J Mol Evol (1995) 41:46–53

JOURNAL OF MOLECULAR EVOLUTION

Caenorhabditis elegans Contains Genes Encoding Two New Members of the Zn-Containing Alcohol Dehydrogenase Family

Jeremy D. Glasner,[1] Thomas D. Kocher,[1,2] John J. Collins[1,3]

[1] Program in Genetics, University of New Hampshire, Durham, NH 03824, USA
[2] Department of Zoology, University of New Hampshire, Durham, NH 03824, USA
[3] Department of Biochemistry and Molecular Biology, University of New Hampshire, Durham, NH 03824, USA

Received: 7 January 1994 / Accepted: 30 September 1994

Abstract. We have characterized two cDNA clones from the nematode *Caenorhabditis elegans* that display similarity to the alcohol dehydrogenase (ADH) gene family. The nucleotide sequences of these cDNAs predict that they encode Zn-containing long-chain ADH enzymes. Phylogenetic analysis suggests that one is most similar to dimeric class III ADHs found in diverse taxa; the other is most similar to the tetrameric forms of ADH previously described only in fungi.

Key words: Alcohol dehydrogenase — *Caenorhabditis elegans* — Evolution — Gene families

Introduction

Alcohol dehydrogenase constitute a large family of related enzymes and isozymes. Numerous ADH genes, cDNAs, and proteins have been sequenced from a phylogenetically diverse group of taxa. Decades of intense study have revealed many of the complexities underlying the evolution of this gene family. There have been many instances of gene duplication and divergence in the evolutionary history of ADH (Cederlund et al. 1991). Variation is observed at every level of protein structure and ADH enzymes exhibit a wide range of substrate specificity (Jornvall et al. 1987). Changes are seen with respect to coenzyme and cofactor binding as well (Sun and Plapp 1992). Diverse mechanisms regulate ADH gene expression. Extensive variation is seen with respect to tissue-specific expression and developmental regulation (Corbin and Maniatis 1990) as well as cellular compartmentalization (Shain et al. 1992). The evolution of these genes and enzymes is further complicated by instances of gene conversion (Shain et al. 1992), gene amplification (Paquin et al. 1992), pseudogene formation (Matsuo and Yokoyama 1990), and functional convergence among distantly related enzymes (Danielsson and Jornvall 1992).

Known ADHs can be divided into three main groups based on the metal cofactors required for catalysis. The first group, represented by *Drosophila* ADHs, are short-chain dehydrogenases that require no metal ions (Jornvall et al. 1981). The second group includes ADH from *Zymomonas mobilis,* a fermentative bacterium, and possibly enzymes from other prokaryotes and yeast (Williamson and Paquin 1987; Conway and Ingram 1989). This group is characterized by the requirement of iron for activity. The third group represents a functionally heterogeneous group of proteins including representative prokaryotic, protistan, fungal, plant, and animal ADHs. Examples include the fermentative enzyme from yeast and the classic liver ADH (Sun and Plapp 1992). A comparison of activities among members of this family reveals that some are distinct enzymes whereas others are simple isozymes. These enzymes all require Zn as a cofactor and are long-chain dehydrogenases.

Correspondence to: J.J. Collins

The three-dimensional crystal structure has been solved at 2.4-Å resolution for the horse liver ADH (Eklund et al. 1976). This information has been used to predict the tertiary structure of other members of the ADH family. Within the Zn-containing long-chain ADHs, major types can be distinguished based on the number of subunits in the active form and the number of Zn atoms bound by each subunit. The plant and animal ADHs characterized to date are all thought to be dimeric enzymes (with two Zn atoms/subunit). Tetrameric forms of ADH (with two Zn atoms/subunit) have been identified in fungi, exemplified by *Saccharomyces cerevisiae* ADHI and ADHII (Jornvall 1977; Russell et al. 1983), but not in plants or animals (Sun and Plapp 1992). In addition, there are several proteins which possess distinct enzymatic activities that share significant sequence identity with Zn-containing ADHs. These include the vertebrate sorbitol dehydrogenases which are tetrameric and bind one Zn atom/subunit and ζ-crystallin from guinea pig lens (Sun and Plapp 1992).

Although ADH genes, cDNAs, and proteins have been sequenced from a wide range of organisms, significant gaps remain in the phylogenetic distribution of the taxa studied. A striking example of such a "gap" exists in the invertebrates. For this immense and diverse animal group, our knowledge is limited to *Drosophila* and cephalopod ADHs. The well-characterized "short-chain" ADHs were thought to be the only insect forms until the recent report of a Zn-containing ADH in *Drosophila* (Danielsson et al. 1994). The Zn-containing ADHs found in cephalopods (Fernandez et al. 1993) and *Drosophila* are typical class III ADHs. The observation of glutathione-dependent formaldehyde dehydrogenases in these invertebrates supports the ubiquitous phylogenetic distribution of class III ADHs. However, members of other classes of Zn-containing ADH have not been reported in invertebrate taxa.

We report here the analysis of two ADH-encoding cDNAs from a well-characterized metazoan invertebrate, the nematode *Caenorhabditis elegans*. This organism has emerged as a premier model organism for molecular genetic analysis of metazoan development (Wood 1988). As one of the most anciently diverged metazoans, this species provides important perspective on the distribution of the ADH gene family in animals.

Methods

ADH cDNA Clones. cm01d5 and cm14h3 were isolated from a sorted cDNA library constructed in bacteriophage lambda vector *SHLX2* by Chris Martin (Palazzolo et al. 1990). They were generously provided to us by R. Waterston. The cDNA inserts were automatically subcloned into the pRATII plasmid vector by growing recombinant phage in an *Escherichia coli* host expressing P1 recombinase (*Cre*), promoting recombination between *loxP* sites flanking plasmid sequences in the *SHLX2* vector. Plasmids were purified using Magic Miniprep columns (Promega) according to the manufacturer's recommended conditions.

Genomic Southern Blot Analysis. Our methods for nematode growth, DNA extraction, and Southern blot analysis have been described elsewhere (Collins et al. 1989). To prepare hybridization probes, cDNA inserts were amplified via PCR from plasmid templates, using oligonucleotide primers specific for SP6 and T7 promoter sequences (see below) that flank inserts in the pRATII vector. Amplification products were radiolabeled by the random primer method (Feinberg and Vogelstein 1983).

