the complete nucleotide sequence, functional characterization of key structural features and the promoter of the mouse hepsin gene. The gene has a size of ≈17 kb, and is composed of 12, 13, or 14 exons depending on alternative intron splicings – one in the 5′-UTR and the other two in the second intron. The latter two, which occur in approximately half of the hepsin transcripts, generate a hepsin mRNA species with an extra exon, which is responsible for producing a hepsin isoform with a unique 20-residue sequence inserted in the cytoplasmic portion of hepsin. Most hepsin transcripts have the 5′-UTR intron spliced, and its splicing can occur independently of the other alternative splicings. The transcriptional initiation site was determined to be 636 bp upstream of the first ATG site in a cytidine-rich region. The 5′-flanking region of hepsin up to nucleotide 274 showed a substantial promoter activity in HepG2 cells, with its expression activity sevenfold higher in the presence of the 5′-UTR intron sequence in comparison to that without the intron sequence. The basal promoter region contains potential binding sites for several transcription factors including SP1, AP2, C/EBP, LF-A1, and E box, which may be responsible for ubiquitous, but liver- and kidney-preferred tissue expression of the hepsin gene.

Keywords: membrane protease; gene organization; alternative splicing; promoter.

Hepsin, a serine protease, was originally found as cDNA clones isolated from a human liver cDNA library [1,2]. Subsequent studies determined that hepsin is a type-II membrane-associated protease of ≈50 kDa with its carboxyl-terminal-half proteolytic subunit at the cell surface [2–4]. Hepsin is also present in subcellular organelle fractions such as nucleus and mitochondria [3]. Recently, the rat hepsin cDNA has been determined [5]. Rat hepsin is composed of 416 amino-acid residues, one residue shorter than human hepsin. Hepsin has been implicated in cell growth [6], development [7], and more recently in cell-surface initiation of blood coagulation [8] as well as possible prostate cancer cell function [9]. These observations strongly suggest that hepsin may play many important roles. However, homologous recombination-based gene inactivation of mouse hepsin showed that deficiency of hepsin does not result in any severe abnormal phenotypes, other than a significant increase in the level of alkaline phosphatase [10]. Unexpectedly, an observed mild hepsin-deficiency phenotype in mice may be due to the possible presence of one or more complementing systems. This situation appears similar to other reported cases where gene inactivation in mice does not necessarily agree with the phenotypic changes predicted or observed with the deficiency of such genes in humans [11–13]. As no human hepsin deficiency is known of to date, it is difficult at the present time to explain the phenotype observed in hepsin knockout animals.

Correspondence to K. Kurachi, Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, 48109-0618, USA. Fax: + 01 734647 3158; E-mail: kkurachi@umich.edu

Abbreviations: DMEM, Dulbecco's modified Eagle's medium; hFIX, human factor IX; RT-PCR, reverse transcription-polymerase chain reaction.

(Received 25 January 1999, revised 23 March 1999, accepted 23 March 1999)

In the present study, we describe the complete nucleotide structure, characteristics of key structural elements, characterization of alternative splicings and functional analyses of the promoter region of the mouse hepsin gene.

EXPERIMENTAL PROCEDURES

Materials

A λ FIXII mouse genomic library constructed with liver DNA from mouse (strain 129) was purchased from Stratagene. Restriction enzymes, calf intestinal alkaline phosphatase, and T_4 DNA ligase were purchased from New England Biolabs. T_4 polynucleotide kinase was obtained from United States Biochemical. Qiagen plasmid kit was purchased from Qiagen Inc. Taq polymerase, SuperScript™II Reverse Transcriptase, SuperScript One-Step™ reverse transcription-polymerase chain reaction (RT-PCR) system, RNase H, pUC18, pBluescript, agarose, Dulbecco's modified Eagle's medium (DMEM), fetal bovine serum, and sheared herring sperm DNA were purchased from GIBCO Life Technologies. DNase I and Fugene™ 6 transfection reagent were obtained from Beohringer Mannheim. β-Galactosidase expression plasmid vector (pCH110) was obtained from Pharmacia P-L Biochemical. Vitamin K (Aqua-Mephyton) was from Merck Sharp & Dohme. TA cloning kit was from Invitrogen. Radioactive nucleotides ([α-^{32}P]dCTP, [γ-^{32}P]ATP, and [α-^{33}P]dNTP), Megaprime labeling kit, and Rapid-Hyb hybridization solution were obtained from Amersham Inc. Maximum strength Nytran filters were from Schleicher & Schell Co. Synthetic oligonucleotides were prepared by using an automated synthesizer (Applied Biosystems, Model 394) at the Molecular Biology Core Facility of this campus. ELISA equipment was from Bio-Rad. Mouse

monoclonal anti-(human factor IX) (hFIX), AHIX5041, was purchased from Hematologic Diagnosis Inc., and polyclonal anti-hFIX IgG–peroxidase conjugate was purchased from Enzyme Research Laboratories Inc. Pooled normal human plasma was obtained from George King Bio-Medical Inc. Top10 F′ cells and X-ray films (X-Omat AR) were purchased from Invitrogen Corp. and Kodak, respectively. All other chemicals used were of analytical grade specified for biochemical and molecular biology use.

Screening of the λFIXII genomic library

Screening of λ phage genomic library was carried out by the standard method [14]. Briefly, the library was propagated in *Escherichia coli* XL-1-Blue MRA, and ≈500 000 independent phages were plated at a density of 5.0×10^4 phages per 150-mm plate. Phage particles transferred to nylon filters were then screened with the human hepsin cDNA [1] labeled with ^{32}P to a specific activity of 1.0×10^9 c.p.m.·μg^{-1} by employing the random priming method using a Megaprime kit (Amersham). Positive clones identified were plaque-purified and amplified on plates to obtain high-titer stocks. Large-scale preparations of recombinant phages were then carried out by the liquid culture method [15]. Phage clones containing the 5′-end region of the mouse hepsin gene were obtained by screening the genomic library with a genomic DNA fragment (667 bp in size) corresponding to the most 5′-end region of the insert of a phage clone mHepλ61 by PCR.

DNA sequence analysis

Phage DNAs were digested with *Not*I, and electrophoresed on an 0.6% agarose gel. The insert DNAs were then recovered by utilizing Geneclean®II (BIO 101 Inc.), subcloned into the pBluescript vector at the *Not*I site, and used for subsequent restriction mapping and sequencing. Southern blot analysis of restriction fragments was carried out with either *Apa*I/*Sty*I (257 bp) or *Bgl*II/*Apa*I (422 bp), which represents the 5′- or 3′-distal portions of the human hepsin cDNA sequence, respectively [1]. Selected restriction fragments of the inserts were subcloned into pUC18 and subjected to sequencing using Thermal Sequenase radiolabeled terminator cycle sequencing kit (Amersham). All sequences were analyzed two (once for each strand) or more times to eliminate inadvertent errors with M13 universal primers and hepsin-specific oligonucleotide primers. DNA sequences were stored and analyzed by the DNASIS program (Hitachi SK) and the GCG program in the VAX computer at the General Clinical Research Center of this campus.

Primer extension analysis

Primer extension analysis was carried out as previously described, with minor modifications [16]. An oligonucleotide primer (19 nucleotides in length), designed to the nucleotide sequence +77 to +95, was labeled with ^{32}P to a specific activity of 5.8×10^8 c.p.m.·μg^{-1} by employing T_4 polynucleotide kinase and [γ-^{32}P]ATP. An aliquot of the probe (2.0×10^6 c.p.m.) was mixed with 10 μg of mouse liver total RNA. A total RNA sample treated with RNase A and a total RNA sample without treatment were included as controls. Sequencing ladders generated from the DNA fragment, which contains the corresponding region as the template, were used as size markers.

