Vol. 205, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1994 Pages 558-564

THE NUCLEOTIDE AND DEDUCED AMINO-ACID SEQUENCES OF A cDNA ENCODING LACTATE DEHYDROGENASE FROM *CAENORHABDITIS ELEGANS*: THE EVOLUTIONARY RELATIONSHIPS OF LACTATE DEHYDROGENASES FROM MAMMALS, BIRDS, AMPHIBIAN, FISH, NEMATODE, PLANTS, BACTERIA, MYCOPLASMA, AND PLASMODIUM

Stephen C.-M. Tsoi and Steven S.-L. Li *

Laboratory of Genetics
National Institute of Environmental Health Sciences
National Institutes of Health
Research Triangle Park, North Carolina 27709

Received October 10, 1994

The nucleotide and deduced amino-acid sequences of a cDNA encoding L-lactate dehydrogenase (LDH) from nematode, *Caenorhabditis elegans*, were reported. This first invertebrate LDH sequence of 333 amino acids, including the initiation methionine, exhibits 63% identity with that of the most primitive vertebrate lamprey. The evolutionary relationships among 36 LDH isozymes from mammals, birds, amphibian, fish, nematode, plants, bacteria, mycoplasma and plasmodium were analyzed. The invertebrate nematode LDH is evolutionarily positioned between plant LDH and mammalian testicular LDH-C isozymes. The mammalian LDH-C isozyme appears to have arisen after the invertebrate LDH, but prior to the divergence of vertebrate LDH-A (muscle) and LDH-B (heart) isozymes as described previously. © 1994 Academic Press, Inc.

L-lactate dehydrogenase (LDH, EC1.1.1.27) catalyzes the interconversion of L-lactate and pyruvate with nicotinamide adenine dinucleotide (NAD$^+$) as coenzyme (1). The LDH enzyme is widely distributed among animals, plants and bacteria. In vertebrates the LDH-A isozyme is best suited for pyruvate reduction in anaerobic tissues (muscle), whereas the LDH-B isozyme is superior for L-lactate oxidation in aerobic tissues (heart). In lower teleost fishes, LDH-C has a generalized tissue distribution, but in advanced teleost fishes it is found

---

* To whom all correspondence should be addressed.
FAX: (919) 541-7593.

Abbreviations: LDH, L-lactate dehydrogenase; NAD$^+$, nicotinamide adenine dinucleotide.

0006-291X/94 $5.00
Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.

either in the liver (e.g. cod) or in the eye (e.g. salmon). In mammals and columbid birds, the LDH-C isozyme is expressed only in mature testes. The relationship of these vertebrate LDH-C isozymes remains to be resolved conclusively. The LDH isozyme system is one of the most extensively studied models used to investigate the origin and evolution of isozymes and regulation of multigene families (2). However, no LDH sequences of invertebrates and reptiles have been determined. Here we report the first invertebrate LDH sequence deduced from a cDNA sequence of nematode, *Caenorhabditis elegans*. We have also analyzed the evolutionary relationships of 36 LDH sequences from mammals, birds, amphibian, fish, nematode, protozoa, plants, and bacteria.

MATERIAL AND METHODS

Cloning and sequencing of nematode LDH cDNA:

Nematode embryo cDNA library in λUni-ZAP XR (Stratagene) was screened using a DNA probe generated by the mixed cDNA templates and PCR-primers (5'-GTCAAGTCGGAATGGCGTGTGGC-3' and 5'-AACACTGGAAT CTCCATGTTCTCC-3') based on the published partial LDH sequence of *Caenorhabditis elegans* (3). This probe was labeled with digoxigenin system, and the positive clones were identified with chemilumininescent detection according to the procedure recommended by the manufacture (Boehringer Mannheim). A clone N104 was identified and purified. The cDNA insert excised to Bluescript SK(-) phagemid was labeled with the Dye Terminator kit, and its nucleotide sequence was determined using an automated DNA sequencer (Applied Biosystems model 373A). Both strands of the inserted DNA were completely determined.

