J. Biochem. 125, 795-802 (1999)

# Genomic Organization and Promoter Analysis of a Mouse Homeobox Gene, Hex[1]

Zaw Myint,*,† Tetsuya Inazu,* Takashi Tanaka,* Kazuya Yamada,* Vincent W. Keng,* Yoshiko Inoue,* Masaru Kuriyama,† and Tamio Noguchi*,‡,[2]

*Department of Biochemistry and †Department of Internal Medicine, Fukui Medical University, Matsuoka, Fukui 910-1193; and ‡Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Chikusa-ku, Nagoya 464-8601

Received November 16, 1998; accepted January 4, 1999

**A homeobox gene, Hex, is mainly expressed in haematopoietic cells and hepatocytes. It is assumed to play a role in the early stage of differentiation of these cells. To understand the mechanisms involved in the regulation of the Hex gene expression in hepatocytes, we cloned and characterized the mouse Hex gene. The gene consists of four exons and three introns, and spans about 5.7 kb. All the exon-intron boundaries are consistent with the "GT-AG" rule. A single transcription start site was identified by primer extension and S1 mapping analyses. Although the 5′-flanking region is G/C rich (69%), it contains probable "TATA and CCAAT" boxes. Potential binding sequences for transcriptional regulatory proteins including Sp1 and AP-2 are also present in this region. Functional analysis of the Hex promoter was performed by transfecting $MH_1C_1$, HeLa, COS-7, and Caco-2 cells with Hex promoter region-luciferase constructs. We found three possible positive regulatory regions, comprising of nucleotides −199 and −172, −154 and −133, and −105 and −68, respectively, required for Hex gene expression in $MH_1C_1$ cells by analyses of a series of 5′-deletion constructs of the fusion genes. The activities of these constructs were extremely low in HeLa, COS-7, and Caco-2 cells suggesting that they possess cell-type specificity. Further analysis revealed two GC boxes, GC box1 and GC box2, at nucleotides −197 to −188 and −176 to −167, respectively, necessary for Hex gene expression. Thus, multiple regulatory elements contribute to the Hex gene expression in hepatocytes.**

**Key words: gene structure and regulation, hepatocyte, Hex, homeobox, tissue specificity.**

Homeobox genes are members of a family of transcription factors that regulate tissue development in many different organisms (*1*). First discovered in *Drosophila* as regulators of body segmentation formation, they share a highly conserved 60 amino acid segment termed the homeobox that mediates DNA binding (*2*). Essential roles for homeodomain proteins in vertebrate development have been demonstrated by many studies in which their expression was modified by targeted disruption of murine homeobox genes (*3, 4*). Homeodomain factors also regulate cell lineage-specific gene expression patterns in vertebrates; liver- (*5*), thyroid- (*6*), and pancreas- (*7*) specific genes are regulated by DNA binding proteins containing homeo-like domains.

The divergent homeobox gene, Hex (*8*), also known as Prh (*9, 10*), is expressed in a range of multipotent haematopoietic progenitor cells and cell lines, and is generally downregulated during terminal cell differentiation (*11*) suggesting a role for it in the early stages of haematopoietic cell differentiation.

Moreover, Hex is preferentially expressed in liver cells (*9, 10, 12*). In *Xenopus laevis*, XHex is first expressed in the dorsal endomesoderm of the gastrula stage embryo. This tissue goes on to contribute to the structures of the embryonic liver and XHex continues to be expressed in the liver throughout development (*13*). Recently, it was reported that the Hex gene exhibited peri-implantation asymmetry in the mouse embryo and was an early transient marker of endothelial cell precursors. Localization of its expression to the thyroid as well as the liver primordia makes it not only one of the earliest markers for these primordia but may also indicate that Hex plays a role in their subsequent development (*14*). We have also reported that Hex expression in mouse embryonic tissues was detected exclusively in the hepatic anlage and thyroid primordium at E (embryonic age) 9.5. At E 12.5 and E 15.5, its expression persisted in the fetal liver and thyroid, and was also detected in the fetal lung (*12*). For the rat homologue, the expression of Hex mRNA was observed in the liver and spleen among adult rat tissues examined and was also detected in the fetal liver from 15 days gestation.

[1] The nucleotide sequence data reported in this paper will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases under the accession numbers, AB017130, AB017131, and AB017132, respectively.

[2] To whom correspondence should be addressed at: Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Chikusa-ku, Nagoya 464-8601. Phone: +81-52-789-4121, Fax: +81-52-789-5050, E-mail: tnoguchi@agr.nagoya-u.ac.jp

Abbreviations: E, embryonic age; RACE, rapid amplification of cDNA ends.

The hepatic expression was only slightly increased after partial hepatectomy but greatly altered at different differentiation stages: the highest expression was observed in freshly isolated hepatocytes, followed by well differentiated hepatoma cells such as $MH_1C_1$, but no Hex mRNA was detected in poorly differentiated hepatoma cells (Tanaka, T., Inazu, T., and Noguchi, T., unpublished results). All these results suggest that Hex may play a role in liver cell differentiation.

