THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1989 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 264, No. 4, Issue of February 5, pp. 2363–2367, 1989
Printed in U.S.A.

Rat Pyruvate Kinase M Gene

ITS COMPLETE STRUCTURE AND CHARACTERIZATION OF THE 5′-FLANKING REGION*

(Received for publication, July 20, 1988)

Masaru Takenaka, Tamio Noguchi, Hiroyasu Inoue‡, Kazuya Yamada, Tamiko Matsuda, and Takehiko Tanaka

From the Department of Nutrition and Physiological Chemistry, Osaka University Medical School, Nakanoshima, Kita-ku, Osaka 530, Japan

Genomic clones containing the rat pyruvate kinase M gene, which encodes the M_1- and M_2-type isozymes, were isolated and their exon sequences were determined. This gene contains 12 exons and 11 introns and is 20 kilobases (kb) long. The sequences specific to the M_1- and M_2-types exist in exons 9 and 10, respectively (Noguchi, T., Inoue, H., and Tanaka, T. (1986) *J. Biol. Chem.* 261, 13807–13812). The seventh intron begins with the GC dinucleotide instead of the consensus GT dinucleotide, but other exon-intron boundaries are consistent with the "GT-AG" rule. S1 mapping analysis showed that M_1- and M_2-type mRNAs had multiple, but the same transcription initiation sites. Thus, the M_1- and M_2-type isozyme mRNAs are concluded to be produced from the same M gene transcript by alternative RNA splicing. RNA blot hybridization analysis indicated that developmental changes of the isozymes in brain and skeletal muscle were regulated at the level of RNA splicing.

The 5′-flanking region of the gene has no "TATA box" or "CAAT box," but contains potential Sp1 binding sites. Bacterial chloramphenicol acetyltransferase assay revealed that a fragment of about 0.5 kb of the 5′-flanking region of the gene was sufficient for promoter activity in the rat hepatoma cell line, dRLh-84. This activity was not present in adult rat hepatocytes, indicating that the 0.5-kb fragment has tissue-specific promoter activity.

A processed-type pseudogene that resembles the M_2-type pyruvate kinase cDNA was also characterized.

Pyruvate kinase (ATP:pyruvate O^2-phosphotransferase, EC 2.7.1.40), a key enzyme in the glycolytic pathway, has four isozymes in mammals: the M_1-, M_2-, L-, and R-types (1–3). The expressions of these isozymes are tissue-specific and are regulated developmentally. The R-type isozyme is expressed only in erythroid cells (1–4). The M_2-type, considered as the prototype isozyme, is the only form detected in early fetal tissues and is present in most adult tissues (2). The L-type is the major isozyme in adult liver and a minor form in kidney and intestine (2). The M_1-type is the major form in adult skeletal muscle, heart, and brain. The M_1- and L-types replace the M_2-type progressively in these tissues during development.

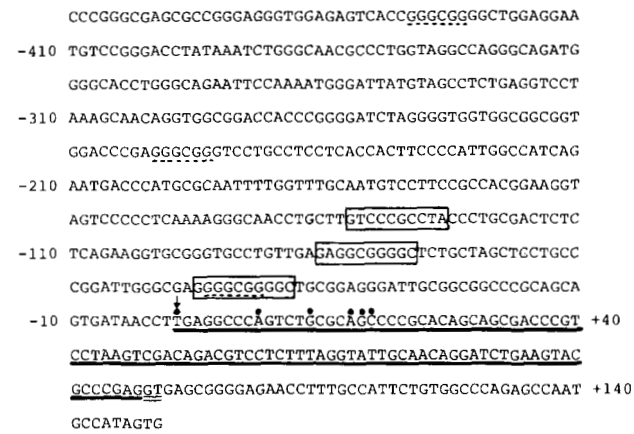

FIG. 4. **Nucleotide sequence of upstream region of the rat pyruvate kinase M gene.** The sequence is that of the first noncoding exon and 5′-flanking region. The first exon is *underlined* and the GT dinucleotide at the exon-intron boundary is shown by a *double line*. The multiple transcription sites shown by S1 nuclease analysis are indicated by *closed circles*. The *arrow* indicates nucleotide 1. The six-base Sp1 binding GC boxes are shown by *dotted lines*, and decanucleotide consensus sequences of the Sp1 binding site are shown by *open boxes*.

To investigate the molecular mechanisms of tissue-specific expression and developmental regulation of pyruvate kinase isozyme genes, we isolated genomic clones encoding the four isozymes. We found that two genes were responsible for production of the L- and R-types, and the M_1- and M_2-types, respectively (5, 6). Alternative use of different promoters of the L gene generates the L- and R-type isozymes (5). On the other hand, results suggested that alternative RNA splicing of the same primary transcript from the M gene produced the M_1- and M_2-types (6). This idea was based on partial sequence analysis of a genomic clone including the M_1- and M_2-specific sequences, but a final conclusion could not be made until the organization of the M gene had been determined.

