Biochem J. (1986) 235, 435–439 (Printed in Great Britain)

Nucleotide sequence of the putative regulatory region of mouse lactate dehydrogenase-A gene

Kayoko M. FUKASAWA and Steven S.-L. LI*
Laboratory of Genetics, National Institute of Environmental Health Sciences, National Institutes of Health, P.O. Box 12233, Research Triangle Park, NC 27709, U.S.A.

The nucleotide sequence of approx. 3 kilobases including the regulatory region, a non-coding exon and the first protein-coding exon from mouse lactate dehydrogenase-A (LDH-A) gene has been determined. The putative initiation sites of transcription and translation were deduced by comparing the nucleotide sequence of mouse LDH-A gene with those of a mouse LDH-A processed pseudogene and the LDH-A full-length cDNAs from rat and human. The tentative TATA and CAAT boxes, and the hexanucleotides CCGCCC have been identified. The sequence of AAATCTTGCTCAA of mouse LDH-A gene has also been found to show striking homology to the cyclic AMP-responsive sequences of eukaryotic genes regulated by cyclic AMP. It has been reported previously that the protein-coding sequence of mouse LDH-A gene is interrupted by six introns and the 3′ untranslated sequence of 485 nucleotides is not interrupted [Li, Tiano, Fukasawa, Yagi, Shimiza, Sharief, Nakashima & Pan (1985) Eur. J. Biochem. **149**, 215–225]. An additional intron of 1653 base-pairs was found in the 5′ untranslated sequence of 101 nucleotides at 24 nucleotides upstream to the translation start site. Thus, mouse LDH-A gene containing seven introns spans approx. 11 kilobases and its length of mature mRNA is 1582 nucleotides, excluding the poly(A) tail.

INTRODUCTION

In mammals and birds, the five isoenzymes of tetrameric lactate dehydrogenase (LDH) are found in various proportions among different somatic tissues and are produced *in vivo* by combination of the A and B subunits, whereas the homotetrameric LDH-C_4 is present only in mature testes and sperms [1]. The LDH-A_4 (muscle), LDH-B_4 (heart) and LDH-C_4 (testis) isoenzymes possess distinct physical, catalytic and immunological properties [2, 3]. The three gene loci coding for the LDH-A, B and C polypeptides are believed to have originated from an ancestral gene during the course of evolution [1, 4]. The expression of these three LDH genes is developmentally regulated and tissue-specific [1].

It is known that human cancer tissues exhibit an increase in LDH activity [5]. Unusual isoenzymes LDH-K [6], LDH-Z [7] and LDH-A_4 of altered conformation [8] were isolated from human cancer. It was also shown that the LDH-A protein is phosphorylated at tyrosine in cells transformed by Rous sarcoma virus [9]. The level of LDH-A mRNA was found to increase following the stimulation of quiescent rat fibroblasts by epidermal growth factor [10]. The synthesis of LDH-A mRNA and polypeptide in rat C6 glioma cells was shown to be induced by isoproterenol or dibutyryl cyclic AMP [11–13].

In order to illustrate the structural and evolutionary relationships of the LDH-A, B and C genes and to study the molecular mechanism(s) of gene regulation, we have undertaken the investigation of protein structure and gene organization of mammalian LDH isoenzymes. We have reported the primary structure of human LDH-A_4, mouse LDH-A_4, mouse LDH-C_4 and rat LDH-C_4 isoenzymes as well as the molecular characterization of mouse and human LDH-A cDNA clones [14–19]. We have also described the genomic organization of protein-coding exons of mouse LDH-A gene [19]. In this paper, we present the nucleotide sequence of the regulatory region of mouse LDH-A gene and discuss its homologies with those of other eukaryotic genes regulated by cyclic AMP.

MATERIALS AND METHODS

The mouse (C57 Bl/10) LDH-A genomic clones were isolated from a Charon 4A genomic library provided by Dr. M. Edgell [20]. The genomic clone λM15 containing LDH-A functional gene was partially characterized [19] and DNA fragments of its regulatory region were purified, further cleaved, and subcloned into M13 mp10/mp11 phages [21, 22]. The DNA isolated from genomic clone λM11 containing LDH-A processed pseudogene was cleaved by *Eco*RI and *Xba*I, subcloned into M13 mp10/mp11 phages, and the M13 phages exhibiting positive hybridization to LDH-A cDNA probe were isolated. (It may be noted that there are no *Eco*RI and *Xba*I sites in the short and long arms of Charon 4A phage.) The nucleotide sequences of genomic DNA fragments purified from M13 phages were determined by the dideoxy chain termination method [23].