Bedford *et al.* isolated the mouse Hex gene but only showed the positions of introns in Hex coding sequence (*8*). However, the genomic sequence including the 5′-flanking sequences as well as the entire genomic structure of the Hex gene have never been reported. Thus, to investigate the molecular mechanisms of the tissue-specific expression and developmental regulation of the mouse Hex gene, we cloned and characterized it. In the present study, we isolated a genomic clone that spans more than 15 kb in length and encodes the entire mRNA as well as the 5′-flanking sequences, and thus determined the structure of the whole Hex gene. Furthermore, we examined the 5′-flanking region of the Hex gene using a transient luciferase expression assay. Here we report the identification of two *cis*-acting elements, GC box1 and GC box2, comprising of nucleotides −197 to −188 and −176 to −167, respectively, that may regulate expression of the Hex gene in hepatocytes.

## EXPERIMENTAL PROCEDURES

*Genomic Library Screening*—One million recombinants from a mouse 129 liver genomic library constructed in λEMBL3 were screened with a random primer-labeled 725 bp rat Hex 3′ cDNA probe (Tanaka, T., Inazu, T., and Noguchi, T., unpublished results) by the plaque-hybridization procedure (*15*). A single positive genomic clone, λmHex1, was plaque-purified, digested with various restriction enzymes, and then analyzed by Southern blotting using either the same probe or a 293 bp long, further upstream 5′ probe (*16*). Further analysis was performed by sequence determination after subcloning into pBluescript-SKII$^+$ (Stratagene). All the exonic regions, exon-intron boundaries, and the 5′ terminus and 3′ terminus regions were sequenced using either universal primers or synthetic Hex gene-specific primers (*17*).

*Primer Extension and S1 Nuclease Protection Analyses*—Total RNA from mouse tissues was isolated using the acid-phenol extraction method (*18*). Then polyadenylated RNA was purified using oligo(dT)-cellulose (Pharmacia LKB Biotechnology). Primer extension was performed as described by Ghosh *et al.* (*19*) using an antisense oligonucleotide primer, 5′ AGCGGCGTGGGCGCATACAG, spanning 71 to 52 nucleotides downstream of the translation start site. The products were analyzed on 7% polyacrylamide/urea gels using a M13 sequence ladder as markers.

S1 nuclease protection was carried out as described (*20*). The probe used was a 394 bp *Xma*I fragment containing the 5′-flanking region of the genomic DNA. The 5′ *Xma*I site was derived from the polycloning site of the pBluescript-SKII$^+$ vector.

*3′ Rapid Amplification of cDNA Ends (3′ RACE)*—We obtained mouse (BALB/c males) liver Marathon ready™ cDNA (Clontech) which has an adaptor sequence ligated to each end. Adaptor primer 1 (5′ CCATCCTAATACGACTCACTATAGGGC) and a Hex gene-specific primer spanning amino acids 134 to 140 were used for the first-round PCR reaction for 35 cycles each at 94°C for 1 min, with annealing at 56°C for 1 min, and extension at 72°C for 2 min using Ex Taq polymerase (Takara Shuzo). Then the two primers were removed and a nested PCR was performed using nested adaptor primer 2 (5′ ACTCACTATAGGGCTCGAGCGGC) and a Hex gene-specific primer spanning amino acids 189 to 195 for 30 cycles each at 94°C for 1 min, with annealing at 58°C for 1 min and extension at 72°C for 2 min. The amplified nested PCR product was gel electrophoresed and Southern hybridized with a *Bgl*II/*Eco*RI digested rat Hex 3′ cDNA probe. About 1.3 kb long hybridized fragment was cloned into the pT7Blue T-vector (Novagen), and the positive clones, which had been hybridized again with the same probe, were sequenced with an ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer).

*Construction of Promoter-Luciferase Constructs*—Promega's luciferase pGL3-Basic vector was digested with *Nco*I, blunt-ended with the Klenow fragment and then digested with *Bgl*II. A 390 bp long *Bam*HI/*Sma*I fragment of the mouse Hex gene (−341 to +46 relative to the cap site of Hex) was isolated and inserted into the above vector to construct pHluc341. For the construction of pHluc67 containing a −67/+46 fragment, the pHluc341 plasmid was digested with *Sma*I in the polylinker region and *Sac*II in the Hex gene, blunt-ended using T4 DNA polymerase and then self-ligated. Plasmids pHluc238, pHluc211, pHluc199, pHluc172, pHluc154, pHluc132, and pHluc105, which contain fragments −238/+46, −211/+46, −199/+46, −172/+46, −154/+46, −132/+46, and −105/+46, respectively, were produced from pHluc341 by 5′-deletion using exonuclease III and mung bean nuclease. Double-stranded oligonucleotides of H-I and H-II were synthesized and separately inserted into the *Nhe*I and *Xho*I sites of the reporter plasmid, pHluc67. The nucleotide sequences of these oligonucleotides are shown in Table I.

*Transfection and Luciferase Expression Assay*—$MH_1C_1$ (rat hepatoma), HeLa (human cervical carcinoma), COS-7 (SV40 transformed African green monkey kidney), and Caco-2 (human colon adenocarcinoma) cells were grown in DMEM (Dulbecco's Modified Eagle Medium) (Nissui Pharmaceuticals) supplemented with 10% fetal calf serum,

TABLE I. **Oligonucleotide sequences used in this study.** M denotes a mutant oligonucleotide, and the numbers represent the 5′ and 3′ ends of the oligonucleotides. The bases mutated are underlined.

| Element | Sequence | Position |
|---|---|---|
| H-I | CTAGCGGGGGCGGGAGCTGGGCCGGT | −199 to −178 |
| H-II | TCGAGGGCCGGTGGGGGCGGATC | −185 to −167 |
| M-196/-191 | CGTGCGTCTCGGACTAGTGAGCTGGGC | −208 to −182 |
| M-175/-170 | CTGGGCCGGTGGACTAGTATCGGCGCGGGG | −187 to −158 |

