Journal of Andrology, Vol. 24, No. 6, November/December 2003

Abundance of Repetitive Sequence Elements in the Mouse Testis-Specific Lactate Dehydrogenase-C Gene[1]

PER G. OLSSON,* HIROSHI TSUJIOKA,† SONOKO NARISAWA,† ERWIN GOLDBERG,‡ AND JOSÉ LUIS MILLÁN*†

*From the *Department of Medical Biosciences, Medical Genetics, Umeå University, Umeå, Sweden; †The Burnham Institute, La Jolla, California; and the ‡Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois.*

ABSTRACT: We have cloned and sequenced the entire mouse *ldhc* gene and mapped it physically in relation to the somatic *ldha* gene. The 2 genes were found to be oriented in head-to-tail fashion with about a 6-kilobase (kb) distance between the 3′ end of *ldha* and the 5′ end of *ldhc.* The *ldhc* gene is composed of 43% repetitive elements compared to only 16% in the *ldha* gene. Despite the close physical distance of mouse *ldha* and *ldhc,* the 2 genes have a very different content of repetitive elements, and this most likely reflects different levels of selective pressure.

Key words: Isozyme, restriction maps.

J Androl 2003;24:918–920

Lactate dehydrogenase (LDH)-C is the LDH isozyme found in the germinal epithelium during spermatogenesis and is most likely adapted to satisfy the metabolic requirements of the differentiating germ cells and functional spermatozoa (Blanco, 1980). It is unclear, however, if LDH-C is essential for spermatogenesis or if other isozymes (LDH-A and LDH-B) would perform the same function(s) in the absence of LDH-C. One way of acquiring insight would be to create animals that lack LDH-C by a targeted disruption of the *ldhc* gene. However, a number of attempts to create a ''knockout'' of the mouse *ldhc* gene have failed. In order to try to understand the reason for these failures, we cloned and sequenced the entire mouse *ldhc* gene.

Materials and Methods

The RPZD mouse p1-derived artificial chromosome (PAC) library was obtained from the Resource Center of the German Human Genome Project (RZPD; Berlin, Germany) and screened using the entire mouse *ldhc* complementary DNA (Wu et al, 1987) as a probe, and 2 confirmed positive clones (RPCIP711P15209Q3 and RPCIP711P01257Q3, here referred to as 209 and 257) were analyzed further. PAC DNA was isolated using a standard protocol (Meier-Ewert et al, 1998), and restriction maps of the parts containing mouse *ldhc* were made. Because the restriction patterns of the 2 PACs differed markedly, the exon content of the 2 PAC clones was investigated using polymerase chain reaction (PCR)-generated exon-specific probes. This showed that PAC 209 contained exons 1–5 while PAC 257 contained exons 5–8 and that the overlap between the 2 clones was about 1 kilobase (kb) (Figure, A). The PACs were subcloned into pBluescript (Stratagene, La Jolla, Calif) using the restriction enzymes *Bam*HI, *Pst*I, *Xba*I, and *Kpn*I. The subclones were screened for their exon content and sequenced by dye terminator cycle sequencing from the ends using exon-specific primers. Where needed, primer walking generated additional sequencing data. Sequences were then assembled using Assemblylign software. Mouse *ldhc* and mouse *ldha* genes have previously been mapped close to each other on mouse chromosome 7 (Stubbs et al, 1994). As it was possible that one of the two PACs that covered the mouse *ldhc* gene also contained parts of the mouse *ldha* gene, we used 3 sets of primers specific for mouse *ldha* (3′, exon4, and 5′) to determine the relative orientation and distance of the 2 genes. PAC 209 was positive for *ldha* in all 3 PCRs, while PAC 257 was negative, placing mouse *ldha* 5′ of mouse *ldhc.*

