BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **238,** 26–32 (1997)
ARTICLE NO. RC977233

Hepatoma-Derived Growth Factor Belongs to a Gene Family in Mice Showing Significant Homology in the Amino Terminus

Yoshitaka Izumoto,[1] Toshifumi Kuroda, Hiroshi Harada, Tadamitsu Kishimoto, and Hideji Nakamura
Department of Medicine III, Osaka University Medical School, Yamada-Oka 2-2, Suita 565, Japan

Received July 31, 1997

Hepatoma-derived growth factor (HDGF) is an acidic polypeptide with mitogenic activity for fibroblasts performed outside the cells despite the presence of a putative nuclear localization signal (NLS). We have now cloned three related mouse cDNAs: one for a mouse homologue of human HDGF and two for additional HDGF-related proteins provisionally designated HDGF-related proteins 1 and 2 (HRP-1 and -2). Their deduced sequences have revealed that HDGF belongs to a new gene family with a highly conserved 98-amino-acid sequence at the amino terminus (*hath* region, for homologous to the amino terminus of HDGF). HRP-1 and HRP-2 proteins are 46 and 432 amino acids longer than mouse HDGF, respectively, and have no conserved amino acid sequence other than the *hath* region. HRP-1 is a highly acidic protein (26% acidic) and also has a putative NLS. HRP-2 protein carries a mixed charge cluster, a sharp switch of positive- to negative-charge residues, which is often found in some nuclear proteins. Northern blotting shows that mouse HDGF and HRP-2 are expressed predominantly in testis and skeletal muscle, to intermediate extents in heart, brain, lung, liver, and kidney, and to a minimal extent in spleen. HRP-1 is expressed specifically in testis. These findings suggest that the HDGF gene family might play a new role in the nucleus especially in testis. © 1997 Academic Press

HDGF was purified from the conditioned medium of human hepatoma-derived cell line, HuH-7, and has growth stimulating activity for fibroblasts and some hepatoma cells (1). Subsequently, the conditioned medium of COS-7 cells expressing copious amounts of HDGF as a result of transfection with the cDNA of human HDGF also showed the growth stimulating activity (2). In both cases, debris from dead and dying cells could not be eliminated as the source of the extracellular protein. The HDGF cDNA was found to lack a signal peptide sequence for secretion and have a basic motif, KRR-AGDLLEDSPKRPK (basic residue underlined), homologous to the reported consensus sequences for bipartite NLSs which consist of two clusters of basic residues separated by 10-12 amino acids including proline residues (2,3). The paradigm for bipartite NLSs is that of *Xenopus laevis* nucleoplasmin, KRPAATKKAGQA-KKKK (4). Of nuclear proteins, 56% were found to contain bipartite motifs of basic residues and only about 5% of nonnuclear proteins contained similar motifs (3). This fact indicates that HDGF might function as a nuclear protein. Recently, it has become evident that some growth factors, subsequent to receptor-ligand internalization, may translocate to the nucleus and directly function in mitogenic processes (5). For example, an FGF-1-encoding cDNA construct lacking the secretion signal is functional in an autocrine system, indicating its intracellular signaling capabilities (6). Induction of mitogenesis by FGF-1 is directly dependent on nuclear localization as conferred by the NYKKPKL sequence, whereby deletion of this NLS abolishes mitogenic activity but does not affect other functions (7). Moreover, the NLS of FGF-1 is shown to be directly involved in the FGF-1-induced nuclear mitogenic signaling. A 28-residue synthetic peptide containing the NLS of FGF-1 can stimulate DNA synthesis after it is delivered into NIH3T3 cells by using a cell-permeable peptide import method (8). Further studies will be required to examine the functional role of the putative NLS of HDGF in HDGF-stimulated mitogenesis.

Although the HDGF cDNA was identified from the human hepatoma-derived cell line, its message is ubiquitously expressed in various normal tissues as well as

[1] Current address: Diagnostic Division, Medical Research Laboratory, Sekisui Chemical Co., Ltd., Hyakuyama 2-1, Shimamoto-cho, Mishima-gun, Osaka 618, Japan.

