Abbreviations section). The BLAST 2 Sequences programme (Tatusova and Madden, 1999) at the NCBI (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html) was used for pairwise sequence alignments. For sequence similarity comparisons in Table 1, multiple sequence alignment was performed with the Clustal W programme (Thompson et al., 1994). The percentage of identical and similar amino acids in Table 1 was obtained using the statistics-report function of the GeneDoc programme (Nicholas et al., 1997). Pairs of similar amino acid residues were defined as those that yield a greater than zero score in the BLOSUM45 scoring matrix (similarity groups: DE, NH, ST, QKR, FYW, LIVM).

Nucleotide sequence accession numbers and the new gene symbol

Sequence data from this article have been deposited with the GenBank database (Benson et al., 2002) (http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html) under accession numbers: AF368232 (U3-55k cDNA), AF368233 (PARP-3 cDNA including alternative exon 1a), AF368234 (*U3-55k – PARP-3* promoter region), AY046316 (PARP-3 splice variant – exon 1b), and AY046317 (PARP-3 splice variant – exon 1c). The International Committee on Standardized Genetic Nomenclature for Mice (http://www.informatics.jax.org/mgihome/nomen/) approved the symbol *Adprt3* for the *PARP-3* gene described in this paper.

Results

cDNA sequences and promoter region of PARP-3 and U3-55k genes

We cloned and sequenced the cDNA of the mouse *PARP-3* (*Adprt3*) gene. The longest open reading frame encodes the protein of 528 amino acid residues and the calculated molecular weight of 59.5 kilodaltons (kDa). The predicted mouse protein sequence is 80% identical to the human PARP-3, the only orthologue characterized so far (Johansson, 1999). We determined sequence similarity between the homologous regions of the human and mouse PARP-1, PARP-2 and PARP-3 proteins (Table 1). These sequence comparisons showed that (i) PARP-1 and PARP-2 proteins are more closely related to each other (41%–42% identity) than to the PARP-3 proteins (29%–33% identity), and (ii) both PARP-3 proteins show a slightly higher similarity to the PARP-2 proteins (33% identity) than to the PARP-1 proteins (29%–31% identity) (Table 1). A different degree of sequence conservation is observed between the pairs of orthologous proteins (human vs. mouse). Thus, human and mouse PARP-1 proteins are 94% identical, PARP-2 proteins from the two organisms show 89% identity, whereas PARP-3 proteins are the least conserved orthologues with 81% identity in the regions compared (see Table 1).

A BLAST homology search of the human sequences in the databases at the NCBI (nr, human genome), using the human PARP-3 cDNA sequence (Johansson, 1999) as query, revealed the human *PARP-3* (*ADPRTL3*) genomic locus (LocusLink LocusID: 10039; Pruitt and Maglott, 2001) within an 81.5-kb PAC clone from the 3p21.1-9 chromosomal region (gb:AC006255, *Homo sapiens* 3p21.1-9 PAC RPCI5-1087L12) and within the *Homo sapiens* chromosome 3 working draft sequence segment (ref:NT_005986) (Lander et al., 2001; Pruitt and Maglott, 2001). Interestingly, the gene encoding human U3 small nucleolar ribonucleoprotein complex-associated 55-kilodalton protein (hU3-55k) (Pluk et al., 1998) is located close to *PARP-3* in a head-to-head orientation (gb:AC006255, ref:NT_005986, LocusLink LocusID: 9136) (Pruitt and Maglott, 2001).

Table 1. Comparison of sequence similarity between the human and mouse PARP-1, PARP-2 and PARP-3 proteins

	HsPARP-1		**MmPARP-1**		**HsPARP-2**		**MmPARP-2**		**HsPARP-3**		**MmPARP-3**	
	%		%		%		%		%		%	
	i	i+s	i	i+s	i	i+s	i	i+s	i	i+s	i	i+s
HsPARP-1	100	(100)	94	(98)	42	(62)	41	(61)	30	(48)	31	(48)
MmPARP-1	94	(98)	100	(100)	41	(61)	41	(60)	29	(48)	29	(48)
HsPARP-2	42	(62)	41	(61)	100	(100)	89	(95)	33	(50)	33	(50)
MmPARP-2	41	(61)	41	(60)	89	(95)	100	(100)	33	(51)	33	(50)
HsPARP-3	30	(48)	29	(48)	33	(50)	33	(51)	100	(100)	81	(89)
MmPARP-3	31	(48)	29	(48)	33	(50)	33	(50)	81	(89)	100	(100)

