

# Conservation of the 3'-untranslated region of the *Rab1a* gene in amniote vertebrates: exceptional structure in marsupials and possible role for posttranscriptional regulation

Niels Wedemeyer<sup>a,1</sup>, Thomas Schmitt-John<sup>a</sup>, Dirk Evers<sup>b</sup>, Cora Thiel<sup>a</sup>, Daniel Eberhard<sup>a</sup>, Harald Jockusch<sup>a,\*</sup>

<sup>a</sup>Developmental Biology and Molecular Pathology, W7, University of Bielefeld, D-33501 Bielefeld, Germany

<sup>b</sup>Faculty of Technology, Practical Computer Science, University of Bielefeld, Bielefeld, Germany

Received 12 April 2000; received in revised form 16 June 2000

Edited by Takashi Gojobori

**Abstract** The YPT1/RAB1 protein, a key regulator of the intracellular vesicle transport in eukaryotes, is highly conserved in function and amino acid sequence. Here we report that the most highly conserved nucleotide sequence of the *Rab1a* gene of amniote vertebrates corresponds to the 3'-untranslated region (3'-UTR) of the mRNA. Sequences of 27 species ranging from mammals to sauropsida are >91% identical in this region. Secondary structure prediction procedures applied to the 3'-UTR sequences between positions 750 and 984 and 1428 (mouse cDNA: Y00094), respectively, of the RAB1a mRNAs revealed families of alternative structures around nucleotide position 800 as recurrent features. The two hairpin loops are also predicted for marsupials, despite of their exceptional extension of the A-rich sequence in between. Yet, sequence conservation is much higher than required to conserve secondary structure. Implications for posttranscriptional regulation and protein binding are discussed. © 2000 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

**Key words:** *Rab1*; 3'-UTR; RNA folding; Sequence conservation; Eutheria; Marsupial; RNA-protein interaction

## 1. Introduction

In the last few years, the non-coding regions of eukaryotic genes were shown to contain essential regulatory information (reviewed in [5]). 3'-Untranslated sequence regions (3'-UTRs) of a number of transcripts are involved in mRNA stability and translational control (reviewed in [15]). In particular, several oncogenes contain specific AU-rich sequences in their 3'-UTRs which are determinants for mRNA degradation and are likely to be involved in the down-regulation of the translation rate ([30]; reviewed in [4]). Another example for the regulation of mRNA stability by specific 3'-UTR sequences is the iron response elements (IREs) in the 3'-UTR of the transferrin receptor (Tfr) mRNA. At a low cellular iron level, the IREs are bound by the *trans*-acting IRE binding protein

IRP1 which leads to a protection of the Tfr mRNA against degradation whereas at a high iron level the unbound Tfr mRNA is rapidly degraded [3]. Moreover, the localization of mRNAs in particular cytoplasmic compartments can be regulated by 3'-non-coding sequences as shown for bicoid and nanos mRNAs in the *Drosophila* egg [9,20] and for  $\beta$ -actin mRNA in myotubes [16]. 3'-UTR sequences may control not only the translation of their own mRNA but also the expression of other genes [27].

A general indication for an important functional role of the 3'-UTRs is the high evolutionary conservation in a number of mRNA species [13,19,29,38]. In nearly a third of the investigated genes, preferentially in housekeeping genes, the 3'-UTRs are more than 70% identical in vertebrate species, over stretches of 100–1450 nucleotides. This implies a strong selective pressure which cannot be explained by protein-RNA interactions only, because a specific binding of regulatory proteins would require less nucleotide conservation over shorter RNA sequences [6].