0006-291X/97 $25.00

(A)

```
                CGCAAACTTGGGCTCGCGCTTCCCGGCTCGGCGCGGAGCCCGGGGCGCCCGCGGCCCCGCC    61
ATGTCGCGATCCAACCGGCAGAAAGAGTACAAGTGCGGAGACCTGGTGTTTGCGAAGATGAAAGGATACCCACAC   136
 M  S  R  S  N  R  Q  K  E  Y  K  C  G  D  L  V  F  A  K  M  K  G  Y  P  H   (25)
TGGCCGGCCCGGATTGATGAGATGCCTGAGGCTGCAGTGAAGTCAACAGCCAACAAATACCAAGTCTTTTTTTTT   211
 W  P  A  R  I  D  E  M  P  E  A  A  V  K  S  T  A  N  K  Y  Q  V  F  F  F   (50)
GGGACCCATGAGACGGCATTCCTGGGCCCCAAAGACCTCTTCCCTTATGAGGAATCCAAGGAGAAGTTTGGCAAG   286
 G  T  H  E  T  A  F  L  G  P  K  D  L  F  P  Y  E  E  S  K  E  K  F  G  K   (75)
CCCAACAAGAGGAAAGGGTTCAGCGAGGGGCTGTGGGAGATCGAGAACAACCCTACAGTCAAGGCCTCTGGCTAC   361
 P  N  K  R  K  G  F  S  E  G  L  W  E  I  E  N  N  P  T  V  K  A  S  G  Y   (100)
CAGTCCTCCCAGAAAAAGAGTTGTGCGGCAGAGCCCGAGGTGGAGCCCGAAGCCCATGAGGGTGACGGTGATAAG   436
 Q  S  S  Q  K  K  S  C  A  A  E  P  E  V  E  P  E  A  H  E  G  D  G  D  K   (125)
AAGGGCAGTGCAGAGGGCAGCAGCGACGAAGAAGGGAAACTGGTGATCGATGAACCAGCCAAGGAGAAGAACGAA   511
 K  G  S  A  E  G  S  S  D  E  E  G  K  L  V  I  D  E  P  A  K  E  K  N  E   (150)
AAGGGCACGCTGAAGAGGAGAGCAGGGGATGTGTTGGAGGACTCCCCTAAACGTCCCAAGGAGTCAGGAGACCAT   586
 K  G  T  L  K  R  R  A  G  D  V  L  E  D  S  P  K  R  P  K  E  S  G  D  H   (175)
GAGGAGGAGGACAAGGAGATAGCTGCCTTGGAGGGTGAGAGGCCCCTGCCTGTAGAGGTGGAGAAGAACAGCACC   661
 E  E  E  D  K  E  I  A  A  L  E  G  E  R  P  L  P  V  E  V  E  K  N  S  T   (200)
CCCTCTGAGCCAGACTCTGGCCAGGGACCTCCTGCAGAGGAAGAAGAGGGAGAGGAAGAGGCTGCCAAGGAAGAG   736
 P  S  E  P  D  S  G  Q  G  P  P  A  E  E  E  E  G  E  E  E  A  A  K  E  E   (225)
GCTGAAGCCCAGGGCGTCAGAGATCATGAGAGCCTGTAGCCACCAATGTTTCAAGAGGAGCCCCTGCCCCGTTCC   811
 A  E  A  Q  G  V  R  D  H  E  S  L  *                                       (238)
TGCTGCTGTCTGGGTGCTACTGGGGAAACTGGCCATGGCCTGCAAACTGGGAACCCTTTCCCACCCTATTTACCC   886
TACTCCCTCACTCACTCTCTCCTCTAAGCCCACTCCTGGAGAGTGTCTTGGCCCTCACCTCCAGCTCCCTTCCTA   961
TATACACCCTGTGCCCCAGGATGAGATGAGGCCTTTGTATCTCTTTACACTTGTTTCCCAGGGTTTCTGCTGGGG  1036
TCTAGGCTGCTGTTTCCACCTCTTGACACCTCTGCCCTGCTGCAGGCATTCTAGACCTTTGGGGTGGATAGTGGG  1111
CAGGAGTGGAGGTGAAAGAATATAAAGGAGTGTGGGTTCATGGATGGCATCGTCTACCTGAGCTCCTGTCTCCAG  1186
CCCCCACACTTATTTTCCCATCTGCCTACATTCAAGAAACAGGACACTGTGGGAGAGAGGCTACCATCCATCCAT  1261
AAATCCTTGTTGATTTTTGGGAACACTTATCCCCCTGACCCCAGGGTTCAAGGAATTGTAGTTTAACATCTAGAC  1336
TTTGGAGTTTCCAAGTTTGGGCCTAGGACCTGGAGGGAGCTAAGAGCTGAAGAATCAACTGATTTGCATTGAGGA  1411
AATGTCTCTTTAGATCTCAGGGCAGAAATGATAACCTGGGGAGACCTGCTGCCTTCATCTACTTCCCAATGCTTG  1486
AGGCCAGCCTGTAGTCAGATATTTCACCCAGACATAAAGGAAAAGACCATTTTTTTTAGGAAATGTTTTTAATAA  1561
AAGCCG                                                                       1567
```

(B)

```
mouse    MSRSNRQKEY KCGDLVFAKM KGYPHWPARI DEMPEAAVKS TANKYQVFFF 50
human    ********** ********** ********** ********** **********

mouse    GTHETAFLGP KDLFPYEESK EKFGKPNKRK GFSEGLWEIE NNPTVKASGY 100
human    ********** ********** ********** ********** **********

mouse    QSSQKKSCAA EPEVEPEAHE GDGDKKGSAE GSSDEEGKLV IDEPAKEKNE 150
human    ********VE ***P****A* *******N** ********** **********

mouse    KGTLKRRAGD VLEDSPKRPK ESGDHEEEDK EIAALEGERP LPVEVEKNST 200
human    **A******* L********* *AENP*G*E* *A*T**V*** **M*******

mouse    PSEPDSGQGP PA--EEEEG- EEEAAKEEAE AQGVRDHESL            237
human    ****G**R** *QEE****DE ****T**D** *P*I******
```

