doi:10.1016/j.jmb.2007.10.041 J. Mol. Biol. (2008) 375, 1–11

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

COMMUNICATION

The *Caenorhabditis elegans* Protein CTBP-1 Defines a New Group of THAP Domain-Containing CtBP Corepressors

Hannah R. Nicholas*, Jason A. Lowry, Tina Wu and Merlin Crossley

School of Molecular and Microbial Biosciences, University of Sydney, Sydney, New South Wales 2006, Australia

Received 2 August 2007; received in revised form 9 October 2007; accepted 16 October 2007
Available online 22 October 2007

The C-terminal binding proteins (CtBPs) play roles in diverse cellular processes including transcriptional regulation, Golgi membrane fission, and synaptic ribbon formation. In the context of transcriptional regulation, they function as corepressors, interacting with promoter-bound transcription factors and recruiting a large protein complex that contains chromatin-modifying enzymes. We recently described the structure of a Thanatos-associated protein (THAP) domain that is found in a new member of the CtBP family, the *Caenorhabditis elegans* CTBP-1 protein. We have identified additional THAP domain-containing CtBPs in the nematode, echinoderm, and cephalochordate lineages. The distribution of these lineages within the animal kingdom suggests that the ancestral form of the animal CtBPs may have contained a THAP domain that was subsequently lost in the vertebrate and arthropod lineages. We also provide functional data indicating that CTBP-1 represses gene expression and homodimerizes and interacts with PXDLS-containing partner proteins, three key features of the previously characterized animal CtBPs. CTBP-1 is therefore the founding member of a new subgroup within the CtBP corepressor family, the THAP domain-containing CtBPs.

Keywords: CtBP; THAP domain; gene regulation; corepressor; *Caenorhabditis elegans*

Edited by J. Karn

Members of the C-terminal binding protein (CtBP) family are multifunctional proteins that have been ascribed roles in transcriptional corepression, Golgi maintenance, and synaptic ribbon formation.[1] As transcriptional corepressors, CtBPs associate with DNA-bound transcription factors, many of which contain a short sequence with the consensus PXDLS.[2] CtBPs function as part of a large corepressor complex that contains chromatin-modifying enzymes including histone deacetylases (HDACs), methyltransferases, and a demethylase.[3,4] CtBPs have been reported to interact with many transcription factors, suggesting that CtBP-mediated gene repression is a widely utilized mechanism.

The CtBPs share primary sequence and structural similarity with the D-isomer-specific 2-hydroxyacid dehydrogenases, which are composed of two functional domains: a substrate-binding domain and a nucleotide-binding domain.[5,6] In addition to an analogous N-terminal substrate-binding domain and central nucleotide-binding domain, the CtBPs possess an intrinsically unstructured C-terminal region.[6,7] The PXDLS-binding cleft through which CtBP contacts most of its partner proteins resides primarily in the substrate-binding domain. The central nucleotide-binding domain contains the nucleotide-binding consensus $G/AXGXXGX_{17}D$ (where X is any amino acid) and mediates interaction with NAD(H) as well as dimerization.[6] Also conserved within the nucleotide-binding domain of the CtBPs is a histidine residue that is required for the catalytic activity of the D-isomer-specific 2-hydroxyacid dehydrogenase enzymes.[5,8] Although CtBPs

**Corresponding author.* E-mail address: hannah@mmb.usyd.edu.au.

Abbreviations used: CtBP, C-terminal binding protein; THAP, Thanatos-associated protein; HDAC, histone deacetylase; AN, ANGUSTIFOLIA; EST, expressed sequence tag; DBD, DNA-binding domain.

have been shown to reduce pyruvate to lactic acid, this enzymatic activity has not been shown to be required for the repression function of the CtBPs.[9–13]

CtBP-related genes have been identified in numerous animal species. Invertebrates including *Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, and *Tribolium castaneum* contain a single CtBP gene.[14–16] In contrast, the studied vertebrates including human, mouse, rat, chicken, quail, zebrafish, and *Xenopus* contain two CtBPs (named CtBP1 and CtBP2). CtBP-like proteins have also been identified in a number of plant species. The best characterized of these is the *Arabidopsis thaliana* protein called ANGUSTIFOLIA (AN).[17–19] Despite its similarity to the animal CtBPs, AN appears to be unable to associate with the PXDLS motif.[20] Furthermore, the nucleotide-binding motif and dehydrogenase catalytic histidine are not conserved in AN.[17,18,20] Consistent with this lack of conservation of key functional elements, phylogenetic analysis indicates that the plant members of the CtBP family form a unique subfamily, suggesting that they may serve functions that are specific to plants.[18]

We recently described the structure of a protein domain, termed Thanatos-associated protein (THAP) domain, contained within a new member of the CtBP family, the CTBP-1 protein from the nematode *Caenorhabditis elegans*.[21] Here, we present evidence that CTBP-1 defines a new subgroup of THAP–CtBP corepressors that extends across animal lineages. Our results suggest that the THAP–CtBPs represent the ancestral form of CtBP, with THAP domains having subsequently been lost in vertebrate and various other lineages.

