The mouse ADP-ribosylation factor-like 4 gene: two separate promoters direct specific transcription in tissues and testicular germ cell

Stephan JACOBS*, Annette SCHÜRMANN*, Walter BECKER*, Tobias M. BÖCKERS†, Neal G. COPELAND‡, Nancy A. JENKINS‡ and Hans-Georg JOOST*†

*Institut für Pharmakologie und Toxikologie, Medizinische Fakultät der Rheinisch-Westfälischen Technischen Hochschule Aachen, Wendlingweg 2, D-52057 Aachen, Germany, †Institut für Anatomie, Westfälische Wilhelms-Universität Münster, Vesaliusweg 2-4, D-48149 Münster, Germany, and ‡Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702-1201, U.S.A.

ADP-ribosylation factor protein 4 (ARL4) is a Ras-related GTPase that has been cloned from the 3T3-L1 preadipocyte cell line as an adipocyte-specific cDNA [Schürmann, Breiner, Becker, Huppertz, Kainulainen, Kentrup and Joost (1994) J. Biol. Chem. 269, 15683–15688]. The ARL4 gene maps to the proximal region of mouse chromosome 12 linked to Lamb1-1, Hlfhbf1 and Sos2. Compared with all other known genes of Ras-related GTPases, the genomic organization of ARL4 is unusual in that its entire coding region, the 3′ untranslated region (UTR) and most of the 5′ UTR are located on a single exon. This structure suggests that ARL4 has evolved by retroposition of an Arf (ADP-ribosylation factor) or Arf-like gene. Isolation of the 5′ UTR by rapid amplification of cDNA ends (RACE)-PCR revealed heterogeneous transcription initiation sites in alternative exons 1. Both 5′-flanking regions exhibited promoter activity when expressed in COS-7 cells, indicating that the expression of Arl4 is directed by two separate promoters. mRNA transcribed under the control of the downstream promoter was isolated by RACE-PCR from all investigated tissues. In contrast, the upstream promoter seems to drive specifically the expression of Arl4 in adult testis. Hybridization of rat testis in situ indicated that Arl4 is expressed in germ cells of puberal and adult testis, but not in prepuberal testis, suggesting that Arl4 is involved in sperm production.

INTRODUCTION

ADP-ribosylation factors (ARFs) represent a subfamily of the Ras-related GTPases comprising six homologous genes (ARF1–6 [1,2]) that are presumably involved in vesicle trafficking [3–5] and signal transduction, e.g. phospholipase D activation [6–8]. In addition to the six ARF isoforms, several ARF-like cDNA species (Arf1–5, Arp) have been cloned and assigned to the extended ARF family [9–14]. Arl4 was found in a PCR-cloning approach designed to identify genes that are differentially expressed in the preadipocyte cell line 3T3-L1 [12]; it was abundant in differentiated 3T3-L1 cells but was undetectable in undifferentiated cells. Thus the function of the GTPase seemed to be related to adipocyte metabolism and, because Arl4 mRNA was also found in muscle and brain, to other cellular pathways present in the differentiated cell.

The expression of most ARF isoforms (ARF1–3) seems to be constitutive [1,2,5]. Striking exceptions are the developmentally regulated expression of Arf2 [16] and the differentiation-dependent expression of Arl4 [12]. The basis of this unusual regulation of expression is of particular interest, and might help to identify the physiological function of ARL4. Here we describe the unique organization of the Arl4 gene, which contains two separate promoter regions directing its tissue-specific expression. The downstream promoter seems to be active in most tissues, whereas the upstream promoter specifically drives a large expression of Arl4 in testicular germ cells, suggesting that the GTPase is involved in spermatid maturation.

MATERIALS AND METHODS

Interspecific mouse backcross mapping

Interspecific backcross progeny were generated by mating (C57BL/6J × Mus spretus) F1 females and C57BL/6J males as described [17]. A total of 205 N1 mice were used to map the ARL4 locus. DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer and hybridization were performed essentially as described [18]. All blots were prepared with Hybond-N+ nylon membrane (Amersham). The probe, a mouse genomic-DNA fragment of the 5′ flanking region (nt –244 to –780; see Figure 2), was labelled with [α-32P]dCTP by using a nick translation labelling kit (Boehringer Mannheim); washing was done to a final stringency of 1.0 × SSCP/0.1 SDS at 65 °C. A fragment of 7.6 kb was detected in SacI-digested C57BL/6J DNA and a fragment of 6.2 kb was detected in SacI-digested M. spretus DNA. The presence or absence of the 6.2 kb SacI M. spretus-specific fragment was followed in backcross mice.