FIG. 1. Nucleotide and translated amino acid sequences of mouse HDGF cDNA. (A) The nucleotide sequence of the mouse HDGF cDNA is shown along with the amino acid sequence of mouse HDGF beginning with the first ATG codon. The sequence of the 98-amino-acid *hath* region is underlined. The putative bipartite NLS is double underlined. (B) Comparison of amino acid sequences of mouse and human HDGF. Asterisks denote the identical amino acid residues. Deletions or insertions are indicated by dashes.

in tumor cell lines (2). Therefore, it not only is significantly related to the proliferation of cancer cells but also has other important physiological functions in normal organs. In this study, characterization of additional HDGF-related cDNAs from mouse testis has now revealed that HDGF belongs to a new gene family with a highly conserved 98-amino-acid sequence that we call '*hath* region'. In addition to delineating three members of the HDGF gene family, we show that mRNAs for the three proteins in mouse are distributed differentially in different organs.

MATERIALS AND METHODS

Cloning and sequence analyses of HDGF-related cDNAs. A 0.9-kb cDNA of human HDGF (2) was used to screen a random-primed BALB/c mouse testis cDNA library constructed in pYEUra3 (Stratagene, La Jolla, CA). A total of one million colonies were screened. From 20 positives, isolated from sequential purification steps, three distinct classes were identified. Clones were subcloned into pBluescript II SK+ (Stratagene, La Jolla, CA) and sequenced manually using Sequenase version 2.0 (U.S. Biochemical Corp.) or by automated DNA sequencing (Applied Biosystems Inc., model 373A). The sequencing was done with T3 and T7 primers and with internal sequence primers. The remaining 5′ sequences of HRP-1 and HRP-

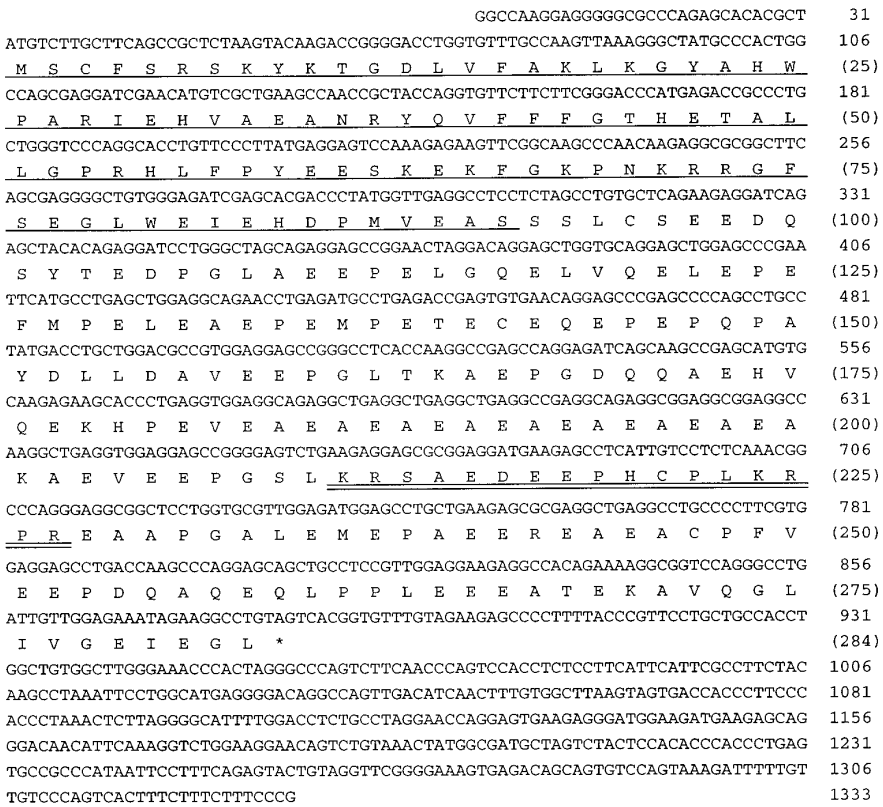

FIG. 2. Nucleotide and translated amino acid sequences of HRP-1 cDNA. Shown is the sequence of sense strand of composite cDNAs and corresponding amino acid sequence beginning at the first initiator methionine at base 32. The sequence of the 98-amino-acid *hath* region is underlined. The putative bipartite NLS is double underlined.

2 were obtained using the 5′-Amplifinder rapid amplification of cDNA ends kit (Clontech, Polo Alto, CA), according to manufacturer's instructions. Mouse testis poly(A)$^+$ RNA (2μg) was reverse-transcribed and then primed with oligonucleotide 5′-TTCTGCCTCCAGCTCAGGCATGAAT-3′ for HRP-1 or 5′-CCGTACTTGTCCTTGCACTTATCAT-3′ for HRP-2. A nested gene-specific primer 5′-CTGGTTCGGCCCATCCTGCACCAGCTCCTGTCCTAGTT-3′ for HRP-1 or 5′-CTGGTTCGGCCCAATGACCCTCCGGGATTCTTTGTCCT-3′ for HRP-2 was used in conjunction with this procedure. Sequence alignments were done using DNASIS program (Hitachi software Engineering Co., Ltd.). The BLAST program of GenomeNet (Japan) was used for database comparisons.