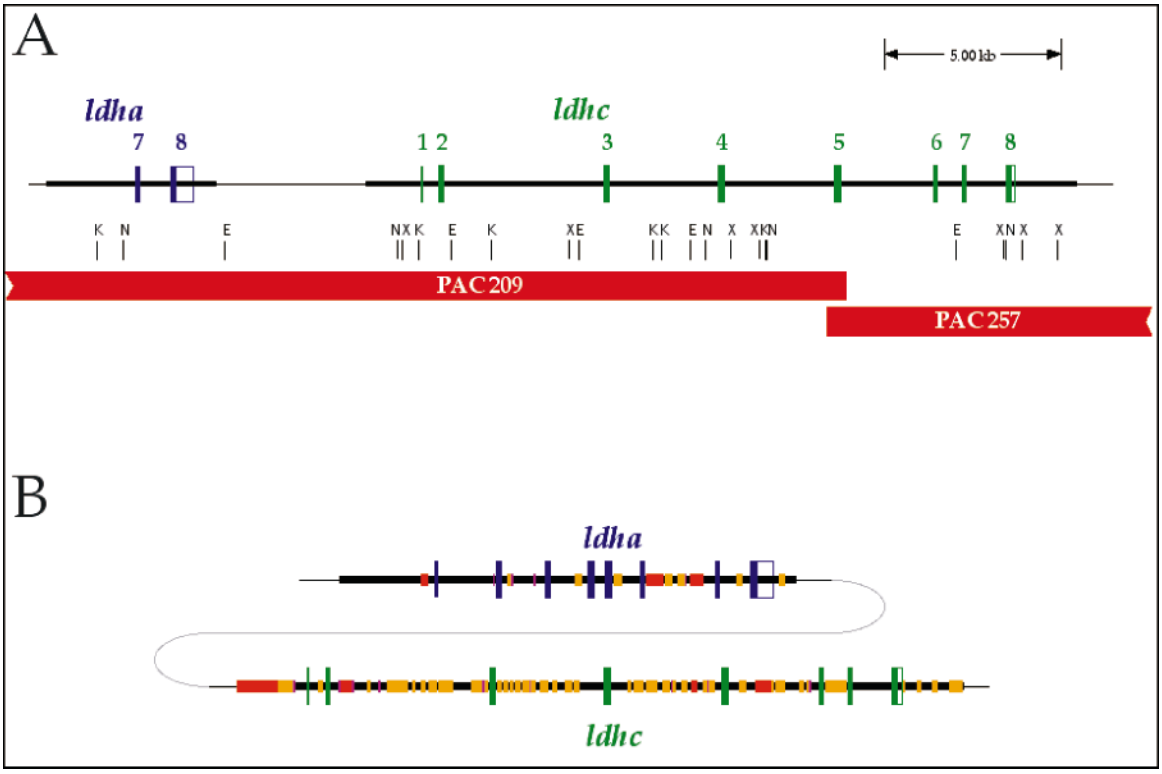
Results

Computer-generated restriction maps of the sequences of the *ldha* and *ldhc* genes were compatible with a head-to-tail orientation of the 2 genes, and this was confirmed by Southern blot analysis using the mouse *ldha* 3′ PCR product and a 134-bp *Sph*I-*Sac*I fragment 5′ of mouse *ldhc* as probes (Figure, A). As predicted, both probes shared the same size of *Nco*I, *Kpn*I, and *Xba*I fragments but had different *Eco*RI fragments, and double-digest using *Eco*RI in combination with *Nco*I and *Xba*I confirmed the predicted map. Thus, the end of the mouse *ldha* exon 8 is

[1] GenBank accession number AF190799.

Correspondence to: Dr José Luis Millán, The Burnham Institute, 10901 N Torrey Pines Rd, La Jolla, CA 92037 (e-mail: millan@burnham.org).

Received for publication April 21, 2003; accepted for publication July 21, 2003.



(A) Map of the *ldhc* locus. Thicker black line denotes the sequenced areas; boxes denote exons (open when untranslated), *ldha* in blue and *ldhc* in green. Restriction enzymes: E = *Eco*RI, K = *Kpn*I, N = *Nco*I, and X = *Xba*I. Primers used for mapping included the following: *ldhc* exon2: 5′-ATA AACTTTCCCGGTGTAAG-3′ and 5′-ATACTAATAGCACACGCCATG-3′, *ldhc* exon3: 5′-GGTTTGGCTGATGAACTTGC-3′ and 5′-CAAAGACGATTT TTGGAGTGC-3′, *ldhc* exon4: 5′-ATCTGCCAACTCCAAACTGG-3′ and 5′-ACTGGGTTAGTGACGATAAT-3′, *ldhc* exon5: 5′-TTGACATACGTGGTT TGGAAG-3′ and 5′-CCATGTTCTCCAAGAACCCA-3′, *ldhc* exon6: 5′-TGCCCATATGGAGTGGTGTA-3′ and 5′-CACCACCTGCTTGTGAACAT-3′, *ldhc* exon7: 5′-CGGCTATGAGGTCCTTAACA-3′ and 5′-CTTAACCAGCGTGGTAACAG-3′, *ldhc* exon8: 5′-GGCTTCCATGGGATAAAGGA-3′ and 5′-CTGTTA CCACGCTGGTTAAG-3′, *ldha* 5′: 5′-TTTCTTTGGGGTGTCGCAG-3′ and 5′-GCCTTAAATGGAAGCTCCG-3′, *ldha* exon4: 5′-CAAGCTGGTCATTAT CACCG-3′ and 5′-GGATTGGAGACGATCAGCAG-3′, and *ldha* 3′: 5′-TCTCGGATGTTGTGAAGGTG-3′ and 5′-TTTCCCCACACCATCTCAAC-3′. **(B)** Genomic organization and location of repetitive elements in the *ldha* gene, derived from the sequence available in GenBank, and of the *ldhc* gene as reported in this paper. Orange boxes denote SINE (short interspersed nuclear element) repeat elements (B1, B1F, B2, B3, B3A, MUSID2, MUSID4, RSINE1, RSINE2, and RSINE2A), red boxes denote LINE (long interspersed nuclear element), transposon, and endogenous retrovirus elements (L1MB6_5, L1P_MA2, MERVL, RMER6A, and URR1), and purple boxes denote short tandem repeats (ie, (GGGGA)n, (CAAAA)n, (CA)n, and (A)n)).

about 6 kb from the 5′ beginning of the mouse *ldhc* exon 1 (Figure, A). The entire mouse *ldhc*, including some flanking regions, was sequenced, and the intron-exon boundaries were identified. The total distance between mouse *ldhc* exons 1 and 8 is about 16 kb, and introns 2, 3, and 4 are all relatively large (3–4 kb in length), while introns 1 and 6 are shorter (<1 kb) (Figure, A).