Construction of evolutionary tree from 36 LDH sequences:

The complete amino-acid sequence of nematode, *Caenorhabditis elegans*, was deduced from the cDNA sequence determined in this investigation as well as the partial sequences reported (3). The unpublished LDH-A sequences from the teleost fish barracuda, *Sphyraena argentea*, and scorpaenid, *Sebastolobus alascanus*, were kindly provided by Dr. G. N. Somero. Thirty-three other LDH sequences were previously reported and were obtained from the databases. The scientific names of the organisms and the accession numbers of the published LDH sequences are as follows: human, *Homo sapiens*, A (X02152), B (Y00711), C (U13680); mouse, *Mus musculus*, A (U13687), B (X51905), C (X04752); rat, *Rattus norvegicus*, A (X01964), B (U07181), C (U07177); pig, *Sus scrofa*, A (U07178), B (U07180); bovine, *Bos taurus*, A (D90143); rabbit, *Oryctolagus cuniculus*, A (M22585); chicken, *Gallus domesticus*, A (X53828), B (P00337); duck, *Anas platyrhynchos*, B (J03869); African frog, *Xenopus laevis*, A (U07179), B (U07176), C (U07175); dogfish (shark), *Squalus acanthias*, A (A00350); killifish, *Fundulus heteroclitus*, B (M33969), C (L07336); lamprey, *Petromyzon marinus*, (M74064); barley, *Hordeum vulgare* (M55685); rice, *Oryza sativa* (D13817); maize, *Zea mays* (Z11754); bacteria, *Thermus aquaticus* (D00585), *Thermus caldophilus* (X04519), *Bacillus subtilis* (A25805), *Bacillus stearothermophilus* (A26053), *Bifidobacterium* *longum* (M33585); pig pathogenic bacterium (without cell-wall), *Mycoplasma hyopneumoniae* (X67286); and human malaria parasite (protozoon), *Plasmodium falciparum* (M93720). The amino-acid sequences of these 36 LDH isozymes were aligned using the pileup program of the Wisconsin GCG8 package based on the method of Feng and Doolittle (4). The evolutionary tree (cladogram) was constructed from the distance matrix of UPGMA method (5).

RESULTS

Nematode LDH sequence:

Nematode LDH cDNA clone N104 was isolated from an embryo cDNA library, and the nucleotide sequence of the insert DNA was completely determined and deposited in the GenBank database (accession no. U15420). The cDNA insert of 1,114 nucleotides includes the protein-encoding sequence of 992 nucleotides, 3′ noncoding region of 105 nucleotides, and a poly(A) tail of 17 nucleotides (Fig. 1). The first seven nucleotides of the protein-