Preparation of probes specific to mouse HDGF, HRP-1, and HRP-2. Oligonucleotide primers were synthesized corresponding to the coding strands of nucleotides 341-364 for mouse HDGF, nucleotides 312-335 for HRP-1, and nucleotides 402-425 for HRP-2; and the noncoding strands of nucleotides 749-772 for mouse HDGF, nucleotides 730-753 for HRP-1, and nucleotides 808-831 for HRP-2. These primers were used in the polymerase chain reactions with plasmids containing mouse HDGF, HRP-1, and HRP-2 cDNAs as templates to prepare fragments that are specific to each gene. Gene-specific probes were prepared by random primer labeling of these fragments after denaturation.

Northern blot analysis. A Northern blot containing 2 μg of poly(A)$^+$ RNA from various mouse tissues was purchased from Clontech. The membrane was sequentially probed with the gene-specific probes for mouse HDGF, HRP-1, and HRP-2. The hybridization conditions were: 50% formamide, 5×SSPE (1×SSPE=0.15M NaCl, 10mM NaH_2PO_4, 1mM EDTA, pH7.4), 5×Denhart's solution, 1%SDS, and 100μg/ml denatured salmon sperm DNA at 42°C for 16 hours. The blot was washed with 1×SSC and 0.1% SDS at 50°C. The RNA blot was also probed with DNA coding β-actin for normalization of results.

