Molecular cloning and nucleotide sequence of the cDNA for sperm-specific lactate dehydrogenase-C from mouse

Ikuya SAKAI, Farida S. SHARIEF and Steven S.-L. LI*

Laboratory of Genetics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, U.S.A.

Mouse sperm-specific lactate dehydrogenase-C (LDH-C) cDNA was cloned and sequenced from λ gtll expression library. The LDH-C cDNA insert of 1236 bp consists of the protein-coding sequence (999 bp), the 5' (54 bp) and 3' (113 bp) non-coding regions, and the poly(A) tail (70 bp). The Northern blot analysis of poly(A)-containing RNAs from mouse testes and liver indicates that the LDH-C gene is expressed in testes but not in liver, and that its mRNA is approx. 1400 nucleotides in length. The nucleotide and amino acid sequences of the mouse LDH-C cDNA show 73% and 72% homologies, respectively, with those of the mouse LDH-A. The Southern blot analysis of genomic DNAs from mouse liver and human placenta indicates the presence of multiple LDH-C gene-related sequences.

INTRODUCTION

In mammals, the homotetrameric lactate dehydrogenase-C (LDH- C_4) isoenzyme is found only in mature testes and spermatozoa, whereas the LDH-A and LDH-B gene products are present predominantly in skeletal muscle and heart tissue, respectively. The LDH-A, B and C polypeptides are encoded by three different genes which appear to have originated from an ancestral gene during the course of evolution [1,2]. The enzymic and immunological properties, amino acid sequence and three-dimensional structure of the mouse LDH-C₄ isoenzyme have been studied extensively [3-8]. The sperm-specific LDH-C₄ isoenzyme is highly immunogenic, not only in females and males of other species, but also in males of the same species. This is because an effective blood-testis barrier isolates LDH-C₄ from the animal's immune system. The active immunization of animals with the $LDH-C_4$ autoantigen has been shown in reduce fertility in females [9,10].

In order to study the molecular mechanism(s) of gene regulation and to understand the structural and evolutionary relationships of the lactate dehydrogenase genes, and also with the hope of developing a contraceptive vaccine using the sperm-specific LDH-C4 autoantigen, we have undertaken the investigation of protein structure and genomic organization of the mammalian LDH-A, B and C genes. We have reported the primary structures of LDH-C₄ isoenzymes from mouse and rat as well as those of LDH-A₄ isoenzymes from mouse and man [6,7,11-13]. We have also described the nucleotide sequence of the LDH-A cDNA and the genomic organization of the LDH-A functional gene and pseudogenes from mouse and man [11-16]. Mouse LDH-C polypeptide has been identified immunologically from the *in vitro* translation products of the poly(A)-containing mRNAs of mouse testes [17]. In this paper we present the molecular cloning and nucleotide sequence of mouse LDH-C cDNA, as well as the genomic complexity of the LDH-C gene-related sequences from mouse and man.

MATERIALS AND METHODS

Total RNAs were obtained from mouse testes (DBA/2J, 15 weeks old, Jackson Laboratory) by the phenol/chloroform and SDS procedure, and poly(A)containing mRNAs were isolated from the oligo(dT)cellulose column [18,19]. Double-stranded cDNA synthesis, λ gtll packaging *in vitro*, and screening of the λ gtll expression library with rabbit antiserum against mouse LDH-C [7] were performed according to the procedure already described [20]. Putative positive clones were plaque-purified and rescreened with rabbit antisera against mouse LDH-C or human LDH-A as well as with human LDH-A cDNA probe labelled with ³²P [19,21]. EcoRI DNA fragments purified from two positive clones, designated mC31 and mC50, were subcloned with M13 mp10 phage, and their partial nucleotide sequences were determined by the dideoxy chain termination



Fig. 1. Restriction endonuclease map and sequencing strategy of mouse LDH-C cDNA clones

The restriction sites given are only those used in the cDNA cloning (EcoRI, E), preparation of the hybridization probe (PstI, P), and nucleotide sequencing (AluI, A; HaeIII, H; Sau3A, S). The direction and length of each sequencing run are indicated by arrows. All nucleotide sequences for both complementary strands of the LDH-C cDNA were determined experimentally. The protein-coding sequence is shown solid. The 5' and 3' non-coding regions are indicated by open boxes with the poly(A) tail hatched.

Abbreviation used: LDH, lactate dehydrogenase.

^{*} To whom correspondence and reprint requests should be addressed.

