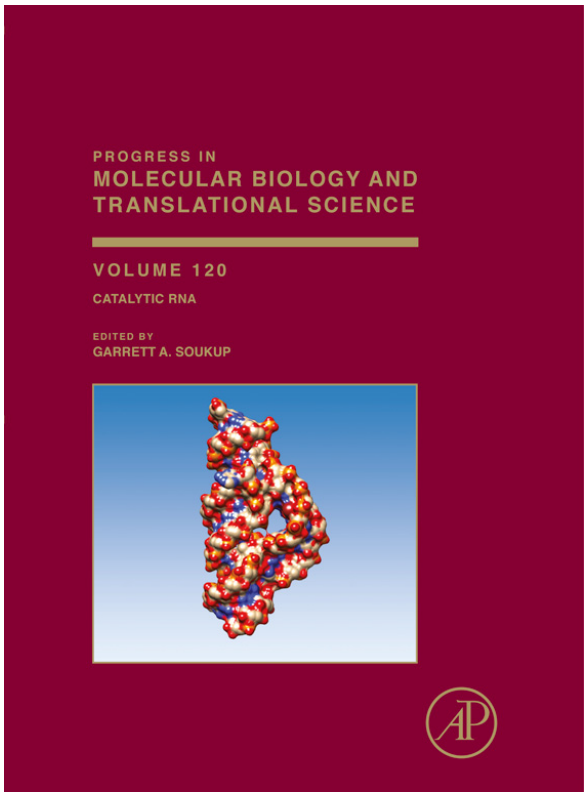

From: Nathan Riccitelli, Andrej Lupták, HDV Family of Self-Cleaving Ribozymes. In Garrett A. Soukup, editor: Progress in Molecular Biology and Translational Science, Vol. 120,
Burlington: Academic Press, 2013, pp. 123-171.
ISBN: 978-0-12-381286-5

CHAPTER FOUR

HDV Family of Self-Cleaving Ribozymes

Nathan Riccitelli[*], **Andrej Lupták**[*,†,‡]
[*]Department of Chemistry, University of California, Irvine, California, USA
[†]Department of Pharmaceutical Sciences, University of California, Irvine, California, USA
[‡]Department of Molecular Biology and Biochemistry, University of California, Irvine, California, USA

Contents

Progress in Molecular Biology and Translational Science, Volume 120
ISSN 1877-1173
http://dx.doi.org/10.1016/B978-0-12-381286-5.00004-4

Abstract

The hepatitis delta virus (HDV) ribozymes are catalytic RNAs capable of cleaving their own sugar–phosphate backbone. The HDV virus possesses the ribozymes in both sense and antisense genomic transcripts, where they are essential for processing during replication. These ribozymes have been the subject of intense biochemical scrutiny and have yielded a wealth of mechanistic insights. In recent years, many HDV-like ribozymes have been identified in nearly all branches of life. The ribozymes are implicated in a variety of biological events, including episodic memory in mammals and retrotransposition in many eukaryotes. Detailed analysis of additional HDV-like ribozyme isolates will likely reveal many more biological functions and provide information about the evolution of this unique RNA.

1. INTRODUCTION

RNAs are responsible for many tasks in the cell, including the transmission of genetic information, ligand binding, and catalysis. RNA enzymes, ribozymes, include the ribosome and a number of other catalytic RNAs found both *in vivo* and *in vitro*. Among the best-understood ribozymes are those RNAs capable of self-scission of their own phosphate backbone.

Self-cleaving ribozymes form a diverse set of RNAs that share the common feature of a 2′ OH-mediated attack on the phosphodiester of the adjoining nucleotide to yield a 2′–3′ cyclic phosphate and a 5′ OH. Although self-cleaving sequences were first identified for their essential role in the life cycles of viroids, viruses, and satellite DNA transcripts, mounting evidence suggests that these sequences are much more widely dispersed in nature and perform a variety of biological roles.[1–9]

The ability to catalyze attack on the phosphodiester backbone is dependent on the RNA's three-dimensional structure. The structure promotes deprotonation of the attacking 2′ OH, orienting it for an in-line attack, stabilization of the phosphorane intermediate, and protonation of the 5′ oxyanion leaving group. Several such catalytic motifs exist, including the hammerhead, hairpin, hepatitis delta virus (HDV), the *Neurospora* Varkud satellite motifs, and the cofactor-dependent *glmS* ribozyme.[10–14] The HDV family of ribozymes is the topic of this chapter.

