Mammalian Genome 12, 456–461 (2001).
DOI: 10.1007/s003350020032

Mammalian Genome
Incorporating Mouse Genome

DQX1, an RNA-dependent ATPase homolog with a novel DEAQ box: expression pattern and genomic sequence comparison of the human and mouse genes

Weizhen Ji,[1] Feng Chen,[2] Trang Do,[2] Anh Do,[2] Bruce A. Roe,[2] Miriam H. Meisler[1]

[1]Department of Human Genetics, 4909 Buhl, Box 0618, University of Michigan, Ann Arbor, Michigan 48109-0618, USA
[2]Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019, USA

Received: 19 December 2000 / Accepted: 26 January 2001

Abstract. DQX1 is a novel gene related to the RNA-dependent ATPases. The gene was classified as a member of the DEAD/H family on the basis of the conserved order and spacing of ten short protein motifs. The unique features of DQX1 include replacement of the signature DEAH motif with DEAQ and the absence of the helicase motif. We determined the coding sequences of human and mouse DQX1, which encode proteins of 717 and 718 amino acids with 84% amino acid sequence identity. The 3.2-kb *Dqx1* transcript has highest expression in muscle and liver. DQX1 is located between AUP1 and HOX11L1 in a gene-dense region of human Chromosome (Chr) 2p13 and mouse Chr 6. Although DQX1 is within the nonrecombinant region for the mouse neuromuscular mutant mnd2, no difference in coding sequence, transcript length, or transcript abundance was observed between normal mice and mnd2 mutant mice. The ubiquitous expression of DQX1 and its close phylogenetic relationship to the yeast pre-mRNA processing (Prp) proteins suggest a role in cellular RNA metabolism.

Introduction

The DEAD/H box proteins mediate ATP-dependent unwinding of double-stranded RNA and belong to RNA helicase superfamily II (SFII). These proteins have been the subject of intensive research because of their role in many aspects of RNA metabolism, including transcription, pre-mRNA splicing, RNA editing, RNA export, translation, and RNA degradation (Schmid and Linder 1992; de la Cruz et al. 1999). They are ubiquitous from bacteria to human and contain several evolutionarily conserved protein motifs, including the ATP-hydrolysis motif II from which they derive their name: DEAD, DEAH, or DExH. The *Drosophila, C. elegans,* and *S. cerevisiae* genomes contain more than 30 DEAD and 10 DEAH box proteins, making this one of the largest gene families in these organisms (Rubin et al. 2000). Site-directed mutagenesis of the yeast proteins identified conserved motifs required for ATP binding and ATP hydrolysis (Lüking et al. 1998). DEAD box proteins also have been implicated in development and differentiation. The *Drosophila* mutants abstrakt, with defective optic nerve pathfinding, and VASA, with defective posterior structure and pole cells, result from mutations in DEAD box proteins (Schmucker et al. 2000; Lasko and Ashburner 1988; Hay et al. 1988). Another recently identified DEAD box protein, dp103, directly interacts with the spinal motor atrophy protein SMN and with viral nuclear proteins, suggesting a role in transcriptional regulation (Grundhoff et al. 1999; Campbell et al. 2000).

The mouse mutant mnd2 arose in 1990 and is characterized by severe muscle atrophy, generalized wasting, and juvenile lethality (Jones et al. 1993). We mapped mnd2 to a 400-kb interval of mouse Chr 6 that corresponds to human Chr 2p13 (Weber et al. 1998). The mnd2 chromosome region contains a high density of genes with conserved gene order in human and mouse (Jang et al. 1999 and manuscript in preparation). Another mouse mutant with a similar phenotype results from mutation of the translation elongation factor Eef1a (Chambers et al. 1998). By comparative genomic sequencing of the human and mouse mnd2 region, we have identified a novel protein with 38% amino acid sequence identity and 58% similarity to the human DEAD/H gene DDX15. We describe the features of DQX1 and its evaluation as a candidate for the mnd2 mutation.

Correspondence to: M. Meisler; E-mail: meislerm@umich.edu

Materials and Methods

RT-PCR. Total RNA (5 μg) from mouse muscle and brain was reverse-transcribed in a volume of 20 μl by using Superscript II reverse transcriptase (Gibco BRL) with random or oligo dT primers, according to the supplier's instructions. One microliter of each RT reaction was amplified by PCR using Taq polymerase (Roche). PCR was carried out with the primer pairs 1F (ATCTG GGCTG CTCGC TTCCT GTTCT T) and 3R (GCAGA AATAG CCTGT CCCAG CAAAA C) or 8F (ATCAC AGCTC TCTGA TCCAG GTCTA TG) and 11R (CTGGT TCAAG AGGTC TCTGC TCTCA CT). The portion of the transcript between exon 3 and exon 8 was amplified by nested RT-PCR with primers 3FA (GTTTT GCTGG GACAG GCTAT TTCTG C) and 10R (GGCAG TTGTC TTTCG AGATG GAGA) for the first round and primers 3FB (GCTTG AGCTG TGTCA GCAGG AAGAG) and 8R (GGCTT CATAG ACCTG GATCA GAGAG) for the second round.