This work deals with a gene encoding a small GTP binding protein belonging to the RAS superfamily and the members of RAB family which are homologous to each other and highly conserved among eukaryotes (reviewed in [21]). The *Rab1a* gene of mouse and human, located on mouse proximal chromosome 11 and human chromosome 2p, respectively, came into the focus of our comparative analysis of these homologous genomic regions [17,28,35]. This gene encodes the small GTP binding protein RAB1 which is found in every eukaryote and is a key regulatory component for the transport of vesicles from the endoplasmic reticulum to the Golgi apparatus (reviewed in [1,25]). Studies on mutations of YPT1, the yeast counterpart of RAB1, showed that this protein is necessary for the budding of vesicles of the endoplasmic reticulum as well as for their transport to, and fusion with, the Golgi apparatus [33,37]. Conservation between mammalian RAB1 and YPT1 was previously shown at different levels. Haubruck et al. [11] used the yeast cDNA as a hybridization probe to isolate the corresponding mouse cDNA. They found 71% identity at the amino acid level. Functional equivalence was demonstrated by the complementation of yeast YPT1 mutations by mouse RAB1a protein [12].

Here we report an extreme sequence conservation of an exonic non-coding region of the *Rab1a* gene, the 3'-UTR, within amniote vertebrates. Secondary structure predictions and binding experiments imply a posttranscriptional cellular function involving protein-RNA interactions.

\*Corresponding author. Fax: (49)-521-106 5654.  
E-mail: h.jockusch@biologie.uni-bielefeld.de

<sup>1</sup> Present address: Institute of Radiobiology, Westfälische Wilhelms-Universität, Münster, Germany.

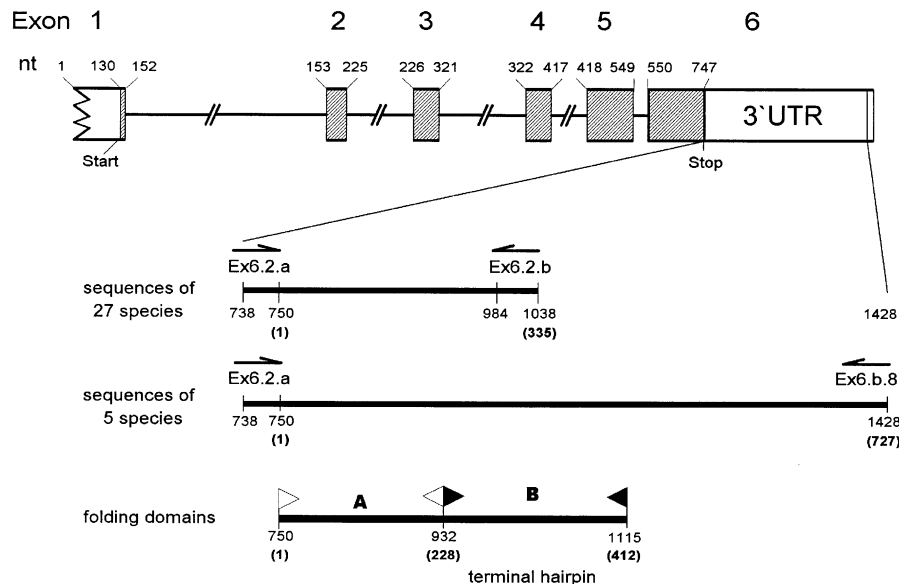


Fig. 1. Structure of the *Rab1a* gene of the mouse, nucleotide numbering system and primers used in this work. Exons and introns of *Rab1a* are shown with protein coding regions shaded. Nucleotide positions (nt) are numbered according to the mouse cDNA sequence of Wichmann et al. [36]. Positions of primers used, and sequences determined, in this work are indicated by half arrows and bold horizontal lines, respectively. Numbers in brackets refer to the nucleotide position in the alignment (see Fig. 2). Folding domains A and B are indicated.