LDH-C -		GAA	CAG	TTA	ITA TCC		111	CII	ACC 3	IO TGT	GCT	GCG	GAG	TCA	GCA GTA		AGG CTC AAG		AAC H-A	Int Met ATG	1 Ser TCC G.A	Thr ACC	Val GTC C	Lys AAG	Glu GAG	Gl n CAG	Leu CTG	lle ATT	Gin CAG GI
10																	-			Ala		Leu		Asp		••••	•••	Val	
10 Asn AAC	Leu CTA T	Val GTT C Leu	Pro CCG AA. Lys	Glu GAA	Asp GAT G Glu	Lys AAA C.G Gln	Leu CTT GCT Ala	Ser TCC C Pro	Ary CGG .A. Gln	20 Cys TGT AAC Asn	Lys AAG	lle ATT	Thr ACT	Val GTG T	Val GTC T	Gly GGA G	Val GTT	Gły GGA T	Asn AAT GC. Ala	30 Val GTG T	Giy GGC	Met ATG	Ala GCG T	Cys TGT	Ala GCT	lle ATT	Ser AGT	lle ATT	Leu TTA
40 Leu CTG A Met	Lys AAG	G1 y GGT .AC Asp	Leu TTG	Ala GCT	Asp GAT	Glu GAA G	Leu CTT	Ala GCC	Leu CTT	50 Val GTT	Asp GAC	Ala GCT .TC Val	Asp GAT ATG Met	Thr ACG GAA Glu	Asn AAC G Asp	Lys AAA 	Leu CTG C	Arg AGG .A. Lys	Gly GGA C	60 Glu GAG	Ala GCA ATG Met	Leu CTG A Met	Asp GAT	Leu CTT C	Leu CTG .A. Gìn	His CAC	Giy GGC	Ser AGT C	Leu CTT C
70 Phe TTC	Leu CTT	Ser AGC .AA Lys	Thr ACT	Pro CCA	Lys AAA	Ile ATC	Val GTC	Phe TTT .CC Ser	Gly GGA A.C Ser	80 Lys AAA	Asp GAT C	Tyr TAC	Asn AAT TG. Cys	Val GTA	Ser TCT A Thr	Ala GCC	Asn AAC	Ser TCC	Lys MA G	90 Leu CTG	Val GTT C	Ile ATT	Ile ATC	Thr ACA	Ala GCT	Gly GGT G	Ala GCA	Arg AGA C.T	Met ATG CA. Gln
100 Val GTG CAA Gln	Ser TCT GAG Glu	Gly GGA	Glu GAA G	Thr ACT .GC Ser	Arg CGC G	Leu CTT C	Asp GAC A Asn	Leu CTG	Leu CTC G Val	110 Gln CAA G	Arg CGT A	Asn AAT	Val GTC	Ala GCT AAC Asn	lle ATC	Met ATG T.C Phe	Lys MA G	Ala GCC TT. Phe	Ile ATT	120 Val GTT A Ile	Pro CCG	Gly GGC AA. Asn	Ile ATT	Val GTC	Gìn CAA A.G Lys	Asn AAC T Tyr	Ser AGT	Pro CCG A	Asp GAC C His
130 Cys TGT C	Lys AAA G	Ile ATA C.G Leu	Ile ATT C Leu	Ile ATC	Val GTC	Thr ACT T.C Ser	Asn AAC T	Pro CCA	Val GTG	140 Asp GAT	Ile ATT C	Leu TTG	Thr ACA	Tyr TAC	Val GTG	Val GTT .C. Ala	Trp TGG	Lys MG	Ile ATA C	150 Ser AGC T	G1y GGC	Phe TTC	Pro CCT	Val GTA AA. Lys	Gly GGC AA. Asn	Ar g CGT 	Val GTG A	Ile ATC	Gìy GGA
160 Ser AGT	Gly GGC	Cys Tgt C	Asn AAC T	Leu CTA G	Asp GAC T	Ser TCA	Ala GCA G	Arg CGT G	Phe TTT C	170 A rg CGT	Tyr TAC	Leu CTG	Ile ATT G Met	Gly GGG A	G]u GAG	Lys AAG .G. Arg	Leu CTG	Gly GGT G	Val GTC T	180 Asn AAC C His	Pro CCT G.G Ala	Thr ACA CTG Leu	Ser AGC	Cys TGC T	His CAC	Giy GGC	Trp TGG 	Val GTT C	Leu CTT
190 Gly GGA	Glu GAA	His CAT	Gly GGG C	Asp GAC	Ser TCC	Ser AGT	Val GTG	Pro CCC T	Ile ATA G.G Val	200 Trp TGG	Ser AGT	Gly GGT	Val GTA G	Asn AAC T	Val GTT	Ala GCT	61 y GGC	Val GTA C	Thr ACT T.C Ser	210 Leu CTG	Lys AAG	Ser TCA T	Leu CTG T	Asn AAC	Pro CCA	Ala GCA .A. Glu	Ile ATA C.G Leu	Gly GGA C	Thr ACT
220 Asp GAC	Ser TCA G Ala	Asp GAT	Lys AAG	Glu GAA G	His CAC G Gln	Trp TGG	Lys AAA G	Asn AAT G.G Glu	Val GTT	230 His CAC	Lys AAG	Gln CAG	Val GTG	Val GTG	Glu GAA C Asp	Gly GGC A.T Ser	Gly GGC .C. Ala	Tyr TAT	Glu GAG	240 Val GTC G	Leu CTT A.C Ile	Asn AAC G Lys	Met ATG C Leu	Lys AAG A	Gly GGC	Tyr TAT C	Thr ACC	Ser TCT	Trp TGG
250 Ala GCT C	Ile ATC T	G1y GGG C	Leu CTG	Ser TCT	Val GTG	Thr ACT G.A Ala	Asp GAT C	Leu CTG T	Ala GCG T	260 Arg CGA GAG G1u	Ser TCC AG.	Ile ATC	Leu TTG A Met	Lys AAG	Asn AAT C	Leu CTT	Lys AAG .G. Arg	Arg AGA C.G	Val GTG	270 His CAT	Pro CCT	Val GTT A Ile	Thr ACC T Ser	Thr ACG C	Leu CTG A Met	Val GTT A Ile	Lys AAG	Gly GGC	Phe TTC C Leu
280 His CAT T Tyr	G1 y GGG A	Ile ATA C	Lys AAG T Asn	Glu GAA G	Glu GAG T Asp	Val GTC	Phe TTC	Leu CTC	Ser AGT	290 Ile ATC G Val	Pro CCT	Cys TGT	Val GTC A	Leu TTG C	Gly GGA	Gln CAA	Ser AGT .A. As	Gly GGT A	Ile ATC	300 Thr ACA T.G Ser	Asp GAC	Phe TTT G Val	Val GTG	Lys AAA G	Val GTC G	Asn AAC .CA Thr	Met ATG C Leu	Thr ACC	Ala GCT C Pro
310 Glu GAG	Glu GAG	Glu GAG	Gly GGT .CC Ala	Leu CTC .G. Arg	Leu CTC	Lys AAG	Lys AAG	Ser AGT	Ala GCG	320 Asp GAC	Thr ACA	Leu CTC	Trp TGG	Asn AAT GGA Gly	Met ATG C Ile	Gìn CAG	Lys AAG	Asp GAT G Glu	Leu CTG	Gìn CAG	331 Leu TTA C Phe	Ter TAA	ACT	CGCC	ACC	TTCG	ACCG	ſG	
TGA	CAGA	TGC	CTGA	TCAC	AT C	ACTG	5 ATCA	0 C GG	CAGT	CCCA	CTG	AAAG	TGT	ттсс	ACAT	CA T	AACA	AAGT	T CA	ATAA i gna	AATT 1	ŦŦĠ	GAAA	сст	113 GTT	A70 Poly	(A)		