DNA sequencing. Large-scale sequencing of BAC and P1 clones was carried out as previously described (Jang et al. 1999). GenBank Accession numbers are AC006544 for human BAC clone 91A19 (DQX1 = nt 678 to 7905), AC005041 for human BAC clone RP11-523H20 (DQX1 = nt 9570 to 16853), and AC007307 for mouse P1-2. For small-scale sequencing, PCR products were gel purified by using QIAEX II (Qiagen), and automated sequencing of both strands was carried out by the University of Michigan Sequencing Core. The GenBank Accession number for the mouse Dqx1 coding sequence is AF318278.

Northern blots. Poly A+ RNA (2 μg) from mouse tissues was separated on a 1.2% agarose-formaldehyde gel and transferred onto a nylon membrane (Zeta-Probe GT, BioRad). The hybridization probe was the 460-bp ^{32}P-labeled RT-PCR product containing exons 8–11 that was amplified with primers 8F and 11R as above. Hybridization was carried out as previously described with a stringent final wash at 65°C in 0.2×SSC, 0.1% SDS.

Sequence analysis. Genomic sequence was analyzed via the NIX website (http://www.hgmp.mrc.ac.uk/NIX), by using RepeatMasker, BLAST, the exon prediction programs Grail, Genemark, Fex, and Hexon, and the

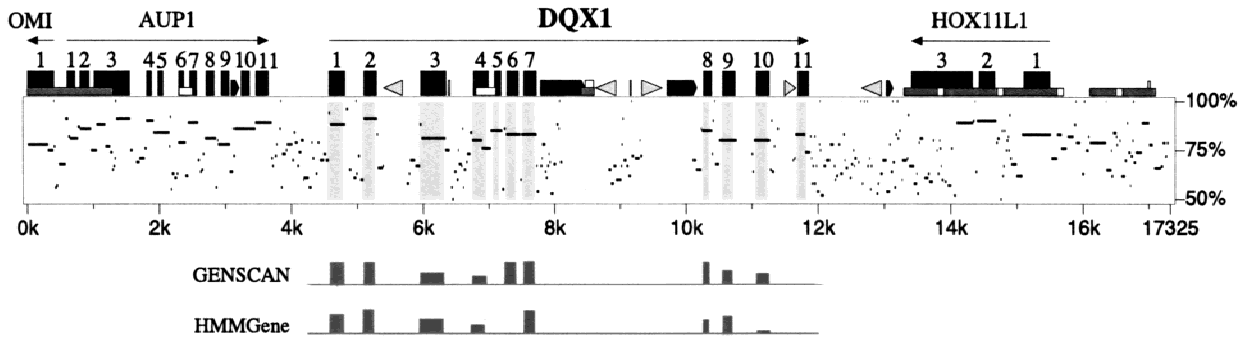

Fig. 1. Percent identity plot (PIP) of aligned human and mouse genomic sequences containing the three genes AUP1, DQX1 and HOX11L1. The human sequence is shown on the horizontal axis, and the percentage identity of the nongapped alignment of human and mouse sequences is plotted in the vertical axis. Large-scale sequencing was carried out as previously described (Jang et al. 1999). Human sequence, GenBank AC005041, 5001-22325; Mouse sequences, AC007307.26, 28487–50426; M75953 (Hox11L1 exon 1) and AB008500 (upstream of Hox11L1). Exons are numbered. The analysis by exon prediction programs is shown below with likelihood represented by height. Light gray triangles represent SINEs; black arrows represent LINE2s; open square boxes represent simple repeats; CpG islands in the human sequence are indicated by short boxes that are open for CpG/GpC $\geq$ 0.6 and dark grey for CpG/GpC $\geq$ 0.75.

Table 1. Exon-intron junctions of the mouse and human DQX1 genes.

Exon				Intron				Exon	
No.	Mouse/ Human	Size (bp)	5′-flanking sequence	5′-Splice donor	No.	Size (bp)	3′-Splice acceptor	3′-flanking sequence	No.
1	M	>228	CACCCAG	gtggagg	1	181	accccag	ATTCCTC	2
	H	>237	CACCCAG	gtgtggt		290	atcccag	ATCCCTC	
2	M	194	TGCTCAG	gtgggcc	2	328	ccaacag	GTTTTGC	3
	H	194	TGCTCAG	gtggggg		701	tcaacag	GTTCTGC	
3	M	388	CGAAGAG	gtaaacg	3	341	ttctcag	GAAATTT	4
	H	385	TGAGGAG	gtaaaaa		385	ttctcag	GAAATTT	
4	M	234	TCGTAGT	gtgagtg	4	89	atgacag	GTTTACA	5
	H	228	CCGAAGT	gtgagtg		91	gcaacag	GTTTACA	
5	M	100	CCACCAG	gtaagag	5	106	ctttcag	GATCCTG	6
	H	100	CCACCAG	gtaagag		104	ccctcag	GATCCTG	
6	M	159	CGGCCTG	gtgagcc	6	80	tcaccag	CACCAGA	7
	H	159	CAGCCTG	gtgagca		84	tcaccag	CTCCAGA	
7	M	192	CTCACTG	gtacatt	7	~2600	cctccag	CTGCTCC	8
	H	192	CTCACAG	gtataaa		2545	cctccag	CTGCCCC	
8	M	120	ATACAAA	gtaagtg	8	150	ctccgag	GTGGAGC	9
	H	120	GTACAAA	gtgagtt		168	ctcctag	GTGGCCA	
9	M	191	CCTCAAG	gttgggg	9	1134	caatgag	GTGGCCC	10
	H	191	TCTCAAG	gttagag		325	caataag	GTGGCCA	
10	M	191	CTGAGAT	gtgagtt	10	152	ctgtcag	GCTGGTG	11
	H	191	CACAGAT	gtgagtt		437	ctattag	GCTGGTG	
11	M	157	TCTGCAG	*tga* + 3′UTR					
	H	154	CCTGCAG	*tga* + 3′UTR					