DNA Sequencing. To obtain templates suitable for DNA sequence analysis we amplified cDNA inserts via PCR using SP6 and T7 primers. Templates were sequenced using the cycle sequencing method with dye-labeled dideoxy terminators according to the method of the manufacturer (Applied Biosystems). Extension products were purified in Centri-sep spin columns (Princeton Separations) and sequence determined using the ABI 373A automated sequencer.

Initial sequence was obtained from the SP6 primer. Complete sequences from each clone were obtained by designing additional primers as sequencing progressed.

Primer sequences are as follows:

SP6 (5′-ATAGAATATGCATCAAGCTGAG-3′)
T7 (5′-ACTATAGGGAGCTAAGCTTGG-3′)
cm01d5-250 (5′-CCTGTGCTATGGGAATGAGA-3′)
cm14h3-406 (5′-GATTTATGCCCGATGGTTCA-3′)
cm01d5-570 (5′-CGTCTTGACGTAGATGAAGC-3′)
cm14h3-739 (5′-CCAAATTCTTCGGAGCCACC-3′)
cm14h3-1094 (5′-TACTCATCGCTGGAACATTG-3′)

Sequence Analysis. Deduced amino acid sequences were derived for both *C. elegans* ADH cDNAs. These sequences were aligned with 22 other members of the Zn-containing ADH family using PILEUP (Devereux et al. 1984), a program that creates a progressive alignment based on pairwise comparisons of sequences (Feng and Doolittle 1987). A phylogenetic tree was constructed from the overlapping portion of 24 aligned ADHs (residues 123–422 from Fig. 2) using the protein parsimony algorithm (Swofford 1993). Subsets of the genes were used to construct trees which were subjected to bootstrap analysis (Felsenstein 1985).

Results

Characterization of ADH-Encoding cDNAs

In their analysis of expressed genes from *C. elegans*, Waterston and co-workers identified two cDNA clones, designated cm01d5 and cm14h3, that encode ADH isoforms (Waterston et al. 1992). BLAST analysis (Altschul et al. 1990) revealed that cm01d5 is most similar to fungal ADHs and cm14h3 is most similar to forms of this enzyme found in animals.

We tested whether these cDNAs are indeed encoded by the *C. elegans* genome (and not the result of reverse transcription of mRNA from a contaminating organism). This concern is especially relevant in this case because of the suggestion that one of these cDNAs exhibits similarity to a form of ADH heretofore identified only in fungi. Such organisms are common contaminants in nematode cultures. To investigate this question, we used each clone as a probe on genomic Southern blots of *C. elegans*

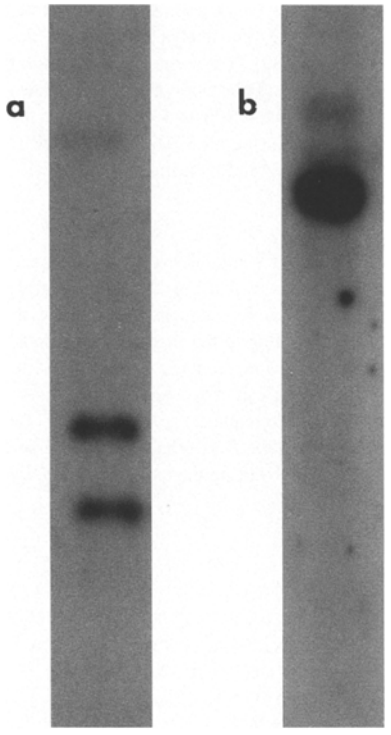

Fig. 1. Southern blots of *C. elegans* genomic DNA. **a** N2 DNA cut with *Eco*RI and probed with cm14h3 cDNA. **b** N2 DNA cut with *Hin*dIII and probed with cm01d5 cDNA.

DNA. The "high stringency" conditions used should not allow hybridization between heterologous sequences. Figure 1a shows a blot of DNA from the wild-type strain Bristol N2 digested with *Eco*RI. The cm14h3 probe hybridized with two fragments. We estimate these fragments to be approximately 3 kb and 4 kb based on migration with respect to size markers (data not shown). This result is consistent with hybridization to a single-copy gene because sequence analysis reveals an *Eco*RI site within the cDNA insert. Hence, we expect the genomic region covered by cm14h3 to consist of at least two *Eco*RI fragments. The blot shown in Fig. 1b contains N2 DNA cut with *Hin*dIII. The cm01d5 probe hybridized strongly with a 7-kb fragment. Faint hybridization to a second fragment of approximately 10 kb is also apparent. This is consistent with hybridization to a single-copy gene since the sequence of cm01d5 reveals a single *Hin*dIII site near the 5′ end of the cDNA insert of this clone. We performed similar analysis using a variety of restriction enzymes (data not shown). In each case the results support the conclusion that cm14h3 and cm01d5 are each derived from single-copy *C. elegans* genes.

C. elegans *Contains Two Distinct Types of ADH*

To determine the relationship of these *C. elegans* ADH genes to each other and to other members of the ADH gene family we determined the nucleotide sequence of the entire cDNA insert for each clone. Our strategy (described in detail in Materials and Methods) involved "walking" across each insert by designing oligonucleotide primers near the 3′ end of each stretch of sequence obtained. The complete sequence of the coding strand was determined for each insert (GenBank accession numbers U18780 and U18781).

The cm14h3 cDNA insert is 1,358 bp and appears to contain the complete coding region. We base this conclusion on the following features: At the 5′ end of the cDNA, a putative translational start (AUG) is preceded by 35 nucleotides that likely represent a stretch of 5′ untranslated region (UTR). This AUG is followed by a long open reading frame (ORF) capable of encoding a protein of 384 amino acids. This ORF ends with a UAA stop codon, followed by 169 nucleotides of 3′ UTR ending with a tract of 20 adenosine residues. The presence of a putative polyadenylation signal 15 nucleotides upstream of this polyadenosine stretch supports the idea that it represents the poly(A) tail of this message.

cm01d5 contains a 903-bp cDNA insert. We conclude that it contains the entire 3′ end of the corresponding mRNA because of the presence of a UAA stop codon followed by 160 bp of 3′ UTR. A putative polyadenylation signal is located near the 3′ end of this region, preceding a poly(A) tract of 20 residues. The 5′ end of this cDNA is not complete; it begins within the coding region of the corresponding transcript. Based on other ADHs of this type (see below) we expect that approximately 108 N-terminal codons (324 nucleotides) are missing. We have information for 239 amino acids out of an anticipated 347 amino acids. These 239 residues were easily aligned to other members of the Zn-containing ADH family and provided a sufficient number of informative characters for phylogenetic analysis.