## 2. Materials and methods

### 2.1. Origin of tissues and GenBank accession numbers

DNAs were prepared from tissues of the following vertebrates: human (*Homo sapiens*, nasal mucosa; AF170935), baboon (*Papio hamadryas*, muscle, Zoo Hellabrunn, Munich, Germany; AF170927), laboratory mouse (*Mus musculus laboratorius*, brain, liver, muscle; Y00094), laboratory rat (*Rattus norvegicus*, kidney; AF170942), horseshoe bat (*Rhinolophus rouxi*, kidney, Zool. Inst. University Munich; AF170928), European mole (*Talpa europaea*, kidney, Bielefeld; AF170937), polecat (*Putorius putorius*, kidney, Bielefeld, Germany; AF170939), serval (*Leptailurus serval*, skeletal muscle, Zoo Hellabrunn, Munich, Germany; AF170943), domestic cat (*Felis catus*, testis, veterinarian, Bielefeld, Germany; AF227734), porpoise (*Phocoena phocoena*, skeletal muscle, Sea Mammal Research Unit, Cambridge, UK; AF170940), fin whale (*Balaenoptera physalus*, stranded male, connective tissue from fluke, Schleswig-Holstein, Germany; AF170934), moose (*Alces alces*, neck muscle, Zoo Hellabrunn, Munich, Germany; AF170932), muntiac (*Muntiacus reevesi*, kidney, Behavioural Sciences, University of Bielefeld; AF170936), cattle (*Bos taurus domestica*, liver; AF227735), sheep (*Ovis aries*, kidney; AF170944), camel (*Camelus bactrianus*, testis, Zoo Hannover; AF170929), Indian elephant (*Elephas maximus*, placenta, Zoo Hagenbeck, Hamburg; AF170933), agile wallaby (*Macropus parma*, kidney, Behavioural Sciences, University of Bielefeld; AF170931), common brush-tailed possum (*Trichosurus vulpecula*, MPI Tübingen; AF227733), Tasmanian devil (*Sarcophilus harrisii*, MPI Tübingen; AF227736), quolls (*Dasyurus byrnei*, MPI Tübingen; AF233389), South American short-tailed gray opossum (*Monodelphis domestica*, MPI Tübingen; AF233388), chicken (*Gallus gallus gallus*, liver; AF170930), zebra finch (*Poephila guttata*, lung, behavioral sciences, Bielefeld; AF170946), ostrich (*Struthio camelus*, liver, ostrich farm via Zoology Department, Tel-Aviv, Israel; AF170938), Mississippi alligator (*Alligator mississippiensis*, kidney, alligator farm via Zoology Department, Tel-Aviv, Israel; AF170926), painted turtle (*Chrysemys scripta elegans*, lung, animal trade; AF170945). Wild animals with

'Bielefeld' given as the origin were either killed by car traffic or by the cat of H. Jockusch.

### 2.2. DNA preparation, amplification and sequencing

DNAs from vertebrate tissues were prepared by SDS phenol extraction or according to Laird et al. [18]. DNAs were amplified by polymerase chain reaction (PCR) using the primers Rab1Ex6.2a biotin 5'-GTACTACCTGCTAAACCGTAGGC-3' and Rab1Ex6.2b 5'-CTTTCCTGGCCTGCTGCTGTGTCC-3' or Rab1Ex6.b.8 5'-GAATGCAAAGCCTATTTCCC-3' (Fig. 1). PCR was performed using 100 ng total DNA, 50 pmol of each primer, 3 U Taq DNA polymerase (Promega, Madison, WI, USA), 200  $\mu$ mol of each nucleotide and buffers as recommended by the suppliers. Cycle conditions were one cycle at 94°C for 4 min and 35 cycles at 60°C (for Rab1Ex6.2a and Rab1Ex6.2b primers) or 56°C (for Rab1Ex6.2a and Rab1Ex6.b.8) for 1 min, 72°C for 90 s, and 90°C for 1 min. 10  $\mu$ l of the amplification product was controlled on an agarose gel and both strands were sequenced.

### 2.3. RNA secondary structure prediction

Secondary structure calculations were performed using Michael Zukers 'mfold 3.0' [22] at 37°C with default parameters and verified using 'RNAfold 1.2.1' [14]. The secondary structure polygonal plots were produced by 'RNA Movies 1.0' [8]. Furthermore the tool was used to compare alternative secondary structures. Covariation methods were not employed because of insufficient input data.