In the present work, we isolated several overlapping genomic clones that extend over a length of more than 20 kb[1] and encode the entire mRNA as well as 5′-flanking sequences, and thus determined the structure of the whole PKM gene. We also showed that the 5′-flanking region of the PKM gene

* This investigation was supported by grants from the Ministry of Education, Science and Culture of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04458 and J04459.

‡ Present address: Dept. of Biochemistry, National Cardiovascular Center Research Institute, Suita, Osaka 565, Japan.

[1] The abbreviations used are: kb, kilobases; PKM, pyruvate kinase M; CAT, chloramphenicol acetyltransferase; bp, base pairs.

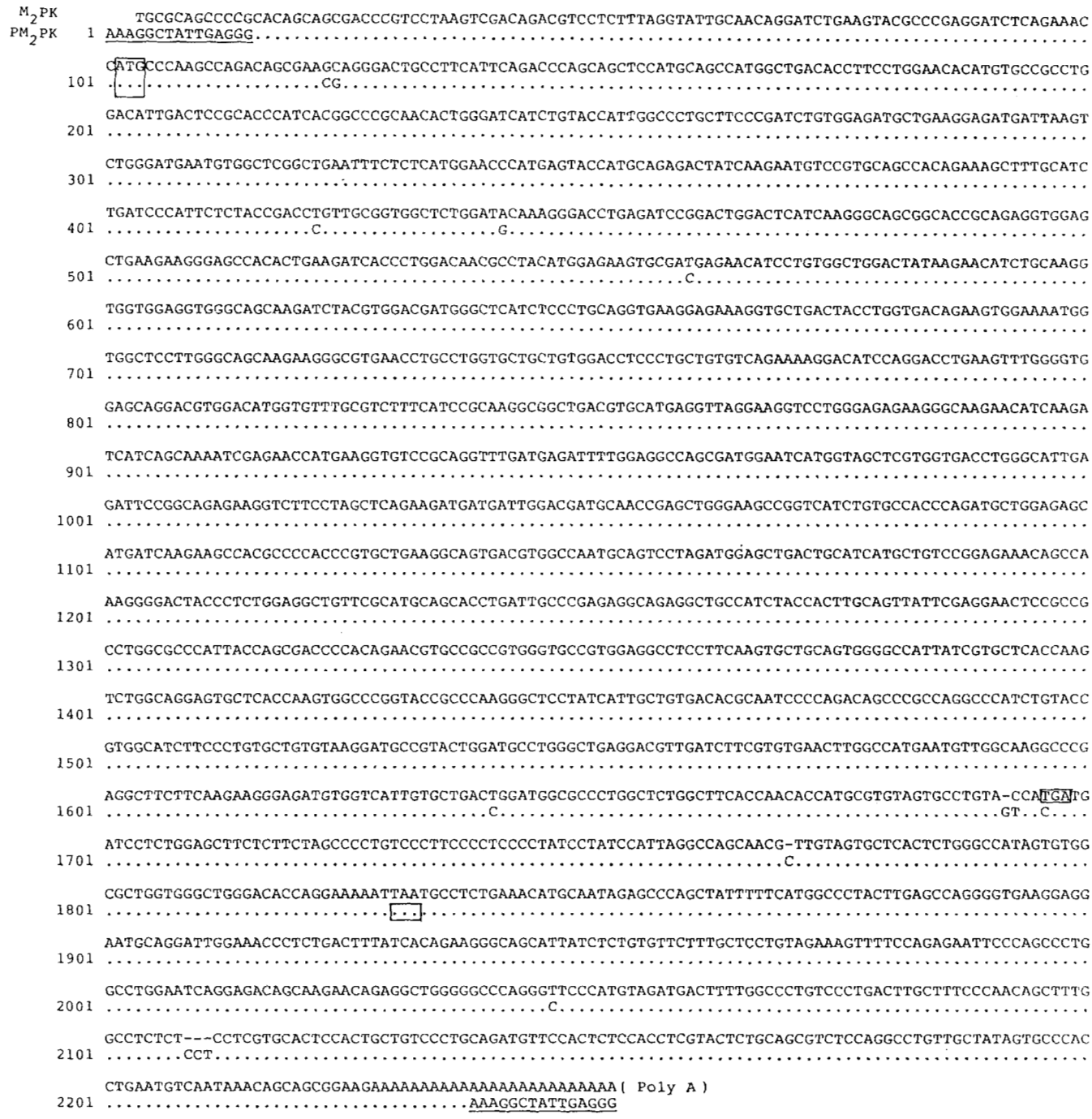

FIG. 6. **Nucleotide sequences of the intronless gene.** The ATG codon and stop codon are shown by *open boxes*. The 15-bp direct repeats that bound the intronless gene are *underlined*. This intronless gene can code for a protein-like M_2-type pyruvate kinase. Its deduced amino acid sequence is 45 amino acids longer than that of the M_2-type pyruvate kinase because of the insertion of a single nucleotide at position 1692. *M_2PK*, M_2-type pyruvate kinase cDNA; *PM_2PK*, intronless gene.

has promoter activity when fused to the chloramphenicol acetyltransferase gene and transfected into a rat hepatoma cell line. In addition, we characterized the genomic intronless gene, and analyzed the M_1- and M_2-type isozyme mRNAs in skeletal muscle and brain during development.