BLAST searches (Altschul et al. 1990) were performed with the predicted amino acid sequences of both cDNAs. Results revealed that cm14h3 encodes an ADH with significant similarity to dimeric ADHs. Similarity was greatest to mammalian class III ADHs (score of 978, associated Poisson probability of 3.4e-133 for *Mus musculus* ADHX). The cm01d5-encoded ADH is most similar to tetrameric ADHs from yeast (score of 301, associated probability of 8.0e-62 for *Kluyveromyces lactis* ADHI). Figure 2 contains the two *C. elegans* sequences aligned with 22 other members of the Zn-containing ADHs. ADH sequences included in the analysis were chosen to represent all major classes of Zn-containing ADHs characterized to date. The alignment illustrates the high degree of overall sequence divergence between members of this enzyme family. At the same time, remarkable conservation is evident for certain functionally important residues in proteins from taxa as different as prokaryotes and humans. Table 1 shows some of the conserved residues found in the *C. elegans* sequences.

Figure 3 shows a phylogenetic tree illustrating the branching relationships between members of the Zn-

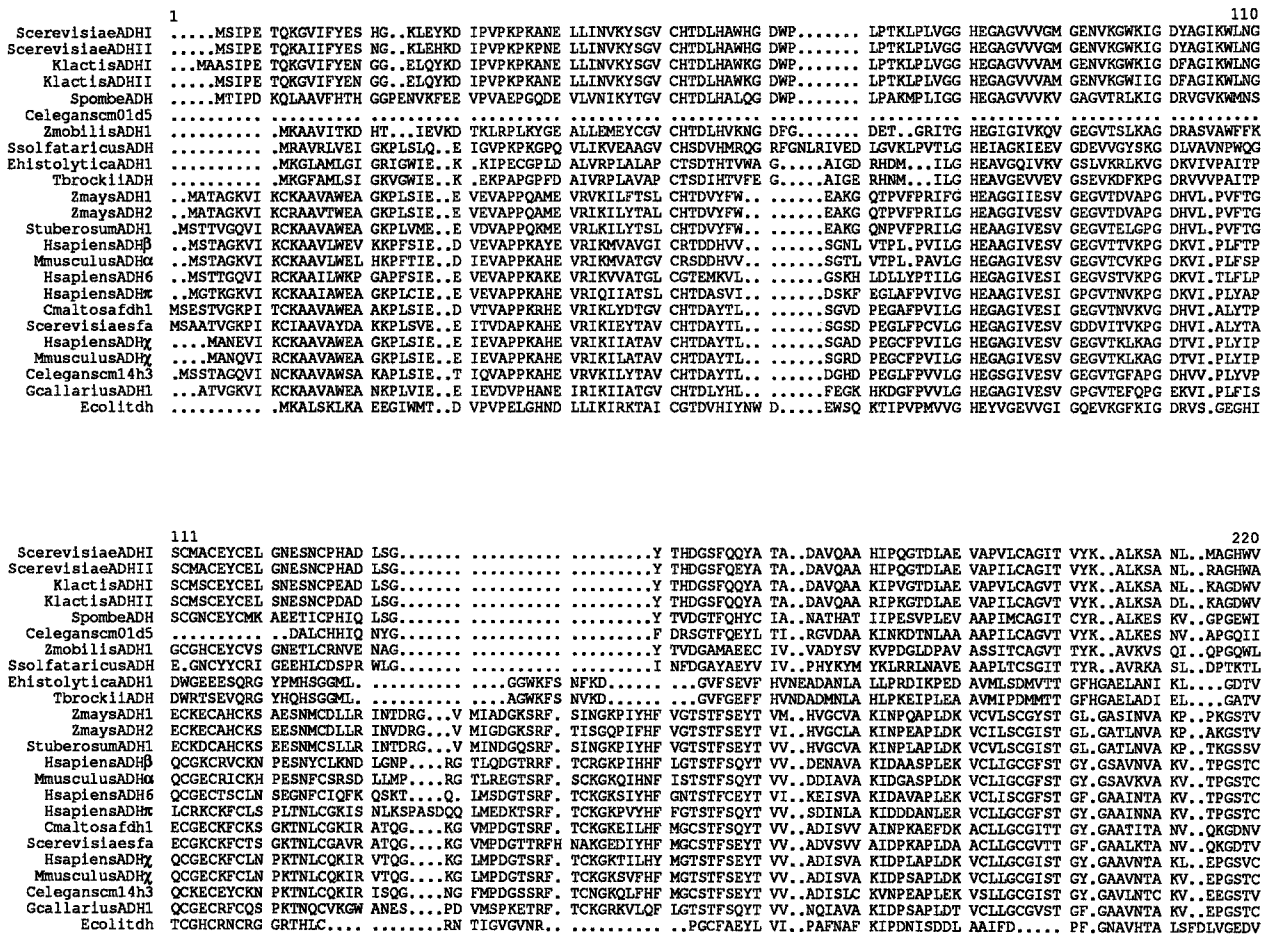

Fig. 2. PILEUP alignment of 21 members of the Zn-containing ADH family, the two predicted amino acid sequences from *C. elegans,* and *E. coli* threonine dehydrogenase. References for most sequences are given in Sun and Plapp (1992). Exceptions include *C. maltosa* FDH (Sasnauskas et al. 1992), *K. lactis* ADHII (Shain et al. 1992), *S. cerevisiae* SFA (Wehner et al. 1993), *S. solfataricus* ADH (Ammendola et al. 1992), and *H. sapiens* ADH6 (Yasunami et al. 1990).

containing ADH family. Several distinct clusters of related enzymes are apparent in the tree. A cluster at the top of the Fig. 3 contains fungal ADHs, two prokaryotic ADHs, and *C. elegans* ADH encoded by cm01d5. The cluster at the bottom of Fig. 3 contains plant and animal ADHs and the *C. elegans* ADH encoded by cm14h3. The presence of two yeast ADHs within this cluster will be addressed in the Discussion. A third cluster contains ADH from *Entamoeba histolytica* and *Thermoanaerobium brockii.* This tree reveals a distinct separation between tetrameric (the "fungal" cluster) and dimeric (the "plant/animal" cluster) forms of ADH.