Note. For each pair of the six PARP protein sequences two numbers characterizing their similarity are shown: i, percentage of identical amino acids; i+s, percentage of identical and similar amino acids (smaller numbers in parentheses). The numbers are based on the multiple sequence alignment - see Material and Methods for the details. The species names are abbreviated as follows: Hs, *Homo sapiens*; Mm, *Mus musculus*. The database accession numbers, amino acid sequence intervals used for the multiple sequence alignment and similarity determination (aa:), and references are as follows: **HsPARP-1** [gb:M32721, aa: 520-1014, (Cherney et al., 1987)]; **MmPARP-1** [emb:X14206, aa: 519-1013, (Huppi et al., 1989)]; **HsPARP-2** [emb:AJ236912, aa: 68-570, (Amé et al., 1999)]; **MmPARP-2** [emb:AJ007780, aa: 61-559, (Amé et al., 1999)]; **HsPARP-3** [gb:AF083068, aa: 35-533, (Johansson, 1999)]; **MmPARP-3** [gb:AF368233 - this work, aa: 33-528].

Consequently, the two genes are transcribed in opposite directions from an approximately 1.5-kb putative bi-directional promoter region. To investigate the organization of the two genes in the mouse genome we cloned and sequenced the cDNA encoding the mouse orthologue of the human U3-55k protein, and the mouse genomic fragment spanning the promoter region and divergently oriented 5'-end exons of the mouse *PARP-3* and *U3-55k* genes (Fig. 1). Both mouse and human U3-55k proteins consist of 475 amino acid residues, are highly conserved throughout their sequence lengths (92% identity), and have similar calculated molecular weights of 52.1 kDa and 51.8 kDa, respectively. Three kinds of protein sequence motifs were identified by Pluk et al. (1998) in the human U3-55k sequence: (i) the N-terminal putative bi-partite nuclear localization signal, (ii) the glutamic acid-rich region near the N terminus, and most significantly, (iii) five WD repeats spanning the major central part of the protein. An additional WD repeat was detected recently by two groups based on further sequence analysis and its comparison with the *Saccharomyces cerevisiae* and *Xenopus laevis* proteins (Lukowiak et al., 2000; Venema et al., 2000). All the sequence motifs are conserved in the mouse U3-55k protein at exactly the same positions as they are in the human sequence. The putative bi-partite nuclear localization signal is located between the residues 8–40, and the glutamic acid-rich stretch between the residues 64–73. The WD-repeat number and arrangement in the mouse U3-55k sequence was analysed using the Protein Sequence Analysis server at the BioMolecular Engineering Research Center, Boston University (http://bmerc-www.bu.edu/psa/) (Stultz et al., 1993; White et al., 1994; Smith et al., 1999; Yu et al., 2000). Six WD repeats were predicted in the mouse U3-55k with the following positions of the first and the last amino acid residue of the conserved WD-repeat core sequence (the so-called GH-WD-core; Smith et al., 1999): 144–174, 197–227, 239–269, 281–311, 322–351, and 374–404.

Alignment of the human PARP-3 transcript (Johansson, 1999) and GenBank EST sequences from the 5' end of the gene with the human genomic sequence of the promoter region revealed two exons located upstream of the ATG initiation codon of *PARP-3* (Fig. 1). In contrast, using a combination of 5'-RACE and RT-PCR analyses we identified three alternative 5'-end exons (designated 1a, 1b and 1c) in the corresponding genomic region of the mouse (Fig. 1). These results are in agreement with mouse EST sequences from the GenBank database (e.g. ESTs dbj:BB591382, gb:BG922859, gb:AW318917, gb:BG964545; Marra et al., 1999; Strausberg et al., 1999; Kawai et al., 2001). We note that each of the three mouse alternative exons contains an in-frame stop codon upstream of the putative ATG initiation codon, indicating that these exons are non-coding (Fig. 1).

Expression of PARP-3, U3-55k, PARP-2 and PARP-1 genes in mouse tissues

The expression patterns of *PARP-3*, *U3-55k*, *PARP-2*, and *PARP-1* genes were determined using a Northern blot with mRNA from various adult mouse tissues and organs (Fig. 2). The highest expression of *PARP-3* was detected in the skeletal muscle mRNA, as estimated from shorter autoradiogram exposures (Fig. 2 and data not shown). High to moderate levels were found in the lung, liver, kidney, ovary, spleen and heart, while thymus, small intestine and colon contained only low levels of the *PARP-3* transcripts. Notably, *PARP-3* expression was barely detectable in mRNA prepared from the brain and testis. In some tissues, e.g. in spleen, lung and small intestine, two PARP-3 mRNA species of approximately 2.7 kb and 2.4 kb were detected. Using the 3'-RACE cloning of the mouse PARP-3 cDNA (data not shown), as well as the alignment of the 3'-end untranslated region of the mouse PARP-3 cDNA (nt 1754 to 2633 in gb:AF368233) with the mouse EST sequences (e.g. ESTs gb:AA267081, dbj:BB218739, dbj:BB223095; Marra et al., 1999; Kawai et al., 2001), we localized an alternative polyadenylation site (nt 2358 in gb:AF368233) of the mouse PARP-3 transcript. The shorter mRNA species may represent transcripts terminated at this site. The estimated lengths of the two mRNA species are in good agreement with the cloned cDNA sequence (2633 bp) and its alternatively terminated variant (2358 bp).