0.03% L-glutamine and 0.015% kanamycin. Transfection was performed using the calcium phosphate precipitation method (*21*) with 5 $\mu$g of various reporter constructs, 0.1 $\mu$g of plasmid pRL-SV40 as an internal control and 2.9 $\mu$g of pBluescriptSKII$^+$ vector. The cells were then incubated at 37°C for 4-6 h, after which the medium was replaced with fresh medium. The cells were harvested after 42-44 h and lysed in 0.4 ml of Passive lysis buffer (Promega). The cell extract was used to measure the firefly luciferase activity using the Dual-Luciferase™ Reporter Assay System (Promega). The results were normalized according to the *Renilla* luciferase (RL) activity of pRL-SV40. The pGL3-Control vector, which contains the SV40 promoter as well as the SV40 enhancer was used as a positive control, and the relative activities of the constructs were expressed as percentages of that of pGL3-Control. All transfections were performed at least three times in duplicate.

*Oligonucleotide-Directed Mutagenesis*—This was essentially performed by the Kunkel method with some modifications (*22*). A Hex promoter fragment, −211/−68, was subcloned into the *Kpn*I-*Sac*II sites of the pBluescriptSKII$^+$ vector. The single-stranded template DNA prepared in the *Escherichia coli* strain, CJ236, was separately annealed to M-196/-191 and M-175/-170, phosphorylated mutagenic oligonucleotides (Table I). The complementary DNA strand was synthesized with T4 DNA polymerase and T4 DNA ligase. The resulting heteroduplex DNA was used to transform DH5$\alpha$ cells. The mutated DNAs were confirmed by sequencing and cloned into the *Kpn*I-*Sac*II sites of pHluc211, and then used for the transient transfection assay.

## RESULTS

*Structure of the Mouse Hex Gene*—On the screening of $10^6$ recombinants with $^{32}$P-labeled cDNA for rat Hex as a probe, one genomic clone, $\lambda$mHex1, was isolated from a mouse 129 genomic library, and analyzed by restriction mapping and Southern blotting. The locations of exons in the mouse Hex gene were determined by Southern blot analysis (data not shown) and DNA sequencing in comparison with the reported mouse Hex coding sequence (*8*). There are three nucleotide differences between the murine Hex cDNA and its gene in the first and second exons. However, these changes do not affect the amino acid sequence. Since no sequences of the 5′ and 3′ untranslated regions of the Hex cDNA were available, the 5′ noncoding region was determined by primer extension and S1 nuclease analyses (see below). The cDNA clones containing the 3′ untranslated region were isolated by 3′ RACE analysis. The longest clone is about 1.3 kb in length and includes 922 bp of the 3′ untranslated region in addition to the protein coding region. The 922 bp sequence perfectly matches that of the Hex genomic DNA. However, the number of poly A residues in this clone was identical to that of dT residues in the primer used for preparation of the cDNA library. In addition, no consensus polyadenylation sequence was found within the 40 bp sequence upstream from the 3′ end of the cDNA. So considering the size of the mouse Hex mRNA, we searched for a potential polyadenylation signal in about 120 bp downstream sequence derived from the Hex genomic clone and found one potential signal sequence, AATATA, at 71 bp downstream from the 3′ end of the cDNA. Thus the length of the mRNA is about 1.85 kb, excluding the poly A sequence, which corresponds to the value of 1.9 kb determined by Northern blot analysis (*8, 12*). Thus, the cloned gene contains the entire sequence of the mouse Hex gene, which is separated into four exons by three introns and spans about 5.7 kb. Figure 1 shows the restriction endonuclease map of the cloned mouse Hex gene. The translation start codon is located within the first exon. The homeobox is in the second and third exons. The exons vary in size from 51 bp for exon 3 to over 1 kb for exon 4, which contains the translation stop codon and the entire 3′ untranslated region. The exon-intron boundaries are listed in Table II. GT and AG residues are present at the 3′ and 5′ boundaries of exons, consistent with the consensus sequence for the splicing of eukaryotic mRNA (*23, 24*).

*Determination of the Transcription Start Site*—We examined the transcription initiation site of the mouse Hex gene by primer extension and S1 nuclease mapping analyses. Figure 2A shows the results of 5′ end analysis of the mouse Hex mRNA by primer extension analysis. The 20 bp synthetic antisense oligonucleotide (71 to 52 nucleotides


Fig. 1. **Genomic structure of a mouse homeobox gene, Hex.** Exons are represented by boxes. Hatched and open boxes represent coding and non-coding regions, respectively. The restriction enzymes used are indicated as follows: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; SII, *Sac*II. The scale for 1 kb is indicated.

TABLE II. **Characteristics of the exon-intron junctions.** The sizes of the exons and introns, along with the exon-intron boundary sequences, are shown. Exon sequences are given in capitals and introns in lower case letters. Each intron begins with a GT and ends with an AG. The consensus sequence was obtained from Sharp (*23*) and Mount (*24*).

| Exon | | Exon/intron junction | | Intron |
|---|---|---|---|---|
| Number | Size (bp) | 5′ Donor | 3′ Acceptor | Size (kb) |
| 1 | 392 | ACCCCCTGG gtaagg. | ctggcccacag GCAAGCCCT | 1.8 |
| 2 | 179 | GAGAGACAG gtcagc. | cttttctgtag GTCAAAACC | 0.2 |
| 3 | 51 | CTGAAACAG gtatcg. | cccatcgccag GAGAATCCT | 1.8 |
| 4 | >1141 | | | |
| | Consensus Sequence | CAG gtaagt. | ttttttttncag G<br>A g ccccccc t | |

downstream from the translation start site) of the mouse Hex cDNA was labeled at the 5′ end and used as a primer. Only one band of 99 bp was detected for liver mRNA, but not for small intestine mRNA, which does not include Hex mRNA. The size of this product indicates that the 5′ untranslated region of the mouse Hex mRNA is 28 nucleotides in length.

