Intronic hammerhead ribozymes are ultraconserved in the human genome

Marcos de la Peña* & Inmaculada García-Robles
Instituto de Biología Molecular y Celular de Plantas, Universidad Politécnica de Valencia-Consejo Superior de Investigaciones Científicas, Valencia, Spain

Small ribozymes have been regarded as living fossils of a prebiotic RNA world that would have remained in the genomes of modern organisms. In this study, we report the ultraconserved occurrence of hammerhead ribozymes in Amniota genomes (reptiles, birds and mammals, including humans), similar to those described previously in amphibians and platyhelminth parasites. The ribozymes mapped to intronic regions of different genes, such as the tumour suppressor RECK in birds and mammals, a mammalian tumour antigen and the dystrobrevin beta in lizards and birds. In vitro characterization confirmed a high self-cleavage activity, whereas analysis of RECK-expressed sequence tags revealed fusion events between the in vivo self-cleaved intron and U5 or U6 small nuclear RNA fragments. Together, these results suggest a conserved role for these ribozymes in messenger RNA biogenesis.

Keywords: ribozyme; intron; splicing; exaptation; retrotransposons

EMBO reports (2010) 11, 711–716. doi:10.1038/embor.2010.100

INTRODUCTION

The term ‘RNA world’ was first used to refer to a hypothetical stage in the emergence of life in which RNA carried out both genetic and catalytic roles (Gilbert, 1986). This hypothesis stimulated research not only on the origin of life but also on new RNA capabilities, ultimately unveiling new features of our ‘modern RNA world’. The discovery of catalytic RNAs in the Group I intron (Cech et al., 1981) and in RNAse P (Guerrero-Takada et al., 1983), the more recent findings of RNA interference (Fire et al., 1998) and the ribozymal nature of peptide bond formation (Nissen et al., 2000) are all hallmarks of the ever-growing biological roles of RNA. In this context, previous data have revealed the presence of small self-cleaving RNA domains in the noncoding regions of mammalian genes, such as the co-transcriptional cleavage ribozyme at the 3′-untranslated region of β-globin (Teixeira et al., 2004), an intronic Hepatitis δ virus-like ribozyme (HDVR) in Cpeb3 (Salehi-Ashtiani et al., 2006), and a discontinuous hammerhead ribozyme (HHR) at the 3′-untranslated region of Clec2 genes (Martick et al., 2008).

In this study, we show the results of a genome-wide search for the HHR domain, a paradigmatic ribozyme initially described to catalyse a transesterification self-cleavage reaction during replication of some viroids and other small RNA plant pathogens (Przybilski et al., 1986). Since then, HHRs have also been observed in a few unrelated eukaryotic genomes, such as those of salamanders (Epstein & Gall, 1987), schistosomes (Ferbeyre et al., 1998, crickets (Rojas et al., 2000), Arabidopsis thaliana (Przybilski et al., 2005) and Clec2 genes of rodents and platypus (Martick et al., 2008). In the latter case, HHRs showed a clear similarity to those observed in schistosomes, suggesting an event of horizontal gene transfer (Martick et al., 2008). On the basis of this similarity, we followed a bioinformatic search among lower metazoan and vertebrate genomes that revealed hundreds of HHR motifs associated with retrotransposable elements. More surprisingly, a few similar ribozymes were also detected as ultra-conserved motifs in the genomes of Amniota spp., indicating ancient exaptation events for this small self-cleaving domain during the evolution of higher vertebrates.

RESULTS

An iterative method for identifying HHR-like motifs

Extensive bioinformatic approaches have been devised previously to search for HHR motifs among genomes (Ferbeyre et al., 1998, 2000; Przybilski et al., 2005; Martick et al., 2008), exclusively considering the 11 catalytically conserved nucleotides as the source of the phylogenetic signal (Fig 1). In this study, we used an iterative bioinformatic approach starting with basic local alignment search tool (BLAST) searches among vertebrate genomes, followed by the implementation of structural and phylogenetic constraints (supplementary information online). Briefly, initial queries (seeds) were composed by naturally occurring Helix II motifs (supplementary Fig S1 online) flanked by the conserved U and P boxes (Fig 1A). Sequence changes in the returned entries were examined to fulfil two conditions: (i) not to affect the conserved U and P boxes and (ii) to be either compensatory within stem II or located in loop 2. Filtered hits were analysed for three
additional criteria to define candidate sequences as HHR motifs: (i) 5′ and 3′ surrounding regions should fold as two helices (I and III), wherein (ii) Helix I had to be either a loop-closed RNA helix (type I/II HHR) or an open helix containing an internal loop (type II/III HHR), and (v) a proper cleavage site triplet must exist (Fig 1B). An extra point of validation was the presence and nature of loops 1 and 2, which allow the required tertiary interactions for in vivo self-cleavage activity (de la Peña et al., 2003, 2009; Khvorova et al., 2003; Martick & Scott, 2006).