Nematode LDH

```
atggcaaGCACTATCAAGGAAGTCTTCGCGGAAATTGCAGCGCCTGTTGAGAACAGTCACGGAAAGGTGACAGTCGTCGGAGTCGGTCAA
M  A  S  T  I  K  E  V  F  A  E  I  A  A  P  V  E  N  S  H  G  K  V  T  V  V  G  V  G  Q    30
GTCGGAATGGCGTGTGCCTACTCGATCCTTCAGCAAAACCTTGCCAACGAGTTGTGCCTTGTTGATGTGGTCGCTGATAAGTTGAAGGGA
V  G  M  A  C  A  Y  S  I  L  Q  Q  N  L  A  N  E  L  C  L  V  D  V  V  A  D  K  L  K  G    60
GAGATGATGGATCTTCAACACGGACTTGCGTTCACCCGCCATTGCACGGTCAAGGCTGATACTGACTACTCTATCACCGCTGGCTCCAAA
E  M  M  D  L  Q  H  G  L  A  F  T  R  H  C  T  V  K  A  D  T  D  Y  S  I  T  A  G  S  K    90
CTGTGTGTAGTGACAGCAGGTGCTCGTCAACGCGAGGGAGAGACCCGTCTCTCGCTGGTCCAGCGAAACGTCGAGATTTTCAAGGGAATC
L  C  V  V  T  A  G  A  R  Q  R  E  G  E  T  R  L  S  L  V  Q  R  N  V  E  I  F  K  G  I   120
ATCCCACAACTTGTCAAGTACTCCCCGGACACTTGCATCCTCGTCGTCTCGAATCCAGTTGATGTTCTCACCTATGTCACCTGGAAGTTG
I  P  Q  L  V  K  Y  S  P  D  T  C  I  L  V  V  S  N  P  V  D  V  L  T  Y  V  T  W  K  L   150
TCTGGACTCCCACGCGAGCGAGTCTTCGGATCCGGAACCAACCTCGATTCGGCTCGTTTCCGCTTCTTGCTCTCCGAGAAGTTGAACATC
S  G  L  P  R  E  R  V  F  G  S  G  T  N  L  D  S  A  R  F  R  F  L  L  S  E  K  L  N  I   180
GCCCCGTCATCGTGCCACGGATGGATTATCGGAGAACATGGAGATTCCAGTGTTGCTGTGTGGTCTGGTGTCAATGTTGCCGGAGTCACA
A  P  S  S  C  H  G  W  I  I  G  E  H  G  D  S  S  V  A  V  W  S  G  V  N  V  A  G  V  T   210
CTTCACGAGATCAAGCCAGACATCGGAGAGAAGACTGACAACGAACACTGGGAGGCTGAGATTCACAAGAAGGTCGTCGACAGCGCCTAC
L  H  E  I  K  P  D  I  G  E  K  T  D  N  E  H  W  E  A  E  I  H  K  K  V  V  D  S  A  Y   240
GAGATCATCAAGCTGAAGGGATACACTTCCTGGGCTATCGGACTTTCTGTTGCAAAGATTGCTCAGGGAATTTTCAGTAATTCTCGCAAC
E  I  I  K  L  K  G  Y  T  S  W  A  I  G  L  S  V  A  K  I  A  Q  G  I  F  S  N  S  R  N   270
GTTTTCGCTCTTTCGACTAACGTGAAGGGCTTCCATGGAATCAACGATGATGTCTACCTTTCTCTTCCAGTCGTTCTTGGATCCGCAGGA
V  F  A  L  S  T  N  V  K  G  F  H  G  I  N  D  D  V  Y  L  S  L  P  V  V  L  G  S  A  G   300
CTCACCCACGTCGTCAAGCAGCAACTTACCGAGGCCGAGGTTCAGAAATTGCACAACTCCGCCAAGGCTCTATTGGAAGTTCAGAATGGA
L  T  H  V  V  K  Q  Q  L  T  E  A  E  V  Q  K  L  H  N  S  A  K  A  L  L  E  V  Q  N  G   330
ATCGTCATGTAAAAATTAAATCATTTAACCTTTTTCCTTCAGTCATCGAACTGTTTATATAGCATACTTCCCCTTGTCACAGAATCTCAT
I  V  M  *                                                                                   333
TGAATAAATAGTTGTATTCCTTTTAAAAAAAAAAAAAAAAA
```

Fig. 1. The nucleotide and deduced amino-acid sequences of nematode LDH.

The sequence of first seven nucleotides in the protein-encoding region, which was truncated in the cDNA of clone N104, was that of clone CM10C10 reported previously (3). The partial sequence of 426 nucleotides from clone CM10C10 is 99% identical to the sequence determined in this investigation. One missense substitution at position 383 may be due to allelic polymorphism. However, two deletions at positions 43 and 107, and two insertions between positions 335-336 and 420-421 in the CM10C10 sequence are probably due to the sequencing errors because they are not in the open-reading frame. The termination codon TAA is indicated by star. The putative polyadenylation signal AATAAA is underlined.

encoding region from clone N104 were truncated, and the sequence presented in Fig. 1 was that of clone CM10C10 reported previously (3). The deduced sequence of 333 amino acids, including initiation methionine, of this invertebrate nematode LDH exhibits 63% identity with the single LDH sequence of the most primitive vertebrate lamprey (6).