S1 nuclease mapping was performed to compare the results of primer extension analysis (Fig. 2B). When assayed using the 394 bp fragment as a probe for S1 mapping, a single band of 47 bp was detected for liver mRNA but not for small intestine mRNA. This result was same as that of primer extension analysis. Thus, we conclude that there is only one transcription initiation site in the mouse Hex gene, which is located at 28 nucleotides upstream of the translation initiation site. The site starts at a G, that is numbered position +1 (Fig. 3).

*Sequence of the 5′-Flanking Region*—We determined 651 bp of the 5′-flanking sequence of the mouse Hex gene. This region is likely to contain at least some elements that may account for the regulated expression of the mouse Hex gene in different tissues and at different stages of development. The sequence that we determined is shown in Fig. 3 in addition to that of a part of exon 1. The 5′-flanking region is moderately G/C rich (69%). A TATA-like promoter sequence (*25*), ATAAAT, was found 30 bp upstream from the putative cap site. Similarly, a CCAAT box-like sequence, CGAAT, was found at position −139 relative to the

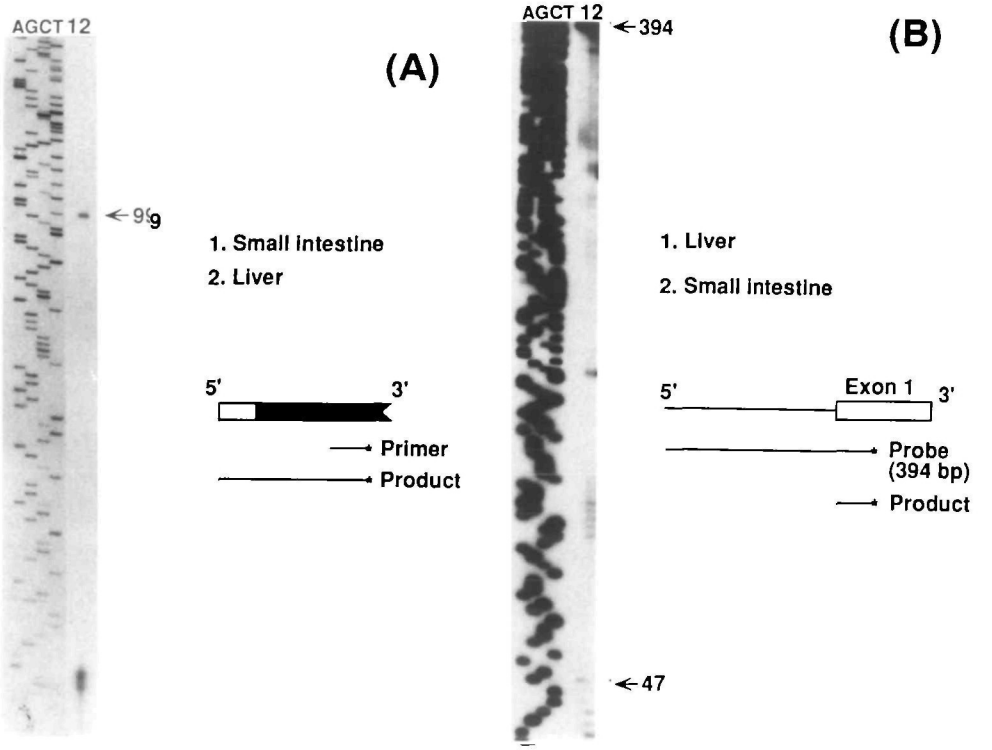


Fig. 2. **Mapping of the 5′ end of the mouse Hex mRNA by primer extension analysis (A), and S1 nuclease protection assay (B).** Polyadenylated RNA from mouse liver or small intestine was hybridized with the $^{32}$P-labeled DNA probes indicated, and then extended by reverse transcriptase (A) or digested with nuclease S1 (B). The products were electrophoresed on 7% polyacrylamide/urea gels using sequence ladders of M13 mp18 DNA as size markers. The lengths of the products are shown on the right. The translated region is indicated by a solid box in (A) and the open box indicates the exon of genomic DNA in (B). The asterisks show the position of $^{32}$P radioactivity.

```
                                                                    -595
            GGAGCAGATTCCGTTAGCATCCAGGGAGGTGGCCGAGCTAGGGGGTCCTCAAATAGG
-594                                                                -529
   TCTGGAACACCTGCAGCCTCGCTGTCAGTACCCGTCTGGGACTAGGACGCAGCCAACGGCTTGGGT
-528                                                  AP-2           -463
   CTTCTCCTTGGAGCGCCGAGCTGCGGAGGTCCCCCGTTTTAGACAGTGCCACCGTTACCCCAAGAA
-462                  MZF1                                           -397
   GCGCAGCTTTCTGGCCGCCCGAAGGTGCCATCCTGCCGGCGGCATCATGGCCCGCAGAGTCCCATC
-396                        AP-2                                     -331
   GCTTCTCCAGGTTCCAGAGGCTGCCCGAGTTTCAGCAGTGGCGCCATCCCGTCAGGATCCGCGCCC
-330                                                  Sp1  AP-2      -265
   CGCGGCGCACGTCTTGTGACCTCCTCTTCGACGGTCGCAGTCGCTCGCTAGGTGCCTGAAACCTCT
-264  AP-2                                                           -199
   GGGGGCGCCTGCGGCGCAGGAACACCCTGGCGGAGGGTCCCCGGAGCCCCACAGCCCGTGCGTCTC
-198  AP-2                                                           -133
   GGGGGCGGGAGCTGGGCCGGTGGGGGCGGATCGGCGCGGGGGTTAGTGGGGGGCGGAGGCGAATCT
-132    Sp1        AP-2 MZF1 Sp1      AP-2      MZF1     Sp1  CTF/NF1  -67
   GAAGCCAGCGCCATTGGCCGCAGGGCCTGGCGGCAGGAAGGGGACCGAGCGCGGCCCCACCCCGCG
-66                 AP-2                                  Sp1  AP-2    -1
   GCTGCCCGAGCCTATCGCCTGGAGCGCAGAGCGCAAATAAATGTAGCGCGGCGCCGCCGGCCGGCA
+1                          ... HNF3                                  +66
   GCTCTCTGCGAGGGGCTGCGGAGCGGCCATGCAGTTCCCGCACCCGGGGCCCGCGGCTGCGCCCGC
```

Fig. 3. **Nucleotide sequences of portions of the 5′-flanking region and the first exon of the mouse Hex gene.** Position +1 refers to the proposed transcription start site. Putative TATA and CCAAT boxes are boxed. Potential binding sites for the indicated transcription factors are underlined and the translation initiation codon is shown by closed circles.