## 3. Results

### 3.1. Evidence for sequence conservation

Genomic PCR was originally applied to the *Rab1a* gene as it was considered a positional candidate for the neuromuscular disease gene *wobbler* (*wr*) of the mouse [35]. Comparative genomic PCR analysis on a variety of mammalian species was

Fig. 2. Comparison of the RAB1 3'-UTR sequences of 27 vertebrate species. For species 1–27, positions 750–985, for species 1, 3, 9, 11 and 12, positions 750–1428 relative to the mouse cDNA sequence (accession # Y00094) are shown, this means that nucleotide 1 corresponds to nucleotide 750 in the mouse cDNA. Sequence variations relative to the human sequence are indicated in black, those not determined by dots. The borders of folding region A are indicated by white, and those of region B by black arrow heads.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
1	human	TC	TG	CG	CT	CG	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC	CA	TC

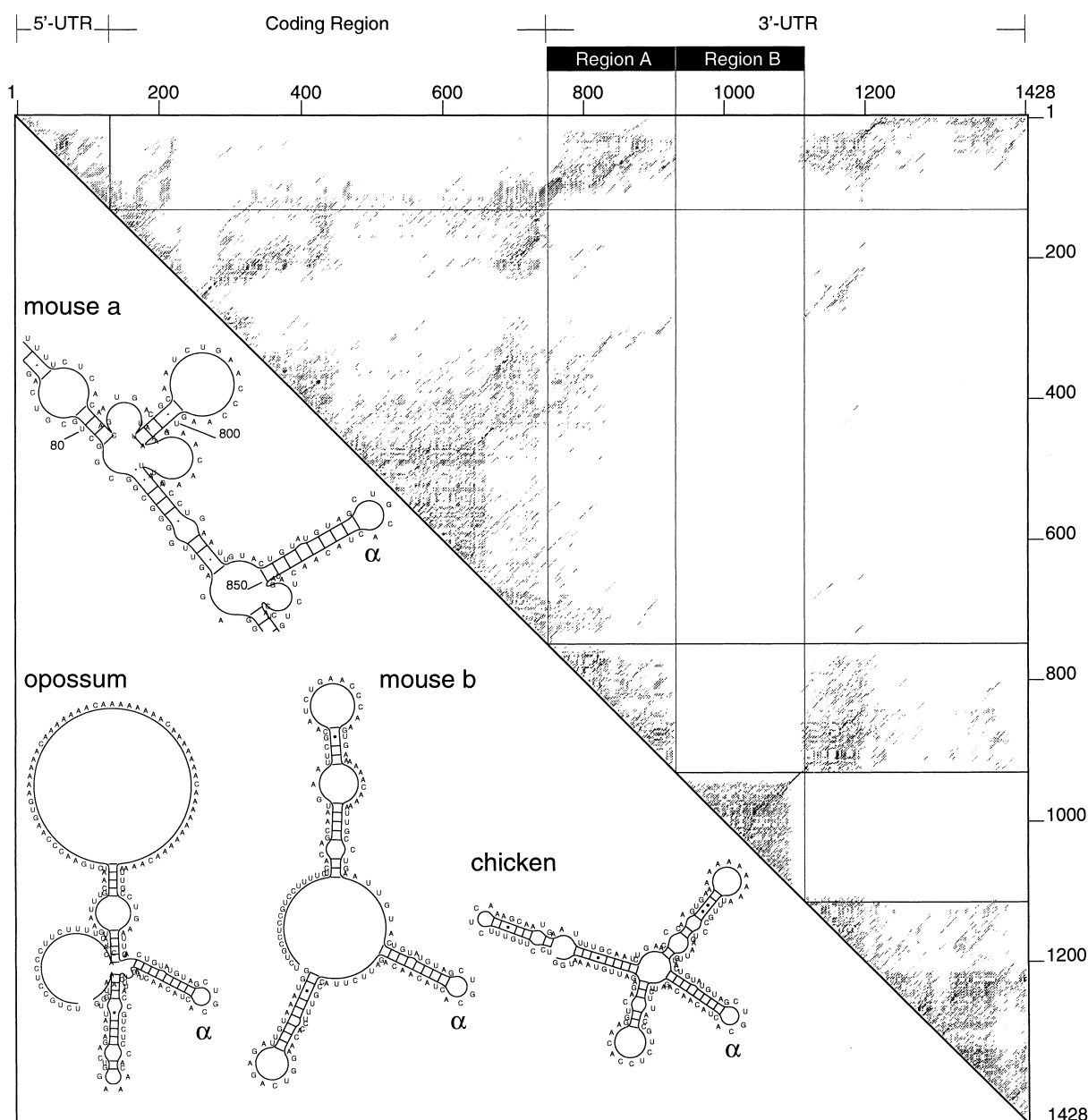


Fig. 3. Predicted secondary structures for RAB1a mRNA and 3'-UTRs. Upper right half: energy dot plot of complete mouse RAB1a mRNA. Numbers at the axes indicate nucleotides according to Figs. 1 and 2. Dots indicate free energy of base pairing interaction, with free energy  $-404.6 < \Delta G \leq -392.6$  kcal/mol. Lower left half: models for secondary structures with participation of region A. Mouse a, detail from folding of total mouse *Rab1a* mRNA (minimal free energy, MFE,  $-404.6$  kcal/mol): backfolding of region A to 5'-UTR. Mouse b, opossum, chicken: interspecies comparison of secondary structures of isolated region A (within a 10% range of the MFE); hairpin  $\alpha$  is visible in all models; marsupial is unique by a large single-stranded A-rich loop.

performed in order to identify conserved sequences. In the course of this work, the high conservation among mammalian species of the 3'-UTR region of the *Rab1a* gene became evident and the analysis was therefore extended to a wide variety of vertebrates, using primers flanking approximately 300 bp/680 bp of sequence between the stop codon and the polyadenylation site (Fig. 1).

The *Rab1a* 3'-UTR sequences of 17 eutherian, five marsupial (wallaby, possum, Tasmanian devil, opossum, quolls), three bird (chicken, zebra finch and ostrich) and two reptile species (alligator and turtle) have been amplified. In the case of marsupials, PCR products approximately 20–40 bp longer

than in the other tested species were obtained. Under stringent conditions, no amplification products were obtained from some lower vertebrates, amphibians (frog), elasmobranchs (shark), teleosts (medaka, zebrafish) as well as deuterostome (sea urchin) and protostome (cockchafer) invertebrates.

