J Hum Genet (2002) 47:681–683

SHORT COMMUNICATION

Jian Xu · Hai Wu · Chengzhi Zhang · Yueqiong Cao
Liu Wang · Li Zeng · Xin Ye · Qihan Wu · Jianfeng Dai
Yi Xie · Yumin Mao

Identification of a novel human *DDX40* gene, a new member of the DEAH-box protein family

Received: August 7, 2002 / Accepted: September 12, 2002

Abstract The DExH/D-box superfamily of RNA helicases seems to play key roles during RNA metabolism, such as pre-mRNA splicing, ribosome biogenesis, and others. We have cloned a new gene of the DEAH-box protein subgroup, designated *DDX40* (DEAD/H-box polypeptide 40 gene). *DDX40* contains 3656 nucleotides and codes for a putative 779-amino-acid protein. Sequence analysis of the cDNA product revealed that it contained a DEAH (Asp-Glu-Ala-His) sequence motif and other conserved motifs. The DDX40 protein shared 53% and 43% amino acid identity with human DDX8 and yeast Drh1, respectively, in the conserved region. Northern blot analysis showed that *DDX40* was expressed ubiquitously in the eight tissues examined, implying a general physiological function of the protein. We speculate that, like other members of the DExH/D-box superfamily, DDX40 may play roles in pre-mRNA splicing, ribosome biogenesis and other RNA processing functions.

Key words pre-mRNA splicing · DExH/D-box proteins · RNA helicase · HRH1

RNA helicases can catalyze the unwinding of double-stranded RNA and play important roles in RNA metabolism, including pre-mRNA splicing, ribosome biogenesis, organellar gene expression, and so on (Jankowsky and Jankowsky 2000; Ono et al. 1994; Tanner and Linder 2001). DExH/D-box proteins are putative ATP-dependent RNA helicases belonging to helicase superfamily 2 (SF2). It has been demonstrated that DExH/D proteins can function as processive and directional molecular motors for unwinding regular RNA duplexes. They may play key roles in the coupling of NTP hydrolysis to RNA conformational changes in macromolecular assemblies such as the spliceosome and ribosome (Jankowsky and Jankowsky 2000). Three members of the DEAH-box protein family, yeast Prp2, Prp16, and Prp22 have been found to be required for distinct steps in the splicing process (Company et al. 1991; Kim et al. 1998; Staley and Guthrie 1998). Previous study has shown that expression of DDX8 in a yeast mutant could partially rescue its phenotype, suggesting that DDX8 is a functional human homolog of the yeast Prp22 protein (Ono et al. 1994). Another member of the DEAH-box protein family, yeast Drh1, has shown different functions. Despite its strong homology to Prp22p and related splicing factors, Dhr1p was found to be required for the structural reorganization of rRNA during ribosome biogenesis rather than to function in pre-mRNA splicing (Colley et al. 2000).

During large-scale cDNA sequencing, a novel human cDNA was cloned from the human fetal cDNA library we constructed. The cDNA library was constructed in a modified pBluescript II SK (+) vector with human fetal brain mRNA purchased from Clontech (Palo Alto, CA, USA). Double-stranded cDNAs were synthesized using a SMART cDNA library construction kit (Clontech). The cDNA inserts were sequenced on an ABI PRISM 377 DNA sequencer (Perkin-Elmer, Shelton, CT, USA) using the BigDye Terminator Cycle Sequencing kit and BigDye Primer Cycle Sequencing kit (Perkin-Elmer). Subsequent editing and assembly of all the sequences from one clone was performed using Acembly (Sanger Centre) (Zhao et al. 2001). The entire cDNA is 3656 nucleotides long and contains an open reading frame (ORF) of 2340 bp from nucleotide 148 to 2487. The nucleotide sequence and the deduced amino acid sequence of this gene are shown in Fig. 1A. The 2340-bp ORF encodes a putative protein of 779 amino acids with a calculated molecular mass of 86 Kda. There is an in-frame stop codon TAA in the 5′-untranslated region (UTR),

J. Xu · H. Wu · C. Zhang · L. Wang · L. Zeng · X. Ye · Q. Wu · J. Dai · Y. Xie · Y. Mao (✉)
State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Sciences, Fudan University, Shanghai 200433, People's Republic of China
Tel. +86-21-65643573; Fax +86-21-65642502
e-mail address: ymmao@fudan.edu.cn

Y. Cao
United Gene Holdings, Ltd, Shanghai, P. R. China

The nucleotide sequence reported in this paper has been submitted to GenBank under accession number AF461690.