PCR analysis of alternative splicings

RT-PCR of mouse liver hepsin mRNA were performed with SuperScript One-Step™ RT-PCR system (GIBCO Life Technologies) to analyze alternative splicings. Aliquots of total liver RNA (1 μg) isolated from C57B/6 mice were reverse transcribed at 50 °C for 30 min with primer f (5′-ATTGGAGCGTGAGGAGCACAGTAG-3′) (Fig. 4A). This primer corresponds to a sequence in exon 6 (nucleotides +11 460 to +11 483) of the mouse hepsin gene. Specific amplification of the regions containing the entire region of alternative splicings was then performed with primer a (5′-CACCCTTGCCTTCCGGGCTGTC-3′, nucleotides +87 to +108 in exon 1) and primer f. Then, PCR was carried out as follows: initial incubation at 94 °C for 2 min, followed by 35 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 1 min, finishing with a 10-min extension at 72 °C. A total RNA sample treated with RNase A was included as a negative control. The resulting PCR products, a mixture of fragments of different size ranging from 300 to 900 bp, were gel-purified and used as the template for subsequent PCR with the common 5′- primer b (5′-TTCCGGGCTGTCCGCTGCTG-3′, nucleotides +97 to +116 in exon 1) and, either primer d (5′-TGCAGTCCGGCCACCCTCCTTC-3′, sequence spanning the junction of exon 2 and exon 4, nucleotides +642 to +649 and nucleotides +6067 to +6080, respectively) or primer e (5′-TGCAGTCCGGCCACCCTTTCCA-3′, sequence spanning the junction of exon 3 and exon 4, nucleotides +4552 to +4559 and nucleotides +6067 to +6080, respectively). PCR reamplification was also performed with primer c (5′-CAGGGTCGGCTGCTCCCTG-3′, nucleotides +484 to +501 in intron 1) and either primer d, e, or f. PCR products were seperated by 2% agarose gel electrophoresis. DNA fragments produced were then cloned into a pCR2.1 vector (TA cloning kit, Invitrogen). Positive identification of the cloned fragments was performed by DNA sequencing analysis.

Cell culture

HepG2 cells, a human hepatoma cell line [17–19], were used for transient expression assay as described previously [20]. Cells were grown and maintained in DMEM supplemented with 10% fetal bovine serum and antibiotics (penicillin and streptomycin) at 37 °C under 5% CO_2 in a humidified incubator. Fetal bovine serum used in HepG2 cell culture for assaying factor IX activity was pretreated with barium sulfate as previously described [21].

Construction of expression vectors

hFIX expression vector p-416FIXm1, which has been previously described [20], was used as the starting vector for constructing two expression vectors containing the hepsin promoter. The hepsin promoter region (nucleotides −274 to +59) was generated by PCR with the hepsin genomic DNA as the templete, using the 5′- and 3′-primers with *Sph*I and *Nhe*I linkers, respectively. The *Sph*I/*Nhe*I fragment generated (333 bp in size) was inserted into p-416FIXm1 at *Sph*I/*Nhe*I sites replacing hFIX promoter sequence (nucleotides −416 to +22), thus generating p-274HSN/FIXm1, which has no 5′-UTR intron. p-274HSNIn/FIXm1 was generated by inserting a PCR-amplified hepsin fragment (890 bp in size, spanning nucleotides −274 to +616), which has the 5′-UTR intron sequence, into p-416FIXm1 in a similar manner as described for the construction of p-274HSN/FIXm1. All the PCR-amplified regions and ligation sites of the expression vectors were sequenced to verify the correct sequences. Large-scale

preparations of p-274HSN/FIXm1 and p-274HSNIn/FIXm1 were performed with Qiagen plasmid kit (Qiagen Inc.), and used for transient expression assays [22]. These expression vectors with the well-defined factor IX reporter gene allow efficient evaluation of promoter strength simply by determining the secreted factor IX.

Transient factor IX expression assays

Transient expression assays of p-274HSN/FIXm1 and p-274HSNIn/FIXm1 were performed as previously described [20] using HepG2 cells and Fugene™ 6 Transfection Reagent (Boehringer Mannheim). hFIX produced into the culture medium was quantified by ELISA [20]. HepG2 cells (60–70% confluency in 6-cm dishes) were cotransfected with a mixture of 8 μg of test-expression vectors and 0.8 μg of pCH110 (internal control for transfection efficiency). Sheared herring sperm DNA and p-416FIXm1 were used for transient assays as the mock and positive controls, respectively. Expression values obtained for the herring sperm DNA mock control were subtracted from those of the factor IX expression vectors. Unless otherwise indicated, all expression vectors were assayed in duplicate in four independent experiments, and averages of the results are presented.

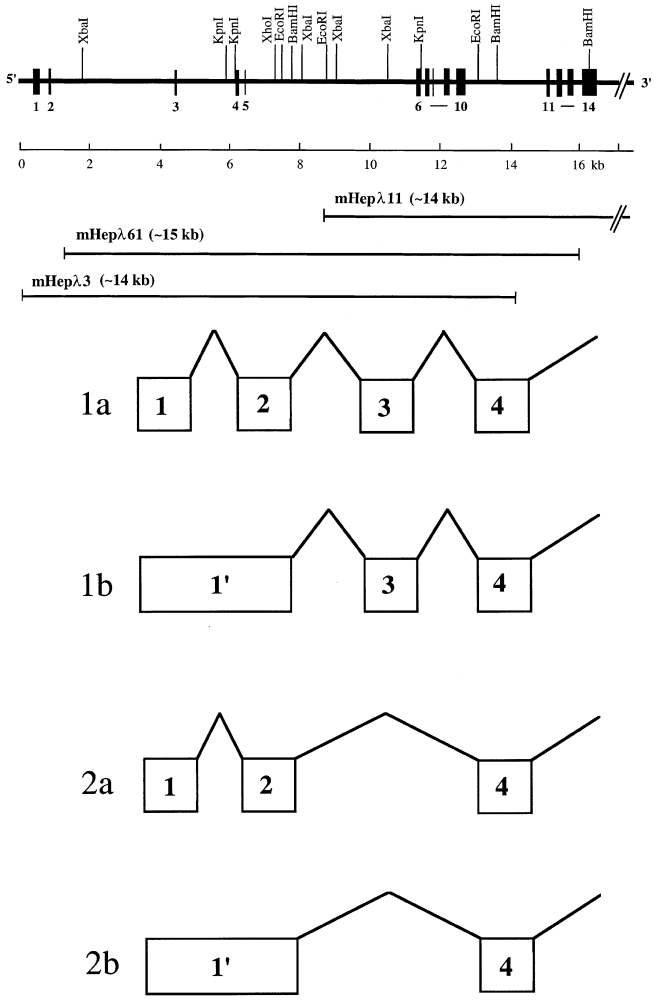

Fig. 1. Exon/intron organization of the mouse hepsin gene (A) and multiple forms of mRNA (B). In panel A, three phage genomic clones, mHepλ11, mHepλ61, and mHepλ3, are shown in relation to the schematic structure of the mouse hepsin gene. Exons are shown by vertical bars. The size scale is in kilobases. In panel B, relevant portions (exons 1–4) of multiple mRNA forms of hepsin generated by alternative splicings are shown. Exons are represented by boxes with numbers and spliced introns are shown by thin lines. Exon 1′ (nucleotides + 1 to +649) is generated by no splicing of the first intron (intron 1), and exon 3 (nucleotides +4500 to +4559) is generated by alternative splicings in the second intron.

RESULTS

Isolation and sequencing of the mouse hepsin gene

Two positive phage clones, designated as mHepλ61 and mHepλ11, were initially isolated from the mouse liver genomic DNA library in λFIXII (5×10^5 independent phage clones) (Fig. 1A). mHepλ61 contained most of the middle portion, and mHepλ11 contained the 3′-half of the gene. By screening $\approx 3 \times 10^5$ independent phages of the same library with the 5′-portion of the insert of mHepλ61 as the screening probe, the third clone mHepλ3, which contains the 5′-end region, was obtained. Inserts of these clones were subjected to sequencing analysis.

The complete contiguous nucleotide sequence for the mouse hepsin gene, which spans ≈17 kb, is shown in Fig. 2. The gene consists of 12–14 exons (1–14) depending on three alternative splicings, which are responsible for generating exon 2 and/or 3. Possible exon/intron organization combinations are shown in Fig. 1B. Exon sequences accounted for 11% of the gene length. The transcription initiation site determined by primer extension analysis was at 636 bp 5′-upstream of the first methionine codon in the 5′-UTR [23] and was designated nucleotide +1 (Fig. 2). Exons size ranges from 41 bp (exon 8) to 349 bp (exon 14), and that of intron from 76 bp (intron 5) to 5417 bp (intron 2). Intron splicing junctions conform to the GT-AG rule [24,25]. The first intron located in the 5′-UTR is subjected to alternative splicing (Fig. 1B, 1a and 1b). Exon 4 encoded most of the transmembrane domain. Two alternative splicings, which might take place in the second intron (Fig. 1B, 1a), could generate an extra exon (exon 3) encoding an in-frame sequence of 20 amino-acids, DEEPGAHRGGSTCSRPQPGK (Fig. 2), which was found in some, but not all, mouse cDNA clone [7]. Exon 5 encodes the C-terminal end region (5 amino-acid sequence) of the transmembrane domain and a stretch of 9 amino acids. Exons 6–8 encode the spacer region between the membrane spanning hydrophobic sequence and catalytic subunit. The serine protease catalytic subunit consisted of 255 amino-acid residues is encoded by six exons (exon 9–14). If the exon/intron organization is that of Fig. 1B, 2a, the predicted size of hepsin mRNA is 1806 bases, in agreement with one of the two distinct mRNA species (≈1.8 kb and 1.9 kb) observed in Northern blot analysis [7]. If no alternative splicing takes place either in the 5′-UTR or in the second intron generating exon 3 as shown in Fig. 1B, 1b, the 1.9-kb mRNA species may be generated.