Figure 4 contains a phylogenetic tree constructed with the two *C. elegans* sequences and six sequences representing major types of ADH. To determine the significance of the placement of the *C. elegans* ADHs, we subjected the tree to bootstrap analysis. Two important points are illustrated. First, the cm01d5-encoded ADH and *Schizosaccharomyces pombe* ADH are more closely related to each other than either is to *Z. mobilis* ADH1 ($P \leq 0.01$). Second, the clustering of the cm14h3-encoded ADH within the animal ADHs is strongly supported ($P \leq 0.01$).

Discussion

We have characterized two cDNAs from *C. elegans* that encode ADH enzymes. cm14h3 encodes an enzyme most like dimeric class III ADHs and cm01d5 encodes an enzyme most like tetrameric ADHs found in fungi. These clones were obtained from a "sorted" cDNA library which reduces the chances of recovering multiple members of a gene family (Waterston et al. 1992). Therefore, it is possible that additional ADH genes exist in *C. elegans.*

Sun and Plapp (1992) aligned 47 members of the Zn-containing ADH family, identified highly conserved residues, and reviewed current thought on the function of these residues in the intact protein. The sequence alignment presented here demonstrates that both *C. elegans* sequences are typical members of the long-chain Zn-containing family of enzymes. Conservation of sequence is seen at several important residues (Table 1).