Fig. 2. Comparison of the nucleotide and amino acid sequences of mouse LDH-C cDNA with those of mouse LDH-A

The deduced amino acid sequence of mouse LDH-C cDNA is given above its nucleotide sequence. The previously reported [6] primary structure determined by direct protein sequencing contains a deletion of Glu at position 14, the reverse sequence of Ile–Val at positions 123–124, the misidentification of Ile as Val at position 134, and the misassignment of 10 acid/amide side chains of Asx and Glx at positions 6, 29, 55, 103, 222, 224, 242, 295, 328 and 330. The deamidations of Asn and Gln residues was probably due to the use of strong alkaline and acidic conditions in order to solubilize the aggregated LDH-C protein and peptides during the peptide purification and sequencing. For mouse LDH-A [13], only those nucleotides and amino acids which differ from those of LDH-C are indicated.

method [22,23]. An *Eco*RI fragment of length 1.2 kb from clone mC50 was isolated from M13 double-stranded DNA and further cleaved with *Sau*3A, *Alu*I or *Hae*III, the subfragments were cloned into M13 mp10 or mp11 phages, and their nucleotide sequences were determined.

A Northern blot of poly(A)-containing RNAs from mouse testes and liver was probed with the 0.4 kb *Eco*RI fragment insert from clone mC31 [19,24]. A Southern blot of genomic DNAs from mouse liver and human placenta was probed with the 1.0 kb *Eco*RI-*Pst*I cDNA for mouse sperm-specific lactate dehydrogenase C

fragment (LDH-C coding sequence) from clone mC50 [19,25].