A description of the probes and restriction fragment length polymorphisms for the loci linked to ARL4, including Lamb1-1, Hlfhbf1 and Sos2, has been reported previously [19]. Recombination distances were calculated using Map Manager, version 2.6.5. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

Library screening and DNA sequencing

Genomic clones of ARL4 were isolated by screening a 129 SvJ mouse genomic library (approx. 3 × 107 plaques) with a PCR-
Figure 1  *Arl4* maps in the proximal region of mouse chromosome 12

*Arl4* was placed on mouse chromosome 12 by interspecific backcross analysis. The segregation patterns of *Arl4* and flanking genes in 143 backcross animals that were typed for all loci are shown at the top of the figure. For individual pairs of loci, more than 143 animals were typed (see the text). Each column represents the chromosome identified in the backcross progeny that was inherited from the intercross (*C57BL/6J × M. spreitus*) F2 parent. Filled boxes represent the presence of a *C57BL/6J* allele, open boxes represent the presence of an *M. spreitus* allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. A partial chromosome 12 linkage map showing the location of *Arl4* in relation to linked genes is shown at the bottom of the figure. Recombination distances (centimorgans) between loci are given at the left, and the positions of loci in human chromosomes, where known, are shown at the right. Chromosome 12 linkage map showing the location of *Arl4* in relation to linked genes is shown at the bottom of the figure. Recombination distances (centimorgans) between loci are given at the left, and the positions of loci in human chromosomes, where known, are shown at the right. References for the human map positions of loci cited here can be obtained from the Genome Data Base (GDB), a computerized database of human linkage information maintained by The William H. Welch Medical Library of The Johns Hopkins University (Baltimore, MD, U.S.A.).

5′ RACE was performed with a kit from Gibco BRL (Eggenstein, Germany) in accordance with the instructions of the manufacturer. Primers for cDNA synthesis and amplification were derived from the *Arl4* cDNA: cDNA synthesis, 5′-CCA TTT CCA GCT GCA AAT G-3′; reverse primer for first amplification, 5′-TCT CAA ACT GAA CGC GT; reverse primer for second amplification, 5′-CGC GTT CTT ATC TGC. PCR conditions: 35 cycles at 1.5 mM MgCl2 and 0.2 mM dNTP; hot start, 1 min at 94 °C, 1 min at 55 °C, 1 min at 72 °C, 7 min at 72 °C in the last cycle. DNA fragments were isolated and subcloned into pUC18 with the Sureclone® kit (Pharmacia, Freiburg, Germany).

Transfection of COS-7 cells and luciferase assay

DNA fragments derived from the 5′-flanking regions of *Arl4* were fused to luciferase cDNA in the vector pGL3 (Promega, Madison, WI, U.S.A.) and the constructs were transfected into COS-7 cells with the aid of the cholesterol derivative DAC-30 (Eurogentec, Seraing, Belgium). Cells were lysed 30 h after transfection; luciferase activity was assayed with a kit from Promega in accordance with the instruction of the manufacturer. Transfection efficiency was determined by co-transfection of a plasmid harbouring *lacZ*, and data were normalized to β-galactosidase activity.

RNA preparation and rapid amplification of cDNA ends (RACE)-PCR

Mouse or rat tissues were homogenized with a Polytron homogenizer in 4 M guanidine thiocyanate. Murine 3T3-L1 fibroblasts (American Type Culture Collection, Rockville, MD, U.S.A.) were differentiated as described [20], frozen in liquid nitrogen and lysed with 4 M guanidine thiocyanate supplemented with 7% 2-mercaptoethanol. The lysates were layered on a CsCl cushion (5.88 M) and centrifuged at 138,000 g (rotor SW40) for 22 h at 20 °C. Pelleted RNA was dissolved in 300 μl of sodium acetate/Tris buffer (0.1 M, pH 9), and was neutralized by the addition of 50 μl of 2 M potassium acetate, pH 5.5.