The *C. elegans ctbp-1* gene encodes a CtBP Homologue with an N-Terminal THAP domain

Early analysis of the genome of *C. elegans* predicted an open reading frame named F49E10.5 encoding a 612-amino-acid CtBP homologue (T34290), which has been cited in sequence comparisons of members of the CtBP family.[11,18] Subsequently, expressed sequence tag (EST) data prompted the merging of F49E10.5 with its neighbor, F49E10.6. We have named this revised predicted gene *ctbp-1*.[21] To verify this prediction, we sequenced products arising from PCR amplification of cDNA derived from RNA extracted from a mixed-stage population of *C. elegans* hermaphrodites.

Five clones were sequenced, two of which corresponded to the predicted cDNA (NM_076582). The remaining clones prompt two amendments to the predicted transcript structure (Fig. 1a). First, there is an alternative 3′ splice site within the 3rd intron, which, when used, results in the inclusion of an additional 9 nucleotides in exon 4. Second, within the predicted 13th exon, there is an intron of 48 nucleotides, retained in some transcripts, which necessitates the definition of a 14th exon. We have not investigated the functional significance of these alternative splice isoforms.

Conceptual translation of the *ctbp-1* cDNA (NM_076582) yields a protein of 727 amino acids in which a search of the National Center for Biotechnology Information Conserved Domain Database identified two domains: amino acids 1–89 comprise a THAP domain, which we have described previously,[21] while amino acids 178–505 show similarity to the D-isomer-specific 2-hydroxyacid dehydrogenases (Fig. 1b). Amino acids 90–177, linking the THAP domain to the dehydrogenase domain, and the C-terminal residues 506–727 do not contain any detectable structural motifs.

Comparison of the dehydrogenase domain of CTBP-1 (residues 178–505) with that from human CtBP1 (residues 27–353) reveals 55% identity and 74% similarity of the amino acid sequences. Furthermore, several features of potential functional importance are conserved in the *C. elegans* protein (Fig. 1c).

Fig. 1. The *C. elegans ctbp-1* gene encodes a protein that contains a THAP domain and is homologous to animal CtBPs. (a) Structure of the *ctbp-1* transcript. Exons are depicted as boxes and introns are depicted as lines. Amendments to the predicted gene structure are shown below the schematic. Splice donor and acceptor sites are shown in bold, intronic sequences are shown in lower case, and exonic sequences are shown in upper case. Underlined in sequence (i) is an alternative 3′ splice site in the 3rd intron. Sequence (ii) shows the 13th intron, which is retained in some transcripts. (b) Schematic representations of the human CtBP1 and *C. elegans* CTBP-1 proteins showing domains identified in the National Center for Biotechnology Information Conserved Domain Database. Dark gray shading indicates the D-isomer-specific 2-hydroxyacid dehydrogenase domain. Light gray shading indicates the THAP domain. (c) ClustalW alignment of representative vertebrate (hCtBP1, NP_001319; hCtBP2, AAH52276) and insect (dCtBP, NP_001014617) CtBPs with *C. elegans* CtBP (CTBP-1, NP_508983). Symbols below the alignment denote the degree of conservation observed in each column. Asterisks indicate identical residues, ":" indicates conservative substitutions, and "." indicates semiconservative substitutions. Features of potential functional importance are indicated above the alignment. "^" indicates residues lining the PXDLS-binding cleft, a line indicates the nucleotide-binding motif $GXGXXGX_{17}D$, and "‡" indicates the His residue that is critical for the catalytic activity of the D-isomer-specific 2-hydroxyacid dehydrogenases. Dark gray shading indicates the D-isomer-specific 2-hydroxyacid dehydrogenase domain. Light gray shading indicates the THAP domain. Methods: Reverse transcriptase DNA amplification reactions were performed to obtain the *ctbp-1* cDNA as follows. Total RNA was extracted from a mixed-stage population of *C. elegans* N2 (wild type) hermaphrodites using TRI Reagent® (Sigma-Aldrich) as described by the manufacturer. First-strand cDNA was then synthesized using the SuperScript™ II First-Strand Synthesis System for RT-PCR (Invitrogen) and oligo$(dT)_{20}$ primers. The *ctbp-1* cDNA was amplified by PCR (5BamcTC, CGGGATCCATGCCGACGACTTGTGGATTTCC; 3CeCtBPcDNA, AGTTGTCACAACCTCGAGAAC), and the product was then used as a template for a nested PCR (5BamcTC; 3EcoflC, CGGAATTCTTATGTGGCCAATGGTTGCTC). The product was cloned into the BamHI-EcoRI pBSSK (Stratagene) vector, and five clones were sequenced. One clone that corresponded to the predicted *ctbp-1* cDNA sequence (NM_076582) was used for all subsequent clonings and is referred to as CTBP-1.