**HHRs are widespread in amphibians and lampreys**

The BLAST searches using the *Schistosoma mansoni* genome (Berriman et al., 2009) revealed the presence of more than 50,000 entries for type II/III HHRs (data not shown). Ribozymes were similar to those described previously in the SMe satellite DNA of these platyhelminth parasites (Fig 2A; Ferbeyre et al., 1998), although their occurrence was not restricted to repetitive DNA. Similar HHRs were also detected in the genome of *Schistosoma japonicum* (Fig 2B), and in a few complementary DNA (cDNA) sequences of the more distant liver fluke *Opisthorchis viverrini* (Fig 2C). Interestingly, one of these cDNAs showed significant homology to an endonuclease reverse transcriptase from *S. japonicum* (supplementary Table S1 online).

Similar type II/III HHRs embedded in 330 bp tandem repeats (satellite 2 DNA) have been described previously in the genomes of several species of newts (Epstein & Gall, 1987) and salamanders (*Caudata* order; Zhang & Epstein, 1996). Our iterative searches revealed new HHRs in the genomes of three more species of salamanders and especially in the genome of the frog *Xenopus tropicalis* (Anura order), suggesting a widespread occurrence among amphibians (Fig 2D–E; supplementary Table S1 online). BLAST-like alignment tool (BLAT; University of California, Santa Cruz) searches using whole HHRs of *X. tropicalis* revealed hundreds of entries dispersed along the frog genome. Some of these ribozymes were observed to be associated with repetitive DNA, but they were also dispersed as single intronic motifs noticeably associated with reverse transcriptase-like coding regions (supplementary Table S1 online).

Type II/III HHRs were also detected in the genomes of *Petromyzon marinus* (marine lamprey) and *Lethenteron japonicum* (arctic lamprey; Fig 2F). BLAT searches further confirmed the presence of hundreds of single HHRs dispersed along the *P. marinus* genome, which in many cases also mapped within intronic regions (supplementary Table S1 online). Although a detailed analysis of HHRs observed in schistosomes, *X. tropicalis* and *P. marinus* will be reported elsewhere (M.D.P., unpublished observations), their occurrence in different intronic regions and their frequent association with reverse transcriptase-like genes was striking.

**A HHR is ultraconserved in a tumour suppressor gene**

Parallel BLAST searches against the human genome were performed with Helix II seeds derived from HHRs of schistosomes and lower vertebrates. For several seeds (supplementary information online), a hit mapping within a contig of the human chromosome 9 was repeatedly obtained (Fig 3A). Surrounding regions of this hit could be folded to give an HHR fold (hereafter referred to as HH9). The HH9 fold showed a 4 bp Helix III capped by a non-palindromic triloop and a 13 bp Helix I containing a 7 nt internal loop (Fig 3B). In vitro characterization of HH9 revealed a *k*_agt_ of 2.43 ± 0.22 min−1 at 25 °C, pH 6.5 and 1 mM Mg²⁺ (supplementary Fig S2 online), which corresponds to a self-cleavage activity similar to those described for most naturally occurring HHRs acting in cis (de la Peña et al., 2003; Khvorova et al., 2003). As reported for the discontinuous HHR in *Clec2* genes (Martick et al., 2008), loop 2 of HH9 and also loop 1 (not included in the searching seeds) were similar to the corresponding loops of the *S. mansoni* HHRs, allowing equivalent loop–loop tertiary interactions (supplementary Fig S2C online; Martick & Scott, 2006). Human HH9 mapped to the middle of the sixth intron (14 kb) of an open reading frame coding for the protein RECK, a factor characterized initially as a tumour suppressor (Takahashi et al., 1998). This membrane glycoprotein is known to inhibit the matrix metalloproteinases involved in remodelling the extracellular matrix, a key step during embryogenesis and vasculogenesis (Oh et al., 2001). The HH9 fold was observed to be ultraconserved (90% identity for 60 nt) among the introns of RECK orthologues for all endothermic vertebrate spp. examined (birds and mammals), but was absent in the orthologues of lower vertebrates and metazoans (Fig 3A; supplementary Fig S3 online). Furthermore, sequence heterogeneity was mostly restrained to the Helix III-loop 3 region and was always compatible with a catalytically active HHR. Sequence alignment of the surrounding intronic regions of HH9 did not show any significant conservation, therefore revealing a strong selection pressure in the ribozyme motif (Fig 3B; supplementary Fig S3 online). Data mining of RECK expressed sequence tags revealed two examples from *Bos taurus* fetal and calf brain tissues, which mapped exactly to the precise 5′-self-cleaved HH9. However, both sequences showed 5′-extensions of 82 and 83 nt that corresponded to fragments of U5 and U6 small nuclear RNAs (snRNAs), respectively (Fig 3C). These two examples reveal that HH9 self-cleaves *in vivo*, and also offers circumstantial evidence that the 5′-OH product could interfere with the spliceosomal machinery.
HHR motifs in trematodes and lower vertebrates. (A) Consensus example of the HHR with some of the natural variability described previously for three schistosomal spp. (Ferbeyre et al., 1998). Examples of HHRs detected in (B) Schistosoma japonicum (variability with respect to Schistosoma mansoni is highlighted) and (C) Opisthorchis viverrini (intraspecies variability is highlighted). Examples of HHRs observed in (D) Caudata (salamanders and newts), (E) Anura (frog) and (F) Agnatha (lampreys) spp. HHR, hammerhead ribozyme.