Several gaps present in the alignment allow discrimination between tetrameric and dimeric forms of ADH. The tetrameric ADHs are, on average, approximately 30 amino acids shorter than the dimeric ADHs. The most

```
                    221                                                                                                      330
  ScerevisiaeADHI   AISGAAGGLG SLAVQYAKAM .GYRVLGIDG GEGKEELFRS IGGEVFID.. ...FTKEKDI VGAVLKAT.N GGAHGVINVS VSEAAIEAST RYVRAN.GTT VLVGMPAGAK
 ScerevisiaeADHII   AISGAAGGLG SLAVQYAKAM .GYRVLGIDG GPGKEELFTS LGGEVFID.. ...FTKEKDI VSAVVKAT.N GGAHGIINVS VSEAAIEAST RYCRAN.GTV VLVGLPAGAK
      KlactisADHI   AISGAAGGLG SLAVQYAKAM .GYRVLGIDA GEEKAKLFKD LGGEYFID.. ...FTKSKNI PEEVIEAT.K GGAHGVINVS VSEFAIEQST NYVRSN.GTV VLVGLPRDAK
     KlactisADHII   AISGACGGLG SLAIQYAKAM .GYRVLGIDT GAEKAKLFKE LGGEYFVD.. ...YAVSKDL IKEIVDAT.N GGAHGVINVS VSEFAIEQST NYVRSN.GTV VLVGLPRDAK
       SpombeADH    CIPGAGGGLG HLAVQYAKAM .AMRVVAIDT GDDKAELVKS FGAEVFLD.. ...FKKEADM IEAVKACT.N GGAHGTLVLS TSPKSYEQAA GFARPG.STM VTVSMPAGAK
   Celeganscm01d5   VLTGAGGGLG SLAIQYACAM .GMRVVAMDH GR.KEAHCKG LGAEWFVD.. ...AFETPDI VSHITKLT.E GGPHGVINFG VARKPMEQAV EYVRKR.GTV VFVGLPKDSK
     ZmobilisADH1   AIYGL.GGLG NLALQYAKNV FNAKVIAIDV NDEQLAFAKE LGADMVIN.. .....PKNED AAKIIQEK.V GGAHATVVTA VAKSAFNSAV EAIRAG.GRV VAVGLPPE.K
 SsolfataricusADH   LVVGAGGGLG TMAVQIAKAV SGATIIGVDV REEAVEAAKR AGADYVIN.. ...ASMQDPL AEIRRITE.S KGVDAVIDLN NSEKTLSVYP KALAKQ.GKY VMVGL.FGAD
 EhistolyticaADH1   CVIGI.GPVG LMSVAGANHL GAGRIFAVGS RKHCCDIALE YGATDIINYK N......GDI VEQILKATDG KGVDKVVIAG GDVHTFAQAV KMIKPGSDIG N.VNYLGEGD
      TbrockiiADH   AVLGI.GPVG LMAVAGAKLR GAGRIIAVGS RPVCVDAAKY YGATDIVNYK D......GPI ESQIMNLTEG KGVDAAIIAG GNADIMATAV KIVKPGGTIA N.VNYFGEGE
        ZmaysADH1   AVFGL.GAVG LAAAEGARIA GASRIIGVDL NPSRFEEARK FGCTEFVNPK DH....NKPV QEVLAEMT.N GGVDRSVECT GNINAMIQAF ECVHDGWGVA VLVGVPHKDA
        ZmaysADH2   AIFGL.GAVG LAAMEGARLA GASRIIGVDI NPAKYEQAKK FGCTEFVNPK DH....DKPV QEVLIELT.N GGVDRSVECT GNVNAMISAF ECVHDGWGVA VLVGVPHKDD
   StuberosumADH1   AIFGL.GAVG LAAAEGARIA GASRIIGVDL NASRFEQAKK FGVTEFVNPK DY....SKPV QEVIAEMT.D GGVDRSVECT GHIDAMISAF ECVHDGWGVA VLVGVPHKEA
     HsapiensADHβ   AVFGL.GGVG LSAVMGCKAA GAARIIAVDI NKDKFAKAKE LGATECINPQ DY....KKPI QEVLKEMT.D GGVDFSFEVI GRLDTMMASL LCCHEACGTS VIVGVPPASQ
    MmusculusADHα   AVFGL.GGVG LSVIIGCKAA GAARIIAVDI NKDKFAKAKE LGATECINPQ DY....SKPI QEVLQEMT.D GGVDFSFEVI GRLDTMTSAL LSCHAACGVS VVVGVPPNAQ
     HsapiensADH6   AVFGL.GGVG LSVVMGCKAA GAARIIGVDV NKEKFKKAQE LGATECLNPQ DL....KKPI QEVLFDMT.D AGIDFCFEAI GNLDVLAAAL ASCNESYGVC VVVGVLPASV
     HsapiensADHπ   AVFGL.GGVG LSAVMGCKAA GASRIIGIDI NSEKFVKAKA LGATDCLNPR DL....HKPI QEVIIELT.K GGVDFALDCA GGSETMKAAL DCTTAGWGSC TFIGVAAGSK
      Cmaltosafdh1  AVFGG.GIVG LSVIQGCAER GAAQIILVDI SDKKEEWGQK LGATAFVNP. T.KLPEGTTI VDKLIEMT.D GGCDFTFDCT GNVGVMRNAL EACHKGWGTS VIIGVAAAGK
   Scerevisiaesfa   AVFGC.GTVG LSVIQGAKLR GASKIIAIDI NNKKKQYCSQ FGATDFVNPK E.DLAKDQTI VEKLIEMT.D GGLDFTFDCT GNTKIMRDAL EACHKGWGQS IIIGVAAAGE
     HsapiensADHχ   AVFGL.GGVG LAVIMGCKVA GASRIIGVDI NKDKFARAKE FGATECINPQ DF....SKPI QEVLIEMT.D GGVDYSFECI GNVKVMRAAL EACHKGWGVS VVVGVAASGE
    MmusculusADHχ   AVFGL.GGVG LAVIMGCKVA GASRIIGIDI NKDKFAKAKE FGASECISPQ DF....SKSI QEVLVEMT.D GGVDYSFECI GNVKVMRSAL EAAHKGWGVS VVVGVAASGE
   Celeganscm14h3   AVWGL.GAVG LAVIMGAKAA GAKKIVGIDL IESKFESAKF FGATECINPK SVELPEGKSF QAWLVEQF.D GGFDYTFECI GNVHTMRQAL EAAHKGWGVS CIIGVAGAGQ
   GcallariusADH1   AVFGL.GAVG LAAVMGCHSA GAKRIIAVDL NPDKFEKAKV FGATDFVNPN D....HSEPI SQVLSKMT.N GGVDFSLECV GNVGVMRNAL ESCLKGWGVS VLVGWTDL.H
         Ecolitdh   LVSGA.GPIG IMAAAVAKHV GARNVVITDV NEYRLELARK MGITRAVNVA ......KENL NDVMAELGMT EGFDVGLEMS GAPPAFRTML DTMNHG.GRI AMLGIPPSDM
```

```
                    331                                                                                   422
  ScerevisiaeADHI   CCSDVFNQVV KSISIVGSYV GNRADTREAL DFFARGLIKS PI.KV..... ...VGLSTLP EIYEKMEKGQ VVG.RYVVDT SK........ ..
 ScerevisiaeADHII   CSSDVFNHVV KSISIVGSYV GNRADTREAL DFFARGLVKS PI.KV..... ...VGLSSLP EIYEKMEKGQ IAG.RYVVDT SK........ ..
      KlactisADHI   CKSDVFNQVV KSISIVGSYV GNRADTREAI DFFSRGLVKA PI.HV..... ...VGLSELP SIYEKMEKGA IVG.RYVVDT SK........ ..
     KlactisADHII   CKSDVFTQVV KSVSIVGSYV GNRADTREAL DFFARGLVHA PI.KI..... ...VGLSELA DVYDKMVKGE IVG.RYVVDT SK........ ..
       SpombeADH    LGADIFWLTV KMLKICGSHV GNRIDSIEAL EYVSRGLVKP YY.KV..... ...QPFSTLP DVYRLMHENK IAG.RIVLDL SK........ ..
   Celeganscm01d5   VTFDTTPFIF NAITIKGSIV GSRLDVDEAI EFVTRGIVKV PL.EL..... ...VKLEDVP AVYQRMLDGK INS.RAGVDF SL........ ..
     ZmobilisADH1   MDLSIPRLVL DGIEVLGSLV GTREDLKEAF QFAAEGKVKP KV.TK..... ...RKVEEIN QIFDEMEHGK FTG.RMVVDF THH....... ..
 SsolfataricusADH   LHYHAPLITL SEIQFVGSLV GNQSDFLGIM RLAEAGKVKP MITKT..... ...MKLEEAN EAIDNLENFK AIG.RQVLIP .......... ..
 EhistolyticaADH1   NIDIPRSEWG VGMGHK.HIH GGLTPGGRVR MEKLASLIST GKLDTSKLIT HRFEGLEKVE DALMLMKNKP ADLIKPVVRI HYDDEDTLH. ..
      TbrockiiADH   VLPVPRLEWG CGMAHK.TIK GGLCPGGRLR MERLIDLVFY KRVDPSKLVT HVFRGFDNIE KAFMLMKDKP KDLIKPVVIL A......... ..
        ZmaysADH1   EFKTHPMNFL NERTLKGTFF GNYKPRTDLP NV..VELYMK KELEVEKFIT HSV.PFAEIN KAFDLMAK.. GEGIRCIIRM EN........ ..
        ZmaysADH2   QFKTHPMNFL SEKTLKGTFF GNYKPRTDLP NV..VEMYMK KELELEKFIT HSV.PFSEIN TAFDLMLK.. GESLRCIMRM ED........ ..
   StuberosumADH1   VFKTHPMNLL NERTLKGTFF GNYKPRSDIP SV..VEKYMN KELELEKFIT HTL.PFAEIN KAFDLMLK.. GEGLRCIITM ED........ ..
     HsapiensADHβ   NLSINPMLLL TGRTWKGAVY GGFKSKEGIP KL..VADFMA KKFSLDALIT HVL.PFEKIN EGFDLLHS.. GKSIRTVLTF .......... ..
    MmusculusADHα   NLSMNPMLLL LGRTWKGAIF GGFKSKDSVP KL..VADFMA KKFPLDPLIT HVL.PFEKIN EAFDLLRS.. GKSIRTVLTF .......... ..
     HsapiensADH6   QLKISGQLFF SGRSLKGSVF GGWKSRQHIP KL..VADYMA EKLNLDPLIT HTL.NLDKIN EAVELMKT.. GKW....... .......... ..
     HsapiensADHπ   GLTIFPEELI IGRTINGTFF GGWKSVDSIP KL..VTDYKN KKFNLDALVT HTL.PFDKIS EAFDLMNQ.. GKSVRTILIF GRCQEQFRIL SD
      Cmaltosafdh1  EISTRPFQLV TGRTWKGAAF GGVKGRSQLP GI..VNNYLD GKLKVEEFIT HRE.PLAAIN KAFEEMHA.. GDCIRAVVDL S......... ..
   Scerevisiaesfa   EISTRPFQLV TGRVWKGSAF GGIKGRSEMG GL..IKDYQK GALKVEEFIT HRR.PFKEIN QAFEDLHN.. GDCLRTVLKS DEIK...... ..
     HsapiensADHχ   EIATRPFQLV TGRTWKGTAF GGWKSVESVP KL..VSEYMS KKIKVDEFVT HNL.SFDEIN KAFELMHS.. GKSIRTVVKI .......... ..
    MmusculusADHχ   EISTRPFQLV TGRTWKGTAF GGWKSVESVP KL..VSEYMS KKIKVDEFVT GNL.SFDQIN QAFDLMHS.. GDSIRTVLKM .......... ..
   Celeganscm14h3   EIATRPFQLV TGRTWKGTAF GGWKSVESVP RL..VDDYMN KKLLIDEFIT HRW.NIDDIN TAFDVLHK.. GESLRSVLAF EKI....... ..
   GcallariusADH1   DVATRPIQLI AGRTWKGSMF GGFKGKDGVP KM..VKAYLD KKVKLDEFIT HRM.PLESVN DAIDLMKH.. GKCIRTVLSL E......... ..
         Ecolitdh   SID......W TKVIFKGLFI KGIYGREMFE TWYKMAALIQ SGLDLSPIIT HRF.SIDDFQ KGFDAMRS.. GQSGKVILSW D......... ..
```