Several different repetitive sequences were found in the introns and flanking regions (Fig. 2). B1 elements of the B-type repeats [26–29] were found in intron 2 (nucleotides +5225 to +5357 in 5′- to 3′- orientation) and intron 4 (nucleotides +10 695 to +10 798 in 3′- to 5′- orientation), while B2 elements were present in intron 9 (nucleotides +13 495 to +13 679 in 5′- to 3′- orientation, and nucleotides +14 507 to +14 699 in 3′- to 5′- orientation). A polypurine tract of the structure $(AGGG)_{13}$ was located in intron 4 (nucleotides +8406 to +8457). Five and a half tandem repeats of 30-bp

```
-274  aggggactgg catggaagaa aaatcgaatg aaggaccaca ttgctacaag catggtggac cggggacagt cagaactttc cacctcattc ctccagtttg
                                                 C/EBP          LF-A1
-174  gaatcttagc cttagggcct ctgagcgccc ccaagtttgt cacagacaca cctgagtccc agtggacgtg ggaggcctgg ctggtaactg agcctgcctg
                                                                                     AP2
                                                                                      +1
 -74  gtccaacccc tgcgggcccc gcccatgag ccaggtggcc tctgtggcag cctgtcctgg ggtccgccct tcccCGCCCC GTCCCCGCCT CCCACCCAGC
                    SP1          E box                                 SP1        SP1
  27  CCCCTCCTCC TCAGGTGAGG CCGCAGGCTC GGCAGGCCCC ACGCTGCTGG CTGCTGCTGC CACCCTTGCC TTCCGGGCTG TCCGCTGCTG TGGGGACAGA
 127  CACCATGCCC CTGCCCAGGC CCGGAGACTA ACCCCAAACC TGCACCATCT CCGGCGAACC CCAGG gtaa ggcacagaat gcgaagcccc tcatcctcca
 226  ccagtctgat gttcagcagc ccgtgtaaga gatcctagaa acccccagac tactccagcc tgctttcact gctaactcag aaatctgagg actcccccaa
 326  gttctcctcc ctcaaaaccc tcaagtcccc tggaagttca acctcgcccg cccggtgttc tgacctgcca ctccaggtgt ccccctggtg gctggtcaca
 426  ccttgccatt attgaccccc ggcccaacct ctgtctttgc atgctctggc cccagcccag ggtcggctgc tccctggatg ggagcctggg actgggggcg
 526  ccaggataga gccaggccag gctcccccag gccatgcctc cttgttcctg tccctag GT TCCGCCCCAG CCCAACAGGT CAACCTGGGA ATCATTAACA
                 1           4
            Met Ala Lys Glu
 625  AGAGTCCCTG AC ATG GCG AAG GAG G  gtgagtcccc cgtccctgaa ccccagcttc agccctacct ggcacccctc ttctgtttcg tctcccagcc
 720  tgtactcagg tcttggtgtc ctgtcagact caagaccacc cgctcaaaat cctcgtctat ctgattaact attaactggc ccttctgtgt ggctgtacct
 820  ccccgctacc ccatatgacc ttactagtgc cttgcctcag tttcccctta ggtctcaggg atttctcagc taaccaacaa catgtcatct tctgggggct
 920  gaggtggggg ctgctgctgt tgccatctct gttaagggta tagagctagt caggttgttt ctgctgtctt ttactgtgtg atcctgggta ggtaacttca
1020  cgtctctggg cttcagtcca cctttccttg cagtcaagta tggatggata ctaatattcc ctccatgagg tgtttgggat catggaacac acaggcaaca
1120  gcttgggcct tggggcatga aatacgtgtt gtaacccaga ggtgttctga ttcaggtcag tcactatgga agcccaagaa ccaaggaccc aagaccccctc
1220  caaatccagg catatagtcc ttcatgtcag ggactcttgg ctttgtgggc aaaagagcag aagggggact ttggacttga cctggacgga tagaggagag
1320  agacagaaaa accaccccag gcgtgaccga gacagcagct cacactcctt ccctacctgg tccgtacatg gggaccaaat ggctaatcag gggcagtccc
1420  caccgcactc tctgccacca ccatccttga ctgccctcta agagtcaggc tagctgattt cagtcctggg cactttctag ctgtgtggcc ttggacatgg
1520  catcctgtct aatcatctgt tttctttgtt aagtggagat ggtcactatg cggggtgggg gatgggggga actcagtggg aagcagtgaa ggtatgggcc
1620  cagggataag caccggcgtc taggctccta gagttcagtg tctgcttata acaactcaaa ctctagagat gagctgtggg cctctgtgaa ccaggctccc
1720  ctcaccttgt gctaagtggt catcaggtga tgggcagagc tgatcccttt gggacagttt ttattttcct gtggagttga tgattgacaa gagatatcag
1820  ggctggggtc tatgagacag ggaagggaag gagtatctgg atggagaatg atgaagaaga cacagagcat ggtccctgca tccctaggca gctctctgtg
1920  ctgggtgagg gcagccctgg gaaaggtgac gggctggaga tgacagctcc attctcatca actgtggcat ggattgaagg gattcaaagg acagtaaagt
2020  gggtggtgtg agagcccaag agcctctgca gcagctgctg atggtgtttg tgagaccgtg atggccatgg aagtgtgcga cactctgggc tccctgtact
2120  gctgctgggg tcctggaatt agaaggagtg atcagggagg aaggaaagag ggaagaggga ggaaggaggc catgccaggc aggggaagat ggagcagggt
2220  gctcagggta gatgtttgat ggagatggtg cctgtggctg acccacctca gaagtgtcct tggcagagtg gggtctctcc acaaggtagt actagaggtt
2320  cctctaaacc ttgttctttt tcttttactc ctgaaattat gaggatcaaa gccaggccca tatacatgct aggcaagagc cctaccactg agccatgccc
2420  ccagcccctc actggaggat tcaatgcagg ggctctacca ctgaactatg cccaagcccc taactgtggg attctaggca gggtctacca ctgagccatg
2520  cccccagccc ctcactgggg gattctaggc tctaccactg agtcgtgtcc ttagccttct tttcttcttt ttttcttctt ttgttttgat acttagtctc
2620  actaagaagc tcaggcttgt ctcaaagtta acaatccttt tgcctccgcc tcctgaaaag ctgtaatgac aggcttctaa tcagcaattt gtgagttttg
2720  tggcttagga taatctgtgg aagttgtttt aggtgtttgc taatgaccaa tgtatatctg agaattagaa gagacctata tggtctctaa aagtcatgct
2820  tattctgatg ggggctggaa atgctgggcc cgggcaccac agaggcctag gggaacgttc ctcatagaga aaggaccact ggggttcata tttacagttc
2920  ccaggtttct gtgtaaaggg gccaccaggg agatgagcca aaatggtatt aagtacacac aatggacggc ggggcttgga gctggccagc cagagcagga
3020  agtgatgagc agataaaggg aggtcctggg ctgaaggctc ctaagacagc tgcctctgtg tggaggagtg gaagtcccag gatgagggct tctgagggca
3120  gggcagaaaa cctagggatg aggggatga agcaggactg ctgtggggag gggaggaggg cctggggatg agggggatga agcaggactg ctgtggggag
3220  tgggggaggg cctggggatg aggggaatga gatggactgc tgtggggagg ggggagggcc tgagaaggag ggggatggac tgcaggggga gggaaggagg
3320  gcctggcaag gagggaaatg tgatggactg ctatggggag gggaagaggg cctgggaacg agacagggct actacaggaa ggagaagagg accctgggat
3420  gagaggctga cacaggacta ctgtggggag gggaggaggg gggacaggct tgggacagca gatgtgttga atgcaaaggg agaagtgtgt agaggcggca