RESULTS

Regulatory sequence of LDH-A functional gene

The genomic clone λM15 was previously shown to contain an LDH-A functional gene in which the

Abbreviations used: LDH, lactate dehydrogenase; PEPCK, phosphoenolpyruvate carboxykinase; TAT, tyrosine aminotransferase; CRP, cyclic AMP receptor protein; kb, kilobases; bp, base pairs.

* To whom all correspondence and reprint requests should be addressed.

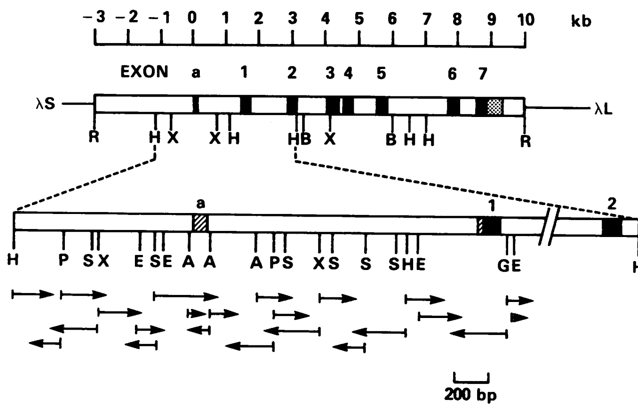

Fig. 1. Restriction endonuclease map of the mouse LDH-A gene and nucleotide sequencing strategy of the regulatory region

The overall map of mouse genomic clone λM15 for restriction endonucleases (R, *Eco*RI; H, *Hin*dIII; B, *Bam*HI; X, *Xba*I) was previously described [19]. The seven protein-coding exons are numbered in Arabic and shown by solid blocks. The 5′ and 3′ untranslated regions are shaded. The *Hin*dIII fragments of 2.4 kb (containing non-coding exon a) and 2.0 kb (containing the first protein-coding exon) were isolated and further cleaved by: P, *Pst*I; S, *Sau*3AI; X, *Xba*I; E, *Hae*III; A, *Alu*I and/or G, *Bgl*II, and subcloned into M13 mp10 or mp11 phages [22]. The nucleotide sequences of the inserted DNAs were determined by the dideoxy chain termination method [23].

protein-coding sequence is interrupted by six introns [19]. The orientation of the LDH-A gene in the genomic clone λM15, the restriction map and the sequencing strategy of the regulatory region are illustrated in Fig. 1. The nucleotide sequence of approx. 3 kb including the regulatory region, a non-coding exon and first protein-coding exon is presented in Fig. 2. The putative initiation sites of transcription and translation of LDH-A gene are deduced by comparing the nucleotide sequence of genomic clone λM15 with those of the LDH-A full-length cDNAs from rat [10] and human [18] as well as a mouse LDH-A processed pseudogene present in genomic clone λM11 (Fig. 3). The most likely transcription start site is at the G residue shown in Fig. 2 as +1. The tentative 'TATA' and 'CAAT' boxes are located at nucleotides −26 and −43, respectively. The regulatory sequence of CCGCCC occurs twice at positions −103 and −137 from the transcription start site. Examination of sequence homologies with several other eukaryotic genes regulated by cyclic AMP revealed a putative cyclic-responsive sequence at position −853 of mouse LDH-A gene (Fig. 4). As to the translation start site, the first methionine codon is located at 101 nucleotides downstream from the 5′ end of the mRNA (Fig. 3). The predicted *N*-terminal sequence is in complete agreement with that of mature LDH-A protein [19]. A sequence comparison of mouse LDH-A functional gene with those of mouse LDH-A pseudogene λM11, rat LDH-A cDNA [10] and human LDH-A cDNA [18] uncovered that the 5′ untranslated sequence of mouse LDH-A gene is interrupted by an intron of 1653 bp at 24 nucleotides upstream from the translation start signal ATG (Fig. 2).