Evolutionary tree of 36 LDH sequences:

The evolutionary relationships among amino acid sequences of 36 LDH isozymes from mammals, birds, amphibian, fish, nematode, plants, bacteria, mycoplasma, and plasmodium were analyzed, and the cladogram usig UPGMA method is presented in Fig. 2. The nematode LDH is evolutionarily positioned between plant LDH and mammalian testicular LDH-C isozymes. The single LDH of lamprey appears to be most closely related to the muscle-type of

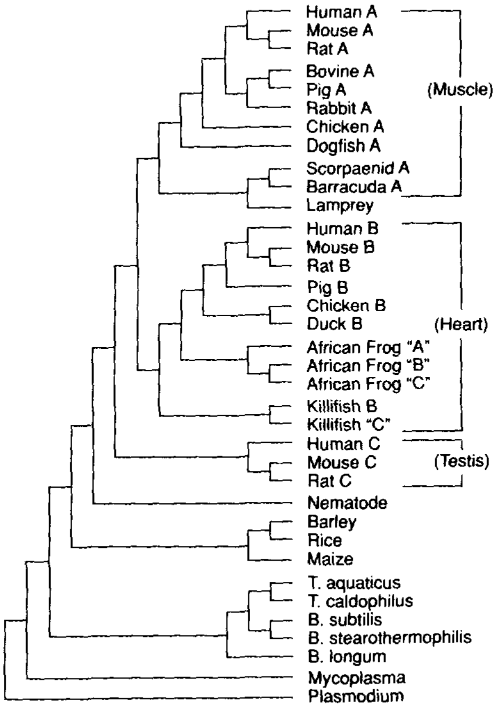

Fig. 2. Evolutionary tree of 36 LDH sequences.

Two types of vertebrate LDH isozymes present in somatic tissues are denoted as LDH-A (muscle) and LDH-B (heart), while the third type of LDH isozyme found in mammalian testis is denoted as LDH-C. The African frog LDH-"A" (muscle), LDH-"B" (liver) and LDH-"C" (oocyte) are most closely related to the heart-type of land vertebrate LDH-B (heart) isoyzmes. The killifish LDH-"C" (eye) is a recent duplication of killifish LDH-B (heart).

vertebrate LDH-A isozymes. The three African frog LDH sequences, "A" (muscle), "B" (liver) and "C" (oocyte), were shown to be very close to each other, and are most closely related to the heart-type of land vertebrate LDH-B isozymes (11). The killifish LDH-"C" (eye) was reported to be derived from a duplication of LDH-B (heart) isozyme (10). Thus, the African frog LDH-"C" (oocyte) and killifish LDH-"C" (eye) isozymes are not orthologous to mammalian LDH-C (testis) isozymes. The LDH from mycoplasma, bacterium without cell-wall, is quite different from those of other five bacteria. The LDH from protozoon parasite plasmodium is most different from all other LDH isozymes.

DISCUSSION

The evolutionary tree shown in Fig. 2 is generally consistent with many previous analyses using different computer programs of LDH sequences, including Li et al. (7), Crawford et al. (8), Hiraoka et al. (9), Quattro et al. (10) and Tsuji et al. (11), in that the mammalian testicular LDH-C isozyme diverged prior to the duplication of the somatic LDH-A and LDH-B isozymes. This evolutionary tree does not support that the mammalian LDH-C arose from either LDH-A or LDH-B after the divergence of LDH-A and LDH-B as suggested by Holmes (12), Markert et al. (2), Millan et al. (13), Hendriks et al. (14), Stock and Whitt (6).

The nematode and all vertebrate LDH isozymes contain an additional amino-terminal arm of approximately 20 amino-acids in comparison with those of LDH enzymes from bacteria, mycoplasma and plasmodium (Fig. 3). Three plant LDH enzymes also possess an amino-terminal arm of approximately 40 amino-acids. All vertebrate and plant LDH isozymes are tetrameric, while bacterial LDH enzymes are dimeric (1). The nematode LDH is likely to be tetrameric, since the primary function of amino-terminal arm apparently is to facilitate the tetramer formation by interacting with the carboxyl-terminus of LDH subunit (7). It should be noted that LDH isozymes from snow crab were shown to be tetrameric (15).