3520  gaagccccgg ggcctagtgt ggtgctcggt gctcggtgct aaggtatcag gagtgcccag ttctcatgct attttcctct gtaagtgaca ctttgtctta
3620  acccactaac agaaactgcc ctggcaagat cctgaaccac tccctgctct cttcctaatg ctgatttggt actgtgacca ttagccatgc tctgggagac
3720  atgggtgaga gggctggggc tcttgcagcc tccggggtac ttcgcagaag gttggggtca gcacaaggtt ttcttagcag ttgtattcgg gtgctaggat
3820  ttctggtcca ccggggagga catccccttc tctctaagct actcctgcca tgcctcccat agcctccccc aggtctcggc actcccactg tttggtgttg
3920  tagccctcta acagaacatc ctcagtcttg cctgctagtc ttagccgtgc agttcctccg acgtgtggct ctgccccttg tcatctctgt ttgctgatct
4020  cagataatct cttcctgggg aatttccccc tttccttcaa aagccccact tgatgcttaa ccctacattg cttaaaggac agatggcaca gagctttatc
4120  atactgacac actgaggctc agagagacag tgacagttgc ccatagcctc acaacaagaa tggcttatca gtgtttatcg aggacctgct gtgtgccggg
4220  cacactggca caacaaaagg cagtttggtt tcttgggcaa ctactttgga acctggatct cagtgtgccc ggggttcagg acatggccgc aggtcgagtc
4320  actggagatt cgtgacgtga tctctcagag ccgtgagaaa agggaaccgt aatgtgggtc ggagctagag acgacgctgg gcacatggca ggtggcagcc
                                                                                          (5)
                                                                                   Asp Glu Glu Pro Gly
4420  gggacttggg tatgaggatg tatttaaaga actctaaagt catttaaaaa gtgagtgcta ggtgcatgct catttaacag AT GAG GAA CCT GGG
                                                                 (24)
      Ala His Arg Gly Gly Ser Thr Cys Ser Arg Pro Gln Pro Gly Lys
4514  GCT CAC AGA GGA GGT TCC ACT TGT TCA AGA CCC CAA CCT GGA AAG G gta aggccgctat tttatatttt atcctttttt tgtatatttt
4603  taatttctta tatgtgtgtg tacatgtgcc catgtgtatg agtatgtgta tgtgtgtgca tgtggggggg tgtgcacatg tgtgtatgag aatgtgtata
4703  tatgtgtgca tgcacatgcg tgtgagcata tgtgtgtgta tatacatggg atgtgcatgt tcacatgtgt gtatgtgtat agatgtgggt gtgtttgtgc
4803  atgcacatgt atgcgtatga atacatgcat gtgcctgtgt ggataccggt gtacacattg gagacggcat gaagaatctt gtatgtccta ccctatcact
4903  ctctgtgtat ttcccatgat ctggggtctc tcactgagcc tgcagctagg ctgatggcca gggagcccca acgaccctcc tgtctgtcca ccccaacagt
5003  gcagggctta aagacctgct ccaccacacc tggcctttta cacagtgctg ggaacctatc tctgaatccc acacttgccc agcaagtgct cttacccatg
5103  gagccatctc tccaaggtca ctatttcaga gtggggttac cccgcatagt gcgatgactg gcattgtagc acttaggaag cacagactag cctgagtatc
5203  atagggagac cttgtctaag ctagtggtgg tgcatgcctt taatcccagc actccagagg caggggccat tggatttctg agttggaagc cagcctggtc
5303  tacagagaga gttccaagac agccagggct acacagagaa accctgtttt gaaaagccaa aaccaaacca gaccaaaata caggcagata tggtgacacg
5403  tgactgccat ccgagcacta agggaggcag gcaggcaaac tctcactagt tcgaggctag cttgtacatt atggtatgtt tgatgtcagc actgttaggt
5503  agtaagacct catctcaaaa tacaaaccaa cccccccccc aaaaaaaaaa atcaaaccaa aaaaaacttg ggttctcaaa ctcccagata caaatttgat
5603  cttggaactt gtgagaatca ggtcttgggg ttaccaagag ggtaaagatc agaggtcaca ggcaccttaa agttgtaaga acttctggaa aggttatcag
5703  aagagaggat ctggggacag gctggccgga agcaacaagg gatatccaga agtgccttgt catcggaaag gagtggccgc cagcactgtg gggggggggg
5803  gcagggggga gcaggaggcc tggccatgcc ccttcccagc tggaaccctt ggggagccag acacctggca atccacacga caggagagag acagccctgc
5903  cagctgccca gccttgggaa atggggctta aggtaccaca ggcccaggcc cgataaaaaa aaaaaacaaa acctgcattc ccgggcaggg gactggccca
                                                                              5(25)
                                                                     Gly Gly Arg Thr Ala Ala Cys Cys Ser Arg
6003  gtgcccaact cttgctgtgc tgaggaccga cctggcagtg gccccagctt ggtgtctgtt gcag GT GGC CGG ACT GCA GCA TGC TGC TCC AGA
                                                                                                   38(58)
      Pro Lys Val Ala Ala Leu Ile Val Gly Thr Leu lLu Phe lLu Thr Gly Ile Gly Ala Ala Ser Trp Ala Ile
6096  CCC AAG GTG GCA GCT CTC ATT GTG GGT ACC CTG CTG TTC CTG ACA GGC ATT GGG GCC GCG TCC TGG GCC ATT G  gtgagaaaag
                                                                                         39(59)
                                                                                    Val Thr Ile Leu Leu Gln Ser
6179  accccccca tccccaagct cccgaggtct cccctacccc agggttacaa cccgcctgtt ctcccctctc cacag TG ACC ATC CTA CTG CAG AGT
                           52(72)
      Asp Gln Glu Pro Leu Tyr Gln
6274  GAC CAG GAG CCA CTG TAC CAA G gt gagtagagca agcctgcatc cctagggtgt ctgggaggag aagcacaatg aggaatgcaa agcattcttg
6368  ggactgataa ctgtatttct actcaatgtc tctcatgctg tttgagccag aattttgtct gcagctcagg ctggcttaga actcatccaa gctggcctca
6468  gatttgtggc aatcctcctg cctcagcttc ccaatgagag gattgcaggc atgaagcaac ttgtccaaat gagaggctgt agtctaagtg actagggaga
6568  ctcccatcca agaggtatag atctgacctg gcttagcttt tgaggtctac ctccgaaggg tatgattgtt gtccaagtcc tagacaaggc agcaacgagt
6668  tttttaagat attttagat atttttcag ttttattttt gagataaggt ctcatgcaac ccaggctggc ctcaaactta ctatggagga ctctgaacct
6768  cctaatcttc caacctccac cttctaagtc ctaggattag tgcatcacca cacctagctg aaacacacat atagacatac atatacatat gtatatatac
6868  atgtatatac atatgtacat acacatacat ataataatat acatattgtg taacacacac atataatata tcatgggatg tggtcaaaca cattccagag
6968  tgacaggatg aactcccagc tctgagaacc tggtagtttc tatctgccct ttatgccaat gtgggacatg aagaccagag aatcactggc ctgtaatgat
7068  ggagccagga acagctctca gagttgtgga cacgtttgca gcacatggaa gatgcagtca caggggaagg actgagcctt ccccggtgac gctccaagct
7168  gctgctctgt gacccagggc agcagaaggg ggtggtgatg gcagggtgga gcatgctttt tcaaagaaat gggattacag atttgaaacc gttggcactg
```