Samples (10 μg) of total RNA were separated by electrophoresis on 1% (w/v) agarose gels containing 3% (v/v) formaldehyde and were transferred to nylon membranes (Hybond; Amersham-Buchler, Braunschweig, Germany). Before transfer, gels were stained with ethidium bromide to ascertain that equal amounts of total RNA had been separated. Probes were generated with the Klenow fragment of DNA polymerase I and [α-32P]dCTP by random oligonucleotide priming [21]. The nylon membranes were hybridized at 42 °C, and blots were washed twice with 0.8 × SSC containing 0.1% SDS.

Hybridization in situ

Rat tissues were removed and frozen immediately in isopentane on solid CO2 at −40 °C. Tissues were cut in 18 μm sections on a cryostat, mounted on Superfrost plus slides (Fisher Scientific, Düsseldorf, Germany) and stored at −70 °C. Oligonucleotides corresponding to bp 2644-2677 of the *Arl4* sequence (see Figure 1) (anti-sense oligonucleotide: 5′-CCA AGG AGT CGA TGA CTC CCA-3′; reverse primer for first amplification, 5′-TCT CAA ACT GAA CGC GT; reverse primer for second amplification, 5′-CGC GTT CTT ATC TGC. PCR conditions: 35 cycles at 1.5 mM MgCl2 and 0.2 mM dNTP; hot start, 1 min at 94 °C, 1 min at 55 °C, 1 min at 72 °C, 7 min at 72 °C in the last cycle. DNA fragments were isolated and subcloned into pUC18 with the Sureclone® kit (Pharmacia, Freiburg, Germany).

S. Jacobs and others
RESULTS
Genomic localization of the mouse Arl4 gene

The mouse chromosomal location of Arl4 was determined by interspecific backcross analysis with progeny derived from matings of [(C57BL/6J × Mus spretus)F1 × C57BL/6J] mice. The 6.2 kb SsoI Mus spretus restriction fragment length polymorphism (see the Materials and methods section) was used to follow the segregation of the Arl4 locus in backcrossed mice. The mapping results indicated that Arl4 is located in the proximal region of mouse chromosome 12 linked to Lamb1-1, Hfhbf1, and Sos2 (Figure 1). The ratios of the total numbers of mice exhibiting recombinant chromosomes to the total numbers of mice analysed for each pair of loci and the most likely gene order are:

<table>
<thead>
<tr>
<th>Locus</th>
<th>Recombinants</th>
<th>Total Mice</th>
<th>Recombinant Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lamb1-1</td>
<td>7.0 ± 1.9</td>
<td>172</td>
<td>1.4 ± 1.0</td>
</tr>
<tr>
<td>Hfhbf1</td>
<td>9.9 ± 2.4</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>Sos2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The proximal region of mouse chromosome 12 shares regions of homology with human chromosomes 7q and 14q (summarized in Figure 1), suggesting that the human homologue of Arl4 also maps to 7q or 14q.

Sequence and organization of the mouse Arl4 gene

A total of five genomic clones of Arl4 were isolated and were characterized with restriction enzymes and probes derived from the 5'to 3'translated region (UTR) of the rat Arl4 cDNA. The

---

Figure 2  Nucleotide sequence of the mouse Arl4 gene

Capital letters denote sequences found in cDNA clones; lower-case letters denote intron sequences. The deduced amino acid sequence of the open reading frame is given in single-letter code above the nucleotide sequence. The starts of exon 1A and exon 1B are deduced from the longest cDNA clones isolated with the RACE procedure (see Figure 3). The polyadenylation signal is highlighted by underlined bold italics; CAAT boxes and Sp1 consensus binding sites (GGGCCGG or CCCGCC) are shown in capital bold characters and are underlined. The EMBL, GenBank accession number for Arl4 is Y12577.
Figure 3 Alternative splicing and transcription initiation sites of Arl4 mRNA