A HHR in a mammalian tumour antigen gene

By using the iterative search strategy described above, a second HHR motif (HH10) was detected in human chromosome 10. The HH10 motif was similar to HH9, with the exception of a larger Helix III (43 nt instead of 11 nt; Fig 4A), and was conserved among most mammalian genomes from platypus to humans, with the exception of rodents (rat and mouse), lagomorphs (rabbit) and cingulata (armadillo; supplementary Fig S4 online). In humans, the motif mapped within the first intron (10 kb) of the uncharacterized C10orf118 gene. The protein encoded by this open reading frame is the CTCL (cutaneous T-cell lymphoma) tumour antigen L14-2, a factor involved in processes of tumorigenesis (Hartmann et al., 2004). The surrounding regions of HH10 Helix I revealed a clear conservation of eight extra base pairs when compared with the HH9 region (Figs 3A, 4A; supplementary Fig S4 online). Furthermore, the 5′-ends of two expressed sequence tags (both from human brain origin) were observed to match with this intronic region. However, the 5′ extremes did not map with the precise self-cleavage site, but with the end of the conserved region of Helix I (supplementary Fig S5 online). Regarding the absence of the HH10 motif among some mammal species, a case of an inactive HHR motif was observed in the C10orf118 orthologue of the lagomorph Ochotona princeps (supplementary Fig S6 online), which suggests the loss of this HHR motif during the evolution of these mammals. Interestingly, a putative gene homologous to C10orf118 (UniProt accession number C5J917) was detected in the S. mansoni genome. Inspection of the corresponding genomic sequence in the trematode revealed the presence of at least two HHR motifs (supplementary Fig S7 online).

A conserved HHR in the DTNB gene of sauropsids

The BLAT searches using the whole HH9 motif revealed a new example of a conserved HHR (HHsau) in genomes of four sauropsids: the Anolis carolinensis lizard and Gallus gallus, Taenopygia guttata and Meleagris gallopavo birds. The structure and sequence of HHsau mostly resembled HH9, although a covariation within stem I matched with the one detected in HH10 (Fig 4B). For the four spp., HHsau mapped within the first intron (9 kb) of a gene encoding the dystrobrevin beta protein (DTNB). DTNB is expressed in brain tissues, where it interacts with dystrophin to form the core of the transmembrane dystrophin glycoprotein complex (DGC; Peters et al., 1997). Although DGC is known to link the extracellular matrix to the intracellular cytoskeleton, the specific presence of DTNB involves the DGC in neuronal signalling (Grady et al., 2006). The mammalian orthologues of DTNB seem to lack the extra first exon/intron present in sauropsids, which would explain the absence of this particular HHR in mammals.

