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

Jeremy D. Glasner, Thomas D. Kocher, John J. Collins

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 Papp 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 Papp 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 ADH and ADHI (Jomvall 1977; Russell et al. 1983), but not in plants or animals (Sun and Flapp 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 Q-crystallin from guinea pig lens (Sun and Flapp 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 (Damelison 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. 200d and 2m14h were isolated from a sorted cDNA library constructed in bacteriophage lambda vector SHL2 by Chris Matta (Paliarekko et al. 1990). They were generously provided to us by R. Wawrzyn. The cDNA inserts were automatically subcloned into the pATIL2 phasmid vector by growing recombinant phage in an E. coli host expressing Pl recombinase (Cevi, promoting recombination between loxP sites flanking plasmid sequences in the SHL2 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 pATIL2 vector. Amplification products were radio-labeled 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 terminations according to the method of the manufacturer (Applied Biosystems). Extension products were purified in Centricon spin columns (Pierce 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'-ATAGAAATGATCGATCAAGCTGAG-3'
T7: 5'-ACATAGGGAGCTAAGCTTGG-3'
cmo1d5: 5'-CTTGTCTATGGAATGGA-3'
cmo1h3: 5'-GATTATGCGCCAGTGTTCA-3'
cmo1d5: 5'-CTTGTCTATGGAATGGA-3'
cmo1h3: 5'-GATTATGCGCCAGTGTTCA-3'
cmo1d5: 5'-CTTGTCTATGGAATGGA-3'
cmo1h3: 5'-GATTATGCGCCAGTGTTCA-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 (Swoford 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 cmo1d5 and cmo1h3, that encode ADH isoforms (Waterston et al. 1992). BLAST analysis (Altschul et al. 1990) revealed that cmo1d5 is most similar to fungal ADHs and cmo1h3 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 herebefore 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
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 EcoRI. 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 EcoRI site within the cDNA insert. Hence, we expect the genomic region covered by cm14h3 to consist of at least two EcoRI fragments. The blot shown in Fig. 1b contains N2 DNA cut with HindIII. 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 HindIII 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 adenose residues. The presence of a putative polyadenylation signal (5' 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 Mas maculatus ADH). The cm01d5-encoded ADH is most similar to tetrameric ADHs from yeast (score of 301, associated probability of 8.0e-62 for Kluyveromyces lactis ADH). 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-
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 cm01b3. The presence of two yeast ADHs within this cluster will be addressed in the Discussion. A third cluster contains ADH from Entamoeba histolytica and Thermoaerobium 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 ADH (P ≤ 0.01). Second, the clustering of the cm14b3-encoded ADH within the animal ADHs is strongly supported (P ≤ 0.01).

**Discussion**

We have characterized two cDNAs from C. elegans that encode ADH enzymes. cm14b3 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
obvious feature common to all tetrameric ADHs, with the exception of the NADP⁺-dependent ADHs from *E. histolytica* and *T. brochu*, is a gap present from amino acids 144-169 of the dimeric amino acid sequence. cm14d contains this gap (see Fig. 2), supporting its placement in the tetrameric ADHs.

cm14d 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* FDIH and *S. cerevisiae* SFA, are formaldehyde dehydrogenases. Notably, these two ADHs resemble the dimeric animal and plant ADHs and not anaerobic fungal ADHs, with respect to gaps (Fig. 2). Consistent with this, the 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 *Saccharomyces* SFA. 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 cm14d within this cluster suggests that it too is a tetrameric enzyme. (2) *E. histolytica* and *T. brochu* ADHs are distinct from the other sequences. These two enzymes bind NAD⁺ 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 cm14d 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 ions per 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.