putative cap site. Comparison of the 5′-flanking sequence with the MatInspector transcription factor binding site profile database (*26*) revealed a number of elements exhibiting homology to known factor binding sites. Among them, potential binding sites for Sp1, AP-2, MZF1, and HNF3 were indicated in this region.

*Localization of Regulatory Elements in the 5′-Flanking Region*—The 5′-flanking region may include regulatory elements necessary for cell-type specific expression of the mouse Hex gene. To elucidate the regulatory region of the mouse Hex gene in hepatocytes, we constructed pHluc341, which contains a fragment comprising of nucleotides −341 to +46 relative to the transcription start site of the Hex gene linked to the firefly luciferase gene. A series of 5′-deletion mutants of this fusion gene were also constructed. These plasmids were transfected into $MH_1C_1$ cells and then the transient expression of luciferase activity was examined. $MH_1C_1$ cells are well-differentiated hepatoma cells which express a similar level of the Hex gene transcript to freshly isolated hapatocytes (Tanaka, T., Inazu, T., and Noguchi, T., unpublished results). The pGL3-Control vector was used as a positive control and its activity was adjusted to 100%. Then the relative luciferase activities of various constructs were measured. The results are shown in Fig. 4. The highest promoter activity was observed with deletion of positions −341 to −211 in $MH_1C_1$ cells. There was no significant reduction in the activity with a further 12 bp deletion. However, deletion up to −172 caused a significant decrease in activity. This suggested that the region from −199 to −172 may contain a positive regulatory element. Another two possible positive regulatory regions are the region covering −154 to −133 and that spanning −105 to −68. Negative regulatory elements may also be present in the −238 to −211 and −172 to −154 regions. To determine whether or not the regulatory elements of the mouse Hex gene show cell-type specificity, the deletion mutants were transfected into HeLa cells (Fig. 4) as well as COS-7 and Caco-2 cells (data not shown), which do not express this gene. In these cells, all transfected plasmids were essentially inactive. These results suggested that the 5′-flanking region up to −199 is necessary for cell-type specific full promoter activity of the Hex gene and contains multiple positive elements.

*Identification of Cis-Acting Positive Elements*—To determine whether or not the −199 to −172 region contains actual positive regulatory elements, the overlapping double-stranded oligonucleotides, called the H-I and H-II boxes, corresponding to the −199 to −178 and −185 to −167 regions, respectively, were synthesized and inserted into pHluc67 separately. Two plasmids were obtained, and both of which contained two copies of the oligonucleotides in the sense direction. These constructs were transfected into $MH_1C_1$ cells and then luciferase activity was measured

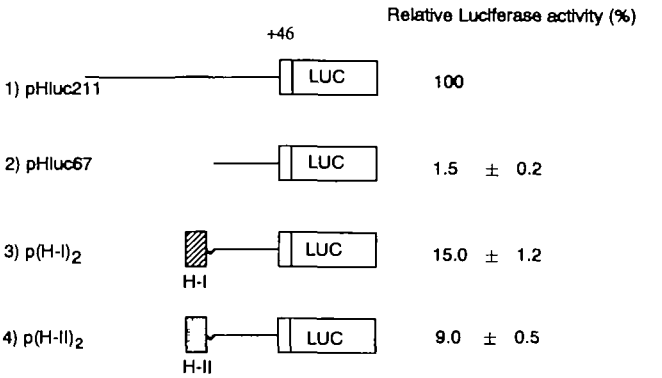


Fig. 5. **Identification of *cis*-acting positive elements.** pHluc211 (1), the activity of which was adjusted to 100%, and pHluc67 (2), which showed almost no promoter activity, are shown at the top two lines. Two copies of the H-I (3) and H-II (4) double-stranded oligonucleotides shown in Table I were ligated to pHluc67. The plasmids were transfected into $MH_1C_1$ cells and the luciferase activities were determined as described in Fig. 4. The data are the means ± SE for three experiments. The straight lines indicate the sequence of the Hex gene and the notches represent that of the polylinker site of pGL3-Basic.