Fig. 2. Continued.

obvious feature common to all tetrameric ADHs, with the exception of the $NADP^+$-dependent ADHs from *E. histolytica* and *T. brockii,* is a gap present from amino acids 144–169 of the dimeric amino acid sequence. cm01d5 contains this gap (see Fig. 2), supporting its placement in the tetrameric ADHs.

cm14h3 most closely resembles dimeric ADHs exemplified by human class III ADH which has been shown to be a glutathione-dependent formaldehyde dehydrogenase. Two fungal ADHs included in our analysis, *Candida maltosa* FDH1 and *S. cerevisiae* SFA, are formaldehyde dehydrogenases. Notably, these two ADHs resemble the dimeric animal and plant ADHs and not tetrameric fungal ADHs, with respect to gaps (Fig. 2). Consistent with this, these ADHs fall within the plant and animal cluster in the phylogenetic tree (Fig. 3).

The phylogenetic tree (Fig. 3) constructed from the aligned sequences clarifies the relationships between different ADHs. The tree was rooted using the *E. coli* threonine dehydrogenase which is considered to be a distantly related member of the Zn-containing ADH family (Aronson et al. 1989). Three distinct clusters are evident: (1) A cluster containing the tetrameric ADHs of fungi, *Z. mobilis* and *Sulfolobus solfataricus.* All enzymes contain the residues required to bind two Zn ions per subunit and the coenzyme NAD^+ (residues indicated in Table 1). The presence of the *C. elegans* clone cm01d5 within this cluster suggests that it too is a tetrameric enzyme. (2) *E. histolytica* and *T. brockii* ADHs are distinct from the other sequences. These two enzymes bind $NADP^+$ rather than NAD^+ (presumably due to the glycine residue at position 249) and lack the four cysteine residues required for binding the noncatalytic Zn. These sequences share several unique gaps such as those present at positions 75–77, 129–144, and 155–163 of the dimeric ADH sequences. (3) This cluster contains *C. elegans* clone cm14h3 and the remaining plant, animal, and microbial members of the Zn-containing ADH family, including the glutathione-dependent formaldehyde dehydrogenases. They contain amino acid residues necessary to bind two Zn/subunit and NAD^+, and are all believed to be active as dimers.

The third cluster can be further broken down into a group containing plant ADHs, a group of mammalian class I and II ADHs, and a cluster of glutathione-dependent formaldehyde dehydrogenases (class III ADHs). A striking feature of this clustering relationship is that the human class III ADH is more similar to *C.*

Table 1. Amino acid residues highly conserved in the Zn-containing, long-chain ADHs are shown[a]

Position	Residue	cm01d5	cm14h3	Suggested role[b]
50	C	?	+	Ligand to catalytic Zn
53	T	?	+	H-bonds with hydroxyl group of alcohol
54	D	?	+	?
80	G	?	+	Part of catalytic domain
81	H	?	+	Ligand to catalytic Zn
85	G	?	+	Part of catalytic domain
91	G	?	+	Part of catalytic domain
94	V	?	+	Part of catalytic domain
100	G	?	+	Part of catalytic domain
101	D	?	+	?
112	C	?	+	Ligand for noncatalytic Zn
115	C	?	+	Ligand for noncatalytic Zn
118	C	?	+	Ligand for noncatalytic Zn
126	C	+	+	Ligand for noncatalytic Zn
130	R	–	+	Involved in glutathione binding[c]
196	C	+	+	Ligand to catalytic Zn
217	G	+	+	Part of coenzyme binding domain
224	G	+	+	Part of coenzyme binding domain
227	G	+	+	Part of coenzyme binding domain
230	G	+	+	Part of coenzyme binding domain
249	D	+	+	Part of coenzyme binding domain
262	G	+	+	Part of coenzyme binding domain