RESULTS AND DISCUSSION

Isolation of Clones for Mouse HDGF and Two Related cDNAs

We previously isolated the cDNA for human HDGF and reported that human HDGF had a single mRNA band of 2.4kb (2). In Northern blot analysis of mouse transcripts, which were hybridized with the human HDGF cDNA fragment, we found quite a broad band in testis (data not shown). This finding most likely indicates either a single transcriptional product that is differentially processed or the transcription of several related but distinct genes. Therefore, we screened a mouse testis cDNA library with the 0.9-kb cDNA of human HDGF. Twenty clones were isolated that could be assigned to three distinct classes, referred to here as mouse HDGF, HRP-1, and HRP-2, based on sequence relationships. Fourteen of the 20 characterized clones

```
      GACTCCCGCCTCTGACTTCGCCTCTCACCGCCGTTGCTGCTTCCCGCGGCCTGGGCCTCCGCCGACAGC    69
ATGCCGCACGCCTTCAAGCCCGGGGACTTGGTGTTTGCCAAGATGAAGGGCTACCCGCACTGGCCGGCCCGGATT   144
 M  P  H  A  F  K  P  G  D  L  V  F  A  K  M  K  G  Y  P  H  W  P  A  R  I    (25)
GATGACATTGCTGATGGTGCCGTGAAGCCCCCACCCAACAAATACCCCATCTTCTTCTTTGGAACGCATGAAACG   219
 D  D  I  A  D  G  A  V  K  P  P  P  N  K  Y  P  I  F  F  F  G  T  H  E  T    (50)
GCCTTCCTGGGACCCAAGGACCTGTTCCCCTATGATAAGTGCAAGGACAAGTACGGAAAGCCCAACAAGAGGAAA   294
 A  F  L  G  P  K  D  L  F  P  Y  D  K  C  K  D  K  Y  G  K  P  N  K  R  K    (75)
GGCTTCAATGAGGGGCTCTGGGAGATCCAGAACAACCCCCATGCCAGCTACAGTGCGCCTCCGCCGGTGAGCTCC   369
 G  F  N  E  G  L  W  E  I  Q  N  N  P  H  A  S  Y  S  A  P  P  P  V  S  S    (100)
TCTGACAGTGAGGCCCCTGAAGCCGACCTGGGTTGTGGCAGCGACGTGGACAAGGACAAAGAATCCCGGAGGGTC   444
 S  D  S  E  A  P  E  A  D  L  G  C  G  S  D  V  D  K  D  K  E  S  R  R  V    (125)
ATGACTGTGACAGCTGTGACCACCACAGCCACCAGTGACAGGATGGAGAGCGATTCTGACTCTGACAAGAGCAGT   519
 M  T  V  T  A  V  T  T  T  A  T  S  D  R  M  E  S  D  S  D  S  D  K  S  S    (150)
GATCACAGTGGCCTGAAGCGGAAGACACCAGTCTTGAAGGTGTCAGTCTCTAAACGAGCTAGAAGAGCTTCCAGT   594
 D  H  S  G  L  K  R  K  T  P  V  L  K  V  S  V  S  K  R  A  R  R  A  S  S    (175)
GACCTGGATCAGGCCAGTGTGTCCCCATCAGAAGAGGACTCTGAGAGCCCATCTGAGTCGGAGAAGACCAGTGAC   669
 D  L  D  Q  A  S  V  S  P  S  E  E  D  S  E  S  P  S  E  S  E  K  T  S  D    (200)
CAGGATTTTACCCCAGAGAAGAAGACAGCGGCCCGGCCGCCACGGCGAGGTCCTCTGGGAGGACGCAAGAAGAAG   744
 Q  D  F  T  P  E  K  K  T  A  A  R  P  P  P  R  R  G  P  L  G  G  R  K  K  K    (225)
AAGGTGCCATCAGCCTCGGATTCAGACTCCAAAGCCGACTCAGATGGGGCCAAGGAGGAGCCTGTGGTTACAGCA   819
 K  V  P  S  A  S  D  S  D  S  K  A  D  S  D  G  A  K  E  E  P  V  V  T  A    (250)
CAGCCGTCTCCATCCTCCTCGTCGTCGTCCTCATCATCATCCTCCTCAGACTCAGATGTGTCTGTCAAGAAACCC   894
 Q  P  S  P  S  S  S  S  S  S  S  S  S  S  S  S  D  S  D  V  S  V  K  K  P    (275)
CCTCGAGGTAGAAAGCCAGCTGAGAAGCCACCCCCAAAACCCCGAGGGCGGAGACCAAAGCCAGAACGACCACCT   969
 P  R  G  R  K  P  A  E  K  P  P  P  K  P  R  G  R  R  P  K  P  E  R  P  P    (300)
TCCACCTCCAGCAGTGACAGTGACAGTGACAGTGGTGAAGTAGACCGCATCAGTGAATGGAAGAGACGTGATGAA  1044
 S  T  S  S  S  D  S  D  S  D  S  G  E  V  D  R  I  S  E  W  K  R  R  D  E    (325)
GAGCGAAGGCGTGAGCTGGAGGCTCGGAGGCGACGGGAGCAGGAGGAAGAGCTCCGGAGACTTCGAGAACAGGAG  1119
 E  R  R  R  E  L  E  A  R  R  R  R  E  Q  E  E  E  L  R  R  L  R  E  Q  E    (350)
AGGGAGGAGAAGGAGCGGCGCAAGGAGCGGGCAGAGCGTGGGGGCAGCAGTGGAGAGGAGCTGGAGGACGAGGAG  1194
 R  E  E  K  E  R  R  K  E  R  A  E  R  G  G  S  S  G  E  E  L  E  D  E  E    (375)
CCTGTGAAGAAGCGTAGCCGAAAGGCTCGAGGCCGAGGCACACCATCCTCCTCTGACTCAGAACCTGAGGGGGAG  1269
 P  V  K  K  R  S  R  K  A  R  G  R  G  T  P  S  S  S  D  S  E  P  E  G  E    (400)
CTTGGGAAAGAGGGGAAGAAGTTGGCCAAGAAGTCCCAACTGCCAGGCTCAGAGTCTGCAAGGAAGCCTGGGCAA  1344
 L  G  K  E  G  K  K  L  A  K  K  S  Q  L  P  G  S  E  S  A  R  K  P  G  Q    (425)
AAGGAGAAGAGAGGGCGGCCAGATGAGAAGCCACGAGCCAGGCCTGTGAAGGTGGAGCGGACGCGGAAGCGGTCA  1419
 K  E  K  R  G  R  P  D  E  K  P  R  A  R  P  V  K  V  E  R  T  R  K  R  S    (450)
GAGGGTCTCTCACTGGAAAGGAAGGGGGAGAAGAAGAAAGAACCCTCCGTAGAGGAGAGACTGCAAAAGCTGCAC  1494
 E  G  L  S  L  E  R  K  G  E  K  K  K  E  P  S  V  E  E  R  L  Q  K  L  H    (475)
AGCGAGATCAAATTTGCACTGAAGGTTGACAACCCAGATGTGAGGAAGTGTCTGAGTGCACTGGAAGAACTGGGG  1569
 S  E  I  K  F  A  L  K  V  D  N  P  D  V  R  K  C  L  S  A  L  E  E  L  G    (500)
ACTCTGCAGGTAACCTCTCAGATTCTTCAGAAGAACACAGATGTGGTAGCAACGCTGAAGAAGATTCGCCGGTAC  1644
 T  L  Q  V  T  S  Q  I  L  Q  K  N  T  D  V  V  A  T  L  K  K  I  R  R  Y    (525)
AAGGCCAACAAGGATGTGATGGCGAAGGCAGCAGAGGTCTACACCCGGCTCAAGTCACGGGTCCTGGGGCCGAAA  1719
 K  A  N  K  D  V  M  A  K  A  A  E  V  Y  T  R  L  K  S  R  V  L  G  P  K    (550)
GTTGAGGCTCTCCAGAAGGTGAACAAGGCTGGAGCCGAGAAGGAACGGGCCGATAACGAGAAGCTAGAGGAGCAG  1794
 V  E  A  L  Q  K  V  N  K  A  G  A  E  K  E  R  A  D  N  E  K  L  E  E  Q    (575)
CCGGGGGAGCAGGCTCCCAGAGAGCTGGCAGAGGATGAGCCCAGCACCGATCGGTCAGCACCTGTGAATGGTGAG  1869
 P  G  E  Q  A  P  R  E  L  A  E  D  E  P  S  T  D  R  S  A  P  V  N  G  E    (600)
GCCACATCCCAGAAGGGGGAGAACATGGAAGACAGAGCACAGGAAGACGGGCAGGACTCTGAAGATGGCCCCAGG  1944
 A  T  S  Q  K  G  E  N  M  E  D  R  A  Q  E  D  G  Q  D  S  E  D  G  P  R    (625)
GGTGGCTCCTCAGAAGAGCTGCATGACAGCCCACGCGACAACTCGGACCCAGCCAAGCCTGGAAATGAGCGTCAG  2019
 G  G  S  S  E  E  L  H  D  S  P  R  D  N  S  D  P  A  K  P  G  N  E  R  Q    (650)
GACCATGAGAGGACACGGCTGGCCTCTGAGTCTGCCAATGATGACAATGAGGACAGCTGAGCTCCAGCCACTGTG  2094
 D  H  E  R  T  R  L  A  S  E  S  A  N  D  D  N  E  D  S  *                    (670)
GCCATGGTGCCATGTCCATCCCCAGCCTCCCAGGGAAGTGCTGCGCTGTTTGTATTTGTGGTCTGGGCTTTTCTG  2169
CCTAGTTCTGTGATTTCAATTGACATGAAATGGCTATAAAGAATTTTGTAATGCCG                      2225
```