gene prediction programs Genefinder, Fgene, GENSCAN and HMMGene. Protein domains or motifs were identified by InterPro (http://www.ebi.ac.uk/interpro/interproscan/ipsearch.html), SMART (http://smart.embl-heidelberg.de/smart), and MOTIF (http://www.motif.genome.ad.jp), which includes the PROSITE searches BLOCKS, PRINTS, and Pfam. Human and mouse genomic sequences were compared by using the Pipmaker program (Schwartz et al. 2000) (http://bio.cse.psu.edu/cgi-bin/pipmaker). Multiple alignments of amino acid sequences were carried out with Clustal W 1.8 (Jeanmougin et al. 1998) or MAP (Huang 1994) from the BCM Search Launcher website (http://www.hgsc.bcm.tmc.edu/searchlauncher), and the shading of multiple alignment was carried out by BOXSHAD server (http://www.ch.embnet.org/software/BOX_form.html). Promoters were predicted by neural network eukaryotic promoter prediction (NNPP/Eukaryotic) (Reese and Eeckman 1995) (http://www.fruitfly.org/seq_tools/promoter.html), and transcription factor binding sites were predicted by MatInspector/TRANSFAC (http://genomatix.gsf.de/cgi-bin/matinspector/matinspector.pl), both through the BCM Search Launcher. Human ESTs were identified by BLAST in libraries from head and neck (AW177873, BE140919, AW178225, AW177910, AW177807, AW177909, AW177808, AW178141, AW178142, AW178143, AW178144, AW178173, AW178226, AW178146); colon (AW062662, AW062664, AW062663, AW062666, AA515987, AI589016); uterine tumor (AI68966); and human fetal liver and spleen (H58606).

Phylogenetic analysis. BLASTP search of the GenBank nonredundant database with the translated sequence of DQX1 identified the 20 most closely related conceptual translated sequences: BAA91754 (AK001556), CG11107 (AE003841), CAB82945 (AL162507), AAB63825 (AC002337), T39615 (AL022104), F56D22.6 (U13644), Prp2 (X55999), Prp16 (M31524), hPRP16 (AF038391), Prp22(X58681), Prp43 (U41851), CG8241 (AE003815), Mog-1 (AF120269), Mog-4 (AF286900), Mog-5 (AF286899), CG10689 (AE003661), DDX8 (D50487), DDX15 (AF279891), Ddx15 (AF017153) and DDX16 (AB001601). The 630 residue core sequence of human and mouse DQX1 was compared with the 20 most closely related proteins by full alignment without manual intervention by using the Clustal W software (http://www2.ebi.ac.uk/clustalw/). The phylogenetic tree was generated from the aligned protein sequences by the neighbor joining method of Saitou and Nei (Ibid). TreeView software was used to display the graphical representation of the unrooted tree (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html) (Page 1996).