Partial sequence of LDH-A processed pseudogene

Mouse genomic clone λM11 containing an insert of 14 kb was isolated and partially characterized by restriction endonuclease mapping and Southern blot analysis (results not shown). A single genomic DNA fragment of 1.6 kb derived from double cleavage with *Eco*RI and *Xba*I exhibited a strong hybridization signal to the LDH-A cDNA probes from mouse [17] and human [18], indicating the presence of the protein-coding sequence as well as the 5′ and 3′ untranslated regions. The partial nucleotide sequences of this DNA fragment from genomic clone λM11 are aligned with those of mouse genomic clone λM15 and the LDH-A cDNAs from rat and human (Fig. 3). The sequences of mouse λM11 LDH-A pseudogene possess the characteristics of a processed pseudogene, that is (*a*) absence of introns at the 5′ untranslated region and between codons 41 and 42, (*b*) the presence of the polyadenylation signal AATAAA [24] followed 12 nucleotides downstream by a poly(A) tail, and (*c*) the presence of exact 9-bp direct repeats, one located just upstream from the 5′ end of the homologous cDNA sequence and the other just downstream from the poly(A) tail. The nucleotide sequence upstream to the 5′ direct repeat of the λM11 LDH-A pseudogene is very different from that of the promoter region of the LDH-A functional gene (λM15). Thus, this LDH-A processed pseudogene is transcriptionally inactive, as it lacks the necessary 5′ regulatory sequences. However, the sequence of the λM11 LDH-A pseudogene has 85% and 98% homologies with the 5′ untranslated region and the sequence coding for first 45 amino acids from mouse LDH-A gene (λM15), respectively. A comparison of the 3′ untranslated sequence near the polyadenylation site from the λM11 LDH-A pseudogene with that of mouse LDH-A gene (λM15) revealed three nucleotide substitutions, deletion of a single nucleotide and insertion of 12 nucleotides.