The nematode LDH cDNA clone N104 was isolated from an embryo cDNA library. In a survey of expressed genes at the mixed stages of nematode, partial sequences of LDH cDNAs from three clones CM10C9, CM10C10, and CM14D3 were determined, and two distinct sites closely located on chromosome II were found to be positively

Species	LDH isozymes	Amino-terminus	Carboxyl-terminus
Human	A	ATLKDQLIYNLLKEEQ--TPQNKITVVG..........	ADTLWGIQKELQF
Mouse	A	ATLKDQLIVNLLKEEQ--APQNKITVVG..........	ADTLWGIQKELQF
Rat	A	AALKDQLIVNLLKEEQ--VPQNKITVVG..........	ADTLWGIQKELQF
Bovine	A	ATLKDQLIQNLLKEEH--VPQNKITIVG..........	ADTLWGIQKELQF
Pig	A	ATLKDQLIHNLLKEEH--VPHNKITVVG..........	ADTLWGIQKELQF
Rabbit	A	AALKDQLIHNLLKEEH--VPQNKITVVG..........	ADTLWGIQKELQF
Chicken	A	SLKDHLIHNVHKEEHA-HAHNKISVVG..........	ADTLWGIQKELQF
Dogfish	A	ATLKDKLIGHLATSQEP-RSYNKITVVG..........	ATTLWDIQKDLKF
Scorpaenid	A	STKEKLISHVMKEEPV-GSRNKVTVVG..........	AETLWGVQKELTL
Barracuda	A	STKEKLIGHVMKEEPI-GSRNKVTVVG..........	AETLWGVQKELTL
Lamprey		ASTKGKLIHEMVPSKERDPPHSKVTIVG..........	SAETLWNVQKELTF
Human	B	ATLKEKLIAPVAEEEAT-VPNNKITVVG..........	SADTLWDIQKDLKDL
Mouse	B	ATLKEKLIASVADDEAA-VPNNKITVVG..........	SADTLWDIQKDLKDL
Rat	B	ATLKEKLIAPVADDETA-VPNNKITVVG..........	SADTLWDIQKDLKDL
Pig	B	ATLKEKLIAPVAEEETT-IPNNKITVVG..........	SADTLWGIQKDLKDL
Chicken	B	ATLKEKLITPVAAGST--VPSNKITVVG..........	SADTLWSIQKDLKDL
Duck	B	ATLKEKLMTPVAAASA--VPSSKITVVG..........	SADTLWSIQKDLKDM
African frog	"A"	ASVQEKLITCVCQDKPA-KPTNKITIVG..........	SAETLWSIQKDLKDL
African frog	"B"	STVQEKLITNVCQDKAA-KPTNKITIVG..........	SAETLWGIQKDLKDL
African frog	"C"	SSVQENLITNVCQDKAA-KPTNKITIVG..........	SSETLWGIQKDLQVL
Killifish	B	SSVLQKLITPLASSSAE-PPRNKVTVVG..........	SADTLWGIQKDLKDL
Killifish	"C"	ASVLHKLITPLACSSPE-PPRNKVTVVG..........	SASTLWDIQKDLRDI
Human	C	STVKEQLIEKLIEDDE--NSQCKITIVG..........	SAETLWNIQKDLIF
Mouse	C	STVKEQLIQNLVPEDK--LSRCKITVVG..........	SADTLWNMQKDLQL
Rat	C	STVKEQLIQNLAPDEK--QSRCKITVVG..........	SCDILWNIQKDLQL
Nematode		ASTIKEVFAEIAAPVE---NSHGKVTVVG..........	SAKALLEVQNGIVM
Barley		HKASSLSELGFDAGGASSGFFRPVADGCPATPTSSAVPH-RRLTKISVIG..........	SAKTLWENCQLLDL
Rice		KKASSLSELGFDAEGASSGFFRPVADG-GSTPTS----HRRRLTKISVIG..........	SAKTLWENCQLLDL
Maize		KKATSLSELGFDAGDASSGFFRPVSGD-SSTPTS--QHHRRRLTKVSVIG..........	SAKTIWENCQLLGL
T. aquaticus		VKVGIVG..........	SAEILKGAASALGF
T. caldophillus		KVGIVG..........	SAEILKEAAFALGF
B. subtilis		NKHVNKVALIG..........	SAGVLKNILKPHFVN
B. stearothermophilus		KNGGARVVVIG..........	SAATLKSVLARAFTR
B. longum		AETTVKPTKLAVIG..........	SAETLKETAAQFGF
Mycoplasma		KPIKIALIG..........	VAIIDKIYQDAIKNI
Plasmodium		APKAKIVLVG..........	EAIAETKRMKALA