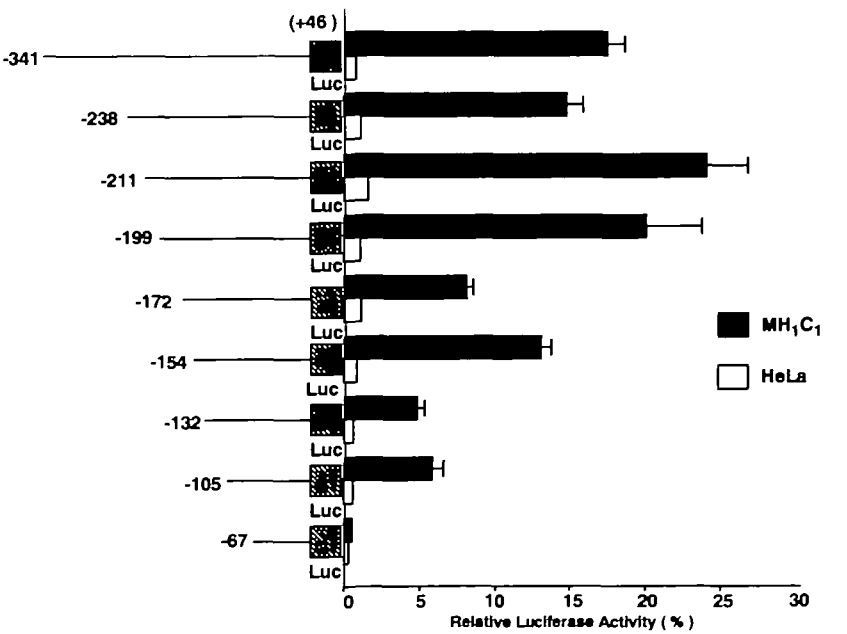


Fig. 4. **Expression of the Hex-luciferase fusion genes in $MH_1C_1$ and HeLa cells.** The series of 5′-deletion constructs shown on the left were transfected into $MH_1C_1$ and HeLa cells with pRL-SV40. The relative luciferase activities of the deletion constructs are expressed as percentages of that of pGL3-Control. The data are the means ± SE for three experiments. Luc, luciferase gene.

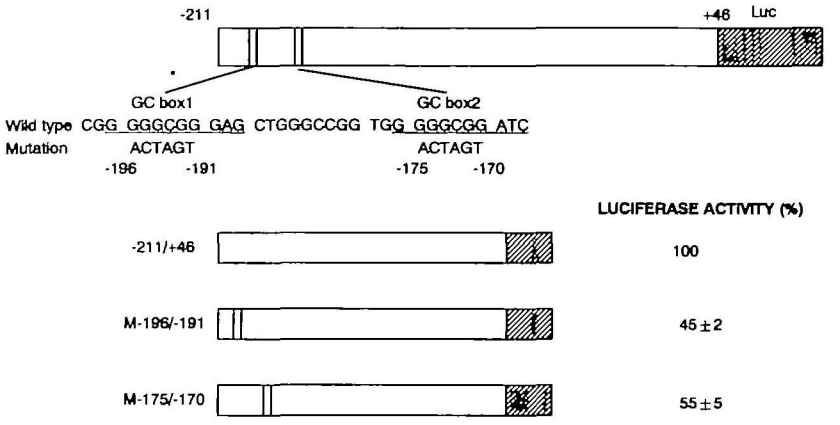


Fig. 6. **Effects of mutations in the two GC boxes of the Hex promoter in $MH_1C_1$ cells.** A schematic representation of the −211/+46 Hex promoter and the locations of the mutated sites are presented. The numbers indicate the 5′ and 3′ ends of each promoter insert, numbered in relation to the transcription start site. Luciferase activity is expressed relative to that of the −211/+46-bp construct. The data are the means±SE for three experiments.

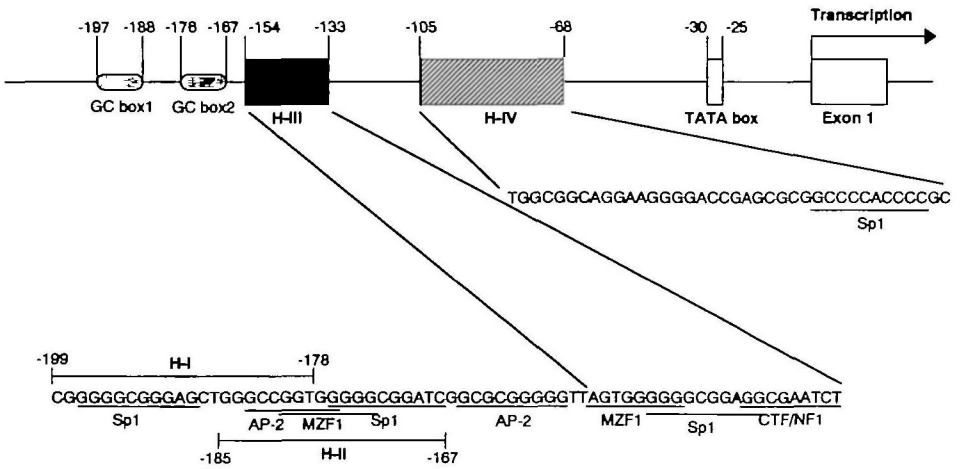


Fig. 7. **Schematic representation of the regulatory regions of the mouse Hex gene.** Nucleotide residues are numbered negatively from the transcription start site of the Hex gene. GC box1 and GC box2 were demonstrated to be involved in regulation of the Hex gene expression. Another two potential positive regulatory regions, called H-III and H-IV, are also shown.