(A) Schematic representation of the exon–intron organization of Arl4 and of its alternative splicing. Boxes represent sequences found in cDNA; the coding region is drawn as a filled box. Alternative splicing was demonstrated by the isolation of 79 cDNA clones that contained either exon 1A (testis, fat) or 1B (3T3-L1 cells, fat, spleen, brain, kidney, muscle, testis). (B) Determination of the transcription initiation sites in different tissues by RACE cloning. The numbers of clones isolated are given below the positions (arrows) of the 5' end of each RACE clone.

clone with the longest 5'-flanking sequence was subcloned and sequenced (Figure 2). A comparison of the genomic sequence with the cDNA sequences of Arl4 (rat and human) indicated that the complete coding region, the 3' UTR including a polyadenylation signal (nt 2093–2098) and a considerable portion of the 5' UTR are located on a single exon. This genomic organization is in striking contrast with that of related genes such as Arf1, Ki-Ras and Rab1 (see Figure 6). The sequence of the coding region (nt 1188–1787) is 97.5% identical with that of the rat (15 nucleotide exchanges) and 100% identical with a partial mouse sequence (connected from the expressed sequence tags MMA03086, MM123891, MMAA9306). It contains four mismatches (99.3% identity) with a database entry of mouse Arl4 cDNA (accession no. U76546).

The genomic sequence of Arl4 predicts that Southern blots of DNA digested with the restriction enzymes BamHI, EcoR1 and NcoI should exhibit single bands hybridizing with the Arl4-specific probe. Indeed, these enzymes generated single bands (5.5, 10 and 12 kb respectively) under low-stringency hybridization and washing conditions (results not shown). This finding is consistent with the conclusion that no highly similar isoforms or pseudogenes of Arl4 exist.

Tissue-specific alternative splicing of the 5' UTR (exon 1) and alternative promoter use in the Arl4 gene

To determine the transcription initiation site of Arl4, RACE-PCR was performed with cDNA from 3T3-L1 cells and various mouse and rat tissues. As anticipated, the sequences of the RACE products corresponded to portions of the genomic sequence (Figure 2). Surprisingly, however, they indicated the presence of two alternatively spliced exons (1A and 1B) encoding the 5' end of the mRNA (Figure 3). All 79 RACE clones analysed contained either exon 1A or 1B; no clone containing both exons was isolated. RACE clones obtained with testis mRNA predominately (16 out of 21) started with exon 1A. Of 13 clones isolated from fat tissue, only two started with exon 1A. All
Separate promoters direct transcription of the Arl4 gene

Figure 5  Localization of Arl4 mRNA in seminiferous tubules of rat testis as detected by hybridization in situ with a labelled oligonucleotide

(A) Testis from adult rat hybridized with anti-sense oligonucleotide; dark-field photomicrograph at 2-fold magnification. (B) Control hybridization with the sense oligonucleotide. (C) Localization of Arl4 mRNA in seminiferous tubules as detected by anti-sense oligonucleotide hybridization at a magnification of ×160. Abbreviations: L, lumen; LS, late spermatids; YS, young spermatids; SG, spermatogonia; IT, interstitial tissue. (D–I) Testis from different developmental stages (D, E, prepuberal; F, G, puberal; H, I, adult rat) hybridized with anti-sense oligonucleotide; magnification ×100. (E), (G) and (I) are darkfield photomicrographs.

RACE clones obtained from 3T3-L1 cells, brain, spleen, kidney and muscle started with exon 1B (Figure 3B). The transcription initiation sites in exon 1B were variable and seemed to cluster around two sites.

To add support to the conclusion that transcripts comprising exon 1A are predominant in testis, Northern blots of testis mRNA were hybridized with probes derived from either exon 1A or 1B. As illustrated in Figures 4(D) and 4(E), a 1.2 kb transcript was detected in puberal and adult rat testis with the exon 1A-specific probe, but not with an exon 1B-specific probe. It should be noted that both probes seemed to hybridize poorly with the Arl4 mRNA because a probe corresponding to the full-length cDNA (Figure 4C) gave a considerably stronger signal. Figure 4(A) also illustrates the tissue distribution of the Arl4 mRNA as assessed with the full-length cDNA probe. mRNA was detected predominantly in testis, and also in spleen and intestine; smaller amounts of Arl4 mRNA were found in brain, heart, total fat, liver, lung and thymus. As we have shown previously [12], Arl4 was expressed in differentiated 3T3-L1 adipocytes but was not detectable in undifferentiated fibroblasts (Figure 4B).