2. HEPATITIS DELTA VIRUS

HDV was first identified in 1977 from clinical isolates of Italian patients carrying the hepatitis B virus (HBV).[15] HDV is not an autonomous virus, as it requires a coinfection by fully functional HBV or other helper

hepadnaviruses.[16–18] Individuals infected with both HDV and an associated helper virus experience an increased rate of liver cirrhosis and an increased chance of developing liver cancer in chronic infections.[18] In these instances of coinfection, a decrease in helper virus particles is observed upon detection of HDV markers.[16]

The HDV genome is a single-stranded circular RNA molecule of approximately 1700 nucleotides (nt).[18] Upon infection, the RNA is transcribed by a rolling-circle mechanism using host machinery to produce its antigenomic form that is subject to one of two pathways (Fig. 4.1).[19] In the first pathway, the RNA undergoes capping and polyadenylation to yield a subgenomic sequence that serves as the mRNA for the sole protein produced from the HDV genome, the delta antigen.[20–22] Two variants of the protein are present *in vivo*, the small and large delta antigens.[23] These two proteins are identical in sequence except at the C-terminus, where

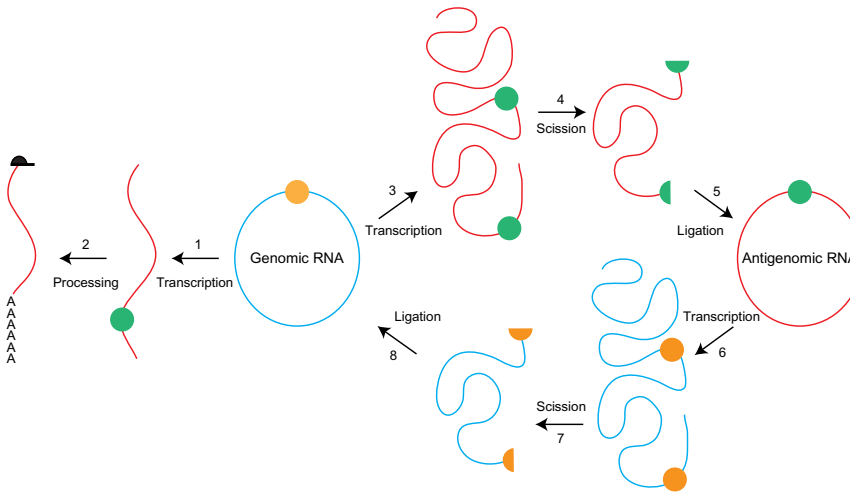

Figure 4.1 The HDV life cycle. Blue lines indicate genomic (+) and red lines indicate antigenomic (−) RNA. Orange and green circles refer to genomic and antigenomic ribozymes, respectively. During an infection cycle, the circular genomic RNA is transcribed into (1) short antigenomic mRNAs coding for the delta antigen that undergo (2) capping (black hat) and polyadenylation prior to translation, or (3) transcription into concatemeric antigenomic RNA. The latter is (4) processed to unit-length antigenomic transcripts by the action antigenomic HDV ribozymes, and (5) ligated using host machinery into an antigenomic circular RNA. This circular antigenome is similarly (6) transcribed into concatemeric genomic RNA, (7) processed into unit-length genomic transcripts by the action of genomic HDV ribozymes, and (8) ligated using host machinery to yield new genomic RNA for packaging into viral particles.

the action of an adenosine deaminase converts the adenosine in the UAG stop codon of the small delta antigen to an inosine. Inosine is read as a guanosine, resulting in a UGG tryptophan codon in lieu of the stop codon, which extends the C-terminus by 19 amino acids, producing the large form of the delta antigen.[24] As no other proteins are produced from the viroid, HDV is completely dependent on the host and helper virus-encoded machinery for replication and packaging into active viral particles.[18,25,26]

In the second pathway, rolling-circle replication produces antigenomic transcripts that are greater than unit-length genomic RNA. Processing of these antigenomic concatemers into unit-length antisense genomes is achieved without the aid of any host, helper virus, or viroid proteins. To process these RNAs, HDV instead employs a catalytic RNA in the form of the antigenomic HDV ribozyme. The ribozyme promotes scission of the phosphate backbone between the nucleotide 5′ to the structure (the leader sequence) and the first nucleotide of the ribozyme.[27,28] Ligation into a circular RNA is achieved with the aid of a host RNA ligase.[29] The viroid also contains a second, genomic version of the ribozyme with a similar secondary structure to the antigenomic form that is responsible for processing of the concatemeric genomic RNA that arises from rolling-circle replication of the antigenomic RNA (Fig. 4.1).

