

Molecular characterization of DDX26, a human DEAD-box RNA helicase, located on chromosome 7p12

A.A. Camargo¹, D.N. Nunes¹,
H.B. Samaia¹, L. Liu²,
V.P. Collins², A.J.G. Simpson¹
and E. Dias-Neto¹

¹Instituto Ludwig de Pesquisa sobre o Câncer, São Paulo, SP, Brasil
²Department of Histopathology, University of Cambridge,
Addenbrooke's Hospital, Cambridge, UK

Abstract

DEAD-box proteins comprise a family of ATP-dependent RNA helicases involved in several aspects of RNA metabolism. Here we report the characterization of the human DEAD-box RNA helicase DDX26. The gene is composed of 14 exons distributed over an extension of 8,123 bp of genomic sequence and encodes a transcript of 1.8 kb that is expressed in all tissues evaluated. The predicted amino acid sequence shows a high similarity to a yeast DEAD-box RNA helicase (Dbp9b) involved in ribosome biogenesis. The new helicase maps to 7p12, a region of frequent chromosome amplifications in glioblastomas involving the epidermal growth factor receptor (EGFR) gene. Nevertheless, co-amplification of DDX26 with EGFR was not detected in nine tumors analyzed.

Key words

- DDX26
- DEAD-box RNA helicase
- 7p12
- Ribosome biogenesis

Correspondence

A.A. Camargo
Instituto Ludwig de Pesquisa
sobre o Câncer
Rua Prof. Antônio Prudente, 109
4º andar
01509-010 São Paulo, SP
Brasil
Fax: +55-11-270-7001
E-mail:
anamaria@compbio.ludwig.org.br

D.N. Nunes and H.B. Samaia are recipients of FAPESP fellowships.

The sequence described was deposited in GenBank under accession No. AF247666.

Received February 1, 2001
Accepted June 28, 2001

Introduction

Helicases, including DNA and RNA helicases, are grouped into two major superfamilies of proteins (SFI and SFII) on the basis of the occurrence of conserved motifs (1). RNA helicases are mostly of the SFII superfamily and can be further classified into families on the basis of particular consensus sequences in the conserved ATP hydrolysis motif. In contrast to DNA helicases, which processively unwind long regular dsDNA structures, most RNA helicases modulate only short duplex regions in RNA molecules in a one-step reaction in the presence of ATP (1).

RNA helicases are involved in various steps of RNA metabolism including tran-

scription, pre-mRNA splicing, ribosome biogenesis, cytoplasmic transport, translation, initiation/elongation, and mRNA decay (1). DEAD-box RNA helicases are characterized by the presence of eight conserved motifs including an RNA interaction domain and an ATP hydrolysis motif containing the core amino acid sequence DEAD (2). In addition to the conserved motifs, RNA helicases contain variable N and C terminal extensions that might confer substrate specificity and/or contain information directing subcellular localization (3).

Chromosome amplifications and overexpression of different DEAD-box RNA helicases have already been reported in different tumors (4-7), suggesting a general role for RNA helicases in tumorigenesis. Here we

report the chromosomal localization, genomic organization and expression of DDX26, a human DEAD-box RNA helicase, with high similarity to an essential yeast RNA helicase (Dbp9b) involved in the assembly of early pre-ribosomal particles. The new helicase maps to 7p12, a region of frequent chromosome amplifications in glioblastomas involving the epidermal growth factor receptor (EGFR) gene. Nevertheless, co-amplification of DDX26 with EGFR was not detected in nine tumors analyzed.

Material and Methods

Computational prediction of gene structure

Gene structure was predicted by tBlastN analysis using the yeast orthologous protein Dbp9b as a query. Proximal and distal exons were then predicted by BlastN analysis using the PAC sequence (AC004938) as a query against human ESTs available in dbEST (AW250279, AI242741). Oligonucleotides corresponding to the 5' and 3' ends of the putative transcript (Figure 1A) were designed and used to amplify cDNAs derived from colon and breast tumors by PCR. Intron/exon boundaries were determined by aligning the cDNA sequence with the genomic sequence from the PAC clone.