EXPERIMENTAL PROCEDURES AND RESULTS[2]

[2] Portions of this paper (including "Experimental Procedures," "Results," Table 1, and Figs. 1–3, 5, and 7) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

DISCUSSION

In this study, we determined the structure of the rat M-type pyruvate kinase gene including that of the 5′-flanking region. The PKM gene consists of 12 exons and 11 introns and is about 20 kb long. The M_1- and M_2-specific sequences are present in exons 9 (M_1) and 10 (M_2) respectively, as reported earlier (6). The first intron was rather long: 8.7 kb. The sequences of all exon-intron boundaries except one are in accord with the consensus sequences for exon-intron boundaries (21, 22). The 5′-end of the seventh intron differs from the consensus GT dinucleotide: it begins with a GC dinucleotide. This change was confirmed by sequencing two

independent clones, λMPK34 and λMPK37. Change from the consensus GT dinucleotide to GC has also been found in the murine αA-crystallin gene (28), the mouse adenine phosphoribosyltransferase gene (29), and the duck (30) and chicken (31) α^D-globin genes. Wieringa *et al.* (32) showed that the mutation converting GT to AT at the 5′-end of an intron of the rabbit β-globin gene led to the use of three normally unused splice sites. Two of the unused splice sites began with a GT sequence and one of them began with a GC sequence. These data suggest that the GC dinucleotide can be used as a splice site in some cases instead of the consensus GT dinucleotide.

The 5′-flanking region of the PKM gene resembles that of other so-called housekeeping genes. It does not have a typical TATA box or CAAT box, but has a high G/C content and three repeats of a hexanucleotide Sp1 binding GC box (GGGCGG). These features reflect the presence of multiple cap sites in this gene (24). Recently, a decanucleotide consensus sequence for the Sp1 binding site ($5'\begin{matrix}GG\\TA\end{matrix}GGCG\begin{matrix}GGGC\\TAAT\end{matrix}$) has been reported (25, 26). These sites were also identified in the 5′-flanking region of the PKM gene (Fig. 4). As the Sp1 transcription factor has been shown to stimulate transcription 10- to 50-fold (25, 26, 33), we assumed that Sp1 may be used in transcription of the PKM gene. CAT assay revealed that a 0.5-kb fragment, which contained the three hexanucleotide GC boxes and the three decanucleotide consensus Sp1 binding sequences, was sufficient for promoter activity in transfected dRLh-84 cells. However, no promoter activity was observed when the fragment was transfected into adult rat hepatocytes. These results indicate that the 0.5-kb fragment of the 5′-flanking region has a tissue-specific regulatory element.

S1 protection assay indicated that M_1- and M_2-type mRNAs have the same transcription initiation sites. As discussed previously (6), gene rearrangement is unlikely to be involved in expression of the two isozymes. Thus, these data demonstrate that the mRNAs of M_1- and M_2-type isozymes have the same primary transcripts, which indicates that the two types are produced by alternative RNA splicing of the same primary transcript (6). Northern blot analysis showed that developmental changes of pyruvate kinase isozymes ($M_1 \rightarrow M_2$) in skeletal muscle and brain were due to alterations in the mRNA levels. Therefore, alternative RNA splicing of the PKM gene transcript is a very important regulatory step for developmental changes of the isozymes. However, its mechanism remains to be determined.

We also characterized a genomic clone that has no introns (Fig. 6). This gene contained the remnants of a poly(A)$^+$ tail in the 3′-untranslated region and was bounded by direct repeats of 15 bp. It may have been derived from a spliced mRNA. The 15-bp direct repeats, which flank the intronless gene, may have been used when reinsertion into the genome occurred (34). The sequence between the direct repeats showed 99% homology with that of the M_2-type mRNA. This suggested that the reinsertion happened very recently. An open reading frame of the intronless gene is capable of coding for a protein-like M_2-type pyruvate kinase, but 45 amino acids longer than the latter. However, no such protein has been reported. The transcript of this gene was also not detected as S1 nuclease digestion did not protect a 59-bp fragment expected from homology between this gene and the probe used (Fig. 3). Furthermore, the 5′-flanking region of the intronless gene did not show any promoter activity in transfected dRLh-84 cells or hepatocytes. These results indicate that this intronless gene is not expressed and is a processed-type pseudogene.