As the products of the self-scission reaction are a 2′–3′ cyclic phosphate, and a 5′ hydroxyl group, the host ligase must utilize these two functional groups to circularize the genomic and antigenomic RNAs. Although host machinery necessary for this step has not been identified, an enzyme with similar activity has been identified in plants and the chordate *Branchiostoma floridae*, and the bacteriophage T4 kinase/ligase system serves as the classical example of such a reaction.[30–32]

An interesting aspect to the rolling-circle replication scheme employed by HDV is how the virus stops the self-cleavage activity of the ribozyme during times when the viral RNA is acting as the template for RNA synthesis. The mechanism that prevents subsequent scission of the HDV ribozymes once they have been ligated with their 3′ termini is not known, but a recent study of a bacterial RNA editing system provides a potential mechanism in which a 2′–3′ cyclic phosphate is a substrate for an enzyme complex that hydrolyzes the phosphate and methylates the 2′ hydroxyl group before ligating the cleaved strands together.[33] Methylation of the 2′ hydroxyl prevents formation of the oxyanion that initiates the attack on the scissile phosphate during self-scission, thus preventing subsequent

cleavage reactions. Such RNA editing could provide the chemical stability necessary for the maintenance of circular genomic and antigenomic RNAs.

3. HDV RIBOZYMES

3.1. Ribozyme structure

Both genomic and antigenomic HDV ribozymes form intricate secondary structures composed of five helical regions (P1, P2, P3, P1.1, and P4) that form a nested double pseudoknot (Fig. 4.2).[28,34–36] The first pseudoknot is formed between the P1 and P2 helices, and within this pseudoknot a second such structure is formed between the P3 and P1.1 regions. This arrangement

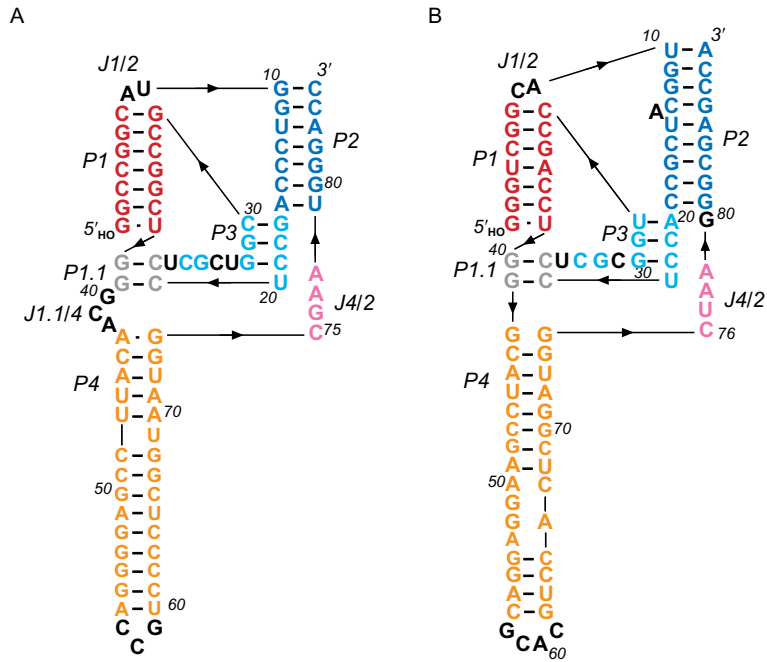

Figure 4.2 Secondary structures of (A) genomic and (B) antigenomic HDV ribozymes. Different helices are colored accordingly: red for P1, dark blue for P2, light blue for P3, gray for P1.1, and orange for P4. The single-stranded portion containing the catalytic cytosine is pink. Black nucleotides denote unconserved, single-stranded joining regions. Arrows connecting nucleotides indicate the direction of the strands, and the 5' OH marks the 5' end of the ribozyme following self-scission.

is further stabilized by helical stacking in which the P1, P1.1, and P4 helices form one coaxial stack and the P2 and P3 helices form a second. Helix length for the P1 and P3 regions is identical between the genomic and antigenomic forms of the ribozyme; however, the P2 helix is several base pairs longer in the antigenomic ribozyme. Less sequence conservation is observed for the P4 helix, and mutational studies indicate that this region can be completely removed with only a 100-fold loss of catalytic activity.[37] This feature was exploited to first crystallize the ribozyme by replacing the P4 helix with a protein binding stem-loop to improve crystal quality.[38]

Linking the helical structures is single-stranded regions that are crucial for the catalytic activity. These regions are named according to the helices that are linked (e.g., J1/2 refers to the "joining" region between the P1 and P2 helices). Whereas sequence conservation between the genomic and antigenomic forms of the ribozyme within the helical regions of the molecule is low, several positions in the joining regions are highly conserved, and mutational studies have shown them to be essential for catalysis (Fig. 4.2).[39,40]