Fig. 2. continued

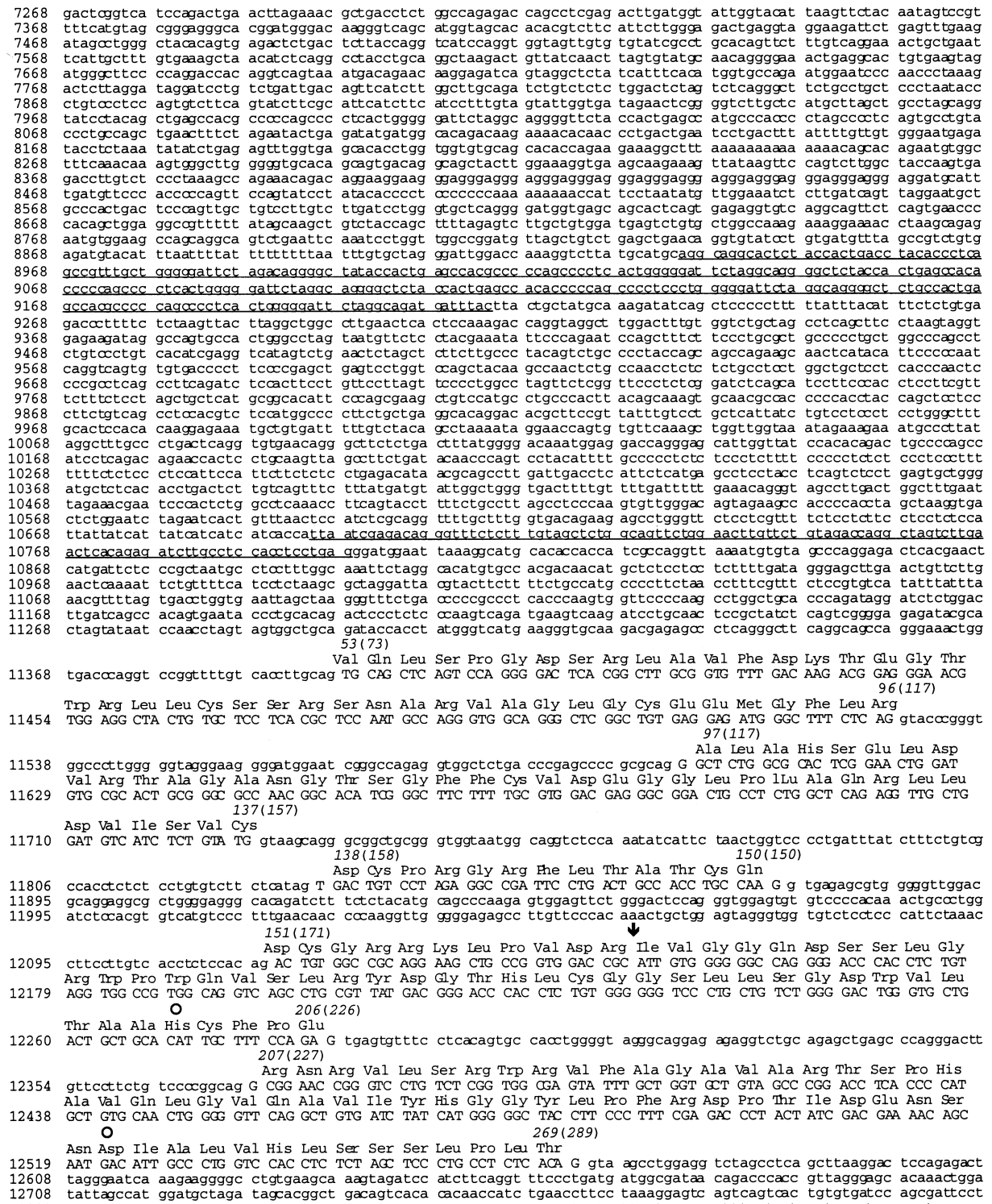

Fig. 2. continued

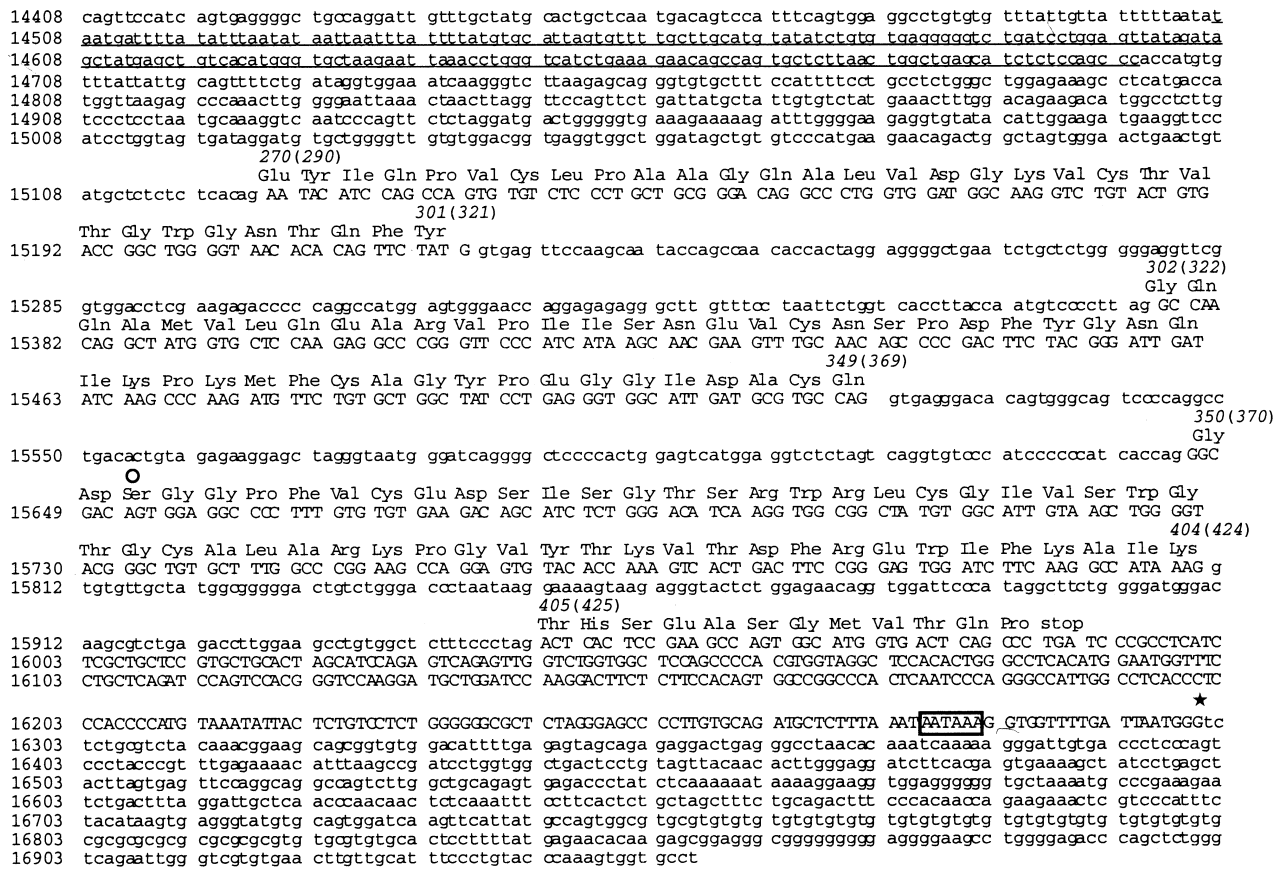

Fig. 2. The complete contiguous nucleotide sequence and encoded amino acid sequences of the mouse hepsin gene. The nucleotide numbering system is based on the definition of the transcription initiation site (C) as nucleotide +1. The 5′ upstream of nucleotide +1 is designated with reverse negative numbers. The site of polyadenylation (nucleotide +16 300) is indicated by an asterisk. Nucleotide sequences of exons are shown with capital letters, whereas those of introns and 5′- and 3′-flanking regions are represented with small letters. The poly(A) signal sequence of AATAAA is boxed. The deduced amino-acid sequence is shown above the corresponding nucleotide sequences. Amino acid numberings in parentheses are of the isoform (Figs 1B, 1a) with the extra amino acid sequence generated by alternative splicings in the second intron. Repetitive sequences including B1, B2, and Alu-like type III elements are underlined, while the sites of putative regulatory elements in the 5′-flanking region are shown by double underlines. The transmembrane domain of the amino acid sequence is indicated by dotted line. The putative zymogen activation site and active site triad forming residues are marked with an arrow and open circles, respectively.

Alu-like type III repetitive sequence [30] were also present in intron 4 (nucleotides +8935 to +9214). This stretch of repeat sequences has a 90% similarity to that found in murine urokinase-type plasminogen activator receptor gene [31]. These repeats may be able to form stem-loop structures. In addition to these, a stretch of dinucleotide repeats $(CA)_7TA(CA)_{13}$ and $(TG)_{23}(CG)_9$, which are capable of forming left-handed Z-DNA structure [32–34], were found in intron 9 (nucleotides +13 423 to +16 464) and in the 3′-flanking region of the gene (nucleotides +16 757 to +16 820), respectively.