DISCUSSION

As described previously [19], the protein-coding sequence of mouse LDH-A gene is interrupted by six introns and the 3′ untranslated sequence of 485 bp is not interrupted. In this report, we have uncovered the presence of an additional intron of 1653 bp in the 5′ untranslated region of the mouse LDH-A gene. While the capping site of mouse LDH-A mRNA has not been determined directly, the putative initiation site of transcription was deduced by aligning the nucleotide sequence of the mouse LDH-A functional gene (λM15) with those of a cloned mouse LDH-A processed pseudogene (λM11), rat LDH-A cDNA [10] and human LDH-A cDNA [18]. It may be noted that both rat and human LDH-A cDNA clones were isolated from cDNA libraries constructed by the method of Okayama & Berg [25], which yields cDNAs of nearly full length. Further, the 5′ end of rat LDH-A mRNA was confirmed experimentally by primer extension [10]. The 5′ untranslated sequence of 101 nucleotides is 62% G+C and it contains 9 CG dinucleotides. The 150-nucleotide region upstream from the transcription initiation site is 74% G+C and it possesses 21 CG dinucleotides. These CG dinucleotides are the potential methylation sites, and under-methylation of the regulatory region is commonly associated with the active expression of eukaryotic genes

```
                                                                                   AAGCTTTGTGATATTAATGGCAGACTGCTCTCCAGGT -1081
CCAACTGTTTGACTTTTTAATTTTTTTAAATTTGCGCTCTGCTGAGGGACTCAGGCCTTGGTGTATGCAAGACAAGTGTCCCTGTCCTGGCCACACTCCTAGGCCTGTTGTTTGTTATAA - 961

                                                                                                           cAMP resp seq
AAGAGATCACAAGGGATACAGACCAAGCGAAAATAAGGGGGGGCTGTGAGCTTCATTCCCTCTCACATGATCCCTGCATATCCAGCACCCTGCAACCAGCTTGTTCAAATCTTGCTCAAG - 841
ACTGTAATAGACCTTAATCTGCAGTGAACACATCCTTCCGGGGAGATGGGTGAGCAGGGTTGAAAGTCACAGTTCTTCCCATCACTAAGGAAATCAAACAGTTGCAAACTCCATTTCACA - 721
TCCTATCAGTGGTGGAGTACCTTAAGATTCACATACGGTTGATCGGGAGCTGGGAGAGGAACAGGAACTGGCCTCACAGCTTAATGAGACCTCTAGAAAGACGTTTAAAGGCAGAGGGGG - 601
TGTGTGAAAACAAGCAAGGGCCCTGATACTCCTTGGTAAGGCTAAACACAAATGCCTGCGGGATGGCATGGGAGAGGGCAGATATGGATGTAAGCTGGCAAGCCATCAGAAGCTGAGCCG - 481
CACCCCCCTCCCCCCATGGTTTGGGAGATGGAAGTGGGGCAGGAGAAAGGCCCATCTGATAGGCTGCTATGGCGGATAGACCGGCACGGTGCTGGCATGTGCTTTCACACAATATTTACT - 361
GAAGGCCTGTTGCTTGCCAGGAGTTGTTCCCGGCGCCGGAACAGCAATGAAGAAAGTGACCAGGGTTTTTTCTGAGTCTCACAAGTTTTCCGGTGAAGGAGGCGGAGGATCGATGCATTT - 241

                                                                                                       CCGCCC
CGGGCTCCTGCTTCTGAGGCTGAGGAGCATGTCGGGTTGGCCTTTCTTTGGGGTGTCGCAGCACACGTGGAGCCACTCTTGCAGGGACATCGTGCTGCGCGCGCCGCCCCGGTCTCGGTG - 121

                 CCGCCC                                            CAAT                       TATA
GCCTAGCCGGCTGGACGCCGCCCCCGGCCCAGCCTACACGTGGGTTCCCGCACGTCCGCTGGGCTCCCACTCTGACGTCAGCGCGGAGCTTCCATTTAAGGCCCCGCCGCGTGCTGCTCT -   1

1        .         .         .         .         .         .         .         .         .         .         .         .
GCGTGCTGGAGCCACTGTCGCCGAGCTCGGCCACGCTGCTTCTCCTCGCCAGTCGCCCCCCCATCGTGCATCTAGCGGTACGGTTGGGCCCCACGCTGCCGGCACAGGGGGTCCTATCCG   120
-----------------------------------------------------------------------------

GGGTGGAGGTGCAGGGTGTTCAGATTTGGGCACGCGTGGGCTACTGTGCTTTGGGGAACGTAGCAGGCGGCCCACCCAGCCTGGGCGGCGACACCCGTGTAAAGAGGACTAAGGGTGGCT   240
GGCCTGAAGGTTGGGAGCCACCGAACGGGGGCGAGGGAGCGGTGCGAAACTTGAGCTCCCGTAACTCGAGCCATGGGGGCCGAAAAAGCCTTGGCACGTCCGGGGCGGGTGCTTCGGAGC   360
AGCCGTGCGGTTTGCATTTTCCTTTGCGTGGGCTCGGTGGAGTGGTTGTTTCTGCAGATTATAGGGCGCTCTTGCCCTAAACGCATTTTGTTAAGAGGCTCGCTCCTGGTGTGACTGGGG   480
ATCGTGCTAAGGCGCTGCGTGCGGAGAGGAAGCGGGAAGAGTCGCAGTCCTTCCTCTGCACCCTAGACGGAAGGAGGAAACCTGTAGCTGAGAGGCCTGCGACACCCATCGGTCGTAGGT   600
GGGAAGTAGAGGCACCCTGTCCTAATAGAGCACAGACTTGAGGTTTGCGTGAAGTAGATTCTCTGAGGGAAGAAGTCCCCACGCCCTGCCCCCAGTCAGCAGAGACCTCACAAGGCTGTC   720
TAGAAATAGCAGCGGTTGTAAGGAGACCGGATTCCTGCTTGATGGCCCTAGCGTGCTTGCCCAGAGATCTTGTCCAGTCCTTTTGCACTTTGGAACGATTTCAAAAATAGACATGGTGCT   840
TGCTGGGAAAGGTGGCCATCGCGGGGGTGGGAGTGGGCTCCAGGCTCAGGCCTCCGCATATTGTATCCCCATTTTGGAACTGAAGAAATTTCCCTTAGCGGCCACATCCTGGGTAATGAG   960
GCCCCGCCTGGTGGTGAAGGCCCTAATCCTGTTAGTGGCATTTTGAGATCTCACTCTGGCACCAGGCTAAGGTGGCTGCCAGCTCCACTTCACCGCGCTTGTGCTTTGGGCTCTAGTAAG  1080
GGCCAAACTTTGACGCCGAACTGCCTGCCAGAGGTCTCATCCATGGCCAGCAGTCGCTAGACCTATATATATGCACCTGATTCTGAAATTACGCGCACTGCCTTCCCCCGCCTGCCAGTC  1200
TTCTGGATCTGGGCTGGTCATACAACTTGGGTTCTTGCGGGGGTGGGGGGGTTAGAAGAAGCTTGCGCGTGCAGGCTTAAGCACGTTGCTATGCCTTGGGGTCGCACCTTGTGGCCGTTA  1320
TTGGCGCCCTCTGCTCTTGATTTTTGGTACTTCCTGGAGCAACTTGGCGCTCTACTTGCTGTAGGGCTCTGGGTGATGGGAGAAGAGCGGGAGGGCAGCTTTCTAACCATATAAGAGGAG  1440
ATACCATCCCCTTTTGGTTCATCAAGATGAGTAAGTCCTCAGGCGGCTACACGTACACGGAGACCTCGGTATTATTTTTCCATTTCAAGGTAGAGCCTTCCTGGTAGAGCCAGAACCACA  1560
TCCTGCCGCTGCTATTCTTGGTTTTCCACTTCTGTTCTTTGTCACATTGTCACTTAATGGAAAGGAGTCCAAGGTAGCAAGTCAGCGTTTTTTTTTTTTTTTTTTAAATACAGGGACTTG  1680
GTGAGTATACCTTGGGCAGGTTACAATGACACACGGTGTATACCCCTTAGGTCTCAAAAGATTCAAAGTCCAAGATGGCAACCCTCAAGGACCAGCTGATTGTGAATCTTCTTAAGGAAG  1800
                                                  ------------------------MetAlaThrLeuLysAspGlnLeuIleValAsnLeuLeuLysGluG

AGCAGGCTCCCCAGAACAAGATTACAGTTGTTGGGGTTGGTGCTGTTGGCATGGCTTGTGCCATCAGTATCTTAATGAAGGTAAGTGGAGATCTTCATGGCCCAAGCTATGGGTGTGGTG  1920
luGlnAlaProGlnAsnLysIleThrValValGlyValGlyAlaValGlyMetAlaCysAlaIleSerIleLeuMetLys #41

TGGGGGAGAGGACATCCCTACATTGTCACATTGTATGTAAAACTATCAAGGTTTGCACACACTCAGTCATCTGTGAAACATTTTGCAACATAATGATACACAAGAAAGGGATTATCCAAA  2040
```