### 3.2. Selected sequences and their alignment

Sequences from amniote vertebrates were aligned (Fig. 2). Even between the species with the greatest phylogenetic distance (man and turtle), the sequences were found to be 92% identical. Within the placental mammals (Eutheria), the lowest observed sequence conservation was 95%.

In addition, the sequence analysis showed the expected conservation within particular orders and families, e.g. sequences specific for rodents (bp 155), carnivores (bp 18), for marsupials (bp 62–63, 70–105) and sauropsida (bp 16–17, 24, 202, 275–279). There were particular sequence identities across phylogenetical boundaries, e.g. alligator and ostrich were identical in positions 21 and 65–69 but different from other bird species and turtle. However, the overwhelming part of the 3'-UTR sequence shows very little variability, except for the A-rich region between bp 56 and 105. The A-rich sequences within the 3'-UTRs of marsupials were extended as compared to all other species investigated.

For five species (human, mouse, cat, fin whale and moose), the RAB1a 3'-UTR was completely sequenced. The results revealed a higher degree of conservation (>97%) as for the shorter region (Fig. 2).

### 3.3. Predicted secondary structure

Standard RNA secondary structure prediction tools were employed to determine the free energy ensemble of the complete mouse RAB1 transcript (Fig. 3). The energy dot plot reveals a region of high stability forming a long helical region of 72 bp (labeled B) in the 3'-UTR (nucleotides 932–1115 of the complete mouse mRNA sequence, accession number Y00094) that overlaps the PCR-amplified region by 52 nucleotides. The amplified region itself approximately coincides with a domain in the ensemble covering nucleotides 750–932 (labeled A) exhibiting mainly local folding and binding to the 5'-UTR. Basepair probability dot plots generated with 'RNA-fold' also accentuate region A (not shown).

Because the coding region of an mRNA may be prevented from folding by bound ribosomes, and since 3'-UTR regions A and B exhibit mainly local folding, we have predicted the secondary structure of the isolated region A. The secondary structure ensembles of the 3'-UTR fragments of all species show broad energy minima leading to families of alternative structures. However, there are recurrent features especially a conserved hairpin structure (labeled  $\alpha$  in Fig. 3) throughout all ensembles.