Nine nucleotide sequence differences were found between the coding sequence reported in the present study and that of the mouse cDNA recently reported by Vu *et al.* [7]. These differences included those at nucleotides 11 423 (A), 11 429 (A), 11 435 (G), 11 450 (T), 11 622 (G), 12 211 (T), 12 506 (T), 12 531 (T), and 15 732 (C). Bases in parentheses were previously reported by Vu *et al.* [7]. The difference at nucleotide 11 435 causes an amino-acid change from Leu65 [7] to Phe65 in the present study. These differences may be due to sequencing errors in the reported cDNA or to mouse strain-dependent polymorphisms. The predicted amino acid sequence of the mouse hepsin showed an overall similarity of 88% or 97% with those of human [1] and rat hepsin [5], respectively.

Transcriptional initiation site

The major transcription initiation site for the mouse hepsin gene was mapped to C at the position, 636 bp 5′-upstream to the first Met residue codon (Fig. 3), and was defined as nucleotide +1 in the numbering system (Fig. 2). No such signals were observed in control lanes containing either a total RNA sample treated with RNase A or a total RNA alone (Fig. 3, lanes 1 and 2).

Characterization of alternative splicings

Comparison of the mouse hepsin genomic sequence with those of human [1] and mouse [7] hepsin cDNAs suggested possible alternative splicings, which were responsible for generating multiple cDNA forms. This possibility was tested by a combination of RT-PCR and subsequent PCR analyses using total mouse liver RNA and pairs of specific primers (Fig. 4A). RT-PCR amplification with a combination of primer a and primer f yielded two bands of 462 and 402 bp, and an unaccounted band of ≈490 bp (Fig. 4B, lane 2). The first two bands corresponded to the products predicted for hepsin cDNA species with or without exon 3 (Figs 1B, 1a and 2a), respectively. This was further confirmed by DNA sequencing. Control RT-PCR performed with RNA samples treated with

RNase A and the same primer set did not result in amplification products, confirming the high specificity of RT-PCR used (Fig. 4B, lane 5). We further confirmed the existence of alternative splicings, generating 1a and 2a, by subsequent PCR using the first RT-PCR products as a template and primer b in combination with splice variant-specific primers d and e, respectively. The resulting PCR products of 105 and 165 bp (Fig. 4B, lanes 6 and 7) corresponded to hepsin isoforms 2a and 1a, respectively. 1a and 2a corresponded to hepsin isoforms with and without a 20 amino acid stretch of extra-sequence in the cytosolic side of hepsin, respectively. No RT-PCR amplification products with expected sizes of 793 and 853 bp for hepsin isoforms 2b and 1b, respectively, which should correspond to hepsin mRNA containing the 5′-UTR intron sequence, were observed (Fig. 4B, lane 2), indicating efficient splicing of the 5′-UTR intron (intron 1) sequence. However, subsequent PCR amplification of the RT-PCR products (lane 2) with primer c and primer f did produce fragments of 395 and 455 bp, which were expected for the presence of the unspliced 5′-UTR sequence, indicating that the intron 1 is mostly, but not completely spliced (Fig. 4B, lane 10). This is in agreement with the observation that no detectable amplified products was obtained for RT-PCR using the same primer sets (Fig. 4B, lanes 3 and 4). In lane 10, an extra band of ≈490 bp corresponding to the unknown band in lane 2 was also observed, suggesting that it is a byproduct of nonspecific priming of primer f. With combinations of primer c and splice variant-specific primers d or e, 2b- and 1b-specific products (181 and 241 bp, respectively) were also amplified (Fig. 4B, lanes 8 and 9). Sequence analysis of these two products confirmed their identity with 2b and 1b, respectively. The alternative splicings in the second intron were estimated to take place in approximately half of the hepsin transcripts, accounting for the similar abundance of 1.8 and 1.9-kb mRNA species [7]. Together, all four hepsin forms (Fig. 1B) were generated from a single hepsin gene by the alternative splicings in the 5′-UTR and the second intron, which independently take place from each other.

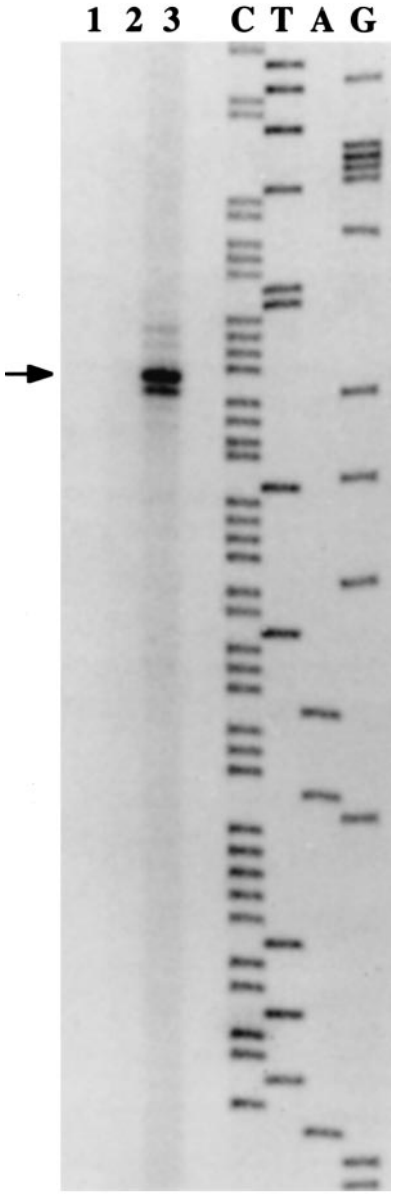

Fig. 3. Transcription initiation site of the mouse hepsin gene. Primer extension analysis was carried out with a radiolabeled 19-mer oligonucleotide, which is designed complementary to the sequence nucleotides +77 to +95. The reference sequence ladders were generated by sequencing reactions with the same primer and corresponding DNA template. Lane 1, primer extension products using total RNA treated with RNase A. Lane 2, same as lane 1 except no [^{32}P]-labeled primer was included. Lane 3, primer extension products generated from mouse liver total RNA. The major transcription initiation site is shown by an arrow.

Promoter activity in HepG2 cells

To examine whether the 5′-flanking sequence of the hepsin gene up to nucleotide −274 actually has promoter activity, we generated two hepsin expression constructs (p-274HSN/FIXm1 and p-274HSNIn/FIXm1), which contain a hFIX minigene reporter gene, FIXm1, linked to the 5′-flanking sequence with and without the 5′-UTR intron (intron 1) sequence. p-416FIXm1, which contains the hFIX promoter linked to FIXm1, was used as a positive control [20]. HepG2 cells transfected with p-416FIXm1 (1×10^6 cells) produced 50–60 ng of recombinant hFIX into the culture medium in 48 h p-274HSN/FIXm1 produced hFIX at a 22% level of that by p-416FIXm1 (Table 1). Inclusion of the 5′-UTR intron sequence in the expression vector (p-274HSNIn/FIXm1) enhanced the hFIX expression level by ≈ sevenfold over that of p-274HSN/FIXm1 (Table 1). These results indicated that the 5′-flanking sequence of the hepsin gene up to nucleotide −274 has a strong promoter activity in HepG2 cells.

DISCUSSION

Hepsin was originally identified from human liver cDNA clones in 1988 [1]. Since then, several important findings have been made, strongly indicating it has potential physiological and pathophysiological roles. However, its major physiological function still remains unknown. We have determined its complete contiguous nucleotide sequence and characterized some important structural properties of the mouse hepsin gene.