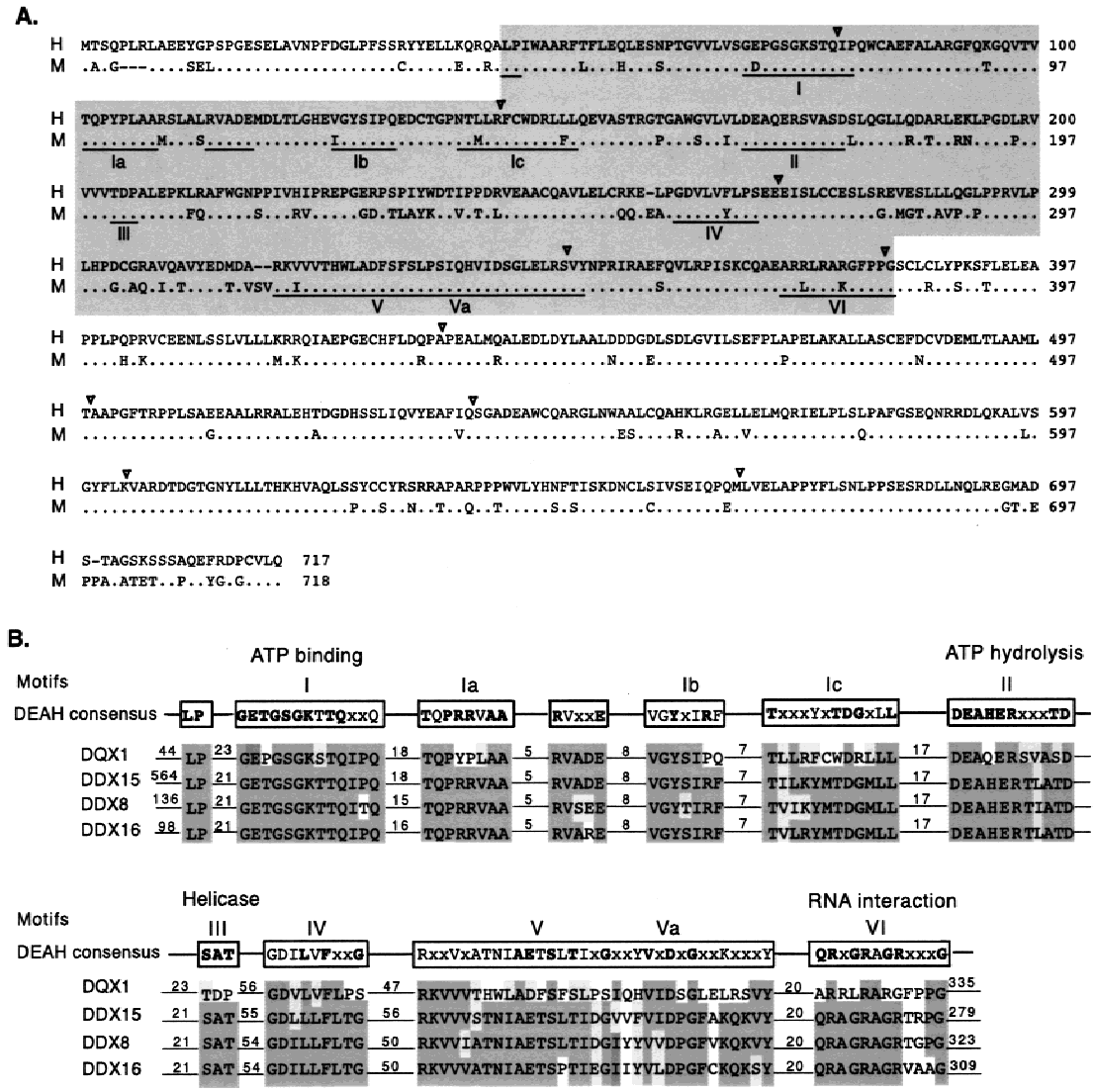

Fig. 2. Amino acid sequence and conserved protein domains of DQX1. **A.** Alignment of human and mouse amino acid sequences. Conserved DEAH protein domains are numbered as below. Dots, identity; dashes, missing residues; arrows, intron/exon junctions. **B.** Conserved protein domains. The shaded region of DQX1 from 2A was aligned with the corresponding regions of the indicated proteins by using Clustal W. Consensus motifs are numbered according to Jankowsky and Jankowsky (2000). Residues with identity to the consensus are shaded darker, and conservative substitutions are shaded lighter.

Results

Identification of DQX1 by comparative genomic sequence analysis and confirmation by RT-PCR. The human and mouse genomic sequences spanning the 10-kb interval between Aup1 and Hox11L1 were obtained as described in Methods and aligned by using the Percent Identity Plot software (Schwartz et al. 2000). Eleven short segments with nucleotide sequence identity greater than 75% were observed (Fig. 1). The programs GENSCAN and HMMGene predicted that several of these conserved segments are exons (Fig. 1). To determine which segments are transcribed, we carried out RT-PCR with the primers described in Materials and methods. Sequencing of three overlapping RT-PCR products demonstrated that portions of all the 11 conserved segments are transcribed. Comparison of the sequences of the RT-PCR products with the genomic sequence identified the intron/exon junctions (Table 1). Exons range from 100 bp to 388 bp in length, and the intron/exon junctions agree with consensus donor and acceptor splice sites (Shapiro and Senapathy 1987). Exons 1, 3, 4, and 11 differ in length in the human and mouse genes by multiples of three nucleotides, resulting in differences in the number of encoded amino acids (Table 1). In addition to RT-PCR products containing the correctly spliced open reading frame, some products contained introns 1 and 4 with in-frame stop codons. Intron 4 was also retained in several human ESTs (GenBank AW178141, AW178144, AW178173, AW178226, BE140919).

DQX1 is located in a conserved gene-dense region of human Chr 2p13 and mouse Chr 6 (Weber et al. 1998; Jang et al. 1999). Gene content and order in this region are identical over more than 500 kb (manuscript in preparation). In both species, DQX1 is located 1 kb downstream of AUP1 and 2 kb upstream of the homeobox gene HOX11L1 (Fig. 1).