DISCUSSION

In this study, we have unveiled the presence of ultraconserved intronic HHRs in the genomes of reptiles, birds and mammals, including humans. To achieve this goal, we first identified a large set of HHRs in the genomes of trematodes, P. marinus and X. tropicalis through iterative bioinformatic searches. The recurrent association of many of these motifs with RT-like coding regions would directly involve these ribozymes in retrotransposable elements, extending previous observations of HHR in the satellite DNA of newts, trematodes and crickets to a more general landscape among eukaryotes and other organisms (this work; de la Peña & ...
Fig 3 | An intronic HHR is ultraconserved in endothermic vertebrates. (A) Top: approximated location of the RECK gene within the human chromosome 9. Below: a scheme of the RECK precursor messenger RNA; exons are marked with vertical lines and introns as lines connecting exons. At the bottom, sequence alignment of the HH9 motifs detected in the intronic regions of RECK orthologues in endothermic organisms. For a more detailed alignment, see supplementary Fig S3 online. (B) Predicted secondary structure of HH9. Underlined nucleotides are invariable among all endothermic species examined. Sequence heterogeneity found in the mouse (Mus musculus), pig (Sus scrofa), platypus (Ornithorhynchus anatinus) and chicken (Gallus gallus) HH9 orthologues is shown in the boxed insets. (C) Two expressed sequence tags from Bos taurus brain tissues mapping at the HHR self-cleaved RECK intron (nuclotides in red) fused with U5 (in blue, left) or U6 (in green, right) snRNA fragments. Minimal free-energy structures of the U5 and U6 snRNAs are depicted for clarity. Nucleotides involved in U6–U4 snRNA interaction (stems I and II) are shown in lighter green. HHR, hammerhead ribozyme; snRNA, small nuclear RNA.
Furthermore, recent work describing the association of the HDVR motif with retrotransposons (Webb et al., 2009) confirms that not only HHR, but self-claving ribozymes in general might function in retrotransposition. Nevertheless, the occurrence of intronic HHRs in Amniota genomes as ultraconserved motifs within specific genes strongly suggests that they have a conserved biological role other than retrotransposition. Previous studies using artificially engineered HHRs into introns did not reveal splicing effects on adjacent exons in wild-type yeast (Lacadie et al., 2006) and human cells (Dye et al., 2006). However, more recently it has been shown that continuous precursor messenger RNAs (pre-mRNAs) seem to be required for proper splicing (Fong et al., 2009). In that work, the authors observed that, although highly active HDVR promotes splicing inhibition, an intronic HHR affects splicing weakly, most probably because of its lower catalytic efficiency. Accordingly, we propose that natural intronic HHRs might not affect pre-mRNA splicing under basal conditions, but either changes in transcription elongation or ribozyme upregulation could interfere with the spliceosomal machinery (for example, fusion with snRNAs), promoting alternative splicing events or even disruption of the pre-mRNA processing. A similar situation is illustrated by the intronic HDVR of the mammalian Cpeb3 gene, which has been described already as a low-activity version of this ribozyme (Salehi-Ashtiani et al., 2006).

Altogether, these findings allow us to conclude that self-cleaving ribozymes present in retrotransposons of lower metazoans would have been exapted or domesticated in the genomes of Amniotes to perform specific functions.

Finally, the resemblance of vertebrate HHRs with those widespread within mobile genetic elements in trematodes suggests that recurrent genomic horizontal transfers could have taken place from parasites to hosts. In fact, increasing evidence for a relationship between carcinogenesis and infections by helminths, specifically trematodes, has accumulated during the last decades (WHO, 1994; Mayer & Fried, 2007), and our present data adds circumstantial support in this direction. Although a deeper molecular characterization is required, a potential relationship can be envisaged between trematode-induced carcinogenesis and ribozyme-containing mobile elements, which would ultimately affect the host genomic stability. In a more unlikely scenario, trematode infection could promote activation of the intronic HHRs of host genes involved in tumour progression.

METHODS

Analysis of data. Genome servers used for the analysis of data and mapping of hits were NCBI-BLAST2 Nucleotide tool at the European Bioinformatics Institute (www.ebi.ac.uk), ENSEMBL.
REFERENCES

Spanish Patent and Trademark Office (OEPM, Madrid, Spain). The authors declare to have a patent pending (no. P201030569) at the CONFLICT OF INTEREST Cajal contract to M.D.P.) and Generalitat Valenciana (GV06/206 to M.D.P.). de Educacio´n y Ciencia of Spain (BFU2008-03154/BMC and Ramo´ny de la Pen˜a M, Dufour D, Gallego J (2009) Three-way RNA junctions with EMBO Berriman M.

ACKNOWLEDGEMENTS We thank W.G. Scott, L. Yenush, S. Delgado and J. de la Peña for advice and critical reading of the paper. This work was supported by Ministerio de Educaci´on y Ciencia of Spain (BFU2008-03154/BMC and Ramo´ny de la Pen˜a M, Dufour D, Gallego J (2009) Three-way RNA junctions with EMBO Berriman M.

CONFLICT OF INTEREST The authors declare to have a patent pending (no. P201030569) at the Spanish Patent and Trademark Office (OEPM, Madrid, Spain).