Compared to *PARP-3*, the other three genes are expressed ubiquitously with less variability of mRNA levels among the individual tissues (Fig. 2). Particularly, U3-55k and PARP-1 transcripts are distributed in a housekeeping gene-like manner. U3-55k mRNA migrates as a single band of approximately 1.6 kb, corresponding well to the size of the cloned cDNA (1513 bp). In addition to the major mRNA species of 2.0 kb, the PARP-2 probe detected a second less prominent 2.5-kb mRNA, in agreement with the previously published expression data (Amé et al., 1999; Berghammer et al., 1999). The estimated length of the PARP-1 transcript of approximately 3.8 kb corresponds well to the previously reported size (Huppi et al., 1989; Ogura et al., 1990; Wang et al., 1995). Also, higher levels of PARP-1 mRNA in the spleen, testis and thymus (Fig. 2) are in agreement with the data of Ogura et al. (1990), Wang et al. (1995) and Schreiber et al. (2002). Overall, *PARP-3* exhibits much more variability of mRNA levels and some degree of tissue-specific expression, compared to *U3-55k*, *PARP-2*, and *PARP-1* genes.

Discussion

Mouse *PARP-3* and *U3-55k* genes are located close to each other in the head-to-head orientation and are linked by an approximately 1.5-kb putative bi-directional promoter region. Comparison with the human