The hepsin gene is ≈17 kb in size, and is composed of 12–14 exons depending on the combinations of the three alternative splicings (Fig. 1B and Fig. 2). Interestingly, its transcription initiation site (nucleotide +1), which is 636 bp 5′-upstream to the first Met codon, is in the middle of cytidine-rich region with no specific TATA box present in the neighboring areas (Fig. 2). Several known repetitive sequences in mouse genes including B1, B2, and 5.5 units of Alu-like type III repetitive elements have been found in introns (Fig. 2). These repeats, particularly Alu-like type III repetitive repeats, which have five potential

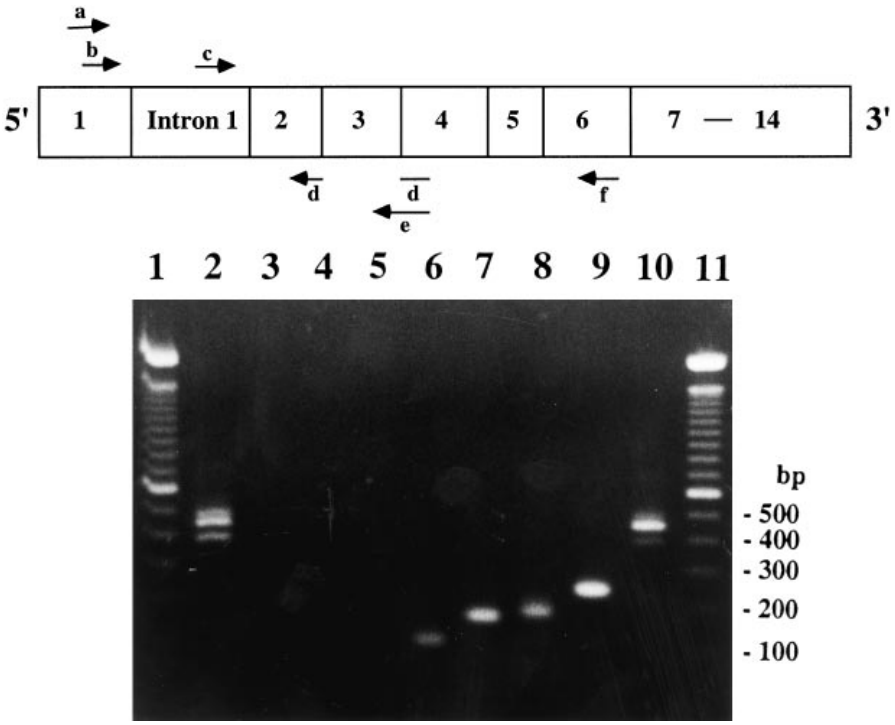

Fig. 4. PCR analysis of alternative splicings. Panel A: schematic drawing of exons and the first intron of hepsin with the relative locations of primers (a–f) used for RT-PCR and PCR. Intron 1 and exon III, which are generated by alternative splicings, are also shown. The drawing does not reflect the actual relative size of the exons. Panel B: agarose gel electrophoresis of PCR products. Total RNA (1 μg) from the mouse liver was subjected to the initial RT-PCR (lane 2), followed by subsequent PCR analyses with various primer sets as described under Experimental procedures. DNA ladder size markers are shown in lanes 1 and 11. Sizes (bp) of some DNA markers are shown on the right side. Lanes 2–4, RT-PCR products generated with different primer sets, a/f, c/d and c/f, respectively. Lane 5, RT-PCR products of mouse liver RNA treated with RNase A (a/f primer set). Lanes 6–10, PCR products obtained from the initial RT-PCR products generated by a/f primer set (lane 2) as the template with primer sets, b/d, b/e, c/d, c/e and c/f, respectively. As confirmed by DNA sequencing, PCR amplified fragments (462 and 402 bp in lane 2) are generated by alternative splicings of intron 2 (Fig. 1B, 1a or 2a). Extra bands of an approximate size of 490 bp seen in lanes 2 and 10 are likely due to nonspecific priming of primer f.

unique stem-loop structures, are of interest in relation to hepsin gene regulation, and warrants further study. The exon/intron organization pattern of the serine protease subunit belongs to the complex type similar to those of plasminogen [35] and prothrombin genes [36], but not to those of trypsin [37] and factor IX [38]. Interestingly, an additional intron exists separating the last 12 amino acids of the C-terminal sequence into the last exon, exon 14 (Fig. 2). The plasminogen and prothrombin genes do not have any corresponding intron in the catalytic subunits. Interestingly, however, the rat elastase gene has an equivalent intron [39], while the overall exon/intron organizations of the hepsin and elastase genes are dissimilar to each other. The exon/intron organization pattern of the N-terminal half of hepsin, where the hydrophobic sequence is contained, is simmilar to those of membrane spanning domains of various other membrane-associated protein genes [40,41].

An alternative splicing takes place in the 5′-UTR of the human hepsin transcripts as previously shown [1] (Fig. 1B, 2a,b). For mouse hepsin cDNAs, only spliced forms of this intron have been observed to date [7]. The present studies, using combinations of RT-PCR, and subsequent PCR demonstrate that the intron in the 5′-UTR is spliced in most, but not all of the hepsin transcripts. Two mRNA species, 1.8 kb and 1.9 kb in size, correspond to spliced and unspliced forms, respectively, resulting from the two alternative splicing sequences in the second intron [7]. Both spliced and unspliced forms are produced in the liver at almost equal amounts as estimated from the Northern blot analysis [7], which is further supported by the present RT-PCR/PCR analyses. Importantly, two alternative splicings in the second intron generate a hepsin isoform containing a unique stretch of 20 extra residues, DEEPGAHRGGSTCSRPQPGK, in frame in the cytosolic side. This finding establishes that the single gene is responsible for generating both forms of hepsin. This unique extra sequence is predicted to have a highly twisted, irregular secondary structure [42]. This isoform may be able to confer a distinct signal transduction, not available to hepsin without this extra sequence, thus endowing it with substantially different biological functions. This and whether the similar alternative splicings take place in tissues other than the liver have yet to be determined. As shown by RT-PCR/PCR analyses, an alternative splicing in the 5′-UTR obviously takes place independently from the other two in the second intron (Fig. 4,A,B).

Table 1. Hepsin promoter activities with the Factor IX reporter gene. Values represent mean hFIX levels ± SD (n = 4). The hFIX level produced by p-416FIXm1 is defined as 100%.

Constructs	Relative factor IX level (%)
p-416FIXm1	100
p-274HSN/FIXm1	22 ± 4.2
p-274HSNIn/FIXm1	155 ± 7.5

The basal hepsin promoter region (nucleotides −274 to +59) can confer strong expression activities, suggesting that the major structural elements required for hepsin expression are contained within this relatively small 5′-flanking region. Elevation of the transcriptional activity (≈ sevenfold) in the presence of an intron when compared to levels in its absence is in agreement with the similar observations for other genes [20]. A computer search predicted several structural elements including SP1, AP2, C/EBP, LF-A1, and E box present within the promoter region (Fig. 2), suggesting that these elements may be in part responsible for hepsin gene expression in the liver and other tissues [3]. Preliminary footprinting analyses using the liver nuclear extracts suggested apparent protein binding to the predicted SP1 site (nucleotides −60 to −42), but not to other sites (data not shown).

Hepsin has been suggested to have potential roles in developement [6,7], both normal and cancer, cell growth [6], and cell-surface initiation of blood coagulation [8]. Interestingly,

however, inactivation of the gene in mice results in an approximate twofold increase in the level of alkaline phosphatase, but gives no other obvious abnormal effects (e.g. during development) and animals are born virtually normal and survive to adulthood [10]. This may be due to a possible presence of redundant hepsin-like activities in mice. Furthermore, absence of hepsin may have detrimental effects on the ageing process or may affect pathological processes, such as cancer growth and metastasis. Indeed, Tanimoto *et al*. [9] recently reported an elevated hepsin expression in ovarian cancer, suggesting a potentially important role in cancer. As several lines of independent studies [3,6–9] strongly support the importance of hepsin, further vigorous studies aiming to determine its biological functions are warranted. Thus, availability of the complete nucleotide sequence and establishment of key functional characteristics of the hepsin gene will provide a dependable foundation for such studies.

ACKNOWLEGEMENTS

This work was supported in part by grants from National Institute of Health (HL38644), University of Michigan Multipurpose Arthritis and Musuloskeletal Diseases Center (NIH5960AR20557), General Research Center (NIH MOIRR00042). S. Kawamura is a recipient of Uehara Memorial Foundation postdoctral fellowship.