DQX1 contains protein domains characteristic of the DEAH family of RNA helicases. The human RT-PCR products contain a 2151-bp open reading frame encoding a predicted protein of 717

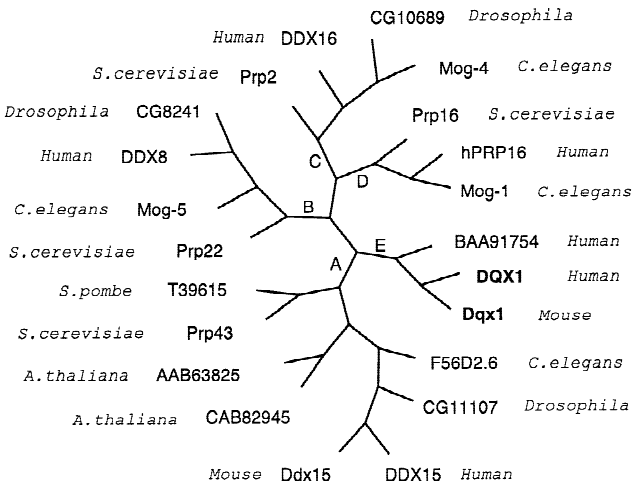

Fig. 3. Phylogenetic relationship of DQX1 and the DEAH protein family. The corresponding 630 residue core region of the indicated proteins was compared. The phylogenetic tree was generated with Clustal W and displayed by TREEVIEW as described in Methods.

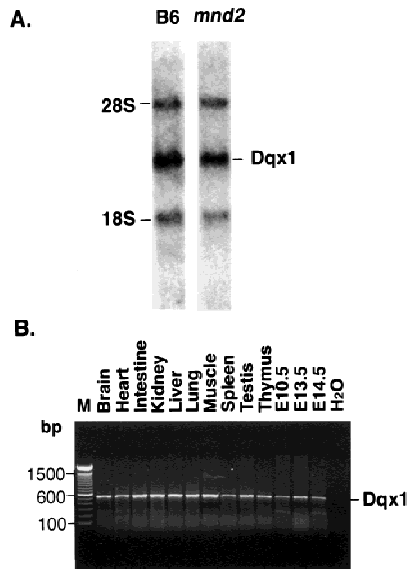

Fig. 4. Ubiquitous expression of DQX1. **A.** The Northern blot containing 2 μg of polyA+ RNA from muscle of C57BL/6J and mnd2 mice was hybridized with a probe containing exons 8–11 of mouse DQX1. One transcript of 3.2 kb was detected. **B.** The predicted 460-bp RT-PCR product was amplified from fetal and adult mouse tissues with primers 8F and 11R.

amino acids, and the mouse ortholog encodes one additional amino acid (Fig. 2A). The predicted amino acid sequences of human and mouse DQX1 exhibit 84% sequence identity and 92% sequence similarity. Translation is predicted to be initiated at the conserved ATG codon at human nucleotide 9570 (GenBank AC005041) and mouse nucleotide 39027 (GenBank AC007307). The sequence context of this ATG conforms to the Kozak consensus GCCAC-CATGG (Kozak 1991), there are no conserved, in-frame, upstream ATG codons, and the mouse gene contains an in-frame stop codon 12 bp upstream of this ATG.

The rare poly-adenylation signal TATAAA, which is found in only 3% of eukaryotic mRNAs (Beaudoing et al. 2000), is present 250 bp downstream of the TGA stop codon in mouse DQX1 and 290 bp downstream of the TGA stop codon of the human gene. The locations of polyA sequences in human DQX1 ESTs indicate that this polyadenylation signal is functional (GenBank H58606, AI689665, AA515987, AI589016, AW510444).

Database searches with the translated sequence revealed homology to the 630 residue core region of the DEAH box proteins, with approximately 35% overall identity and 55% similarity to many proteins in this gene family. The DEAH box proteins contain 10 small sequence motifs separated by conserved distances (Jankowsky and Jankowsky 2000; Fig. 2B, top line). The corresponding sequence of human DQX1 (shaded region, Fig. 2A) was aligned with the consensus for these 10 motifs (Fig. 2B). DQX1 contains the characteristic ATP binding and hydrolysis motifs I, Ia, and II, but lacks the helicase motif III. In place of the DEAH or DEAD sequence in motif II, however, DQX1 contains the sequence DEAQ. This unique glutamine residue (Q) in the signature sequence is the basis for the gene symbol DQX1. Overall, the 10 motifs in DQX1 (83 amino acids) exhibit 63% amino acid sequence identity with the consensus (Fig. 2B). In addition, the four C-terminal residues of DQX1 constitute a potential prenylation site, CVLQ, suggesting that the protein may be membrane-localized (Sinensky 2000).

Alignment of the 630-residue core region of DQX1 with three human DEAH box RNA helicase family members (DDX8, DDX15, and DDX16) is shown in Fig. 2B. There is 55% amino acid sequence identity and 70% similarity between DQX1 and DDX15 in this 630 residue core. The distances between protein motifs in DQX1 are also consistent with the spacing in the other proteins. The combination of highly conserved and divergent motifs indicates that DQX1 represents a new subfamily of the RNA-dependent ATPases.

Phylogenetic relationships between DQX1 and the DEAH gene family. The 630-residue core region of DQX1, between amino acids 38 and 667, was compared with the corresponding regions of 20 related proteins identified by BLAST searches. The proteins include four yeast RNA helicases (Prp2, Prp16, Prp22, and Prp43); three mog genes from *C. elegans*, the human and mouse DEAH proteins DDX8, DDX15, and DDX16; and eight predicted proteins of unknown function. A phylogenetic tree was generated with Clustal W software and displayed using TREEVIEW (Fig. 3). The four Prp genes were distributed into four distinct branches, A to D, in agreement with the analysis of Puoti and Kimble (2000). Each branch contains one mammalian gene, and three branches contain a *C. elegans* mog gene. DQX1 is located on a separate branch and is most closely related to the human predicted protein BAA91754 (Fig. 3). BAA91754 contains an ATP binding site, ATP hydrolysis site, the sequence DDIH in place of DEAQ, and overall 42% amino acid sequence identity and 62% similarity to DQX1 within the 630-residue core region.