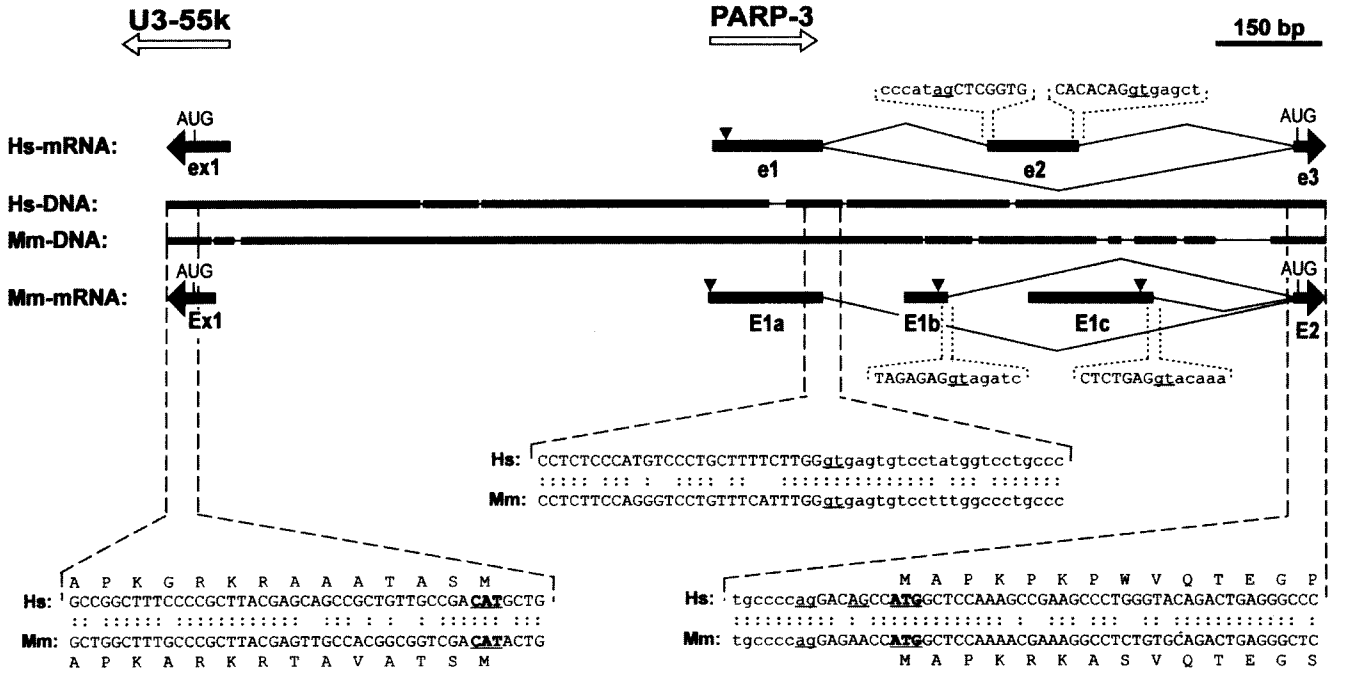

Fig. 1. Schematic comparison of the human and mouse promoter regions of *PARP-3* and *U3-55k* genes. The two parallel thick horizontal lines represent the alignment of the human (Hs-DNA; nt 52161 to 50592 [complementary strand] in gb:AC006255) and mouse (Mm-DNA; nt 1 to 1454 in gb:AF368234 – this work) genomic DNA regions encompassing the 5' ends of *PARP-3* and *U3-55k* genes. (The short interruptions of these lines filled with the thin line represent small gaps introduced into the genomic sequences in order to optimize the alignment.) The open arrows on the very top of the figure indicate the direction of transcription of the two genes. The 5'-end exons of the human and mouse U3-55k and PARP-3 transcripts are depicted as black boxes and black arrows (5'-end portions of AUG translation initiation codon-containing exons). (For simplicity, the alignment gaps are not indicated in the exon boxes.) Exon symbols: ex1, human U3-55k; Ex1, mouse U3-55k; e1, e2 and e3, human PARP-3; E1a, E1b, E1c and E2, mouse PARP-3. The human U3-55k exon ex1 is according to the reported cDNA sequence (emb:AJ001340; Pluk et al., 1998) and EST sequences (e.g. EST gb:BG825193; Strausberg et al., 1999), the mouse U3-55k exon Ex1 is according to gb:AF368232 (this work). Angled lines indicate the alternative splicing patterns of the human and mouse PARP-3 exons. The human PARP-3 splice variant e1-e2-e3 is according to EST sequences (gb:BG818516, gb:BG702085; Strausberg et al., 1999), while the splice variant e1-e3 is present in the published full-length cDNA sequence (gb:AF083068; Johansson, 1999) and several EST sequences (e.g. ESTs gb:BI554046, gb:BG913289; Strausberg et al., 1999). The mouse PARP-3 exons E1a, E1b, and E1c are alternatively spliced onto exon E2 (gb:AF368233, gb:AY046316, and gb:AY046317 – this work). We note that the splice donor site located downstream of exon E1b appears to be leaky, as a portion of exon E1b transcripts extend through this site into exon E1c and are spliced at the donor site downstream of E1c (data not shown). The vertical arrowheads above the PARP-3 exons point at the positions of the stop codons that are in frame with the downstream AUG initiation codons according to the published human and mouse cDNA sequences (Johansson, 1999, and this work, respectively). The nucleotide sequences of the exon-intron boundaries of human e2 and mouse E1b and E1c exons are shown above and below these exons, respectively, enclosed in dotted-line braces. The upper-case nucleotide symbols indicate the exon sequences, and the lower-case letters stand for the intron sequences. The conserved dinucleotides of the splice donor (gt) and splice acceptor (ag) sites are underlined. The lower part of the figure shows in a blown up detail three human (Hs) versus mouse (Mm) aligned sequence regions (enclosed in dashed line braces). Colons between the two sequences indicate identical nucleotides. The nucleotide sequences are written in the conventional 5'-to-3' left-to-right direction. *Bottom left.* Alignment of the 5' ends of the coding regions of U3-55k. Because of the leftward orientation of U3-55k transcription in this scheme, the two nucleotide sequences are given as mRNA-complementary strands and the amino acid sequences deduced from them are written, accordingly, from the right to the left. The complements of the translation initiation codons are in boldface and underlined. *Lower middle.* Conserved splice donor site of the human e1 and mouse E1a exons. The exon-intron boundary is marked as above. *Bottom right.* Alignment of the 5' ends of the human e3 and mouse E2 exons. The intron-exon boundary is marked as above. The ATG translation initiation codons are in boldface and underlined. The deduced human and mouse amino acid sequences are depicted above and below the DNA sequences, respectively. Note that there is an alternative splice acceptor site in the human exon e3, located four nucleotides upstream of the translation initiation ATG (underlined upper-case dinucleotide AG), which is used in some of the splice variant e1-e3 EST sequences (e.g. EST gb:BG913289; Strausberg et al., 1999). This alternative splice acceptor site does not appear to be conserved in the mouse sequence.