Fig. 2. Nucleotide sequence of the regulatory region of the mouse LDH-A gene

The nucleotide sequence of mouse LDH-A gene is numbered from the putative initiation site of transcription, and the number of the last nucleotide in each row is given at the right-hand side. The tentative TATA and CAAT boxes, the hexanucleotides CCGCCC and the cyclic AMP-responsive sequence (cAMP resp seq) are identified above the DNA sequence. The 5′ untranslated sequence underlined is interrupted by an intron of 1653 nucleotides. The initiating methionine and predicted sequence of 41 amino acids from the first protein-coding exon are given below the DNA sequence.

[26]. However, the effect of DNA methylation on the transcription of the mouse LDH-A gene remains to be determined. The signals of 'TATA' and 'CAAT' are tentatively identified as indicated in Fig. 2. The sequence of CCGCCC occurs twice in the G+C-rich region of 150 nucleotides upstream from transcription start site. It has previously been shown that the TATA box is essential for correct transcription initiation by eukaryotic RNA polymerase II and that the CAAT box and the CCGCCC sequence are involved in regulating the level of transcription of many eukaryotic genes [27, 28]. In the simian virus 40, the CCGCCC sequence is present twice within each of the 21-bp repeats of the early promoter [29] and is known to bind the transcription factor Spl [30]. In herpes simplex virus, the CCGCCC sequence is involved in the modulation of IE gene expression [31] and is apparently essential for maintaining high levels of transcription of the thymidine kinase gene [32].

Cyclic AMP has been shown to regulate the synthesis of several proteins including LDH-A [11–13], PEPCK [33–36], TAT [37], and prolactin [38]. In rat C6 glioma cells, either isoproterenol or dibutyryl cyclic AMP can