(Fig. 5). The activities of $p(H\text{-}I)_2$ and $p(H\text{-}II)_2$ were 10-fold and 6-fold higher than that of pHluc67, respectively, indicating that these regions, now termed H-I and H-II, respectively, indeed contain the positive *cis*-acting elements required for Hex gene expression.

According to the results in Fig. 5, we assumed that two GC boxes, termed GC box1 and GC box2, at positions −197 to −188 and −176 to −167, respectively, may contribute to the positive activities of the respective H-I and H-II regions. Thus, we introduced 6-bp block mutations at the −196/−191 site for GC box1 and the −175/−170 site for GC box2 separately, and then examined the effects of these perturbations in the context of the −211/+46-bp promoter (Fig. 6). These separate mutations at −196/−191 and −175/−170 caused 55 and 45% decreases in activity, respectively, compared with that of the wild type plasmid, pHluc211. Therefore, we conclude that GC box1 and GC box2 contribute to the positive regulation of the Hex gene expression in hepatocytes.

## DISCUSSION

In this paper, we report the genomic organization of the mouse Hex gene, defining the exon-intron boundaries and portions of the 5′-flanking DNA. We also performed functional analysis of its promoter region. The Hex gene consists of four exons and three introns, and is about 5.7 kb long. The second intron is rather short, comprising only 200 bp. The sequences of all exon-intron boundaries are in accord with the consensus sequences for exon-intron boundaries of eukaryotic genes (*23, 24*). Although the 5′-flanking region has a high G/C content and three repeats of a hexanucleotide Sp1 binding GC box (GGGCGG), it is unlikely that it resembles that of other so-called housekeeping genes (*27*) because of the presence of TATA and CCAAT box-like promoter sequences and the identification of the only one cap site. However, the functional significance of these sequences remains to be determined. In addition, we identified the putative polyadenylation signal considering the size of Hex mRNA. But this sequence, AATATA, is one base different from the consensus sequence. Thus, further studies are required to confirm this.

Accumulating evidence indicates that the 5′-flanking region of the gene contains the *cis*-acting DNA elements responsible for tissue-specific expression (*28, 29*). Thus, the region upstream from the cap site of the Hex gene may include the DNA sequences responsible for tissue-specific expression of that gene. The transient luciferase assay revealed that the 5′-flanking region up to −199 was necessary for full promoter activity in transfected $MH_1C_1$ cells. However, very low promoter activity was observed when the reporter gene containing that region was transfected into HeLa, COS-7, and Caco-2 cells. These results indicate that the 5′-flanking region up to −199 is responsible for cell-type specific expression of the Hex gene. We found multiple possible positive regulatory regions re-

quired for expression of the Hex gene in hepatocytes. Then, we identified two positive regulatory elements, designated as GC box1 and GC box2, in the nucleotide −197 to −167 region by using double-stranded oligonucleotides and site-specific mutagenesis.

The positions and sequences of the two GC boxes in addition to two possible regulatory regions named H-III and H-IV are shown schematically in Fig. 7. It is very likely that members of the transcription factor Sp1 family bind to the two GC boxes to regulate the Hex gene expression. These members include Sp2, Sp3, and Sp4 in addition to Sp1 (*30-32*). They function as transcriptional activators or repressors depending upon the gene in question (*31, 33*). Sp1 binding sites are also present in two other possible H-III (−154 to −133) and H-IV (−105 to −68) regulatory regions. However, the functional significance of these sites remains to be determined. In addition, potential binding sites for other transcription factors such as Ap-2, MZF1, and CTF/NF1 are also included in the region from −199 to −133. Among them, MZF1 is unlikely to be involved in the Hex gene expression in hepatocytes since it is expressed almost exclusively in early myeloid progenitor cells (*34*). However, the Hex gene is also expressed in these cells. Thus, MZF1 may be involved in expression of this gene in the progenitor cells.

Although the computer analysis showed a potential binding site for liver-enriched transcription factor HNF3 (*35*) at nucleotides −36 to −26, it may not be involved in regulation of Hex gene expression in hepatocytes since this site containing pHluc67 exhibits very low promoter activity. Transcription factors such as the GC box binding Sp1 family, AP-2 (*36*) and CTF/NF1 (*37*) are expressed ubiquitously. Therefore, it is difficult to explain the cell-type specific transcription of the Hex gene only by GC box binding factors, although these boxes are known to be involved in this gene expression. Further studies are required to solve this problem.