RT-PCR and sequencing

Total RNA was extracted from breast and colon tumors using Trizol reagent (Gibco-BRL, Gaithersburg, MD, USA) following manufacturer's instructions. Two micrograms of total RNA was reverse-transcribed into first-strand cDNA using SuperScript II and oligo dT in a final volume of 20 μ l. PCR was carried out in a 10- μ l reaction mixture containing 1 μ l of cDNA, 1X Taq DNA polymerase buffer, 200 μ M dNTPs, 2 pmol of each primer and 1 U Taq DNA polymerase (Gibco-BRL). Cycling conditions were 94°C for 5 min, followed by 40 cycles

of 94°C for 1 min, 60°C for 1 min, 72°C for 2 min and a final extension at 72°C for 10 min. PCR products were analyzed on ethidium bromide-stained agarose gels before sequencing. PCR products were used directly in sequencing reactions with Big Dye terminator mix on an ABI377 sequencer according to manufacturer's instructions. Sequencing reactions were performed with the same primers used for RT-PCR and internal primers as indicated in Figure 1.

Northern and dot blot hybridization

A commercial "Human Master Blot" (Clontech, Palo Alto, CA, USA) containing normalized loading of polyA⁺ RNA from 75 different human tissues and a "Cancer Cell Lines Northern blot" (Clontech) containing 2 μ g of polyA⁺ RNA was used for evaluating DDX26 expression. A ³²P-labeled probe corresponding to the 3' end of the transcript was hybridized in 5 ml ExpressHyb solution (Clontech) at 65°C overnight. The membranes were washed twice with 2X SSC/0.1% SDS at room temperature for 5 min and then twice with 2X SSC/0.1% SDS for 20 min at 65°C before exposure at -70°C for 24 h.

Southern blotting and densitometry

Allele dosage at the DDX26 locus was quantitated by Southern blotting in a panel of nine glioblastomas with known 7p12 amplifications. Southern blotting, hybridization and PhosphorImage analysis were performed as previously described (8). Probes used for hybridizations were: WI-15640, SGC34528, D7S2542, WI-6548, EGFR, SGC32574, WI-18503, WI-16990, SHGC11828 and a 1.2-kb fragment from the 3' end of DDX26 transcript.

Results

As part of the Ludwig/FAPESP Human

Cancer Genome Project (9), we identified an EST with significant similarity to a DEAD-box RNA helicase from yeast and various other organisms as well as to a finished human genomic sequence corresponding to the PAC clone DJ0971C03 (AC004938). Matches with previously known human helicases, however, were found to be only marginal, indicating that this EST is a new member of the DEAD-box family not annotated in the PAC sequence.

We have used the yeast orthologous gene to search for the human genomic sequence and to predict gene structure. Primers for RT-PCR amplification were designed for validation and sequencing of the predicted transcript. As expected from the computational prediction, a 1,889-bp fragment was obtained by RT-PCR and fully sequenced by primer walking (Figure 1A). This gene was named DDX26 after consulting the HUGO Genome Nomenclature Committee and deposited in GenBank (AF247666). The gene consists of 14 exons distributed over an extension of 8,123 bp of genomic sequence. Exons and introns vary in size from 74 to 275 bp and from 92 to 1,594 bp, respectively (Figure 1B). Computer analysis of the PAC sequence using Genscan failed to predict the correct gene structure due to the apparent existence of other genes in the same region and to the small size of exon 13 that was not detected by using this program.

Complete sequencing of this helicase revealed a single putative open reading frame coding for a protein of 548 amino acids (Figure 1A). Analysis of the predicted protein sequence using Pfam (<http://pfam.wustl.edu/>) revealed the presence of two highly conserved domains corresponding to a DEAD/DEAH-box helicase domain (20-234 amino acids) and a helicase C-terminal conserved domain (272-380 amino acids). Figure 2 shows the alignment of the predicted amino acid sequence of DDX26 with RNA helicases from *Drosophila*, yeast, *Can-*

*did*a and the human DDX6 using CLUSTAL W (10). As for the first EST, a high similarity was observed with an essential yeast RNA helicase (Dbp9b) involved in the assembly of early pre-ribosomal particles (43% identity and 64% similarity). The putative function of this protein in ribosome biogenesis was recently confirmed by Zirwes and colleagues (11) during the preparation of this manuscript.

To determine the pattern of expression of DDX26 in different human tissues, a 1.5-kb cDNA fragment corresponding to the 3' portion of the mRNA was hybridized against commercial filters (Clontech) (Figure 3). While dot blot analysis indicated a ubiquitous expression of DDX26 in all tissues analyzed, a more detailed analysis showed a higher expression in cerebellum, pituitary gland, fetal kidney and fetal lung (Figure 3A). The specificity of the dot blot signal was confirmed by Northern blot analysis which showed the presence of a band of the expected size in all tumor cell lines tested. The highest expression level was observed in the melanoma cell line G-361 (Figure 3C). No alternative splicing forms were evident by Northern blot or Blast analysis against human ESTs.