Fig. 3. Amino- and carboxyl-terminal sequences.

The amino-acid sequences of 36 LDH isozymes were aligned using Wisconsin GCG8 package. Only amino- and carboxyl-terminal sequences are presented. The gaps are denoted by hyphen. The dots indicate the sequences not presented. The initiation methionine is post-translationally removed, and the amino-terminus of vertebrate LDH isozymes is then acetylated.

hybridized using the LDH cDNA probe of clone CM14D3 (3). The functional expression of these two loci and the number of LDH isozymes in the nematode remain to be investigated.

ACKNOWLEDGMENTS

We thank Professor G. N. Somero for providing the unpublished amino acid sequences of two fish LDH-A isozymes, Drs. R. H. Waterston and F. Schachat for informing us the partial sequences of nematode LDH, Ms. E. W. Hou for assisting automatic DNA sequencing, Dr. B. K. Choudhury for computer analyses of LDH sequences, and Drs. F. M. Johnson and J. E. Welch for reading the manuscript.

REFERENCES

1. Holbrook, J.J., Liljas, A., Steindel, S.J., and Rosmann, M.G. (1975) in The Enzyme: Oxidation-Reduction, eds. Boyer, P. (Academic Press, NY), Vol. 11, part A, pp.191-192.
2. Markert, C.L., Shaklee, J.B., and Whitt, G.S. (1975) Science 189, 102-114.
3. Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J., and Sulston, J. (1992) Nature Genetics 1, 114-123.
4. Feng, D., and Doolittle, R.F. (1987) J. Mol. Evol. 25, 351-360.
5. Swofford, D.L., and Olsen, G.J. (1990) in Molecular Systematics, eds. Hillis, D.M. & Moritz, C. (Sinauer, Sunderland, MA), pp. 411-501.
6. Stock, D.W., and Whitt, G.S. (1992) Proc. Natl. Acad. Sci. USA 89, 1799-1803.
7. Li, S.S.-L., Fitch, W.M., Pan, Y.-C.E., and Sharief, F.S. (1983) J. Biol. Chem. 258, 7029-7032.
8. Crawford, D.L., Consteantino, H.R., and Powers, D.A. (1989) Mol. Biol. Evol. 6, 369-383.
9. Hiraoka, B.Y., Sharief, F.S., Yang, Y.-W., Li, W.-H., and Li, S.S.-L. (1990) Eur. J. Biochem. 189, 215-220.
10. Quattro, J.M., Woods, H.A., and Powers, D.A. (1993) Proc. Natl. Acad. Sci. USA 90, 242-246.
11. Tsuji, S., Qureshi, M.A., Hou, E.W., Fitch, W.M., and Li, S.S.-L. (1994) Proc. Natl. Acad. Sci. USA 91, 9392-9396.
12. Holmes, R.S. (1972) FEBS Lett. 28, 51-55.
13. Millan, J.L., Driscoll, C.E., LeVan, K.M., and Goldberg, E. (1987) Proc. Natl. Acad. Sci. USA 84, 5311-5315.
14. Hendriks, W., Mulders, J.W.M., Bibby, M.A., Slingsby, C., Bloemendal, H., and de Jong, W.W. (1988) Proc. Natl. Acad. Sci. USA 85, 7114-7118.
15. Angers, A., Pothier, F., Sevigny, J.-L., and Sainte-Marie, B. (1994) Comp. Biochem. Physiol. 108B, 385-395.