The catalytic core of the ribozyme is formed around the P1.1 and P3 helices and is composed of the L3 and J4/2 regions. In the genomic structure of the ribozyme, the P3 helix is extended by invariant U20 and G25 nucleotides that form a reverse wobble pair and stack on another conserved nucleotide, C24.[36] Substitutions to either U20 or G25 that force Watson–Crick base pairing result in ribozymes with markedly decreased activity.[39] Some variability is permitted in the L3 core region, as the number and composition of nucleotides flanking the C24 and G25 residues differs between the genomic and antigenomic HDV ribozymes.

The J4/2 region of the ribozyme can range somewhat in length, but requires a "CNRA" motif within it to support self-cleavage.[41–44] When this motif is properly positioned in the genomic ribozyme, the N4 of the conserved C75 residue forms two hydrogen bonds, one with a nonbridging oxygen atom of the phosphate backbone linking the C21 and C22 residues that form the 5′ side of the P1.1 helix, and a second to the 2′ OH of U20.[34–36] These interactions lock the J4/2 region into the active site of the ribozyme. Mutating the genomic C75 or antigenomic C76 residue to a U or G results in a loss of catalytic activity.[39,40,44–46]

The J4/2 strand is further stabilized by the A78 residue of the genomic ribozyme via an A-minor tertiary interaction with the C18 and G29 residues that form the middle base pair of the P3 helix.[34–36,47] A purine nucleotide is required 5′ to A78 to allow for hydrogen bonding between its Watson–Crick face and the 2′ OH of the C19 nucleotide at the base of P3.[34–36]

The interactions result in a sharp turn in the genomic ribozyme structure from nucleotides G74 to A77 that forces C75 into close proximity with the cleavage site while simultaneously jutting the nucleobase of the unconserved residue at position 76 into solution.

3.2. Ribozyme folding

The HDV ribozymes form a tightly packed core structure that is solvent-inaccessible in both *cis*- and *trans*-acting constructs.[48] Kinetics of folding the genomic and antigenomic HDV ribozymes into the nested double pseudoknots have not been described in detail, but the fast self-cleavage kinetics and thermodynamic stability imply that the compaction is rapid. The stability of the ribozymes is particularly striking given that in the absence of Mg^{2+}, the ribozymes, particularly the genomic version at 0.4 M NaCl, do not reach a plateau of their melting curve at 95 °C.[49] At low concentration of NaCl and in the absence of Mg^{2+}, the ribozymes do reach a melting curve plateau at 95 °C, indicating that the ribozymes can be fully thermally denatured in low ionic strength solvents.[50] The HDV ribozymes are active in the presence of high concentrations of denaturants (18 M formamide or 8 M urea), which likely help resolve misfolded ribozymes, as a higher fraction of the ribozymes self-cleave at high urea concentration.[2,50–52] In a measurement of secondary structure content using ethidium fluorescence, both the precursor and product forms of the antigenomic HDV sequence have been shown to retain 50% secondary structure in 18 M formamide (95% vol), whereas a single-stranded DNA of the same sequence showed much lower stability.[53] Similarly, studies have shown that the HDV ribozymes are more active above physiological temperature, with cleavage rate constants increasing up to 55 °C even for *trans*-acting ribozymes.[54]

Ribozyme activity at elevated temperatures and denaturant concentrations can be attributed to the denaturation of misfolded molecules. Such kinetically trapped alternative conformations can form through improper base-pairing interactions of flanking sequences with the ribozyme core or even within the cores.[50,55–61] Alternative conformations result in slow overall cleavage rates, biphasic kinetics, or high fractions of uncleaved ribozymes. Mutations that destabilize these ribozyme-disrupting interactions tend to result in faster cleaving ribozymes. These results, however, are artifacts of sample preparation, whereas ribozymes studied under cotranscriptional conditions or purified without denaturation tend to exhibit fast, monophasic self-scission and uniform hydrodynamic radii, respectively.[62,63] Overall,

these data suggest that the ribozymes are properly folded and highly active under *in vivo*-like conditions, but can sample many inhibitory conformations resulting from alternative base-pairing interactions.

4. HDV RIBOZYMES' REACTION MECHANISM

Like other small self-cleaving RNAs, the HDV ribozymes promote self-scission through a nucleophilic attack by a 2′ hydroxyl group of the −1 nucleotide on the adjacent phosphate. The reaction yields a 2′–3′-cyclic phosphate on the upstream (−1) nucleotide and a 5′ OH moiety on the downstream product (formally the first nucleotide of the ribozyme).