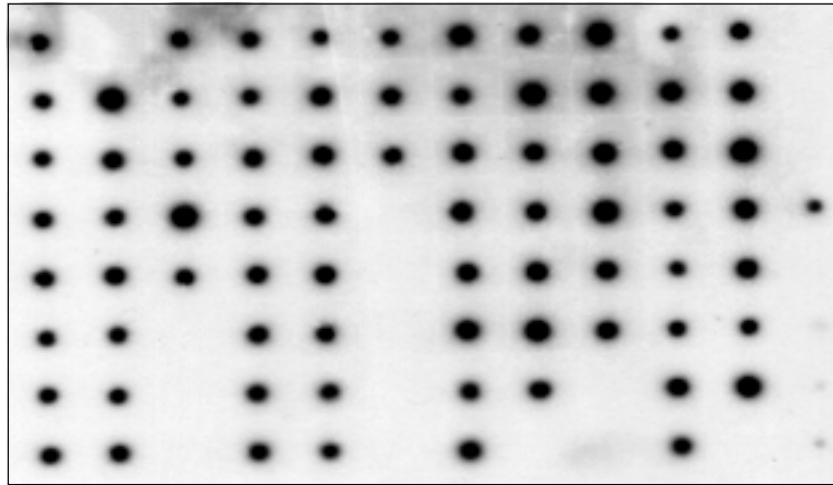
The PAC clone was mapped to human chromosome 7p12, a region of frequent chromosome rearrangements in glioblastomas (12). The EGFR gene is located in this same region which is amplified in 40% of glioblastomas together with extensive amplicons including a number of adjacent loci in some of these tumors. Five percent of 7p12 amplicons do not include EGFR, suggesting that other still unidentified genes at this locus may be amplified in glioblastomas (8). Southern blotting and densitometry were used to test the possible amplification of this human RNA helicase in tumors. As shown in Table 1, of nine glioblastoma analyzed, eight have amplifications of the EGFR locus, but co-amplification of DDX26 was not detected in any of them.

B

Exon	Acceptor site	Donor site	Exon size (bp)	PAC position
1		GCTCCTTCAG gt acacgcga	94 bp	15275-15368
2	ggtctect ag GCTGTCACCG	TAGGAAGGCG gt gggtaccg	162 bp	15460-15621
3	attcttac ag ACAGGTCGG	TCTCTCAGAG gt gggtaaaa	161 bp	16154-16314
4	ccacctcc ag AGCTGTGCTG	GTCTCCTCT gt aaggcaga	171 bp	16460-16630
5	ttcatttc ag TCACTTGCCC	ACATAACCOG gt aagaggca	91 bp	16767-16857
6	tctcccc ag GTTACCCTTA	TGCGCTCCAG gt tctgccaca	245 bp	17468-17712
7	ttggctgc ag GTGCCACATC	AAGGGGACAA gt gagtccat	119 bp	18327-18445
8	atctctgc ag GGCCTCTGAT	GAGCTGGCAG gt tagtagtgt	124 bp	19075-19188
9	tccettgc ag GACAGCAOGC	CTCAGTGGAG gt aagagcct	95 bp	19300-19394
10	ttctccac ag AGAACAGGGG	TCGCTGCAGG gt gagctgct	74 bp	20000-20073
11	ccaccccc ag GATGCCATGC	GAAGCTTAAG gt gagtggat	90 bp	20211-20300
12	ccccggc ag ACATACTTTG	GACTAOCTGG gt gagtgtgg	106 bp	20980-21085
13	tctctcc ag TTCTCCTGTC	GAAGGCCAAG gt acggetcc	77 bp	22679-22755
14	gcctcett ag AGAGCAAAGT		275 bp	23124-23398

Figure 1. A, cDNA (AF247666) and predicted protein sequence of DDX26. Primers used for amplification and sequencing are marked in bold and the putative polyadenylation site is underlined. The putative initiation methionine was considered to be the first ATG codon in the sequence. B, Intron/exon junctions of DDX26. Upper and lower case letters correspond to the exon and intron sequences, respectively. The conserved bases in the 5'donor and 3'acceptor splice site junctions are shown for each exon in bold. Exon size and position, based on the PAC DJ0971C03 sequence, are indicated.