RESULTS AND DISCUSSION

Cloning and sequence of LDH-C cDNA

Eight putative LDH-C cDNA clones were identified and plaque-purified from a screen of approx. 120000 recombinant phages in a λ gtll expression library constructed from the poly(A)-containing RNAs of mouse testes. All of these clones were positive to anti-(mouse LDH-C) serum but negative to anti-(human LDH-A) serum. Further, four clones exhibited hybridization to a human LDH-4 cDNA probe containing the conserved protein-coding sequence. The *Eco*RI fragments from two LDH-C cDNA clones, mC50 showing cross-hybridization to LDH-A cDNA probe and mC31 exhibiting no such hybridization, were isolated and subcloned into M13 mp10/mp11 phages, and their complete nucleotide sequences were determined (Figs. 1 and 2). Clone mC50 contains a cDNA insert of 1236 bp,



Fig. 3. Northern blot analysis of poly(A)-containing RNAs from mouse testes and liver

Samples of poly(A)-containing RNAs (20 μ g) from mouse testes and liver, and RNA size markers, were denatured with glyoxal, electrophoresed on 1% agarose, and blotted onto nitrocellulose membrane. The Northern blot was probed with the ³²P-labelled 0.4 kb *Eco*RI fragment from clone mC31. The hybridization was performed in a solution containing 50% (v/v) formamide, 5 × SSC, 5 × Denhardt's reagent, 50 μ g of salmon sperm DNA/ml, 1 μ g of poly(dA)/ml, 10% (w/v) dextran sulphate and 0.1% SDS at 42 °C overnight, and the membrane was washed twice with 2 × SSC/0.1% SDS and once with 0.1% SDS at 50 °C for 30 min each.



Fig. 4. Southern blot analysis of genomic DNAs from mouse liver and human placenta

The high- M_r DNAs from mouse liver (m) and human placenta (h) were isolated and cleaved by restriction endonucleases *Eco*RI, *Bam*HI or *Hin*dIII. After electrophoresis on 0.75% agarose, the DNA fragments were transferred to a nylon membrane (Genescreen; New England Nuclear) and probed with the protein-coding sequence (*Eco*RI-*Pst*I fragment, 1.0 kb) from clone mC50. The hybridization was carried out in a solution containing 50% (v/v) formamide, $10 \times$ Denhardt's reagent, 10% (w/v) dextran sulphate, 20 μ g of salmon sperm DNA/ml, 0.1% sodium pyrophosphate, 1.0 m-NaCl and 0.05 m-Tris buffer (pH 7.5) at 42 °C overnight, and the membrane was washed twice with 0.3 m-NaCl/1% SDS/2 mM-EDTA/ 0.06 m-Tris buffer (pH 8.0) at 65 °C for 30 min.

while clone mC31 possesses a shorter cDNA insert (372 base pairs). The nucleotide sequence of clone mC31 is identical with the corresponding sequence of clone mC50, indicating that both cDNA inserts were derived from the transcripts of the same LDH-C gene. The nucleotide sequence of the clone mC50 LDH-C cDNA contains a single 331-amino-acid open reading frame between the first ATG codon triplet of translation initiation and the termination codon triplet, TAA. The amino acid sequence predicted from the LDH-C cDNA is consistent with that previously determined by direct protein sequencing [6]. The LDH-C cDNA insert of clone mC50 consists of the 5' (54 nucleotides) and 3' (113 nucleotides) non-coding regions, and the poly(A) tail (70 nucleotides), in addition to the protein-coding sequence

(999 nucleotides, including initiation and termination codons). The Northern blot analysis of poly(A)containing RNAs from mouse testes and liver indicates that the LDH-C gene is expressed in testis but not in liver, and that its mRNA is approx. 1400 nucleotides in length (Fig. 3). Thus, clone mC50 contains a LDH-C cDNA insert of nearly full length, although the exact initiation site for transcription remains to be determined. The absence of an in-phase termination codon in the 5' non-coding region of clone mC50 permits the contiguous translation of the fused β -galactosidase-LDH-C polypeptide. It will be of interest to compare the enzymic properties of the LDH-C polypeptide between the fused protein and the native isoenzyme. The 3' non-coding region of mouse LDH-C mRNA contains a single polyadenylation signal (AAUAAA) located 16 nucleotides 5' to the poly(A) addition site [26]. This region is much shorter than that of mouse LDH-4 (484 nucleotides), and there was no obvious sequence homology between them. The nucleotide and amino acid sequences of mouse LDH-C, when compared with those of mouse LDH-A (Fig. 2), show 73% and 72% homologies, respectively. Furthermore, only 55% of the 265 nucleotide differences resulted in amino acid substitutions. The functional significance of these differences in amino acid sequence between LDH-C and LDH-A proteins remains to be investigated through the use of site-directed mutagenesis. The LDH-C cDNA insert of clone mC31 possesses the coding sequence for amino acids 252-331, the 3' non-coding region, and the poly(A) tail of 19 nucleotides. Clone mC31 showed strong reaction to anti-LDH-C serum, indicating the presence of (an) antigenic epitope(s). Indeed, two antigenic determinants were predicted previously in the 80 amino acids from the C-terminus of LDH-C protein [7].

