A database search for hammerhead ribozyme motifs

S. Gräf, R. Przybilski, G. Steger and C. Hammann

Abstract

The hammerhead ribozyme is the smallest naturally occurring RNA endonuclease. It is found in subviral plant pathogens and transcripts of satellite DNA from a limited number of organisms. We have performed a database search for novel examples of this catalytic RNA, taking into consideration the recently defined structural requirements for an efficient cleavage under physiological magnesium ion concentrations. In this search, we find, apart from the known examples, several hundreds of motifs in organisms of all kingdoms of life. In a first set of experiments, we analysed hammerhead ribozymes from Arabidopsis thaliana. We found that these sequences are tissue-specifically expressed and that they undergo self-cleavage in planta. Furthermore, their activity under physiological magnesium ion concentrations depends on functional loop-loop interactions, as shown by the lack of activity of appropriate mutants.

Introduction

The hammerhead ribozyme is a small catalytic RNA motif found in a variety of subviral pathogens, including viral satellite RNAs and some viroids. It is also found in transcripts from satellite DNA in various amphibians, in schistosomes and in crickets [1–4]. In their natural context, hammerhead ribozymes perform a site-specific cleavage of their own phosphodiester backbone by means of a transesterification reaction (cis cleavage) (summarized in [5]).

For more than a decade, research on the hammerhead motif has, for a variety of reasons, mainly been performed on minimized (truncated) versions [6,7]. These structures are characterized by a core of highly conserved nucleotides from which three helical stems radiate. Here, usually only one of the three helical stems is closed by a hairpin loop, allowing for the application of these hammerheads as trans ribozymes. In naturally occurring hammerhead ribozymes, which promote cis cleavage, on the other hand, two helices are closed by a hairpin loop structure.

Only recently, it was realized that the truncated versions of the hammerhead represented suboptimal catalytic entities [8,9] with a folding behaviour that deviates from that of naturally occurring hammerhead ribozymes [10,11]. More specifically, the interaction of loop sequences at the ends of two helical stems of the hammerhead motif is essential for efficient intracellular cleavage. To allow for this interaction, stems I and II should contain 5–7 and 4–5 nt respectively, and loop sizes are 3–7 and 4–8 nt for loops L1 and L2 respectively, with sequences that allow for an interaction [8,9]. These requirements were shown recently for naturally occurring hammerhead ribozymes of type III; for sequences of type I, a similar interaction was proposed earlier, which is mediated between an internal loop in helix I and the sequence of loop L2 [12].

A database search for hammerhead ribozyme motifs

Using the well-known structural features necessary for efficient cis-hammerhead cleavage as restriction parameters, we have searched the EMBL database (rel.78) for examples of hammerhead ribozyme motifs. For this search, we used a pattern description language [13], as in a recent search [14] for double-stranded RNA.

For type III hammerheads, the result of this search was a total of 302 sequences, the majority of which (180) corresponded to annotated hammerhead ribozymes from a variety of satellite RNAs and some viroids. A further 94 sequences stemmed from synthetic or patented constructs. The remainder included novel sequences from Homo sapiens, Mus musculus, Arabidopsis thaliana, Danio rerio and Xenopus tropicalis. To address whether the new hits were coincidental, we calculated the probability of a type III hammerhead (defined by our search string) to occur in a random sequence: the probability of occurrence of a hammerhead sequence is the product of the probabilities of its discrete pattern units, under consideration of RNA base-pairing rules for helical elements of the hammerhead, allowing for U-G and G-U wobbles; alternatives in a discrete pattern unit are summed up. From these calculations, type III hammerheads are predicted to occur in a random sequence once in $6 \times 10^7$ nt. This number is larger than any of the genomes in which we find type III sequences. In other words, genomes would have to be between 6 times (M. musculus) and 50 times (A. thaliana) larger than they are to account for random occurrences.
Database searches for type I hammerhead ribozymes resulted in 1129 hits. Of these, 255 independent sequences are derived from annotated hammerheads, mainly from various amphibians and schistosomes. The remainder originated from database entries of 41 different eukaryotic organisms. Statistical analysis predicts the occurrence of a type I hammerhead ribozyme once in every $3 \times 10^7$ nt in a random sequence. This implies that in the genome of *Fugu rubripes* or *Drosophila melanogaster*, for example, 22 or 9 type I hammerhead motifs are expected in the case of a random sequence. We find, however, for these organisms only one and two motifs respectively, indicating that the motifs are under-represented in these organisms when compared with their genome size. The same situation is also found for a number of other organisms, including *Anopheles gambiae*, *Caenorhabditis briggsae* or *Dictyostelium discoideum*. Conversely, type I hammerhead ribozymes occur up to twice as frequently as predicted for *H. sapiens* or *D. rerio*. Given this asymmetric distribution, we conclude that hammerhead ribozyme motifs do not occur by chance in eukaryotic genomes.

**Genomically encoded hammerhead ribozymes are expressed in thale cress**

We set out to analyse some examples of these newly identified hammerheads. Initially, we concentrated on two sequences of type III from *A. thaliana*. Both these sequences are encoded on chromosome IV and are approx. 5 kb apart. They are identical throughout the hammerhead domain with the exception of two, probably compensating nucleotide exchanges in loops L1 and L2. Their sequence context is conserved 230 nt upstream and 26 nt downstream of the hammerhead domain. When transcribed in vitro, both these sequences cleave to completeness during transcription. However, the addition of an antisense oligonucleotide in the transcription reaction, which would prevent folding of the hammerhead structure, allows for the preparation of full-length transcripts. Cleavage of these can be induced at submillimolar magnesium ion concentrations, as has been shown for hammerhead ribozymes with interacting loops [11]. To prove the importance of the loop L1–loop L2 interaction for activity under these conditions, the natural sequence of loop L2 was replaced by an unrelated sequence, resulting in a complete loss of cleavage activity at submillimolar magnesium ion concentrations, whereas, at 10 mM, cleavage was observed, indicating that the RNA then behaved like a truncated hammerhead ribozyme [6]. By reverse-transcriptase–PCR, we find tissue-specific expression of hammerhead ribozyme-containing sequences and S1 nuclease protection assays revealed the presence of both cleaved and uncleaved species in the respective tissues.

We are currently performing experiments addressing the expression and cleavage behaviour of novel hammerhead ribozymes from other organisms.

**References**


Received 14 January 2005