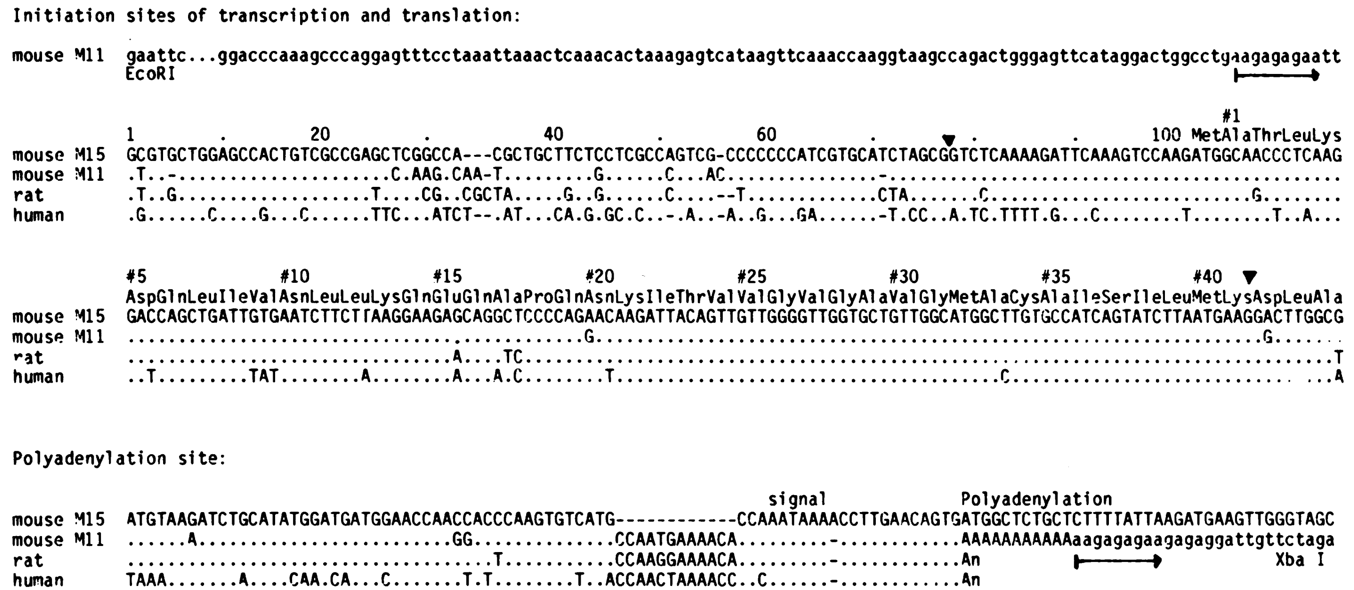

Fig. 3. Nucleotide sequences near the initiation sites of transcription and translation as well as the polyadenylation site of a mouse LDH-A processed pseudogene

The 5′ and 3′ flanking sequences of mouse λM11 LDH-A pseudogene are given in lower case, and the 9-bp direct repeat is denoted by arrows. The partial nucleotide sequences of mouse λM11 LDH-A pseudogene, and LDH-A cDNAs from rat [10] and human [18] are compared with that of mouse λM15 LDH-A functional gene, and only those nucleotide differences are given. Identical residues are indicated by dots, and deletions are denoted by hyphens. The transcription starts at G numbered as +1, and the translation initiates at the methionine just before the alanine at residue # 1 of mature LDH-A protein. The solid triangle indicates the position of introns between nucleotides 77 and 78 and codons 41 and 42 of mouse LDH-A gene.

induce the synthesis of LDH-A mRNA and polypeptide by increasing the rate of transcription, the stability and the translation efficiency of LDH-A mRNA [11–13]. Jungmann and his associates have proposed a model in which either isoproterenol or cyclic AMP induces phosphorylation of cyclic AMP-dependent protein kinase and this results in the phosphorylation of RNA polymerase II and/or histones [39]. Recently, rat liver LDH-A_4 isoenzyme has been found to possess DNA-binding property [40]. Chicken LDH-A protein has also been shown to be phosphorylated by tyrosine kinase [9]. It would be of great interest to determine the extent of phosphorylation of the LDH-A_4 isoenzyme found in the nucleus and the effect of the phosphorylated LDH-A protein in gene expression and DNA replication. On the other hand, it is also conceivable that the molecular mechanism of cyclic AMP induction involves a cyclic AMP-binding protein and the cyclic AMP–receptor protein complex interacts directly with the regulatory region of eukaryotic genes, including the LDH-A gene, independent of protein kinase. This would be analogous to the catabolite regulatory protein (CRP) system in prokaryotic cells [41]. The sequence of AAATCTTGCTCAA found in the mouse LDH-A gene is strikingly similar to the regulatory sequences of rat PEPCK [33, 34], chicken PEPCK [35, 36], rat TAT [37], rat prolactin [38] and the consensus CRP sequence found in prokaryotic genes regulated by cyclic AMP [41]. However, the functional significance of the cyclic AMP-responsive sequence of mouse LDH-A gene remains to be demonstrated.