FIG. 3. Nucleotide and translated amino acid sequences of HRP-2 cDNA. Shown is the sequence of sense strand of composite cDNAs and corresponding amino acid sequence beginning at the first initiator methionine at base 70. The sequence of the 98-amino-acid *hath* region is underlined. The mixed charge cluster is double underlined.

were assigned to mouse HDGF. Complete sequencing and comparison of the clones revealed 100% homology and the longest clone (Y26) contained a 714-bp open reading frame (ORF) and 786-bp 3′-untranslated trail (Fig.1A). The translational start assumed corresponds to that determined for human HDGF. Comparison of the mouse HDGF cDNA with the human HDGF cDNA revealed an 88% identities at the DNA level over the coding region and 88% homology at the amino acid level. Especially, the amino terminal 108 amino acids are in 100% match (Fig.1B). In this region, all of the observed base changes (18 bases) occurred at the third position of codons and do not result in a change of the amino acid residue. The putative bipartite NLS of human HDGF is conserved in mouse (residues 155-170) (Fig.1A). This high homology is also extended in the

FIG. 4. Homology of amino acid sequence among mouse HDGF, HRP-1, and HRP-2. HRP-1 and HRP-2 proteins show significant homology to the N-terminal region of mouse HDGF (72 and 68%, respectively, over the 98-amino-acid *hath* region). Shaded areas show sequence identity between at least two of the proteins. Gaps introduced to generate this alignment are represented by dashes. Amino acid residues for each protein are numbered from the initiation methionine.

5′- and 3′-untranslated regions. The 5′-untranslated region of mouse HDGF (61 bp) is 95% homologous to that of the human HDGF gene. In the 3′-untranslated regions of these mRNAs a 92-bp region immediately after the termination codon is 97% identical.

Only one out of the 20 clones were assigned to HRP-1 and the remaining 5 clones fell into the third class, HRP-2. Both HRP-1 and HRP-2 represent novel HDGF related proteins. A 1.5-kb HRP-1 clone (K67) was sequenced in both directions and found to contain an 849-bp ORF encoding a protein of 283 amino acids with a calculated *M*r of 31,665 and a pI of 4.12. The 5′-cDNA was obtained using 5′ rapid amplification of cDNA ends (5′RACE) procedure. An 406-bp DNA fragment was amplified by this technique and then subcloned into pBluescript and sequenced. A novel sequence of 25 bp was added to the 5′ end of K67 (Fig.2). The inserts from two separate HRP-2 clones (Y313 and K12) were completely sequenced on both strands and were both found to be missing 5′ ends, as was the case for all remaining clones in the class. With the help of 5′ RACE procedure, an 459-bp DNA fragment was amplified and then subcloned into pBluescript, sequenced, and found to contain a start codon with a consensus Kozak sequence (9) followed by 350 bp that were identical to that of the 5′ end of Y313. The entire ORF of HRP-2 is 2007 bp encoding a protein of 669 amino acids with a predicted *M*r of 74,287 and a pI of 9.05 (Fig.3).

The nucleotide sequence data reported in this paper have been submitted to the DDBJ, Genbank, and EMBL Data Bank with accession numbers D63707 for mouse HDGF, D63663 for HRP-1, and D63850 for HRP-2.

The Amino Terminal 98-Amino-Acid Region Is Conserved in the Three HDGF-Related Proteins

A comparison of the amino acid sequences for the HDGF-related proteins revealed that these three proteins shared a highly homologous amino terminal region (Fig.4). Except for two short variable stretches (residues 1-10 and 32-41, numbering refers to mouse HDGF), this region has only 2 positions which are not shared between at least two of the three proteins. Many of the differences represent conservative changes. This 98-amino-acid region will be referred to as the *hath* region, for homologous to the amino terminus of the HDGF. This conservation in the N-terminal region is striking compared to the variation in the remainder of the coding region of these proteins. For example, mouse HDGF is 46 and 432 amino acids shorter than HRP-1 and HRP-2 respectively and has no conserved sequence in addition to the *hath* region. The strong conservation suggests that the hath region may have a common biological function. Thus far, since the HDGF-related cDNAs were isolated by nonfunctional screening and the *hath* region does not resemble motifs of any previously identified protein, we have few clues as to what the *hath* region might represent.