A



B

whole brain		substantia nigra	heart	esophagus	transverse colon	kidney	lung	liver	leukemia, HL-60	fetal brain	yeast total RNA
cerebral cortex	right cerebellum	nucleus accumbens	aorta	stomach	descending colon	skeletal muscle	placenta	pancreas	HeLa S3	fetal heart	yeast tRNA
frontal lobe	corpus callosum	thalamus	left atrium	duodenum	rectum	spleen	bladder	adrenal gland	leukemia, K-562	fetal kidney	E. coli rRNA
parietal lobe	amygdala	pituitary gland	right atrium	jejunum		thymus	uterus	thyroid	leukemia, MOLT-4	fetal liver	E. coli DNA
occipital lobe	caudate nucleus	spinal cord	left ventricle	ileum		peripheral blood lymphocytes	prostate	salivary gland	Burkitt's lymphoma, Raji	fetal spleen	poly r(A)
temporal lobe	hippocampus		right ventricle	ileocecum		lymph nodes	testis	mammary gland	Burkitt's lymphoma, Daudi	fetal thymus	human C ₀ t-1DNA
posterior girus of cerebral cortex	medulla oblongata		inter-ventricular septum	appendix		bone marrow	ovary		colorectal adenocarcinoma, SW480	fetal lung	human DNA 100 ng
pons	putamen		apex of the heart	ascending colon		trachea			lung carcinoma, A-549		human DNA 500 ng

C

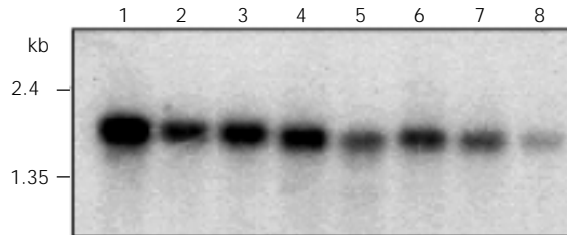


Figure 3. The expression pattern of DDX26 in normal tissues and tumor cell lines. A ³²P-labeled probe was hybridized under stringent conditions to a Human Master Blot (Clontech) containing normalized loadings of polyA⁺ RNA from 75 different human tissues (A) as well as to a Cancer Cell Lines Northern blot (Clontech) containing 2 µg of polyA⁺ RNA (C). B, Location of the various RNAs on the dot blot filter is shown schematically. C, For the Northern blot membrane, the cancer cell lines are: 1, melanoma, G-361; 2, lung carcinoma, A-549; 3, colorectal adenocarcinoma, SW480; 4, Burkitt's lymphoma, Raji; 5, lymphoblastic leukemia, MOLT-4; 6, chronic myelogenous leukemia, K-562; 7, HeLa S3, and 8, promyelocytic leukemia, HL-60. Molecular markers (kb) are indicated.

Table 1. Amplification of different markers located on chromosome 7p12 in glioblastoma tumors.

Tumor	WI-15640	SGC34528	D7S2542	WI-6548	EGFR	SGC32574	WI-18503	WI-16990	SHGC11828	cda1bc08	DDX26
GB175	A	A	A	A	A	N	A	A	A	N	N
GB178	A	A	A	A	A	N	A	A	N	A	N
GB168	A	A	A	A	A	N	N	N	N	N	N
GB199	A	A	A	A	A	N	N	N	N	N	N
GB176	A	N	A	A	A	A	A	A	N	A	N
GB197	A	A	N	A	A	A	N	N	N	N	N
GB188	N	N	A	N	A	N	A	A	N	N	N
GB180	N	N	N	N	A	N	A	A	N	A	N
GB177	N	N	N	A	N	N	N	N	nd	nd	nd

N = no amplification detected; filled squares = amplification detected; nd = not determined.

Discussion

DDX26, a human RNA helicase from the DEAD-box family, was cloned and characterized based on similarity to a yeast orthologous gene *Dbp9b*. Cross-species sequence comparison was shown to be a useful tool for identification of exons due to their strong conservation during evolution when compared with random genomic sequences. Accordingly, we have used the yeast orthologous gene to search the human genomic sequence and to predict gene structure. Primers for RT-PCR amplification were designed for validation and sequencing of the predicted transcript.