<table>
<thead>
<tr>
<th>Position</th>
<th>Residue</th>
<th>cm01d5</th>
<th>cm14h3</th>
<th>Suggested rolea</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>C</td>
<td>?</td>
<td>+</td>
<td>Ligand to catalytic Zn</td>
</tr>
<tr>
<td>53</td>
<td>T</td>
<td>?</td>
<td>+</td>
<td>H-bonds with hydroxyl group of alcohol</td>
</tr>
<tr>
<td>54</td>
<td>D</td>
<td>?</td>
<td>+</td>
<td>Part of catalytic domain</td>
</tr>
<tr>
<td>80</td>
<td>G</td>
<td>?</td>
<td>+</td>
<td>Part of catalytic domain</td>
</tr>
<tr>
<td>81</td>
<td>H</td>
<td>?</td>
<td>+</td>
<td>Ligand to catalytic Zn</td>
</tr>
<tr>
<td>85</td>
<td>G</td>
<td>?</td>
<td>+</td>
<td>Part of catalytic domain</td>
</tr>
<tr>
<td>91</td>
<td>G</td>
<td>?</td>
<td>+</td>
<td>Part of catalytic domain</td>
</tr>
<tr>
<td>94</td>
<td>V</td>
<td>?</td>
<td>+</td>
<td>Part of catalytic domain</td>
</tr>
<tr>
<td>100</td>
<td>G</td>
<td>?</td>
<td>+</td>
<td>Part of catalytic domain</td>
</tr>
<tr>
<td>101</td>
<td>D</td>
<td>?</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>112</td>
<td>C</td>
<td>?</td>
<td>+</td>
<td>Ligand for noncatalytic Zn</td>
</tr>
<tr>
<td>115</td>
<td>C</td>
<td>?</td>
<td>+</td>
<td>Ligand for noncatalytic Zn</td>
</tr>
<tr>
<td>118</td>
<td>C</td>
<td>?</td>
<td>+</td>
<td>Ligand for noncatalytic Zn</td>
</tr>
<tr>
<td>125</td>
<td>C</td>
<td>+</td>
<td>+</td>
<td>Ligand for noncatalytic Zn</td>
</tr>
<tr>
<td>130</td>
<td>R</td>
<td>-</td>
<td>+</td>
<td>Involved in glutathione binding</td>
</tr>
<tr>
<td>196</td>
<td>C</td>
<td>+</td>
<td>+</td>
<td>Ligand to catalytic Zn</td>
</tr>
<tr>
<td>217</td>
<td>G</td>
<td>+</td>
<td>+</td>
<td>Part of coenzyme binding domain</td>
</tr>
<tr>
<td>224</td>
<td>G</td>
<td>+</td>
<td>+</td>
<td>Part of coenzyme binding domain</td>
</tr>
<tr>
<td>227</td>
<td>G</td>
<td>+</td>
<td>+</td>
<td>Part of coenzyme binding domain</td>
</tr>
<tr>
<td>230</td>
<td>G</td>
<td>+</td>
<td>+</td>
<td>Part of coenzyme binding domain</td>
</tr>
<tr>
<td>249</td>
<td>D</td>
<td>+</td>
<td>+</td>
<td>Part of coenzyme binding domain</td>
</tr>
<tr>
<td>262</td>
<td>G</td>
<td>+</td>
<td>+</td>
<td>Part of coenzyme binding domain</td>
</tr>
</tbody>
</table>

*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 120 in this protein. The amino acid residues are given as single-letter abbreviations.

*b Reviewed in Sun and Flipp (1992)

*C. 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 α-hydroxymethylglyoxaldehyde in glutathione-dependent formaldehyde dehydrogenase activity (Engel et al. 1993). The sequence conserva-

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.

ation among the phylogenetically diverse glutathione-dependent ADHs and the discovery of a related enzyme in E. coli (Guthel 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 (Daniellsson 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...
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 cm14d3 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 Penna and Clyde Dennis for critical discussion. This is scientific contribution number 1843 from the New Hampshire Agricultural Experiment Station.

References


Conway T, Ingram L (1990) Similarity of *Escherichia coli* propionated dehydrogenase (isoCO product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis* and *Saccharomyces cerevisiae*. J Bacteriol 171:3754-3759


Jormalvi H, Pearson M, Jeffrey 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


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