A search of the expressed sequence tag (EST) databases identified two mouse and three human EST entries of Arl4 cDNA with a 5′ UTR. Both mouse ESTs started with exon 1B. However, of the three human EST entries, one entry (HS84754)
started with a different 5' UTR. Thus we tentatively assume that the transcription of the human ARL4 gene also generates two isoforms with alternative exons 1.

Promoter activity of the 5'-flanking regions of exons 1A and 1B of Arl4
The alternative splicing of the 5' UTR suggested that the expression of Arl4 is controlled by two separate promoters. Indeed, both 5'-flanking regions of the exons 1A and 1B exhibited promoter activity when expressed in COS-7 cells as luciferase fusion constructs (Table 1). We had expected that COS-7 cells lacked the components required to produce a large activation of the testis-specific promoter A. Accordingly, the upstream promoter (A) seemed less active than the downstream promoter (B). Furthermore Table 1 demonstrates that the activity of promoter B is not dependent on the presence of promoter A, because B exhibited the same activity as a construct comprising both promoters.

Both promoters contain G/C-rich regions (73 % in promoter A, nt –403 to 1; 74 % in promoter B, nt 97 to 515) and elements characteristic of promoter regions (Figure 2). Two CAAT boxes and four consensus sequences for the binding of Sp1 are present in promoter B. Promoter A lacks a CAAT box in the proximity of the transcription start but contains six Sp1 consensus binding sites, two motifs matching presumed testis-specific elements of the proacrosin gene [22], a CAMP response element (CRE)-like motif [23], and a palindromic motif flanked by CAAT boxes (Figure 2). These sequence characteristics are consistent with the conclusion that the 5'-flanking regions of both exons 1 might function independently as promoters.

Localization of Arl4 mRNA in rat testis
Northern blot analysis of mRNA from rat tissues indicated the highest levels of Arl4 mRNA to be in testis (Figure 4A). Thus, to test a possible involvement of ARL4 in spermatogenesis, we compared mRNA levels of Arl4 in testis from rats of different ages. As illustrated in Figure 4(C), a large expression of Arl4 was detected in testis from adult rats (8 and 16 weeks of age), but not in testis from prepuberal rats (2 weeks of age). Because this pattern seemed to parallel the time course of spermatogenesis, hybridization in situ of tissue slices from rat testis was performed with a 33-mer oligonucleotide corresponding to nt 2644–2677 of the Arl4 sequence shown in Figure 1. As is illustrated in Figure 5, the hybridization in situ revealed an intense labelling of all seminiferous tubules per section. The highest density of silver grains seemed to be associated with spermatogonia and/or young spermatids (Figure 5C); no Arl4 mRNA was detected in interstitial tissue. The lack of a specific signal in testis from prepuberal rats indicates that Arl4 expression depends on sperm production. Thus the location of the tracer and the time course of expression in testis are consistent with a specific expression of Arl4 mRNA in germ cells.

DISCUSSION
The organization of the Arl4 gene is unique among the known Ras-related GTPases in that its coding region, its 3' UTR and a considerable portion of the 5' UTR are located on a single exon (Figure 6). This unusual genomic organization suggests that the gene has evolved by retroposition and that its evolution is a relatively recent event. The dendrogram of the multiple sequence alignment of ARL4 with ARF, Ras and Rab1 suggests that the precursor of ARL4 is ARF or an ARF-like isoform. The coding regions of all known genes of Ras-related GTPases, including three mammalian ARF isoforms [15,16,24] and the Plasmodium falciparum ARF [25], are interrupted by several introns (Figure 6). Thus it is unlikely that the Arl4 gene has originated from these genes by duplication of genomic DNA, unless one assumes that the intronless structure was conserved over a very long period. Furthermore a database search failed to identify a yeast homologue of ARL4 (the closest relative in Saccharomyces cerevisiae is the ARL1 homologue SCZ36033 with 43.7 % identical amino acids), which is consistent with the conclusion that the evolution of ARL4 is a relatively recent event.