### 3.4. Preliminary evidence for protein binding

A reason for the conservation of a non-coding RNA sequence and a defined folding pattern might be the recognition by protein(s) with a conserved RNA binding site. Therefore, gel-shift experiments with cell extracts were performed. PCR products derived from mouse and chicken were cloned, transcribed and <sup>32</sup>P-radiolabeled in vitro. Subsequently, the labeled RNAs were incubated with total soluble protein extracts from mouse brain, liver and skeletal muscle. Complexes were electrophoretically separated under non-denaturing conditions. A shift of the position of radiolabeled RNAs was observed upon incubation of mouse and chicken 3'-UTRs with any of the mouse extracts, indicating interaction within a species and across species boundaries (not shown). After covalent cross linking and RNase treatment, a 200 kDa polypeptide(s)/RNA complex was detected on gel electrophoresis under denaturing conditions (not shown).

## 4. Discussion

We describe an extreme sequence conservation among the 3'-UTRs of the *Rab1a* genes among 27 amniote vertebrate

species. Although 3'-UTRs from several other genes have been shown to be conserved from mammals to lower vertebrates [6,7], the sequence homologies, e.g. between mammals and sauropsida, were not as high as with *Rab1a*. In the coding sequences (618 bp) of the *Rab1* genes, the following nucleotide identities relative to humans were found in the database: dog 96% (603 nucleotides of 618), mouse 95% (592/618), rat 93% (577/618). In comparison, the lowest sequence conservation we identified in the *Rab1a* 3'-UTR of placental mammals was 95%, the average 98%. Thus, the sequence conservation of the RAB1a 3'-UTR is higher than that of the protein coding region. The exceptional extension of the poly-A-rich sequence in marsupials underlines the uniqueness of this group relative to the placental animals. It would suggest a monophyletic origin as accepted in the recent literature [31].

This result raises three questions:

1. What is the relation of the 3'-UTRs sequenced to possible isoforms of the *Rab1* transcripts?
2. What is the nature of the presumed protein binding site within the *Rab1a* 3'-UTR?
3. What might be the cellular function of the *Rab1a* 3'-UTR and of *Rab* 3'-UTRs in general?

### 4.1. Ad 1: relation of the 3'-UTRs sequenced to possible isoforms of the RAB1 transcripts

In mammals, two *Rab1* RNA species with 1.6 and 3.2 kb have been found in a variety of tissues [21,37]. Haubruck and colleagues [12] suggested that both RNA species are expressed, by differential splicing, from the same gene and differ only in the lengths of their 3'-UTRs. The same phenomenon was recently described for the human *Rab1a* gene [10]. On the protein level, two functionally interchangeable *Rab1* isoforms, a and b, have been identified with 92% amino acid sequence identity [26]. In the rat, the 3'-UTR amplified from the *Rab1a* gene shows no homology to the 3'-UTR sequence of the *Rab1b* isoform. Mapping studies using the *Rab1a* 3'-cDNA as a probe revealed the identity of a further *Rab1*-like gene [34,39]. In all vertebrate species analyzed in this work, except for rat and man, very little is known about *Rab1* transcriptional or genomic isoforms. However, based on the experimental evidence on the better analyzed species, especially the isoform specificity of the 3'-UTRs, it seems likely that our amplified products are indeed genetic orthologs to the mouse and human *Rab1a* standard isoforms.

### 4.2. Ad 2: the nature of the binding site within the RAB1a 3'-UTR

The secondary structure prediction gives rise to two hypotheses: (1) the A-rich variable region between nucleotides 56 and 103 (as in Fig. 2) is single-stranded. (2) The region 113–136 (as in Fig. 2) forms a stable hairpin structure.

For several 3'-UTRs, specific binding proteins are known, e.g. the U-rich element RNA binding protein (AUF1), the hnRNP L protein which interacts specifically with *VEGF* mRNA in hypoxic cells, the IRP1 in the iron regulatory system and the stem-loop binding protein(s) which stabilize(s) the histone mRNA during the DNA synthesis [23,24,32]. Specific RNA foldings have been predicted, and according to these, most of the conserved sequences are located not in the base-paired but in the single-stranded regions [37].