## REFERENCES


1. Levine, M. and Hoey, T. (1988) Homeobox proteins as sequence-specific transcription factors. *Cell* **55**, 537-540
2. Gehring, W.J. (1987) Homeoboxes in the study of development. *Science* **236**, 1245-1252
3. Chisaka, O. and Capecchi, M.R. (1991) Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene hox-1.5. *Nature* **350**, 473-479
4. Joyner, A.L., Herrup, K., Auerbach, B.A., Davis, C.A., and Rossant, J. (1991) Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the En-2 homeobox. *Science* **251**, 1239-1243
5. Frain, M., Swart, G., Monaci, P., Nicosia, A., Stampfli, S., Frank, R., and Cortese, R. (1989) The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. *Cell* **59**, 145-157
6. Guazzi, S., Price, M., DeFelice, M., Damante, G., Mattei, M.-G., and DiLauro, R. (1990) Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. *EMBO J.* **9**, 3631-3639
7. Karlsson, O., Thor, S., Norberg, T., Ohlsson, H., and Edlund, T. (1990) Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. *Nature* **344**, 879-882
8. Bedford, F.K., Ashworth, A., Enver, T., and Wiedemann, L.M. (1993) HEX: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human. *Nucleic Acids Res.* **21**, 1245-1249
9. Crompton, M.R., Bartlett, T.J., MacGregor, A.D., Manfioletti, G., Buratti, E., Giancotti, V., and Goodwin, G.H. (1992) Identification of a novel vertebrate homeobox gene expressed in haematopoietic cells. *Nucleic Acids Res.* **20**, 5661-5667
10. Hromas, R., Radich, J., and Collins, S. (1993) PCR cloning of an orphan homeobox gene (PRH) preferentially expressed in myeloid and liver cells. *Biochem. Biophys. Res. Commun.* **195**, 976-983
11. Manfioletti, G., Gattei, V., Buratti, E., Rustighi, A., De Iuliis, A., Aldinucci, D., Goodwin, G.H., and Pinto, A. (1995) Differential expression of a novel proline-rich homeobox gene (Prh) in human hematolymphopoietic cells. *Blood* **85**, 1237-1245
12. Keng, V.W., Fujimori, K.E., Myint, Z., Tamamaki, N., Nojyo, Y., and Noguchi, T. (1998) Expression of Hex mRNA in early murine postimplantation embryo development. *FEBS Lett.* **426**, 183-186
13. Newman, C.S., Chia, F., and Krieg, P.A. (1997) The XHex homeobox gene is expressed during development of the vascular endothelium: overexpression leads to an increase in vascular endothelial cell number. *Mech. Dev.* **66**, 83-93
14. Thomas, P.Q., Brown, A., and Beddington, R.S.P. (1998) Hex: a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors. *Development* **125**, 85-94
15. Benton, W.D. and Davis, R.W. (1977) Screening lambdagt recombinant clones by hybridization to single plaques *in situ*. *Science* **196**, 180-182
16. Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517
17. Sanger, F., Nicklen, S., and Coulsen, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
18. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159
19. Ghosh, P.K., Reddy, V.B., Piatak, M., Lebowitz, P., and Weissman, S.M. (1980) Determination of RNA sequences by primer directed synthesis and sequencing of their cDNA transcripts. *Methods Enzymol.* **65**, 580-595
20. Berk, A.J. and Sharp, P.A. (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* **12**, 721-732
21. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*, pp. 1633-1636, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
22. Kunkel, T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**, 488-492
23. Sharp, P.A. (1981) Speculations on RNA splicing. *Cell* **23**, 643-646
24. Mount, S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459-472
25. Breathnach, R. and Chambon, P. (1981) Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**, 349-383
26. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **23**, 4878-4884
27. Lowy, D.R. and Willumsen, B.M. (1993) Function and regulation of ras. *Annu. Rev. Biochem.* **62**, 851-891
28. Edlund, T., Walken, M.M., Barr, P.J., and Rutter, W.J. (1985) Cell-specific expression of the rat insulin gene: evidence for role of two distinct 5′ flanking elements. *Science* **230**, 912-916
29. Widen, S.G. and Papaconstantinou, J. (1986) Liver-specific expression of the mouse alpha-fetoprotein gene is mediated by *cis*-acting DNA elements. *Proc. Natl. Acad. Sci. USA* **83**, 8196-8200
30. Hagen, G., Muller, S., Beato, M., and Suske, G. (1992) Cloning by recognition site screening of two novel GT box binding pro-

teins: a family of Sp1 related genes. *Nucleic Acids Res.* **20**, 5519-5525
31. Hagen, G., Muller, S., Beato, M., and Suske, G. (1994) Sp1-mediated transcriptional activation is repressed by Sp3. *EMBO J.* **13**, 3843-3851
32. Kingsley, C. and Winoto, A. (1992) Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. *Mol. Cell. Biol.* **12**, 4251-4261
33. Majello, B., De Luca, P., and Lania, L. (1997) Sp3 is a bifunctional transcriptional regulator with modular independent activation and repression domains. *J. Biol. Chem.* **272**, 4021-4026
34. Hromas, R., Collins, S.J., Hickstein, D., Raskind, W., Deaven, L.L., O'Hara, P., Hagen, F.S., and Kaushansky, K. (1991) A retinoic acid-responsive human zinc finger gene, MZF-1, preferentially expressed in myeloid cells. *J. Biol. Chem.* **266**, 14183-14187
35. Cereghini, S. (1996) Liver-enriched transcription factors and hepatocyte differentiation. *FASEB J.* **10**, 267-282
36. Mitchell, P.J., Wang, C., and Tjian, R. (1987) Positive and negative regulation of transcription *in vitro*: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. *Cell* **50**, 847-861
37. Santoro, C., Mermod, N., Andrews, P.C., and Tjian, R. (1988) A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. *Nature* **334**, 218-224