Acknowledgment—We wish to thank Dr. Hiromu Nakajima for helpful discussions.

REFERENCES

1. Tanaka, T., Haranao, Y., Sue, F., and Morimura, H. (1967) *J. Biochem. (Tokyo)* **62,** 71–91
2. Imamura, K., and Tanaka, T. (1972) *J. Biochem. (Tokyo)* **71,** 1043–1051
3. Nakashima, K., Miwa, S., Oda, S., Tanaka, T., Imamura, K., and Nishina, T. (1974) *Blood* **43,** 537–548
4. Imamura, K., Tanaka, T., Nishina, T., Nakashima, K., and Miwa, S. (1973) *J. Biochem. (Tokyo)* **74,** 1165–1175
5. Noguchi, T., Yamada, K., Inoue, H., Matsuda, T., and Tanaka, T. (1987) *J. Biol. Chem.* **262,** 14366–14371
6. Noguchi, T., Inoue, H., and Tanaka, T. (1986) *J. Biol. Chem.* **261,** 13807–13812
7. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
8. Maniatis, T., Fritsche, E., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
9. Sanger, F., Nicklen, S., and Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 2460–2464
10. Messing, J., Crea, R., and Seeburg, P. H. (1981) *Nucleic Acids Res.* **9,** 309–321
11. Berk, A. J., and Sharp, P. A. (1977) *Cell* **12,** 721–732
12. Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982) *Mol. Cell. Biol.* **2,** 1044–1051
13. Sato, J., Yabe, T., and Chikata, E. (1965) *Jpn. J. Exp. Med.* **35,** 491–511
14. Sato, J., Yabe, T., and Chikata, E. (1970) *Acta Med. Okayama* **24,** 559–57
15. Chu, G., Hayakawa, H., and Berg, P. (1987) *Nucleic Acids Res.* **15,** 1311–1326
16. Seglen, P. O. (1976) *Methods Cell Biol.* **13,** 29–83
17. Crabb, B. W., and Dixon, J. E. (1987) *Anal. Biochem.* **163,** 88–92
18. Lehrach, H., Diamond, D., Wozney, J. M., and Boedtker, H. (1977) *Biochemistry* **16,** 4743–4751
19. Noguchi, T., Inoue, H., Chen, H.-L., Matsubara, K., and Tanaka, T. (1983) *J. Biol. Chem.* **258,** 15220–15223
20. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13
21. Sharp, P. A. (1981) *Cell* **23,** 643–646
22. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472
23. Noguchi, T., and Tanaka, T. (1982) *J. Biol. Chem.* **257,** 1110–1113
24. Breathnach, R., and Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383
25. Briggs, M. R., Kadonaga, J. T., Bell, S. P., and Tjian, R. (1986) *Science* **234,** 47–52
26. Kadonaga, J. T., Jones, K. A., and Tjian, R. (1986) *Trends Biochem. Sci.* **11,** 20–23
27. Imamura, K., Noguchi, T., and Tanaka, T. (1986) in *Markers of Human Neuroectodermal Tumors* (Staal, G. E. J., and van Veelen, C. W. M., eds) pp. 191–122, CRC Press Inc., Boca Raton, FL
28. King, C. R., and Piatigorsky, J. (1983) *Cell* **32,** 707–712
29. Dush, M. K., Sikela, J. M., Khan, S. A., Tischfield, J. A., and Stambrook, P. J. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 2731–2735
30. Erbil, C., and Niessing, J. (1983) *EMBO J.* **2,** 1339–1343
31. Dodgson, J., and Engel, J. D. (1983) *J. Biol. Chem.* **258,** 4623–4629
32. Wieringa, B., Meyer, F., Reiser, J., and Weissmann, C. (1983) *Nature* **301,** 38–43
33. Dynan, W., and Tjian, R. (1985) *Nature* **316,** 774–778
34. Rogers, J. H. (1985) *Int. Rev. Cytol.* **93,** 187–279

Continued on next page.

SUPPLEMENTARY MATERIAL TO:
Rat pyruvate kinase M gene. Its complete structure and characterization of the 5' flanking region.
Masaru Takenaka, Tamio Noguchi, Hiroyasu Inoue, Kazuya Yamada, Tamiko Matsuda, and Takehiko Tanaka.

EXPERIMENTAL PROCEDURES

Materials - T4 DNA ligase, T4 polynucleotide kinase, S1 nuclease, Klenow DNA polymerase, a M13 sequence kit and some restriction endonucleases were from Takara Shuzo. T4 DNA polymerase and other restriction endonucleases were from Toyobo. Sequenase,a sequence kit, was from United States Biochemical. Acetyl coenzyme A was from Sigma. Ham's F12 medium was obtained from Nissui Seiyaku and Williams' medium E was from Flow Laboratories. Nitrocellulose filters were from Schleicher and Schuell. Oligo (dT)-cellulose was from Collaborative Research. [alpha-^{32}P]dCTP (3000Ci/mmol), [alpha-^{32}P]dATP (3000Ci/mmol), [gamma-^{32}P]ATP (3000Ci/mmol), [^{14}C]chloramphenicol (60mCi/mmol) and Colony/Plaque Screen were from New England Nuclear.