In the case of *Rab1a* 3'-UTRs, the sequence homology is

certainly higher than required to maintain a given secondary structure. Furthermore, except for the  $\alpha$ -hairpin, there is no clearly defined stable overall secondary structure. Thus, some hitherto unknown biological function must reside in the primary sequences themselves.

#### 4.3. *Ad 3: possible physiological function of the highly conserved 3'-UTR of the RAB1a transcript*

Nothing is known about the protein(s) we have found to bind to the 3'-UTR of the RAB1a transcript. Therefore we can only speculate on the significance of this interaction. Two functions of 3'-UTRs are well documented in the literature: regulation of mRNA stability [4] and localization of mRNAs to specific regions of the cytoplasm [2]. In contrast to the *Rab1* coding region which is highly homologous among all members of the *Rab* family, the *Rab1a* 3'-UTR shows no significant homology to that of other genes, including other members of the *Rab* gene family. Even between isoforms RAB1A and RAB1B the 3'-UTRs are dissimilar. All RAB proteins are involved in intracellular vesicle transport but can be clearly distinguished regarding the cellular compartment in which they are active (reviewed in [21]). Therefore it is likely that the function of the *Rab1a* 3'-UTR is to localize the message to specific regions of the cytoplasm, so that translation would occur close to the localization of the respective RAB protein.

**Acknowledgements:** We thank Deutsche Forschungsgemeinschaft (SFB 223/C4 and Graduiertenkolleg 'Strukturbildungsprozesse', project Dr. R. Giegerich) for financial support. We are grateful to Professor H. Wiesner (director), Drs. med. vet. Ch. Lendl and J. Hector (Zoo Hellabrunn), Dr. med. vet. P. Brandt (Zoo Hannover), Prof. G. Neuweiler and Dr. H. Schweizer (Zoological Institute, University Munich), Dr. M. Walton (Sea Mammal Research Unit, Cambridge, UK), Dr. med. vet. Schulze (Husum), Professors F. Trillmich and H.H. Hendrichs (Behavioural Sciences, University of Bielefeld, Germany), Dr. med. vet. S. Kozik (Bielefeld), Dr. med. vet. M. Flügge (Zoo Hagenbeck, Hamburg, Germany); Professor U. Oron (Zoology Department, Tel-Aviv, Israel) and Dr. Tichy (Max-Planck-Institute for Biology, Tübingen, Germany) for providing, or helping to obtain, tissue samples for DNA preparation or DNA samples. Some DNAs have been prepared in the course of Dr. C. Witt's diploma thesis. Technical assistance of M. Ronsiek and B. Mersch is gratefully acknowledged.