Ubiquitous expression of DQX1. polyA+ RNA was prepared from mouse tissues and examined by Northern blotting and hybridization with a DXQ1 cDNA probe spanning exon 8 to exon 11. A transcript of approximately 3.2 kb was detected in muscle (Fig. 4A) and liver (not shown). The abundance of the DQX1 transcript in other tissues was below the level detectable by Northern blot. When RT-PCR was carried out with the forward primer from exon 8 (8F) and a reverse primer from exon 11 (11R), the predicted 460-bp product was amplified from all fetal and adult tissues tested (Fig. 4B).

Twenty two human DQX1 ESTs were identified in the public databases. Of these, 14 were isolated from head and neck tumors and 1 from a uterine tumor, suggesting that DQX1 expression may be elevated in human cancers.

Dqx1 transcripts in mnd2 mice. Since mouse Dqx1 is located within the nonrecombinant region for the mnd2 mutant gene, its expression in mnd2 mice was examined. As shown in Fig. 4A, the size and abundance of the Dqx1 transcript in muscle did not differ between mnd2 and normal mice. The coding region of Dqx1 was

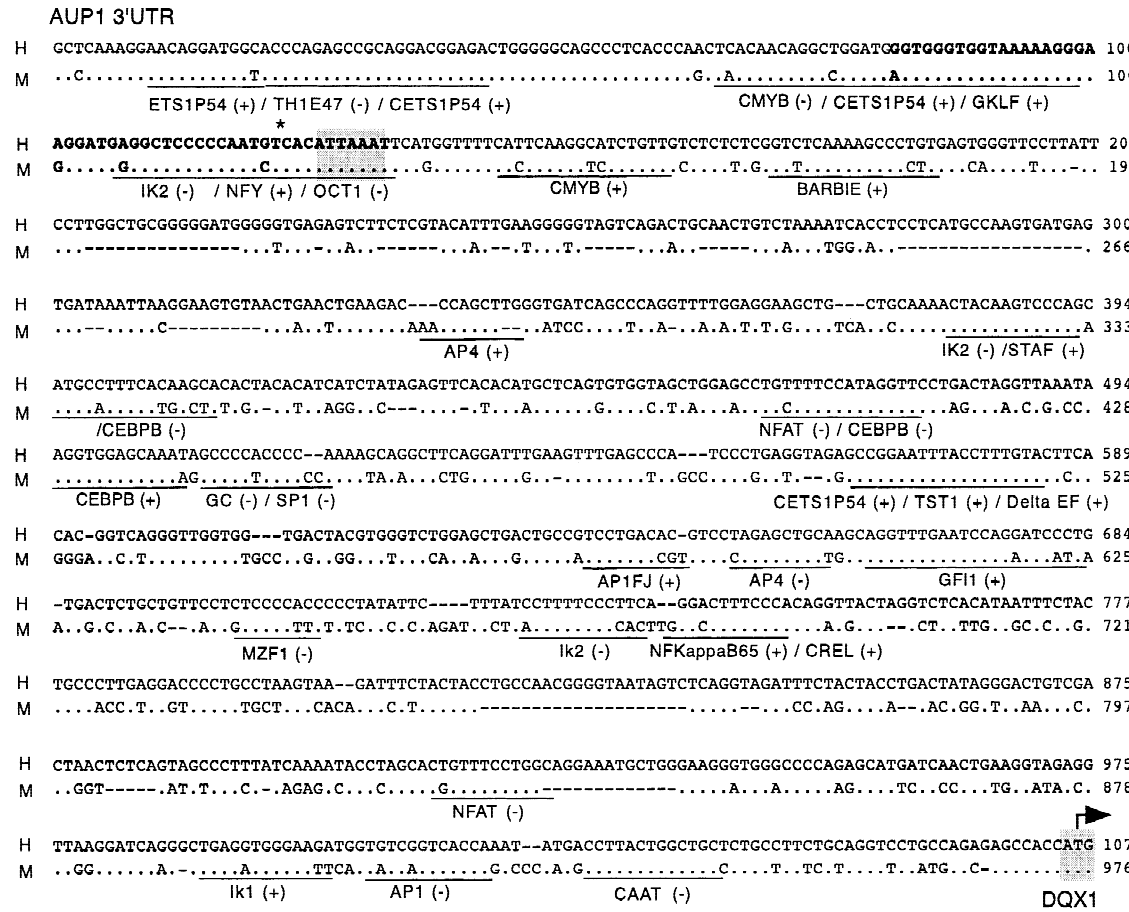

Fig. 5. Predicted promoter and transcription factor binding sites upstream of DQX1. The 1-kb intergenic region between AUP1 and DQX1 was analyzed with the promoter prediction program NNPP/Eukaryotic. Transcription factor binding sites that are conserved in the human and mouse sequences were identified by MatInspector version 2.2 with Transfac 3.3 matrices. The polyadenylation site of AUP1 and the first codon of DQX1 are marked by shading. The predicted promoter is shown in boldface. The predicted start site for transcription is marked by *. The evolutionarily conserved transcription factor binding sites are underlined.

amplified by RT-PCR from mnd2 muscle and brain and sequenced, but no differences from the background strain, C57BL/6J, were detected (data not shown). These data indicate that Dqx1 is not mutated in mnd2 mice.