[a] Column 1 contains the position of each residue in the alignment in Fig. 2. Column 2 contains the conserved amino acid residue. The presence (+) of each residue in the deduced *C. elegans* proteins is shown in columns 3 and 4. Note: cm01d5 cDNA is truncated at the 5′ end; hence the uncertainty for positions preceding 126 in this protein. The amino acid residues are given as single-letter abbreviations

[b] Reviewed in Sun and Plapp (1992)

[c] Engeland et al. (1993)

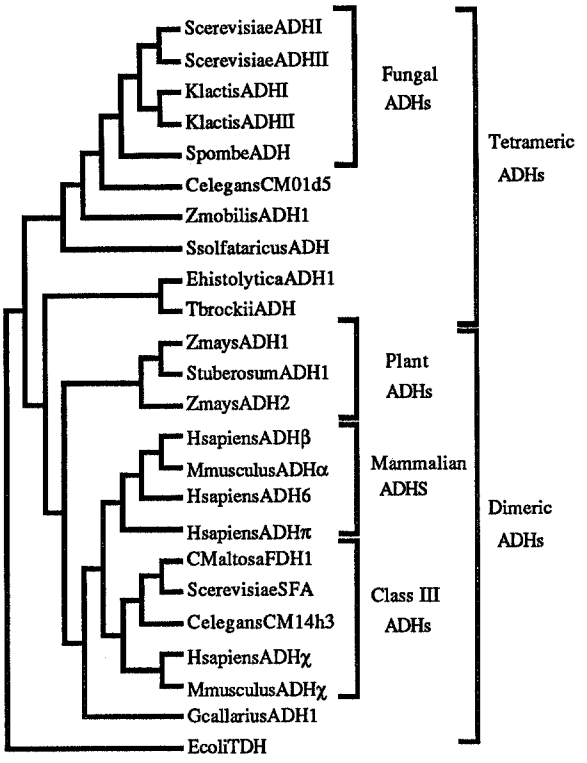

Fig. 3. Phylogenetic tree constructed from residues 123–422 of the aligned sequences in Fig. 2 using protein parsimony (Swofford 1993). The tree was rooted using *E. coli* threonine dehydrogenase.

elegans cm14h3 and yeast formaldehyde dehydrogenases than to other classes of human ADH. All class III proteins contain an arginine residue at position 130. This arginine is thought to be part of the binding site for activating fatty acids and of s-hydroxymethylglutathione in glutathione-dependent formaldehyde dehydrogenase activity (Engeland et al. 1993). The sequence conservation among the phylogenetically diverse glutathione-dependent ADHs and the discovery of a related enzyme in *E. coli* (Gutheil et al. 1992) indicate that the presence of this class of ADH predates the divergence of prokaryotes and eukaryotes and lends support to the hypothesis that the class I enzymes arose from the duplication of a functional class III gene (Danielsson and Jornvall 1992).

The tree shown in Fig. 4 illustrates the strength of statistical support for the clustering relationships. A high bootstrap value separates the cluster containing tetrameric fungal ADHs and *C. elegans* cm01d5-encoded ADH from the dimeric plant and animal ADH cluster. This underscores the similarity between this nematode ADH and the yeast ADHs. The placement of the *C. elegans* cm14h3-encoded ADH within the cluster of dimeric animal class III ADHs and separate from the plant ADHs is also supported by a high bootstrap value.

The discovery of a *C. elegans* ADH with a high degree of similarity to tetrameric ADHs from fungi is somewhat surprising since no precedent exists for fungal-like ADH sequences among metazoans. The identification of an ADH of this type in the invertebrate metazoan *C. elegans* might predict the discovery of similar (fungal-like) ADHs in other animals or plants. Indeed, the lack of information regarding ADHs in invertebrates leaves this possibility open. On the other hand, it seems unlikely that this form exists in humans. Intensive study of ADH in this species would likely have led to its discovery by now. How can we explain the apparent presence of fungal-like tetrameric ADH in some animals but

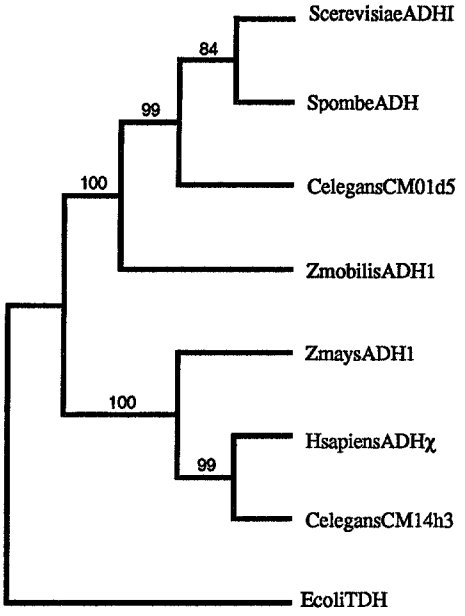

Fig. 4. Phylogenetic tree constructed from eight representative ADH sequences and subjected to bootstrap analysis. The sequences chosen for inclusion in the bootstrap analysis represent all major clades identified in Fig. 3 and are sufficient to define the position of the *C. elegans* ADHs. *Numbers* indicate the proportion of 100 bootstrap samples in which a particular clade was found. The tree was rooted with *E. coli* threonine dehydrogenase.

not others? This form of ADH may have been lost once or multiple times in lineages giving rise to modern plants and animals. Alternatively, the similarity may be a result of convergent evolution.

Crude protein extracts from *C. elegans* have been shown to contain ADH activity (Williamson et al. 1991). The *C. elegans* extract is active on ethanol and displays a preference for longer, primary alcohols. The structure of two putative *C. elegans* ADHs we have characterized predicts that the cm01d5-encoded ADH may represent the ethanol-active form of ADH detected in these extracts. This is based on its similarity to ethanol-utilizing yeast ADHs. The primary structure of the ADH encoded by cm14h3 predicts a glutathione-dependent formaldehyde dehydrogenase activity for this enzyme. This form of ADH would not be expected to be active on ethanol.

Our analysis of ADH from *C. elegans* has uncovered a predicted protein most similar to tetrameric fungal ADHs. This represents the first example of a tetrameric ADH in a metazoan. Characterization of additional ADHs from *C. elegans* and other invertebrates will fill one of the most conspicuous gaps in our knowledge about ADH from different taxa. This information will contribute to our understanding of the evolution of the ADH enzyme family.