REFERENCES

1. Leytus, S.P., Loeb, K.R., Hagen, F.S., Kurachi, K. & Davie, E.W. (1988) A novel trypsin-like protease (Hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells. *Biochemistry* **27**, 1067–1074.
2. Kurachi, K., Torres-Rosado, A. & Tsuji, A. (1994) Hepsin. *Meth Enzymol*. **244**, 100–114.
3. Tsuji, A., Torres-Rosado, A., Arai, T., Le Beau, M.M., Lemons, R.S., Chou, S.-H. & Kurachi, K. (1991) Hepsin, a cell membrane-associated protease. Characterization, tissue distribution, and gene localization. *J. Biol. Chem.* **266**, 16948–16953.
4. Zhukov, A.A., Hellman, V. & Ingelman-Sundberg, M. (1997) Purification and characterization of hepsin from rat liver microsomes. *Biochem. Biophys. Acta* **1337**, 85–95.
5. David, F., Françoise, R. & Hanspeter, N. (1993) Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase. *Biochem. Biophys. Acta* **1173**, 350–352.
6. Torres-Rosado, A., O'Shea, S.K., Tsuji, A., Chou, S.-H. & Kurachi, K. (1993) Hepsin, a putative cell-surface serine protease, is required for mammalian cell growth. *Proc. Natl Acad. Sci. USA* **90**, 7181–7185.
7. Vu, T.-K.H., Liu, R.W., Haakasma, C.J., Tomasek, J.J. & Howard, E.W. (1997) Identification and cloning of the membrane-associated serine protease, hepsin, from mouse preimplantation embryos. *J. Biol. Chem.* **272**, 31315–31320.
8. Kazuma, Y., Hamamoto, T., Foster, D.C. & Kisiel, W. (1995) Hepsin, a putative membrane-associated serine protease, activates human factor VII and initiates a pathway of blood coagulation on the cell surface leading to thrombin formation. *J. Biol. Chem.* **270**, 66–72.
9. Tanimoto, H., Clarke, J.Y.Y., Korourian, S., Shigemasa, K., Parmley, T.H., Parham, G.P. & O'Brien, T.J. (1997) Hepsin, a cell surface serine protease identified in hepatoma cells, is overexpressed in ovarian cancer. *Cancer Res.* **57**, 2884–2887.
10. Wu, Q., Yu, D., Post, J., Halks-Miller, M., Evan Sadler, J. & Morser, J. (1998) Generation and characterization of mice deficient in hepsin, a hepatic transmembrane serine protease. *J. Clin. Invest.* **101**, 321–326.
11. Bugge, T.H., Suh, T.T., Flick, M.J., Daugherty, C.C., Rømer, J., Solberg, H., Ellis, V., Danø, K. & Degan, J.L. (1995) The receptor for urokinase-type plasminogen activator is not essential for mouse developement or fertility. *J. Biol. Chem.* **270**, 16886–16894.
12. Mcmahon, H.T., Bolshakov, V.Y., Janz, R., Hammer, R.E., Siegelbaum, S.A. & Südhof, T.C. (1996) Synaptophysin, a major synaptic vesicle protein, is not essential for neurotransmitter release. *Proc. Natl Acad. Sci. USA* **93**, 4760–4764.
13. Krezel, W., Dupé, V., Mark, M., Dierich, A., Kastner, P. & Chambon, P. (1996) RXR gamma null mice are apparantly normal and compound RXR alpha +/–/RXR beta –/–RXR gamma –/–mutant mice. *Proc. Natl Acad. Sci. USA* **93**, 9010–9014.
14. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular cloning: A Laboratory Manual, 2nd edn.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
15. Silhavy, T.J., Berman, M.L. & Enquist, L.W. (1984) *Experiments with Gene Fusion.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
16. Kurachi, S., Furukawa, M., Salier, J.-P., Wu, C.-T., Wilson, E.J., French, F.S. & Kurachi, K. (1994) Regulatory mechanism of human factor IX gene: protein binding at the Leyden-specific region. *Biochemistry* **33**, 1580–1591.
17. Knowles, B.B., Howe, C.C. & Aden, D.P. (1980) Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. *Science* **209**, 497–499.
18. Fair, D.S. & Bahnak, B.R. (1984) Human hepatoma cells secrete single chain factor X, prothrombin, and antithrombin III. *Blood* **64**, 194–204.
19. Darlington, G.J., Kelly, J.H. & Buffone, G.J. (1987) Growth and hepatospecific gene expression of human hepatoma cell in a defined medium. *In Vitro Cell. Dev. Biol.* **23**, 349–354.
20. Kurachi, S., Hitomi, Y., Furukawa, M. & Kurachi, K. (1995) Role of intron I in expression of the human factor IX gene. *J. Biol. Chem.* **270**, 5276–5281.
21. Yao, S.-N. & Kurachi, K. (1992) A simple treatment of serum for precise determination of recombinant factor IX in the culture media. *Biotechniques* **12**, 524–526.
22. Briggs. M.R., Yokoyama, C., Wang, X., Brown, M.S. & Goldstein, J.L. (1993) Nuclear protein that binds stenol regulatory element of low density lipoprotein receptor promotor. I. Identification of the protein and delineation of its target nucleotide sequence. *J. Biol. Chem.* **268**, 14490–14496.
23. Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids. Res.* **12**, 857–872.
24. Breathnach, R. & Chambon, P. (1981) Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**, 349–383.
25. Mount, S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459–472.
26. Kramerov. D.A., Grigoryan, A.A., Ryskov, A.P. & Georgiev, G.P. (1979) Long double-stranded sequences (dsRNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: evidence from DNA cloning experiments. *Nucleic Acids Res.* **6**, 697–713.
27. Ryskov, A.P., Ivanov, P.L., Kramerov, D.A. & Georgiev, G.P. (1983) Mouse ubiquitous B2 repeat in polysomal and cytoplasmic poly (A) +RNAs: unidirectional orientation and 3′-end localization. *Nucleic Acids Res.* **11**, 6541–6558.
28. Kalb, V.F., Glasser, S., King, D. & Lingrel, J.B. (1983) A cluster of repetitive elements within a 700 base pair region in the mouse genome. *Nucleic Acids Res.* **11**, 2177–2184.
29. King, D., Snider, L.D. & Lingrel, J.B. (1986) Polymorphism in an androgen-regulated mouse gene is the result of the insertion of a B1 repetitive element into the transcription unit. *Mol. Cell. Biol.* **6**, 209–217.
30. Hurst, H.C. & Parker, M.G. (1984) Rat prostatic steroid binding protein: characterization of the Alu element upstream of the C3 genes. *Nucleic Acids Res.* **12**, 4313–4322.
31. Suh, T.T., Nerlov, C., Danÿ, K. & Degan, J.L. (1994) The murine urokinase-type plasminogen activator receptor gene. *J. Biol. Chem.* **269**, 25992–25998.
32. Hamada, H., Petrino, M.G., Kakunaga, T., Seidman, M. & Stollar, B.D. (1984) Characterization of genomic poly (dT-dG). poly (dC-dA) sequences: structure, organization, and conformation. *Mol. Cell. Biol.* **4**, 2610–2621.
33. Hamada, H., Seidman, M., Howard, B.H. & Gorman, C.A. (1984)

Enhanced gene expression by the poly (dT-dG). poly (dC-dA) sequence. *Mol. Cell. Biol.* **4**, 2622–2630.

34. Naora, H. & Deacon, N.J. (1982) Relationship between the total size of exons and introns in protein-coding genes of higher eukaryotes. *Proc. Natl Acad. Sci. USA* **79**, 6196–6200.
35. Petersen, T.E., Martzen, M.R., Ichinose, A. & Davie, E.W. (1990) Characterization of the gene for human plasminogen, a key proenzyme in the fibrinolytic system. *J. Biol. Chem.* **265**, 6104–6111.
36. Degen, S.J. & Davie, E.W. (1987) Nucleotide sequence of the gene for human prothrombin. *Biochemistry* **26**, 6165–6177.
37. Craik, C.S., Choo, Q.L., Swift, G.H., Quinto, C., MacDonald, R.J. & Rutter, W.J. (1984) Structure of two related rat pancreatic trypsin genes. *J. Biol. Chem.* **259**, 14255–14264.
38. Yoshitake, S., Schach, B.G., Foster, D.C., Davie, E.W. & Kurachi, K. (1985) Nucleotide sequence of the gene for human factor IX (Antihemophilic factor B). *Biochemistry* **24**, 3736–3750.
39. Swift, G.H., Craik, C.S., Stary, S.J., Quinto, C., Lahaie, R.G., Rutter, W.J. & MacDonald, R.J. (1984) Structure of the two related elastase genes expressed in the rat pancreas. *J. Biol. Chem.* **259**, 14271–14278.
40. Vihinen, T., Auvinen, P., Alanen-Kurki, L. & Jalkanen, M. (1993) Structural organization and genomic sequence of mouse syndecan-1 gene. *J. Biol. Chem.* **265**, 17261–17269.
41. Wissenbach, V., Bosse-Doenecke, E., Freise, D., Ludwig, A., Murakami, M., Hofmann, F. & Flocherzi, V. (1998) The structure of the murine calcium channel gamma-subunit gene and protein. *Biol. Chem.* **379**, 45–50.
42. Chou, P.Y. & Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymed. Relat. Areas Mol. Biol.* **47**, 45–41.