Promoter sequences of human and mouse DQX1. The first coding exon of DQX1 is located only 900 bp downstream of the polyadenylation signal for AUP1 (Figs. 1 and 5). The small intergenic region exhibits 66% nucleotide sequence identity in human and mouse. The neural network promoter prediction program NNPP (Reese and Eeckman 1995) predicted one transcription start site in the intergenic region (associated false positive rate <0.1%) (Fig. 5, asterisk). The predicted site is located 5 bp upstream of the polyadenylation signal for AUP1 (Jang et al. 1996), suggesting that the AUP1 and DQX1 transcripts may overlap. The position of the predicted transcription start site is consistent with the length of the 3.2-kb transcript detected by Northern blot (Fig. 4A), taking into account the coding region of 2.1 kb and the location of the polyadenylation site 0.3 kb downstream of the stop codon. The program MatInspector/TRANSFAC (Quandt et al. 1995) predicted several conserved transcription factor binding sites in the human and mouse intergenic region (underlined in Fig. 5).

Discussion

A novel member of the RNA-dependent ATPase gene family with motifs characteristic of the DEAH proteins has been identified. The new gene was designated DQX1 to indicate its relationship to the mammalian DDX genes and the replacement of the DEAH motif with DEAQ. BLAST analysis demonstrated that DQX1 is most closely related to the yeast pre-mRNA processing (*Prp*) genes, the *C. elegans* oocyte differentiation (*mog*) genes, and the mammalian genes DDX8, DDX15, and DDX16, with which it shares 35% amino acid sequence identity and 55% sequence similarity. The yeast *Prp* genes encode the DEAH box proteins Prp2, Prp16, Prp22, and Prp43 that are involved in different stages of nuclear pre-mRNA-processing (Lüking et al. 1998). Mammalian DDX8, DDX15, and DDX16 are homologs of the Prp proteins and can complement the yeast mutants in vivo (Ono et al. 1994; Imamura et al. 1997, 1998; Gee et al. 1997). Mouse DDX15 specifically complements Prp43 (Gee et al. 1997), consistent with their close relationship indicated in the phylogenetic tree (Fig. 3). The *mog* genes of *C. elegans* (masculinization of the germ line) are required for the switch from sperm to oocyte differentiation (Graham et al. 1993). Three of these genes encode DEAH proteins (Puoti and Kimble 1999, 2000). The *mog* genes may function as post-transcriptional regulators that interact with the 3′ UTR of specific mRNAs.

A search of the protein databases identified nine predicted proteins with the tetrapeptide sequence DEAQ. All of these also contain ATP binding and ATP hydrolysis domains and exhibit significant similarity to the DNA-dependent ATPases in the SNF2 family of DEGH or DEAH containing proteins. One of the predicted proteins, yeast INO80, appears to be expressed, since the

null mutant exhibits inositol dependence, sensitivity to ethanol, and reduced expression of several other genes (Birkmann and Schuller 1999) (http://genome-www4.stanford.edu).

The conservation of the open reading frame in human and mouse indicates that DQX1 is a functional gene. The degree of conservation of human and mouse DQX1, 84% nucleotide sequence identity in the coding region, 66% identity in the 5′-UTR, and 58% in the 3′-UTR, is typical for human and rodent genes (Makalowski and Boguski 1998). DQX1 contains protein motifs related to RNA binding, ATP binding, and ATP hydrolysis, but lacks the RNA helicase domain found in many family members. Another recently described family member that lacks the helicase domain, benign gonial cell neoplasm, functions as a germ line differentiation factor in *Drosophila* (Ohlstein et al. 2000). The ubiquitous expression of DQX1 and its close relationship to the DEAH protein family suggest a biological role in cellular RNA metabolism.

Acknowledgments. This work was supported by National Institutes of Health grants GM24872 and HG02153 and the Muscular Dystrophy Association.