Mouse	LDH-A	-853	A A A T C T T G C T C A A
Rat	PEPCK	-262	A A A G T T T A G T C A A
Chicken	PEPCK	-144	A A A G T G T G G T T A A
Rat	TAT	-282	A A A G T T A A - T C A C
Rat	Prolactin	- 37	A A G G T T T A - T A A A
E. coli	CRP consensus		A A N T G T G A N T N N N

Fig. 4. Homologies of the cyclic AMP-responsive sequences in the regulatory region of genes induced by cyclic AMP

The putative cyclic AMP-responsive sequences of the genes for LDH-A, PEPCK, TAT and prolactin as well as the consensus sequence of CRP in *E. coli* are aligned to give maximum homologies. The hyphen denotes the gap and N means A, C, G or T. The negative number indicates the position of the first nucleotide upstream from the transcription start site of the corresponding gene. See the Discussion section for details.

The translation initiation signal AUG of mouse LDH-A mRNA is the very first methionine codon from its 5′ end, and the predicted *N*-terminal sequence is in complete agreement with that of mouse LDH-A_4 isoenzyme [19]. The size of the 5′ untranslated sequence of LDH-A mRNA was found to be very similar among mouse (101 nucleotides), rat (103 nucleotides) and human (100 nucleotides), and in this region, mouse LDH-A mRNA contains 19 and 39 differences from that of rat [10] and human [18], respectively.

The mouse LDH-A intronless pseudogene present in

genomic clone λM11 appears to contain a cDNA-like sequence of full length flanked by a 9-bp direct repeat at both 5′ and 3′ ends. The nucleotide sequences 5′ and 3′ to the polyadenylation signal AATAAA of the mouse LDH-A pseudogene (λM11) appear to be more similar to those of the LDH-A cDNAs from rat [10] and human [18] than that of mouse LDH-A functional gene (λM15). Thus, the insertion of a single nucleotide and deletion of 12 nucleotides occurred in the mouse LDH-A functional gene (λM15) only after the divergence between mouse and rat and the appearance of the LDH-A processed pseudogene (λM11). The functional significance of these differences is presently not known.

In short, the mouse LDH-A gene, containing seven introns, spans approx. 11 kilobases, and its total length of eight exons is 1582 bp. The information on the structure of the mouse LDH-A functional gene will facilitate the molecular characterization of genetic mutations affecting the expression and structure of mammalian LDH-A_4 isoenzymes [42–44]. The future investigation of genomic organization and regulatory sequences of mammalian LDH-B and LDH-C genes will certainly yield insights on the mechanisms of tissue-specific expression and gene regulation by cyclic AMP.

We thank Mr. H. F. Tiano for technical assistance, Dr. M. Edgell for the mouse genomic library, Dr. B. H. Judd for his interest and support to this project, and Drs. A. Sugino, L. Searles and F.-Z. Chung for discussion and reading the manuscript.