Isolation and Characterization of Genomic DNA Clones - As described previously (6), we obtained several clones after screening a Charon 4A rat genomic library constructed from partial HaeIII digests of rat liver DNA using EcoRI linkers (obtained from Dr. J. Bonner, California Institute of Technology). Although these clones were characterized by restriction mapping, Southern blotting (7) and sequence determination, none of them contained the entire 5' non-coding region. To obtain a clone containing the 5' non-coding exon and 5' flanking region, we prepared a genomic library from rat liver DNA in EMBL4 using completely digested HindIII fragments with EcoRI linkers (8). After screening this library with a ^{32}P-labeled probe, one independent clone, λMPK5, was obtained which was characterized further.

Phage DNAs of λMPK37, λMPK34 and λMPK5 were subcloned into pUC18 or 19 vectors. Nucleotide sequence analysis was carried out by the chain termination method after subcloning the restriction fragments into M13 mp vectors (9,10).

During the screening of a Charon 4A rat genomic library, we found the intronless gene, λMPK14, and characterized it as well.

S1 Nuclease Mapping - The appropriate probe for S1 mapping was isolated, dephosphorylated with bacterial alkaline phosphatase and labeled at the 5' end with [gamma-^{32}P]ATP by T4 polynucleotide kinase. S1 analysis was performed as described by Berk and Sharp (11). S1-protected fragments were analyzed by electrophoresis on 8% polyacrylamide/8M urea gel using sequence ladders of M13 mp18 DNA as markers.

Plasmid Constructions and CAT Expression - We constructed a CAT vector that was useful for inserting the test promoter by ligating a HindIII/BamHI fragment containing the CAT gene isolated from pSV0cat (12) into the HindIII site of pUC18. This plasmid was named pUC0cat. An AccI/BamHI fragment containing the

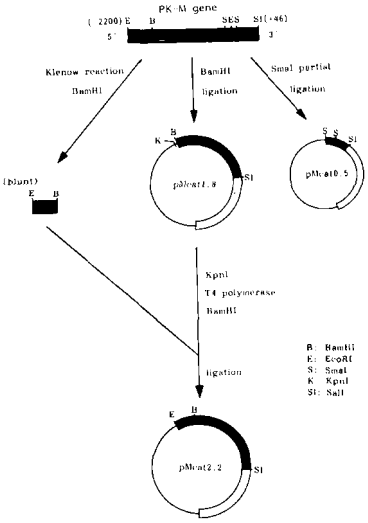

Fig. 1. Plasmid construction for chloramphenicol acetyl transferase assay. - Fragments of about 2.2kb, 1.8kb and 0.5kb from the SalI site at nucleotide +46 in the first exon of the PKM gene were introduced into pUC0cat as described under "Experimental Procedures". Closed boxes represent the fragments of the 5' flanking region of the PKM gene. Open boxes represent the CAT gene, the polyadenylation site and small t intron from SV40. Lines represent the pUC18.

SV40 early enhancer, promoter and CAT gene, was isolated from pSV2cat (12) and ligated into the AccI/BamHI sites of pUC19. This plasmid was named pUC2cat. A BamHI/SalI fragment (nucleotides +46 to about -1800) containing the 5' flanking region of the PKM gene was introduced into pUC0cat by ligation at the BamHI and SalI sites (pMcat1.8), as shown in Fig. 1. This BamHI/SalI fragment was partially digested with SmaI to isolate a 0.5kb SmaI/SalI fragment. This 0.5kb fragment was ligated to the SmaI/SalI sites of pUC0cat (pMcat0.5). A 1.8 kb EcoRI/EcoRI fragment from λMPK5 was isolated, treated with Klenow fragment to make a blunt end, and digested with BamHI to isolate an EcoRI/BamHI fragment of about 400 bp. The pMcat1.8 was digested with KpnI, treated with T4 DNA polymerase to generate a blunt end, and digested with BamHI. The EcoRI/BamHI fragment of about 400 bp was inserted into this pMcat1.8 and the construct was named pMcat2.2. An EcoRI/SalI fragment of about 1.5kb of the 5' flanking region of the intronless gene λMPK14 was introduced into the SalI site of pUC0cat digested with BamHI/SalI, treated with Klenow fragment to make a blunt end, and ligated. We named this construct pMPcat. These plasmid DNAs (20μg) were used to transfect the rat hepatoma cell line, dRLh-84 (13,14), and adult rat hepatocytes by the electroporation method (15). The rat hepatoma cell line, dRLh-84, was provided by the Japanese Cancer Research Resources Bank, and hepatocytes were prepared as described by Seglen (16). After transfection, dRLh-84 cells and hepatocytes were cultured in Ham's F12 medium and Williams' medium E with 10% fetal calf serum, respectively, for 48 hours, and harvested for determination of CAT activity as described (12,17). pUC2cat and pUC0cat were used as positive and negative controls, respectively.