A

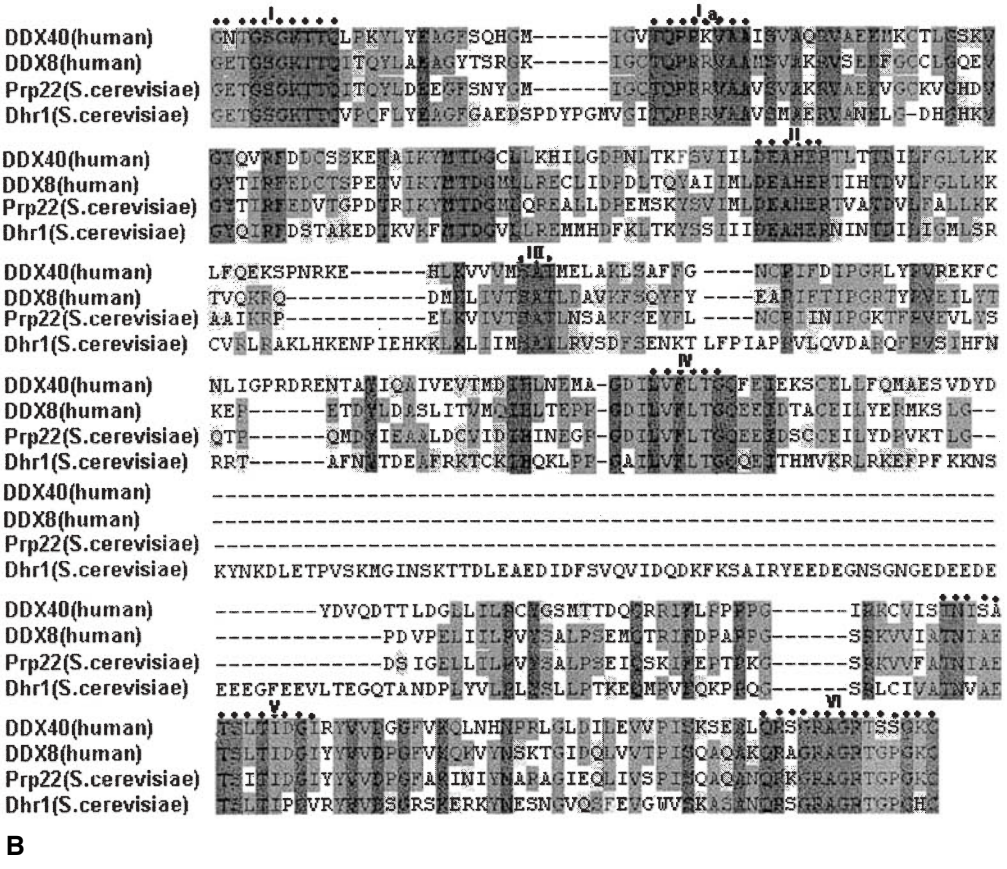
Fig. 1. A The nucleotide and deduced amino acid sequences of the *DDX40* gene and gene product. The cloned cDNA gave one extensive reading frame coding for a protein of 779 amino acids. The start and stop codons bordering the open reading frame are marked in *bold italic type*. The in-frame stop codon is *underlined*. *Dark shading* highlights the seven well-conserved motifs (motifs *I*, *Ia*, *II*, *III*, *IV*, *V*, and *VI*). The four polyadenylation signals are indicated by boxes. The sequence data reported here have been deposited in GenBank (accession No. AF 461690). **B** Alignment of conserved motifs of DDX40 and several DEAH-box proteins. The corresponding GenBank numbers for these proteins are AF461690, D50487, X58681, Q04217, from top to bottom. Identical residues are shown by *black background*; *gray shading* indicates conserved residues. Gaps (−) are introduced to achieve maximum homology. The seven well conserved motifs are indicated by *bold dots*

and there are four polyadenylation signals downstream of the stop codon TAA (nucleotides 2485–2487). By scanning the deduced amino acid sequence against the PROSITE database (http://www.expasy.ch/tools/scanprosire/), we have found that DDX40 has a well-conserved DEAH-box region (motifs I, Ia, II, III, IV, V, and VI), which is known to be characteristic of the DEAH-box protein family. Alignment analysis has shown that it has high amino acid identity with human DDX8, yeast Prp22, and Dhr1 (Fig. 1B), indicating that DDX40 may have similar functions in RNA processing. We term this gene *DDX40* (DEAD/H-box polypeptide 40 gene) in agree-ment with the HUGO Nomenclature Committee (http://www.gene.ucl.ac.uk/nomenclature/).