Complete sequencing of the 1,889-bp PCR fragment revealed a single putative open reading frame coding for a protein of 548 amino acids with high similarity to an essen-

tial yeast RNA helicase (*Dbp9b*) involved in the assembly of early pre-ribosomal particles. The corresponding transcript is expressed in a broad range of tissues in agreement with the putative functional role in ribosome biogenesis.

DDX26 is located on chromosome 7p12, a region of frequent chromosome rearrangements in glioblastomas (12). The *EGFR* gene is located in this same region which is amplified in 40% of glioblastomas together with extensive amplicons including a number of adjacent loci in some of these tumors. Recently, Wang et al. (13) described a new gene coding for a putative tyrosine kinase substrate (GBAS) that was co-amplified with *EGFR* in 2 of 12 tumors and in 2 of 3 cell lines analyzed.

Chromosome rearrangements, amplifications and overexpression of different DEAD-

box RNA helicases have already been reported for different tumors (4-7). Nevertheless, co-amplification of DDX26 was not detected in eight tumors with amplification at the EGFR locus, excluding the involvement of this helicase in glioblastoma development.

Acknowledgments

We thank Dr. Ricardo R. Brentani for continuous support and for a critical reading of the manuscript. We also thank Elaine Pereira Guimarães for technical assistance.

References

- de La Cruz J, Kressler D & Linder P (1999). Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends in Biochemical Sciences*, 24: 192-198.
- Schmid SR & Linder P (1992). D-E-A-D protein family of putative RNA helicases. *Molecular Microbiology*, 6: 283-391.
- Wang Y & Guthrie C (1998). PRP16, a DEAH-box RNA helicase, is recruited to the spliceosome primarily via its nonconserved N-terminal domain. *RNA*, 4: 1216-1229.
- Akao Y, Seto M, Yamamoto K, Iida S, Nakazawa S, Inazawa J, Abe T, Takahashi T & Ueda R (1992). The RCK gene associated with t(11;14) translocation is distinct from the MLL/ALL-1 gene with t(4;11) and t(11;19) translocations. *Cancer Research*, 52: 6083-6087.
- Amler LC, Schurmann J & Schwab M (1996). The DDX1 gene maps within 400 kbp 5' to MYCN and is frequently coamplified in human neuroblastoma. *Genes, Chromosomes and Cancer*, 15: 134-137.
- Arai Y, Hosoda F, Kobayashi H, Arai K, Hayashi Y, Kamada N, Kaneko Y & Ohki M (1997). The inv(11)(p15q22) chromosome translocation of de novo and therapy-related myeloid malignancies results in fusion of the nucleoporin gene, NUP98, with the putative RNA helicase gene, DDX10. *Blood*, 89: 3936-3944.
- Ishiguro T, Nakajima M, Naito M, Muto T & Tsuruo T (1996). Identification of genes differentially expressed in B16 murine melanoma sublines with different metastatic potentials. *Cancer Research*, 56: 875-879.
- Liu L, Ichimura K, Pettersson EH & Collins VP (1998). Chromosome 7 rearrangements in glioblastomas; loci adjacent to EGFR are independently amplified. *Journal of Neuropathology and Experimental Neurology*, 57: 1138-1145.
- Dias-Neto E, Garcia CR, Verjovski-Almeida S, Briones MR, Nagai MA, da Silva WJ, Zago MA, Bordin S, Costa FF, Goldman GH, Carvalho AF, Matsukuma A, Baia GS, Simpson DH, Brunstein A, de Oliveira PS, Bucher P, Jongeneel CV, O'Hare MJ, Soares F, Brentani RR, Reis LF, de Souza SJ & Simpson AJ (2000). Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proceedings of the National Academy of Sciences, USA*, 97: 3491-3496.
- Thompson JD, Higgins DG & Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22: 4673-4680.
- Zirwes RF, Eilbracht J, Kneissel S & Schmidt-Zachmann MS (2000). A novel helicase-type protein in the nucleolus: protein NOH61. *Molecular Biology of the Cell*, 11: 1153-1167.
- Chaffanet M, Chauvin C, Laine M, Berger F, Chedin M, Rost N, Nissou MF & Benabid AL (1992). EGF receptor amplification and expression in human brain tumours. *European Journal of Cancer*, 28: 11-17.
- Wang XY, Smith DI, Liu W & James CD (1998). GBAS, a novel gene encoding a protein with tyrosine phosphorylation sites and a transmembrane domain, is co-amplified with EGFR. *Genomics*, 49: 448-451.