RNA Blot Hybridization Analysis - Total brain and muscle RNA (25μg) were separated under denaturing conditions as described (18), and transferred to a nitrocellulose filter. The probes used were ^{32}P-labeled complementary strands synthesized from M13 mp DNAs. DNA sequences specific to the M_1-type (88-bp AluI/AluI fragment in exon M_1) and M_2-type (85-bp TaqI/PstI fragment in exon M_2) (6) were subcloned into M13 mp vectors. After synthesis of the complementary strand in the presence of [alpha-^{32}P]dCTP, the ^{32}P-DNA was digested with appropriate restriction endonucleases and the ^{32}P-labeled probe was isolated by electrophoresis on a 5% polyacrylamide/urea sequencing gel. Hybridization was carried out with the ^{32}P-probe in the presence of 50% formamide at 50°C for 2 days as described previously (19).

Other Methods - Preparation of plasmid DNA, phage DNA and liver DNA, electrophoresis of DNA, and restriction mapping were carried out as described by Maniatis et al. (8). Hybridization probes were labeled with [alpha-^{32}P]dCTP or [alpha-^{32}P]dATP with random oligonucleotide primers (20).

RESULTS

Isolation and Nucleotide Sequence of Genomic Clones of the M Gene - As described previously (6), we isolated several genomic clones from a Charon 4A

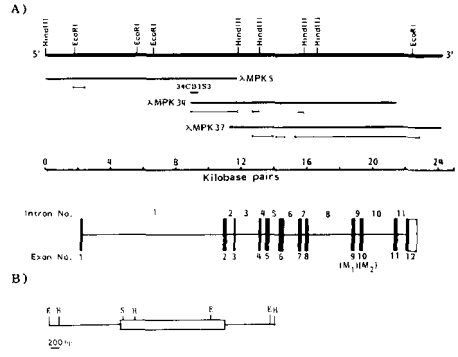

Fig. 2. Structure of the pyruvate kinase M gene and the intronless gene.
A) The three overlapping genomic clones used to characterize the gene. The sequenced regions are shown below the genomic clones. The 34CB1S3 fragment was used as the screening probe. Exons are represented by boxes. Closed and open boxes represent coding and non-coding exons, respectively. Exons 9 and 10 encode the M_1- and M_2-type specific sequences, respectively.
B) Restriction map of the intronless gene. The open box indicates the sequences that resemble those of the M_2-type cDNA: E,EcoRI; S,SalI; H,HindIII.

Table 1

Sequences of the Exon-Intron Boundaries

Intron Number	Exon	Sequence of Exon-Intron Junctions 5'	Intron	3' Exon
1	TACGCCCGAG	gtgagc	.....	atctccgggatccag GATCTCAGAA
2	TGTACCATTG	gtgagt	.....	aatgtttcttggtag GCCCTGCTTC
3	AACCCATGAG	gtgagt	.....	catggaacccatgag TACCATGCAG
4	CATCAAGGGC	gtgagt	.....	ctgtctttaatctag AGCGGCACCG
5	AAGGAGAAAG	gtatgt	.....	atcctttcacacag GTGCTGACTA
6	GTGTCCGCAG	gtgaag	.....	ctttctttttatcag GTTTGATGAG
7	CGCCACCCAG	gcatgt	.....	tttgtcccctcctag ATGCTGGAGA
8	GCAGCACCTG	gtgagt	.....	ccctaaaccttacag ATAGCTCGAG
9	AGTCTGGCAG	gtaggg	.....	atgttgctcccctag ATTGCCCGAG
10	AGTCTGGCAG	gtagga	.....	ttcccaccttctcag GAGTGCTCAC
11	ATGAATGTTG	gtatgt	.....	tccttctcccctcag GCAAGGCCCG
Consensus Sequence	$^{C}_{A}$AG	$gt^{a}_{g}agt$	.....	$ttttttttttn^{c}_{t}ag$ G

The 5'-end of the seventh intron which is underlined differs from the consensus GT dinucleotide. The consensus sequence was obtained from Sharp (21) and Mount (22).