Since previous attempts to generate knockout mice appear to have failed because of a high proportion of repetitive sequences in constructs containing the 5′ end of the gene, the PAC clones were initially screened for repetitive elements using mouse Cot-1 (Invitrogen Corp, Carlsbad, Calif) hybridization. Surprisingly, most of the Cot-1–positive bands were the same as the ones positive for mouse *ldhc*, indicating that while much of the area surrounding the gene is relatively free from repeats, the mouse *ldhc* gene appears to contain a large number of repetitive elements. We compared the 20-kb *ldhc* gene sequence to the ''Censor'' (Jurka et al, 1996) repeat database, and the results indicated that, indeed, 43% of the gene sequence was composed of repetitive elements. Only a 1.3-kb area around exon 4 and the entire intron 7 were free of repetitive elements. The adjacent mouse *ldha* gene was more compact (10 kb between exons 1 and 8) (Fukasawa and Li, 1987) than the mouse *ldhc* gene, and, when analyzed for repetitive elements, its content was significantly lower—16%. This is in agreement with our Cot-1 hybridization result, which indicated that the region adjacent to the mouse *ldhc* gene contained fewer repetitive elements than the gene itself (Figure, B).

Discussion

The 3 LDH isozymes from the mouse (and other mammals) share about 70% amino acid homology with each other. LDH-A and LDH-B have been conserved stringently during evolution; LDH-C has diverged more be-

tween species. While the mammalian LDH-B subunits show a conservation of the amino acid identity around 95% and the LDH-A subunits show 90%, the LDH-C homology drops to 75% conservation between mammalian species, which means that LDH-C proteins are as similar to each other as to the other 2 isozymes. Even at larger evolutionary distances, such as those that exist between the mouse and *Xenopus laevis,* the LDH-A and LDH-B proteins are more than 80% identical, which is higher than the identity between mouse and human LDH-C—74%. This means that the *ldhc* gene is allowed to diverge at a much higher rate than *ldha/ldhb* and must thus be subject to a different and less stringent selective pressure.

The tandem orientation of *ldha* and *ldhc* genes is also found in humans. The evolution of the *ldh* genes in vertebrates through fishes indicates that they stem from a common ancestor that was duplicated before the speciation of mammals. Indeed, the evolutionary tree suggested by Tsuji et al (1994) indicates that the duplication of *ldha* and *ldhc* occurred before *ldhb* existed. On the other hand, Markert et al (1975) proposed that the duplication of *ldha* gave rise to *ldhb* and that, in fishes, it was the duplication of the *b* gene that led to the appearance of *ldhc* in the eye or liver, depending on the species. We suggested (Millán et al, 1987) that the orientation of *ldha* and *ldhc* genes and the relative homology of the coding regions of the 3 LDH subunits indicate that *ldhc* represents a duplication of *ldha.* Despite the close physical distance of mouse *ldha* and *ldhc,* the 2 genes have a very different content of repetitive elements, and this most likely reflects the different levels of selection under which they have operated. The accumulation of repetitive elements in the mouse *ldhc* introns has doubled the size of the gene compared to *ldha.* The reason why the 2 genes have been kept in close proximity is unclear; it may be pure chance, but there may also be functional constraints that keep the 2 genes together, possibly common regulatory elements situated between the 2 genes.

Acknowledgments

This work was supported in part by grant HD05863 from the National Institutes of Health and a grant from the Swedish Medical Research Council.

References

Blanco A. *Johns Hopkins Med J.* 1980;146:231–235.

Fukasawa KM, Li SS. *Genetics.* 1987;116:99–105.

Jurka J, Klonowski P, Dagman V, Pelton P. *Comput Chem.* 1996;20:119–122.

Markert CL, Shaklee JB, Whitt GS. *Science.* 1975;189:102–114.

Meier-Ewert S, Schalkwyk L, Francis F, Lerach L. In: Spurr NK, Young BD, Bryant SP, eds. *ICRF Handbook of Genome Analysis.* Oxford: Blackwell Science Ltd; 1998:367–420.

Millán JL, Driscoll CE, LeVan KM, Goldberg E. *Proc Natl Acad Sci U S A.* 1987;84:5311–5315.

Stubbs L, Rinchik EM, Goldberg E, Rudy B, Handel MA, Johnson D. *Genomics.* 1994;24:324–332.

Tsuji S, Qureshi MA, Hou EW, Fitch WM, Li SS. *Proc Natl Acad Sci U S A.* 1994;91:9392–9396.

Wu KC, Chan K, Lee CY, Lau YF. *Biochem Biophys Res Commun.* 1987; 146:964–970.