References

Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D (2000) Patterns of variant polyadenylation signal usage in human genes. Genome Res 10, 1001–1010

Birkmann ER, Schuller HJ (1999) The product of the SNF2/SWI2 paralogue IN080 of *Saccharomyces cerevisiae* required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex. Mol Microbiol 32, 741–751

Campbell L, Hunter KM, Mohaghegh P, Tinsley JM, Brasch MA, Davies KE (2000) Direct interaction of Smn with dp103, a putative RNA helicase: a role for Smn in transcription regulation? Hum Mol Genet 9, 1093–1100

Chambers DM, Peters J, Abbott CM (1998) The lethal mutation of the mouse wasted (wst) is a deletion that abolishes expression of a tissue-specific isoform of translation elongation factor 1-alpha, encoded by the Eef1a2 gene. Proc Natl Acad Sci USA 95, 4463–4468

de la Cruz J, Kressler D, Linder P (1999) Unwinding RNA in *Saccharomyces cerevisiae:* DEAD-box proteins and related families. Trends Biochem Sci 24, 192–198

Gee S, Krauss SW, Miller E, Aoyagi K, Arenas J, Conboy JG (1997) Cloning of mDEAH9, a putative RNA helicase and mammalian homologue of *Saccharomyces cerevisiae* splicing factor Prp43. Proc Natl Acad Sci USA 94, 11803–11807

Graham PL, Schedl T, Kimble J (1993) More mog genes that influence the switch from spermatogenesis to oogenesis in the hermaphrodite germ line of *Caenorhabditis elegans.* Dev Genet 14, 471–484

Grundhoff AT, Kremmers E, Türeci Ö, Glieden A, Gindorf C et al. (1999) Characterization of DP103, a novel DEAD box protein that binds to the Epstein-Barr virus nuclear proteins EBNA2 and EBNA3C. J Biol Chem 274, 19136–19144

Hay B, Jan LY, Jan YN (1988) A protein component of Drosophila polar granules is encoded by vasa and has extensive sequence similarity to ATP-dependent helicases. Cell 55, 577–587

Huang X (1994) On global sequence alignment. Comp Appl Biosci 10, 227–235

Imamura O, Sugawara M, Furuichi Y (1997) Cloning and characterization of a putative human RNA helicase gene of the DEAH-box protein family. Biochem Biophys Res Commun 240, 335–340

Imamura O, Saiki K, Tani T, Ohshima Y, Sugawara M, Furuichi Y (1998) Cloning and characterization of a human DEAH-box RNA helicase, a functional homolog of fission yeast Cdc28/Prp8. Nucleic Acids Res 26, 2063–2068

Jang W, Weber JS, Bashir R, Bushby K, Meisler MH (1996) Aup1, a novel gene on mouse chromosome 6 and human chromosome 2p13. Genomics 36, 366–368.

Jang W, Hua A, Spilson SV, Miller W, Roe BA, Meisler MH (1999) Comparative sequence of human and mouse BAC clones from the *mnd2* region of chromosome 2p13. Genome Res 9, 53–61

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

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. Trends Biochem Sci 23, 403–405

Jones JM, Albin RL, Feldman EL, Simin K, Schuster TG et al. (1993) mnd2: a new mouse model of inherited motor neuron disease. Genomics 16, 669–677

Kozak M (1991) An analysis of vertebrate mRNA sequences: intimations of translation control. J Cell Biol 115, 887–903

Lasko PF, Ashburner M (1988) The product of the Drosophila gene vasa is very similar to eukaryotic initiation factor-4A. Nature 335, 611–617

Lüking A, Stahl U, Schmidt U (1998) The protein family of RNA helicases. Critical Rev Biochem Mol Biol 33, 259-296

Makalowski W, Boguski MS (1998) Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. Proc Natl Acad Sci USA 95, 9407–9412

Ohlstein B, Lavoie CA, Vef O, Gateff E, McKearin DM (2000) The drosophila cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles. Genetics 155, 1809–1819

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

Page RD (1996) TreeView: an application of display phylogenetic trees on personal computers. Comp Appl Biosci 12, 357–358

Puoti A, Kimble J (1999) The *Caenorhabditis elegans* sex determination gene mog-1 encodes a member of the DEAH-Box protein family. Mol Cell Biol 19, 2189–2197

Puoti A, Kimble J (2000) The hermaphrodite sperm/oocyte switch requires the *Caenorhabditis elegans* homologs of PRP2 and PRP22. Proc Natl Acad Sci USA 97, 3276–3281

Quandt K, Frech K, Karas H, Wingender E, Werner T (1995) MatInd and MatInspector—new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res 23, 4878–4884

Reese MG, Eeckman FH (1995) Novel neural network algorithms for improved eukaryotic promoter site recognition. The Seventh International Genome Sequencing and Analysis Conference. Abstract. (http://www.fruitfly.org/seq tools/promoter.html).

Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, et al. (2000) Comparative genomics of the eukaryotes. Science 287, 2204–2215

Schmid SR, Linder P (1992) D-E-A-D protein family of putative RNA helicases. Mol Microbiol 6, 283–291

Schmucker D, Vorbrüggen G, Yeghiayan P, Fan HQ, Jäckle H, Gaul U (2000) The *Drosophila* gene *abstrakt*, required for visual system development, encodes a putative RNA helicase of the DEAD box protein family. Mech Dev 91, 189–196

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C et al. (2000) PipMaker—a web server for aligning two genomic DNA sequences. Genome Res 10, 577–586

Shapiro M, Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15, 7155–7174

Sinensky M (2000) Recent advances in the study of prenylated proteins. Biochim Biophys Acta 1484, 93–106

Weber JS, Jang W, Simin K, Lu W, Yu J, Meisler MH (1998) High-resolution genetic, physical and transcript map of the *mnd2* region of mouse chromosome 6. Genomics 54, 107–115