rat genomic library with ^{32}P-labeled M_2-type cDNA as a probe, and analyzed the clone λMPK37 by partial sequence determination. The results showed that this clone contained sequences specific to the M_1- and M_2-types as two distinct and adjacent exons in addition to a few exons which encode sequences common to the two isozymes. Since these features coincide with those of the active genes, we analyzed clone λMPK37 further and found that it did not contain a 5'-terminal region including the coding sequence. We therefore analyzed another overlapping clone, λMPK34, which contained a sequence extending about 2.5kb in the 5' direction. The 5'-terminal coding sequence and part of the noncoding sequence were identified in this extended portion, but the remaining 5'-untranslated sequence and 5' flanking region were not found in this clone. To obtain a genomic clone containing these sequences, we analyzed restriction digests of liver DNA by Southern blotting using two probes. One was a 5'-terminal fragment (34CB1S3,439bp EcoRI/Sau3AI) of λMPK34; the other was a 5'-noncoding cDNA fragment (100bp PstI/Sau3AI, dG/dC tail plus nucleotides -97 to -32 in Ref. 6). As both probes hybridized to a HindIII fragment of about 11.3kb, we isolated HindIII fragments of 9 to 14kb from liver DNA and constructed a genomic library in EMBL4 using these fragments with EcoRI linkers. After screening this library with the two probes described above, we isolated clone λMPK5. EcoRI digestion of λMPK5 DNA yielded four fragments of 5, 3.8, 1.8 and 1 kb named E1, E2, E3 and E4, respectively. These fragments were subcloned into a pUC vector. Southern blot analysis showed that the E2 fragment (pλMPK5E2) had a 5' non-coding exon (data not shown). The E2 fragment was sequenced after subcloning into M13 mp vector. The sequenced regions and restriction maps of these clones are shown in Fig. 2. The PKM gene has 12 exons and 11 introns extending over a length of about 20kb. Exons 9 and 10 contained the sequences unique to the M_1- and M_2-types, respectively, as reported (6), while the other exons encoded sequences common to both isozymes. Table 1 shows the sequences at the exon and intron boundaries. All but one boundary are consistent with the consensus splicing sequence. The 5' end of seventh intron begins with GC instead of the GT consensus dinucleotide (21, 22). This was confirmed by determining the sequences of two independent clones, λMPK34 and λMPK37.

Identification of the Transcription Initiation Sites - We examined the transcription initiation site of the PKM gene by S1 nuclease mapping analysis. As rat skeletal muscle contains only the M_1-type pyruvate kinase (2, 23, Fig. 7), we used it as a source of the M_1-type mRNA. For a similar reason, we chose AH-130 Yoshida ascites hepatoma cells as a source of the M_2-type mRNA (2, 23). We employed a 361 bp Sau3AI/Sau3AI fragment (nucleotides -279 to +82 in Fig. 4) from subclone pλMPK5E2V2 labeled at the 5' end as a probe for S1 mapping. This subclone was derived from pλMPK5E2. When this fragment was hybridized with mRNAs of rat skeletal muscle and AH-130 Yoshida ascites hepatoma cells and subjected to S1 nuclease digestion, the same six protected fragments were detected. These fragments were 82, 74, 69, 65, 64, and 63 nucleotides in length (Fig. 3). The

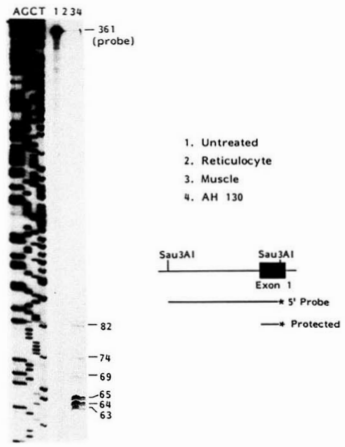

Fig. 3. Determination of the 5' termini of the M_1- and M_2-type isozyme mRNAs. - Thirty μg of poly(A)$^+$ RNA was hybridized with ^{32}P-labeled DNA probe. The RNA-DNA hybrids were digested with nuclease S1, and the products were electrophoresed on 8% polyacrylamide/8M urea gels using sequence ladders of M13 mp18 DNA as size markers. Six protected fragments are detected and their lengths are shown on the right of the autoradiogram. The solid box indicates the exon of genomic DNA. The Sau3AI/Sau3AI DNA fragment used as a probe is shown by a solid line. The asterisk shows the position of ^{32}P-label.

major protected band was 64 nucleotides in length. The specificity of the reaction was confirmed by demonstrating the absence of protected bands when rat reticulocyte mRNA containing neither the M_1- nor M_2-type was used. These results show that the M gene has multiple transcription initiation sites and that these are common to the M_1- and M_2-types.

The nucleotide sequence of the upstream region of the PKM gene and the transcription initiation sites assigned from the results of S1 mapping are shown in Fig. 4. This region lacked a "TATA box" and a "CAAT box" (24). But a 6 bp sequence, GGGCGG, was repeated three times at positions -427, -252 and -47. And three decanucleotide Sp1 binding consensus sequences (25, 26) were located at positions -133, -86, and -48. The G/C content was rather high (64%).