A nucleic acid sequence comparison of mHDGF, HRP-1, and HRP-2 cDNAs suggests that these cDNAs are products of three distinct but related genes, rather than the products of alternative splicing. For example, in the stretches of greatest homology, the amino acid sequence is more highly conserved than the nucleic acid sequence.

Amino Acid Analyses of HRP-1 and -2

Both HRP-1 and HRP-2 have no putative amino terminus signal peptide sequence for secretion and reveal compositional biases, particularly of charged amino acids. HRP-1 shows a significant high negative net charge (Lys + Arg:8.4%, Glu + Asp:26.4%) and is rich in proline residues (9.5%). HRP-2 protein carries a significantly high level of basic and acidic residues (Lys + Arg:20.6%, Glu + Asp:19.7%) and is rich in proline and serine redidues (8.5% and 12.4%, respectively). Previous statistical studies have identified invariants and contrasts with respect to amino acid usage for various human protein subclasses (10,11). Thus, human nuclear proteins emphasize hydrophilic residues compared with cellular enzymes and glycoproteins in which hydrophobic residues are foremost. Interestingly, proline and serine residues are also rich in the nuclear proteins (11). These characters raise the possibility of

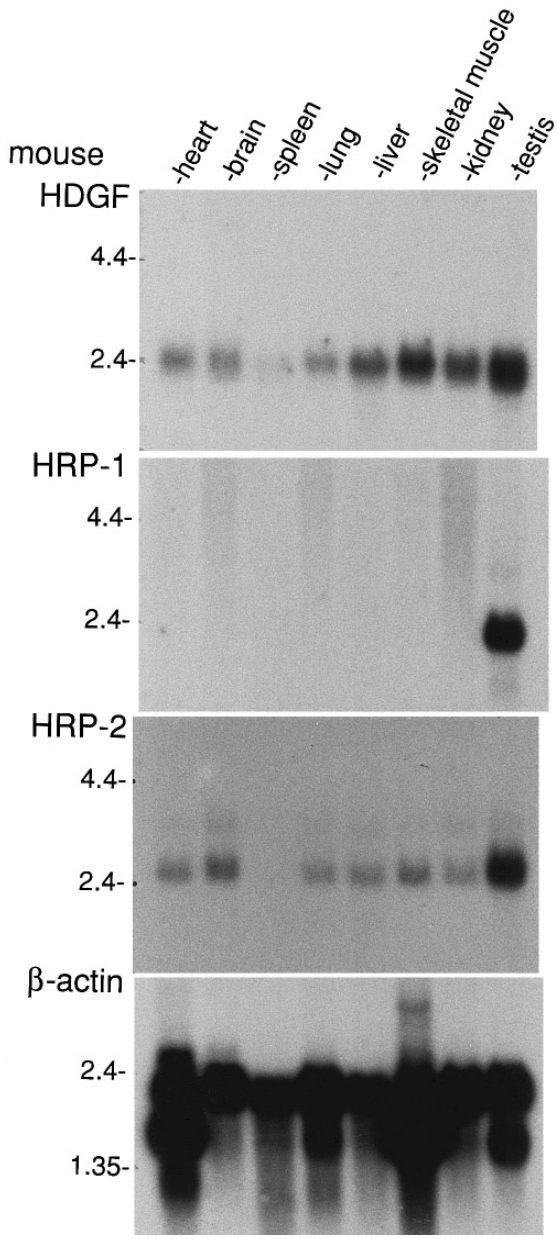

FIG. 5. Northern blot analysis of HDGF family mRNAs in various mouse tissues. Poly(A)$^+$ mRNAs from various tissues of a BALB/c mouse were hybridized with α-^{32}P-labeled gene-specific probes derived from the cDNAs of three different HDGF-related proteins (mouse HDGF, HRP-1, and HRP-2) as well as β-actin. RNA size markers (in kilobases) are indicated at the left of each panel.

nuclear sites of action for HRP-1 and 2. The presence of a putative bipartite type NLS of HRP-1, KRSAEDEEPHCPLKRPR (residues 211-227), supports this idea (Fig. 2). One striking feature of HRP-2 is the presence of a significant charge configuration in which a percentage of charged residues is extremely high (residue 321-363; 20 basic and 16 acidic residues in 43; called 'a mixed charge cluster') (Fig.3). Statistical studies have revealed several associations between significant charge configurations and protein functions (12-17). In these studies, a charge cluster is defined as a short protein segment (25-75 residues) with significant high charge content relative to the charge composition of the whole protein. In particular, a positive (or negative) charge cluster is a segment with positive (or negative) net charge, and a mixed charge cluster is a segment high in charged residues of both signs. Charge clusters are characteristic of eukaryotic reguratory proteins, including transcription and replication factors, development control proteins, high molecular weight heat shock proteins (hsp), and many G protein-coupled receptors (14, 15, 16). By contrast, cellular proteins generally lacking distinctive charge configurations include cytoplasmic structural proteins (myosin, actins), ribosomal proteins (with the exception of the acidic proteins P0, P1, P2), and most enzymes (hydrolases, oxidoreductases, transferases, nontransmembrane kinases) (17). It is plausible, therefore, that HRP-2 protein may function as a regulatory protein.