Genomic complexity of LDH-C gene-related sequences

The Southern blot analysis of genomic DNAs from mouse liver and human placenta probed with the protein-coding 1.0 kb EcoRI-PstI fragment of mouse LDH-C cDNA demonstrates the presence of multiple LDH-C gene-related sequences (Fig. 4), as was the case with the mouse and human LDH-A genes and pseudogenes [14]. The chromosomal locations of the LDH-C gene and pseudogenes from man, mouse and other mammals remain to be mapped experimentally. The availability of nearly full-length mouse LDH-C cDNA should facilitate (a) the structural analysis of mammalian LDH-C genes, (b) studies on regulation of LDH-C gene expression during spermatogenesis, and (c) the development of a contraceptive vaccine using genetically engineered LDH-C protein.

We thank Dr. M. Okabe for preparation of rabbit antiserum against mouse LDH-C protein, Dr. B. H. Judd for his interest and support of this project, and Drs. P. Bainbridge, C. Teng and L. Wright for reading the manuscript.

REFERENCES

- 1. Market, C. L., Shakelee, J. B. & Whitt, G. S. (1975) Science 189, 102–114
- 2. Li, S. S.-L., Fitch, W. M, Pan, Y.-C. E. & Sharief, F. S. (1983) J. Biol. Chem. 258, 7029-7032
- 3. Goldberg, E. (1972) J. Biol. Chem. 247, 2044-2048
- 4. Goldberg, E. (1977) Isozymes: Curr. Top. Biol. Med. Res. 1, 79–124
- Wright, L. C. & Swofford, J. H. (1984) Scand. J. Immunol. 19, 247–254
- Pan, Y.-C. E., Sharief, F. S., Okabe, M., Huang, S. & Li, S. S.-L. (1983) J. Biol. Chem. 258, 7005–7016
- Li, S. S.-L., Feldmann, R. J., Okabe, M. & Pan, Y.-C. E. (1983) J. Biol. Chem. 258, 7017–7028
- Musick, W. D. L. & Rossman, M. G. (1979) J. Biol. Chem. 254, 7611–7620
- 9. Goldberg, E. (1973) Science 181, 458-459
- Wheat, T. E. & Goldberg, E. (1983) Isozymes: Curr. Top. Biol. Med. Res. 7, 113–130
- 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
- 12. Tsujibo, H., Tiano, H. F. & Li, S. S.-L. (1985) Eur. J. Biochem. 147, 9-15
- 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
- Chung, F.-Z., Tsujibo, H., Bhattacharyya, U., Sharief, F. S. & Li, S. S.-L. (1985) Biochem. J. 231, 537-541
- 15. Fukasawa, K. M. & Li, S. S.-L. (1986) Biochem. J. 235, 435-439
- Fukasawa, K. M., Li, W.-H., Yagi, K., Luo, C.-C. & Li, S. S.-L. (1986) Mol. Biol. Evol. 3, 330–342
- Okabe, M., Akai, K. & Li, S. S.-L. (1982) Int. J. Biochem. 14, 371–375
- Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 1408–1412
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory, New York
- Young, R. A. & Davis, R. W. (1985) in Genetic Engineering (Setlow, J. & Hollaender, A., eds.), vol. 7, pp. 29–41, Plenum, New York
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) J. Mol. Biol. 113, 237–251
- Messing, J., Crea, R. & Seeberg, P. H. (1981) Nucleic Acids Res. 8, 309–321
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467
- Stark, G. R. & Williams, J. G. (1979) Nucleic Acids Res. 6, 195–203
- 25. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517
- 26. Proudfoot, N. J. & Brownlee, G. G. (1976) Nature (London) 263, 211-214

Received 19 September 1986/1 December 1986; accepted 16 December 1986