Evidence of Promoter Activity - CAT assay was performed to ascertain whether this 5' flanking region contained a functionally active promoter. pMcat2.2, pMcat1.8 and pMcat0.5, which contain fragments of about 2.2kb, 1.8kb and 0.5kb, respectively, from the SalI site in the first exon, were constructed as described under "Experimental Procedures". These plasmid DNAs were transfected into the rat hepatoma cell line, dRLh-84, by the electroporation method (15). The dRLh-84 cells express only the M_2-type pyruvate kinase (data not shown). When CAT activity was determined (12, 17) 48 h after transfection, extracts from these three transformants showed similar CAT activities, whereas the extract from cells transfected with pUC0cat had negligible activity (Fig 5). The pUC2cat DNA which we used as a positive control produced relatively high levels of chloramphenicol acetyltransferase activity. We also transfected pMcat2.2 and pMcat0.5 DNAs into primary cultures of adult rat hepatocytes as these cells contain only the L-type isozyme (27). Hepatocytes transfected with these DNAs showed no chloramphenicol acetyltransferase activity, whereas pUC2cat DNA produced high levels of enzyme activity.

Pseudogene of the Pyruvate Kinase M Gene - During screening of a Charon 4A rat genomic library, we found and sequenced a clone (ΔMPK14) that lacked introns (Fig. 6). Its sequence was very similar to that of the M_2-type mRNA and contained the remnants of a poly(A) tail in the 3'-untranslated region. This gene is bounded by 15bp direct repeats (the sequence: 5' AAAGGCTATTGAGGG). An open reading frame of the intronless gene is capable of coding for a protein like M_2-type pyruvate kinase, but 45 amino acids longer than the latter. We introduced an EcoRI/SalI fragment of about 1.5kb containing the 5' flanking region of the intronless gene into pUC0cat to examine whether this region contained functional promoters. This pMPcat DNA was transfected into dRLh-84 cells and adult rat hepatocytes by electroporation. However, the fragment did not show any promoter activity in these cells (Fig. 5).

Northern Blot Analysis - We examined the developmental regulation of expression of the M gene by Northern blot analysis (Fig. 7). In brain and muscle, the mRNA of M_2-type pyruvate kinase was the major type in the early stage of development, but its level decreased gradually during gestation concomitant with increase in the level of M_1-type mRNA. These results were consistent with those of enzymological studies (2, 27).

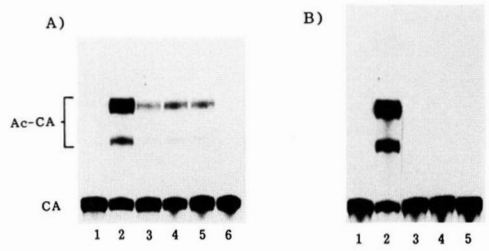

Fig. 5. Chloramphenicol acetyl transferase activity in extracts of dRLh-84 cells and rat hepatocytes transfected with the fusion genes. - Four kinds of plasmid, pMcat0.5, pMcat1.8, pMcat2.2 and pMPcat were transfected into dRLh-84 cells and adult rat hepatocytes. After 48h, CAT activity was determined. CA and Ac-CA show [^{14}C]chloramphenicol and its acetylated form respectively. A) CAT activity in the transfected dRLh-84 cell lysates. Lane 1, pUC0cat; lane 2, pUC2cat; lane 3, pMcat0.5; lane 4, pMcat1.8; lane 5, pMcat2.2; lane 6, pMPcat. B) CAT activity in the transfected rat hepatocyte lysates. Lane 1, pUC0cat; lane 2, pUC2cat; lane 3, pMcat0.5; lane 4, pMcat2.2; lane 5, pMPcat.

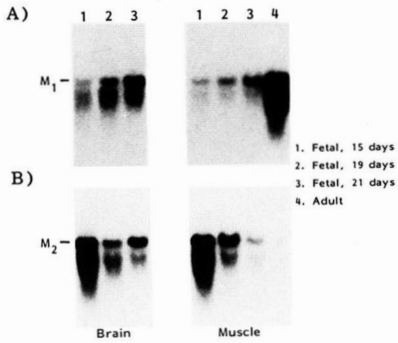

Fig. 7. Northern blot analysis of rat M_1- and M_2-type pyruvate kinase mRNA in fetal brain and muscle. - Samples of 25μg of total RNA isolated from the indicated tissues were separated on 0.8% agarose gel containing 2.2M formaldehyde, transferred to a nitrocellulose filter and hybridized with the ^{32}P-probe. M_1- and M_2-specific probes were prepared as described under "Experimental Procedures". A) ^{32}P-labeled M_1-specific probe. B) ^{32}P-labeled M_2-specific probe.