One of the few known retroposons encoding a functional protein is the testis-specific isoform of glucose-6-phosphate

Figure 6  Comparison of the genomic organization of Arl4 with that of other Ras-related GTPases
The dendrogram of a multiple alignment of amino acid sequences (left panel) was performed with the CLUSTAL program. Genomic organization of the GTPases (right panel) was obtained from references [24] (Arf8, Arf, [37] (Rab1), [38] (Ki-Ras) and [25] (Plasmodium falciparum Arf). Boxes represent exons, filled boxes denote coding regions. The sizes of exons and introns are given above and below the structures respectively. The main GTP-binding motifs PM1 (GKT or GKS), PM3 (DVGG or DTAG) and G2 (NKQD or NKCD) are indicated above the structures at their approximate positions within the coding region.
dehydrogenase-2 [26]. It has probably evolved, because the glucose-6-phosphate dehydrogenase-1 gene is X-chromosomal and cannot be transcribed in post-meiotic male germ cells. Thus we speculated that Arl4 originated in a similar way. However, the Southern blot analysis of mouse DNA gave no indication that a highly similar Arl4 isomorph exists.

The heterogeneous transcription start with alternative exons 1 indicates that two separate promoters induce the expression of the Arl4 gene. Many other genes are directed by separate promoters that control an independent, tissue-specific (e.g. of glucococinase [27]) or a developmentally regulated expression (e.g. of Pax6 [28]). The Arl4 gene is unique in that its two promoters seem to be separated by only approx. 100 bp, whereas in other genes the distance between dual promoters is usually much larger. On the basis of the preliminary analysis of the 5’-flanking regions with luciferase fusion constructs, we assume that the upstream promoter resides in the 5’-flanking region of exon 1A and that the second promoter is located in the 419 bp intron separating exon 1A and 1B. Both presumed promoter regions of the mouse Arl4 gene are TATA-less but contain elements characteristic of promoter regions. The presumed downstream promoter is G/C-rich, contains two CAAT boxes and three Sp1 consensus binding sites and one CRE-like motif. The presumed downstream promoter is G/C-rich, contains two CAAT boxes and three Sp1 consensus binding sites and one CRE-like motif. The expression of fat-specific genes [29]. Thus the specific expression of Arl4 in the differentiated 3T3-L1 cells might be induced by the transcription factor C/EBP. However, the possibility cannot be excluded that the differential expression of Arl4 mRNA reflects a repression of the gene in the 3T3-L1 fibroblasts rather than its induction in the adipocytes.

The upstream promoter contains a G/C-rich region with six Sp1 consensus binding sites and one CRE-like motif. The expression of many testis-specific transcripts is regulated by CRE-like motifs, namely angiotensin-converting enzyme [23,30], phosphoenolpyruvate carboxykinase [30], calserpin [31], proprotein convertase 4 [32], lactate dehydrogenase [33] and NO synthase [34]. The transcription factor responsible for the large expression of these genes in testis seems to be CREM-τ [23], a testis-specific isoform of the CRE-binding protein CRE modulator (CREM). Expression of this transcription factor is repressed in premeiotic germ cells and is induced by the follicle-stimulating hormone in postmeiotic spermatids [35,36]. Thus the lack of expression of Arl4 in prepuberal testis and its large expression in germ cells of adult rat testis is consistent with the presence of the CRE-like element in the presumed upstream promoter of Arl4. In the CRE motif, two other elements considered to be responsible for the testis-specific expression of the proacrosin gene [22] are present in the upstream promoter of Arl4. Thus the sequence characteristics of the upstream promoter of Arl4 are consistent with our finding that it is specifically activated in testicular germ cells.

We have compared our interspecific map of chromosome 12 with a composite mouse linkage map that reports the map location of many uncloned mouse mutations (provided by Mouse Genome Database, a computerized database maintained at The Jackson Laboratory, Bar Harbor, ME, U.S.A.). Arl4 mapped in a region of the composite map that lacks mouse mutations with a phenotype that might be expected for an alteration in this locus. On the basis of the present data we can only speculate on the cellular function of ARL4. However, because of its specific expression in testicular germ cells, it seems reasonable to assume that the GTPase is involved in sperm production.