By searching against the Unigene and human genome databases, we found the *DDX40* gene represented by 164 expressed sequence tags (ESTs) and a genomic clone (NT_010740) from chromosome 17q23.3-q23.3. These ESTs are from various tissues, showing almost the same distribution as discussed below. Comparison of the cDNA sequence of *DDX40* with the genomic sequence revealed that the *DDX40* gene spanned more than 43 Kb of genomic DNA and consisted of 18 exons. We could thus determine the complete exon–intron structure of the *DDX40* gene; all sequences of the exon–intron junctions were consistent with the AG-GT rule.

A multiple-tissue Northern blot analysis (Clontech) was made to determine the size and tissue distribution of *DDX40* mRNA in humans. The result showed that *DDX40* was expressed ubiquitously in the eight tissues examined, with a major band corresponding to about 3.6 Kb (Fig. 2).

Fig. 1. Continued



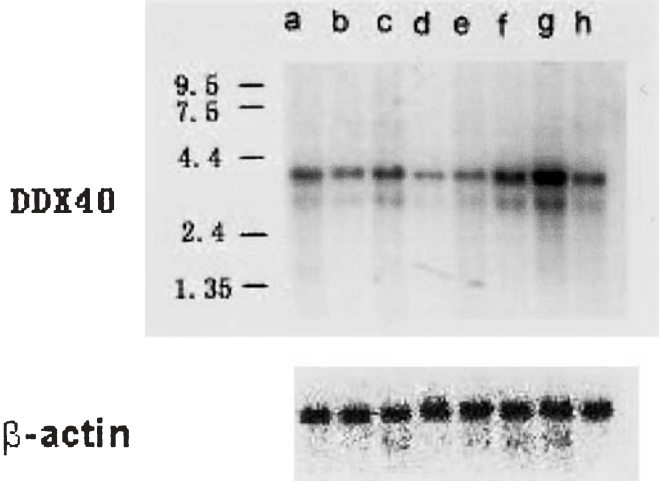


Fig. 2. Northern blot analysis of the *DDX40* gene in eight adult human tissues. *Upper*: Multiple tissue Northern blots containing 2μg of poly(A^+) RNA per lane were probed with the open reading frame of the *DDX40* cDNA. Numbers on the left indicate the molecular mass markers. *Lower*: The blot was stripped and reprobed with β-actin as an indicator of RNA loading. The tissues were *a*, heart; *b*, brain; *c*, placenta; *d*, lung; *e*, liver; *f*, skeletal muscle; *g*, kidney; *h*, pancreas

This result is consistent with the tissue distribution determined by searching ESTs. The wide expression of *DDX40* in many tissues and organs reveals that it may play fundamental roles in cells.

In this study, we report the molecular cloning and functional characterization of *DDX40*, a novel human DEAH-box protein. It has well-conserved motifs, indicating that it may also function in the process of splicing and ribosome biogenesis, like other members of the DEAH-box subgroup. Northern blot analysis showed that *DDX40* was expressed in a wide variety of tissues, implying a general physiological function of the protein. Further studies may be required to elucidate whether DDX40 has helicase activity and to identify the cofactors and substrate of DDX40.

Acknowledgments This research is a part of Project 30170345 supported by the National Natural Science Foundation of China.

References

Colley A, Beggs JD, Tollervey D, Lafontaine DLJ (2000) Dhr1p, a putative DEAH-Box RNA Helicase, is associated with the Box C+D snoRNP U3. Mol Cell Biol 20:7238–7246

Company M, Arenas J, Abelson J (1991) Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. Nature 349:487–493

Jankowsky E, Jankowsky A (2000) The DExH/D protein family database. Nucleic Acids Res 28:333–334

Kim JL, Morgenstern KA, Griffith JP, Dwyer MD, Thomson JA, Murcko MA, Lin C, Caron PR (1998) Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure 6:89–100

Ono Y, Ohno M, Shimura Y (1994) Identification of a putative RNA helicase (HRH1), a human homolog of yeast Prp22. Mol Cell Biol 14:7611–7620

Staley JP, Guthrie C (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell 92:315–326

Tanner NK, Linder P (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol Cell 8:251–262

Zhao W, Tang R, Huang Y, Wang W, Zhou ZX, Gu SH, Dai JL, Ying K, Xie Y, Mao YM (2001) Cloning and expression pattern of human NDRG3 gene. Biochim Biophys Acta 1519:134–138