REFERENCES

1. Markert, C. L., Shakelee, J. B. & Whitt, G. S. (1975) Science **189**, 102–114
2. Everse, J. & Kaplan, N. O. (1973) Adv. Enzymol. **28**, 61–133
3. Holbrook, J. J., Liljas, A., Steindel, S. J. & Rossmann, M. G. (1975) Enzymes 3rd Ed. **11**, 191–292
4. Li, S. S.-L., Fitch, W. M., Pan, Y.-C. E. & Sharief, F. S. (1983) J. Biol. Chem. **258**, 7029–7032
5. Goldman, R. D., Kaplan, N. O. & Hall, T. C. (1964) Cancer Res. **24**, 389–399
6. Anderson, G. R. & Kovacik, W. P., Jr. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 3209–3213
7. Siciliano, M. J., Bordelon-Riser, M. E., Freedman, R. S. & Kohler, P. O. (1980) Cancer Res. **40**, 283–287
8. Okabe, K., Hayakawa, T., Hamada, M. & Koike, M. (1968) Biochemistry **7**, 79–90
9. Cooper, J. A., Esch, F. S., Taylor, S. S. & Hunter, T. (1984) J. Biol. Chem. **259**, 7835–7841
10. Matrisian, L. M., Rautmann, G., Magun, B. E. & Breathnach, R. (1985) Nucleic Acids Res. **13**, 711–726
11. Derda, D. F., Miles, M. F., Schweppe, J. S. & Jungmann, R. A. (1980) J. Biol. Chem. **255**, 11112–11121
12. Miles, M. F., Hung, P. & Jungmann, R. A. (1981) J. Biol. Chem. **256**, 12545–12552
13. Jungmann, R. A., Kelley, D. C., Miles, M. F. & Milkowski, D. M. (1983) J. Biol. Chem. **258**, 5312–5318
14. Chang, S.-M. T., Lee, C. Y. & Li, S. S.-L. (1979) Biochem. Genet. **17**, 715–729
15. Pan, Y.-C. E., Sharief, F. S., Okabe, M., Huang, S. & Li, S. S.-L. (1983) J. Biol. Chem. **258**, 7005–7016
16. Li, S. S.-L., Feldmann, R. J., Okabe, M. & Pan, Y.-C. E. (1983) J. Biol. Chem. **258**, 7017–7028
17. Akai, K., Yagi, H., Tiano, H. F., Pan, Y.-C. E., Shimizu, M., Fong, K., Jungmann, R. A. & Li, S. S.-L. (1985) Int. J. Biochem. **17**, 645–648
18. Tsujibo, H., Tiano, H. F. & Li, S. S.-L. (1985) Eur. J. Biochem. **147**, 9–15
19. Li, S. S.-L., Tiano, H. F., Fukasawa, K. M., Yagi, K., Shimizu, M., Sharief, F. S., Nakashima, Y. & Pan, Y.-C. E. (1985) Eur. J. Biochem. **149**, 215–225
20. Weaver, S., Comer, M. B., Jahn, C. L., Hutchison, C. A. III & Edgell, M. H. (1981) Cell **24**, 403–411
21. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory, New York
22. Messing, J., Crea, R. & Seeberg, P. H. (1981) Nucleic Acids Res. **9**, 309–321
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
24. Proudfoot, N. J. & Brownlee, G. G. (1976) Nature (London) **263**, 211–214
25. Okayama, H. & Berg, P. (1983) Mol. Cell. Biol. **3**, 280–289
26. Doerfler, W. (1983) Annu. Rev. Biochem. **52**, 93–124
27. Breathnach, R. & Chambon, P. (1981) Annu. Rev. Biochem. **50**, 349–383
28. Cordingley, M. G., Campbell, M. E. M. & Preston, C. M. (1983) Nucleic Acids Res. **11**, 2347–2365
29. Everett, R. D., Baty, D. & Chambon, P. (1983) Nucleic Acids Res. **11**, 2447–2464
30. Gidoni, D., Dynan, W. S. & Tijian, R. (1984) Nature (London) **312**, 409–413
31. Kristie, T. M. & Roizman, B. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 4065–4069
32. McKnight, S. L. & Kingsbury, R. (1982) Science **217**, 316–324
33. Yoo-Warren, H., Monahan, J. E., Short, J., Short, H., Bruzel, A., Wynshaw-Boris, A., Meisner, H. M., Samols, D. & Hanson, R. W. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3656–3660
34. Wynshaw-Boris, A., Lugo, T. G., Short, J. M., Fournier, R. E. K. & Hanson, R. W. (1984) J. Biol. Chem. **259**, 12161–12169
35. Hod, Y., Morris, S. M. & Hanson, R. W. (1984) J. Biol. Chem. **259**, 15603–15608
36. Hod, Y., Yoo-Warren, H. & Hanson, R. W. (1984) J. Biol. Chem. **259**, 15609–15614
37. Shinomiya, T., Scherer, G., Schmid, W., Zentgraf, H. & Schutz, G. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 1346–1350
38. Maurer, R. A. (1981) Nature (London) **294**, 94–97
39. Lee, S.-K., Schweppe, J. S. & Jungmann, R. A. (1984) J. Biol. Chem. **259**, 14695–14701
40. Williams, K. R., LoPresti, M. B., Stone, K. L. & Patel, G. (1984) Fed. Proc. Fed. Am. Soc. Exp. Biol. **43**, 1541 (abstr.)
41. Ebright, R. H. (1982) in Molecular Structure and Biological Function (Griffen, J. & Duax, W., eds.) pp. 91–102, Elsevier/North-Holland, New York
42. Soares, E. R. (1979) Environ. Mutagen. **1**, 19–25
43. Charles, D. J. & Pretsch, W. (1981) Biochem. Genet. **19**, 301–309
44. Kanno, T., Sudo, K., Kitamura, M., Miwa, S., Ichiyama, A. & Nishimura, Y. (1983) Current Top. Biol. Med. Res. **7**, 131–150

Received 27 September 1985/19 November 1985; accepted 10 December 1985