#### References

- [1] Balch, W.E. (1990) Trends Biochem. Sci. 15, 473–477.
- [2] Bashirullah, A., Cooperstock, R.L. and Lipshitz, H.D. (1998) Annu. Rev. Biochem. 67, 335–394.
- [3] Binder, R., Horowitz, J.A., Basilion, J.P., Koeller, D.M., Klausner, R.D. and Harford, J.B. (1994) EMBO J. 13, 1969–1980.
- [4] Chen, C.-Y.A. and Shyu, A.-B. (1995) Trends Biochem. Sci. 20, 465–470.
- [5] Decker, C.J. and Parker, R. (1995) Curr. Opin. Cell Biol. 7, 386–392.
- [6] Duret, L., Dorkeld, F. and Gautier, C. (1993) Nucleic Acids Res. 21, 2315–2322.
- [7] Eastman, E.M. and Gilula, N.B. (1989) Gene 79, 151–158.
- [8] Evers, D. and Giegerich, R. (1999) Bioinformatics 15, 32–37.
- [9] Gavis, E.R., Curtis, D. and Lehmann, R. (1996) Dev. Biol. 176, 36–50.
- [10] Gromov, P.S., Celis, J.E., Hansen, C., Tommerup, N., Gromova, I. and Madsen, P. (1999) FEBS Lett. 429, 359–364.
- [11] Haubruck, H., Disela, C., Wagner, P. and Gallwitz, D. (1987) EMBO J. 6, 4049–4053.
- [12] Haubruck, H., Prange, R., Vorgias, C. and Gallwitz, D. (1989) EMBO J. 8, 1427–1432.
- [13] Hill, K.E., Lloyd, R.S. and Burk, R.F. (1993) Proc. Natl. Acad. Sci. USA 90, 537–541.
- [14] Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M. and Schuster, P. (1994) Mon. heft. Chem. 125, 167–188.
- [15] Jackson, R.J. (1993) Cell 74, 9–14.
- [16] Kislauskis, E.H., Li, Z., Singer, R.H. and Taneja, K.L. (1993) J. Cell Biol. 123, 165–172.
- [17] Korthaus, D., Wedemeyer, N., Lengeling, A., Ronsiek, M., Jockusch, H. and Schmitt-John, T. (1997) Genomics 43, 242–244.
- [18] Laird, P.W., Zijderveld, A., Linders, K., Rudnicki, M.A., Jaenisch, R. and Berns, A. (1991) Nucleic Acids Res. 19, 4293.
- [19] Määttä, A., Bornstein, P. and Penttinen, R.P. (1991) FEBS Lett. 279, 9–13.
- [20] MacDonald, P.M. and Struhl, G. (1988) Nature 336, 595–598.
- [21] Martinez, O. and Goud, B. (1998) Biochim. Biophys. Acta 1404, 101–102.
- [22] Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) J. Mol. Biol. 288, 911–940.
- [23] McCarthy, J.E.G. and Kollmus, H. (1995) Trends Biochem. Sci. 20, 191–197.
- [24] Nakamaki, T., Imamura, J., Brewer, G., Tsuruoka, N. and Koefler, H.P. (1995) J. Cell Physiol. 165, 484–492.
- [25] Nuoffer, C. and Balch, W.E. (1994) Annu. Rev. Biochem. 63, 949–990.
- [26] Nuoffer, C., Davidson, H.W., Matteson, J., Meinkoth, J. and Balch, W.E. (1994) J. Cell Biol. 125, 225–237.
- [27] Rastinejad, F. and Blau, H.M. (1993) Cell 72, 903–917.
- [28] Resch, K., Korthaus, D., Wedemeyer, N., Lengeling, A., Thiel, C., Ronsiek, M., Baer, K., Jockusch, H. and Schmitt-John, T. (1998) Mamm. Genome 9, 893–898.
- [29] Rouault, J.P., Samarut, C., Duret, L., Tessa, C., Samarut, J. and Magaud, J.P. (1993) Gene 129, 303–306.
- [30] Shaw and Kamen (1986) Cell 46, 659–667.
- [31] Shoshani, J. and McKenna, M. (1998) Mol. Phylogenet. Evol. 9, 572–584.
- [32] Shih, S.C. and Claffey, K.P. (1999) J. Biol. Chem. 274, 1359–1365.
- [33] Tisdale, E.J., Bourne, J.R., Khosravi-Far, R., Davidson, H.W., Der, C.J. and Balch, W.E. (1992) J. Cell Biol. 119, 749–761.
- [34] Vielh, E., Touchot, N., Zahraoui, A. and Tavitian, A. (1989) Nucleic Acids Res. 17, 1770.
- [35] Wedemeyer, N., Lengeling, A., Ronsiek, M., Korthaus, D., Baer, K., Wuttke, M. and Jockusch, H. (1996) Genomics 32, 447–454.
- [36] Wichmann, H., Disela, C., Haubruck, H. and Gallwitz, D. (1989) Nucleic Acids Res. 17, 6737–6738.
- [37] Wilson, B.S., Nuoffer, C., Meinkoth, J.L., McCaffery, M., Feramisco, J.R., Balch, W.E. and Farquhar, M.G. (1994) J. Cell Biol. 125, 557–571.
- [38] Yaffe, D., Nudel, U., Mayer, Y. and Neuman, S. (1985) Nucleic Acids Res. 13, 3723–3737.
- [39] Zahraoui, A., Touchot, N., Chardin, P. and Tavitian, A. (1989) J. Biol. Chem. 264, 12394–12401.