Acknowledgments. We would like to thank Chris Martin and Bob Waterston for providing cDNA clones and partial sequences, and Nicole Perna and Clyde Denis for critical discussion. This is scientific contribution number 1843 from the New Hampshire Agricultural Experiment Station.

References

Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Ammendola S, Raia C, Caruso C, Camardella L, D'Auria S, De Rosa M, Rossi M (1992) Thermostable NAD(+)-dependent alcohol dehydrogenase from *Sulfolobus solfataricus:* gene and protein sequence determination and relationship to other alcohol dehydrogenases. Biochemistry 31:12514–12523

Aronson B, Somerville R, Epperly B, Dekker E (1989) The primary structure of *Escherichia coli* L-threonine dehydrogenase. J Biol Chem 264:5226–5232

Cederland E, Peralba J, Pares X, Jornvall H (1991) Amphibian alcohol dehydrogenase: the major frog liver enzyme. Relationships to other forms and assessment of an early gene duplication separating vertebrate class I and class III alcohol dehydrogenases. Biochemistry 30:2811–2816

Collins J, Forbes E, Anderson P (1989) The Tc3 family of transposable genetic elements in *Caenorhabditis elegans.* Genetics 121:47–55

Conway T, Ingram L (1989) Similarity of *Escherichia coli* propanediol oxidoreductase (fucO product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis* and *Saccharomyces cerevisiae.* J Bacteriol 171:3754–3759

Corbin V, Maniatis T (1990) Identification of cis-regulatory elements required for larval expression of the *Drosophila melanogaster* alcohol dehydrogenase. Genetics 124:637–646

Danielsson O, Atrian S, Luque T, Hjelmqvist L, Gonzalez-Duarte R (1994) Fundamental molecular differences between alcohol dehydrogenase classes. Proc Natl Acad Sci USA 91:4980–4984

Danielsson O, Jornvall H (1992) "Enzymogenesis": classical liver alcohol dehydrogenase origin from the glutathione-dependent formaldehyde dehydrogenase line. Proc Natl Acad Sci USA 89: 9247–9251

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12(1): 387–395

Eklund H, Nordstrom B, Zeppezauer E, Soderberg G, Tapia O, Branden C, Akeson A (1976) Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. J Mol Biol 102: 27–59

Engeland K, Hoog J, Holmquist B, Estonius M, Jornvall H, Vallee B (1993) Mutation of Arg-115 of human class III alcohol dehydrogenase activity and fatty acid activation. Proc Natl Acad Sci USA 90:2491–2494

Feinberg A, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Feng D-F, Doolittle R (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J Mol Evol 25:351–360

Fernandez M, Jornvall H, Moreno A, Kaiser R, Pares X (1993) Cephalopod alcohol dehydrogenase: purification and enzymatic characterization. FEBS Lett 328:235–238

Gutheil G, Holmquist B, Vallee B (1992) Purification, characterization, and partial sequence of the glutathione-dependent formaldehyde dehydrogenase from Escherichia coli: a class III alcohol dehydrogenase. Biochemistry 31(2):475–481

Jornvall H (1977) The primary structure of yeast alcohol dehydrogenase. Eur J Biochem 72:425–442

Jornvall H, Persson M, Jeffery J (1981) Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. Proc Natl Acad Sci USA 78:4226–4230

Jornvall H, Persson M, Jeffery J (1987) Characteristics of alcohol/polyol dehydrogenases. Eur J Biochem 167:195–201

Matsuo Y, Yokoyama S (1990) Cloning and sequencing of a processed

pseudogene derived from a human class III adh gene. Am J Hum Genet 46:85–91

Paquin C, Dorsey M, Crabble S, Sprinker K, Sondej M, Williamson V (1992) A spontaneous chromosomal amplification of the ADH2 gene in *Saccharomyces cerevisiae.* Genetics 130:263–271

Palazzolo M, Hamilton B, Ding D, Martin C, Mead D, Mierendorf R, Raghavan K, Lipshitz H (1990) Phage lambda cDNA cloning vectors for subtractive hybridization, fusion protein synthesis and Cre-*loxP* automatic plasmid subcloning. Gene 88:25–36

Russell D, Smith M, Williamson V, Young E (1983) Nucleotide sequence of the yeast alcohol dehydrogenase II gene. J Biol Chem 258:2674–2682

Sasnauskas K, Jomantiene R, Januska A, Lebediene E, Lebedys J, Janulaitis A (1992) Cloning and analysis of a *Candida maltosa* gene which confers resistance to formaldehyde in *Saccharomyces cerevisiae.* Gene 122:207–211

Shain D, Salvadore C, Denis C (1992) Evolution of the alcohol dehydrogenase (ADH) genes in yeast: characterization of a fourth ADH in *Kluyveromyces lactis.* Mol Gen Genet 232:479–488

Sun H, Plapp B (1992) Progressive sequence alignment and molecular evolution of the Zn-containing Adh family. J Mol Evol 34:522–535

Swofford D (1993) PAUP: phylogenetic analysis using parsimony, version 3.1 Computer program distributed by the Illinois Natural History Survey, Champaign, IL

Waterston R, Martin C, Craxton M, Hunynh 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, Sulston J (1992) A survey of expressed genes in *Caenorhabditis elegans.* Nature Genet 1:114–123

Wehner E, Rao E, Brendel M (1993) Molecular structure and genetic regulation of SFA, a gene responsible for resistance to formaldehyde in *S. cerevisiae* and characterization of its encoded protein. Mol Gen Genet 237:351–358

Williamson V, Long M, Theodoris G (1991) Isolation of *C. elegans* mutants lacking Adh activity. Biochem Genet 29:313–323

Williamson V, Paquin C (1987) Homology of Saccharomyces cerevisiae ADH4 to iron-activated alcohol dehydrogenase from *Zymomonas mobilis.* Mol Gen Genet 209:374–381

Wood W (1988) The nematode *Caenorhabditis elegans.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Yasunami M, Chen C, Yoshida A (1990) A human alcohol dehydrogenase gene (ADH6) encoding an additional class of isozyme. Proc Natl Acad Sci USA 88:7610–7614