HDGF Family Members Are Differentially Expressed

To determine the tissue distribution of mouse HDGF, HRP-1, and HRP-2 mRNAs, the corresponding radiolabeled ORF DNAs without the *hath* region were used to probe Northern Blot of poly(A)$^+$ RNA from various mouse tissues (Fig.5). Mouse HDGF and HRP-1 transcripts are expressed in similar pattern; being highly abundant in testis and skeletal muscle; at intermediate levels in heart, brain, lung, liver, and kidney; and minimal or undetectable levels in spleen. For mouse HDGF, in agreement with our previous study of human HDGF, a single transcript of 2.5kb was identified. For HRP-2, two different mRNA bands were detected in each of these tissues. The major band was 2.5kb, whereas the minor band was 3.5kb. The molecular basis and significance of the two transcripts remain to be determined. In contrast, HRP-1 (transcript size: 2.2kb) was highly expressed only in testis.

The selective enrichment of the three HDGF-related genes mRNAs in testis suggests potential roles for the three genes in testis. The 5′-untranslated region of mouse HDGF, HRP-1, and HRP-2 cDNAs contained GC-rich nucleotide sequences (GC content, 83%, 74% and 72%, respectively). Some genes specifically expressed in testis or in embryonic development are reported to have a high content of GC nucleotide sequences at the 5′ region of the cDNA and its promoter sequence (18, 19). The GC-rich sequence might be one of the common functional sequences used for specific gene expression in male germ cell development and is concerned with DNA methylation, chromatin conformation, or translational control (20, 21, 22). Future studies including genomic searches and gene knockout

analyses are needed to sort out biological roles of the each member of this gene family.

REFERENCES

1. Nakamura, H., Kambe, H., Egawa, T., Kimura, Y., Ito, H., Hayashi, E., Yamamoto, H., Sato, J., and Kishimoto, S. (1989) *Clin. Chim. Acta* **183,** 273–284.
2. Nakamura, H., Izumoto, Y., Kambe, H., Kuroda, T., Mori, T., Kawamura, K., Yamamoto, H., and Kishimoto, T. (1994) *J. Biol. Chem.* **269,** 25143–25149.
3. Dingwal, C., and Laskey, R. A. (1991) *Trends Biol. Sci.* **16,** 478–481.
4. Robbins, J., Dilworth, S. M., Laskey, R. A., and Dingwall, C. (1991) *Cell* **64,** 615–623.
5. Jans, D. A. (1994) *FASEB J.* **8,** 841–847.
6. Jaya, M., Lyall, R. M., Mudd, R., Schlessinger, J., and Sarver, N. (1988) *EMBO J.* **7,** 963–969.
7. Imamura, T., Engleka, K., Zhan, X., Tokita, Y., Forough, R., Roeder, D., Jackson, A., Majer, J. A. M., Hla, T., and Macing, T. (1990) *Science* **249,** 1567–1570.
8. Lin, Y., Yao, S., and Hawiger, J. (1996) *J. Biol. Chem.* **271,** 5305–5308.
9. Kozak, M. (1989) *J. Cell Biol.* **108,** 229–241.
10. Brendel, V., Bucher, P., Nourbakhsh, I. R., Blaisdell, B. E., and Karlin, S. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89,** 2002–2006.
11. Karlin, S., Blaisdell, B. E., and Bucher, P. (1992) *Protein Engineering* **5,** 729–738.
12. Karlin, S., Bucher, P., Brendel, V., and Altschul, S. F. (1991) *Annu. Rev. Biophys. Biophys. Chem.* **20,** 175–203.
13. Karlin, S., Blaisdell, B. E., and Brendel, V. (1990) *Methods Enzymol.* **183,** 388–402.
14. Karlin, S. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90,** 5593–5597.
15. Brendel, V., and Karlin, S. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86,** 5698–5702.
16. Karlin, S., and Brendel, V. (1990) *Oncogene* **5,** 85–95.
17. Karlin, S. (1990) *in* Structure and Methods. Volume 2: DNA Protein Complexes and Proteins (Sarma, R. H., and Sarma, M. H., Eds.), pp. 171–180, Adenine, Albany, NY.
18. Rappold, G. A., Stubbs, L., Labeit, S., Crkvenjakov, R. B., and Lehrach, H. (1987) *EMBO J.* **6,** 1975–1980.
19. Ariel, M., McCarrey, J., and Cedar, H. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88,** 2317–2321.
20. Gardine-Garden, M., and Frommer, M. (1987) *J. Mol. Biol.* **196,** 261–282.
21. Antequera, F., Boyes, J., and Bird, A. (1990) *Cell* **62,** 503–514.
22. Kozak, M. (1990) *J. Cell Biol.* **115,** 887–903.