4.1. Active-site groups

The mechanism of HDV ribozyme self-scission has been one of the most intensely studied enzymatic reactions of the last decade. Initial mutagenesis, crosslinking, and phosphorothioate interference studies implicated J4/2, particularly the genomic C75 (C76 in antigenomic ribozyme) residue, J1/4, and P3/L3 as key regions for catalysis.[40,48,64,65] Crystal structure of the genomic ribozyme product afforded the first glimpse of the active site at atomic resolution and provided two surprises: lack of electron density corresponding to a magnesium ion in the active site, and a nucleobase (C75) located within potentially hydrogen bonding distance of the G1 5′ hydroxyl group, which is the leaving group of the reaction.[34] The authors thus concluded "that C75 acts as the general base that activates the 2′-hydroxyl group of nucleotide −1 for nucleophilic attack … the negative electrostatic potential in the region of C75 (resulting from the trefoil turn), could perturb the pK_a values of C75, making it basic."[34] Besides C75, the active site consists of the L3 ribose-phosphate backbone between C22 and U23 and the U20–G25 pair, which in the most recent and highest resolution crystal structure of the inhibited ribozyme precursor forms a reverse wobble pair and coordinates a divalent metal ion.[36]

The proposal that a nucleobase could act directly in proton transfer during a transesterification reaction motivated a large number of studies aimed at deciphering the catalytic mechanism of these ribozymes. All of the studies implicated the cytosine in proton transfer during the rate-limiting step of the mechanism. This was the case for both *cis*- and *trans*-cleaving constructs, as well as ribozymes acting on nonnatural substrates including 2′–5′ phosphodiesters and 5′ phosphorothiolates.[66,67] The arguments for a direct role of C75 or C76 in the self-scission of genomic or antigenomic ribozymes,

respectively, can be summarized as follows: (1) in all crystal structures, the nucleobase is located within several Ångstroms of the scissile phosphate[34–36]; (2) the C75/76U and C75/76G variants do not support catalysis and C75/76A shows only partial activity[45,46]; (3) the kinetic pH profile of the C75/76A variant shows an apparent pK_a shifted by a number of pH units corresponding to the difference between the pK_a values of free cytidine and adenosine[46]; (4) substitution of the active-site cytosine with cytosine analogs similarly shifts the apparent kinetic pK_a by approximately the same amount as the pK_a difference between the free bases[67,68]; (5) self-scission of the inactive C75/76U or deleted C75/76 at this position can be rescued by exogenous nitrogenous bases including free cytosine, imidazole, and other analogs with the kinetic pK_a values tracking the pK_a values of the free bases[45,69,70]; and (6) proton inventory experiments and solvent kinetic isotope effects of the C75, C75A, and imidazole-rescued inactive ribozymes support a proton transfer from this position in the ribozyme (Fig. 4.3).[46,70,71]

For the majority of studies conducted on the HDV ribozymes, the active-site cytosine could be assigned a proton-transferring role as either an acid or a base. However, three experiments strongly implied that the nucleobase acts as a general acid in the mechanism. First, a sulfur substitution of the 5′ bridging oxygen of G1 produced an activated leaving group that no longer required the participation of the nucleobase in catalysis.[67] Second, a crystal structure of an inhibited ribozyme located the C75 residue within hydrogen bonding distance of the leaving group where it would act as a general base, as would be the case with the C75U mutant, rather than in proximity to the 2′ OH nucleophile.[35,36] Third, the rate of self-scission in the absence of Mg^{2+} results in a pH profile and a kinetic solvent–isotope effect that support transfer from a protonated cytosine.[46,49,71–73]

Kinetic pH profiles, proton inventory experiments, and Brønsted analysis indicated that at least one proton is transferred during the rate-limiting step of the reaction, most likely the one from a protonated C75/76 to the oxyanion leaving group of G1.[69,71,74] In order for the cytosine to efficiently transfer a proton to the 5′ oxyanion leaving group, its pK_a needs to be shifted from approximately 4.2 toward neutrality. Several biophysical experiments have been designed to measure the equilibrium pK_a of the active-site cytosine in various ribozyme constructs. Raman crystallography detected the protonation state of inhibited precursor forms of the ribozyme (2′-deoxy, 2′-methoxy, or 2′-fluoro substitutions at the −1 position) and yielded a pK_a of 6.1–6.2 and 6.4 in 20 and 2 mM Mg^{2+}, respectively, for the 2′-methoxy inhibited ribozyme.[75] The RNA construct used was similar

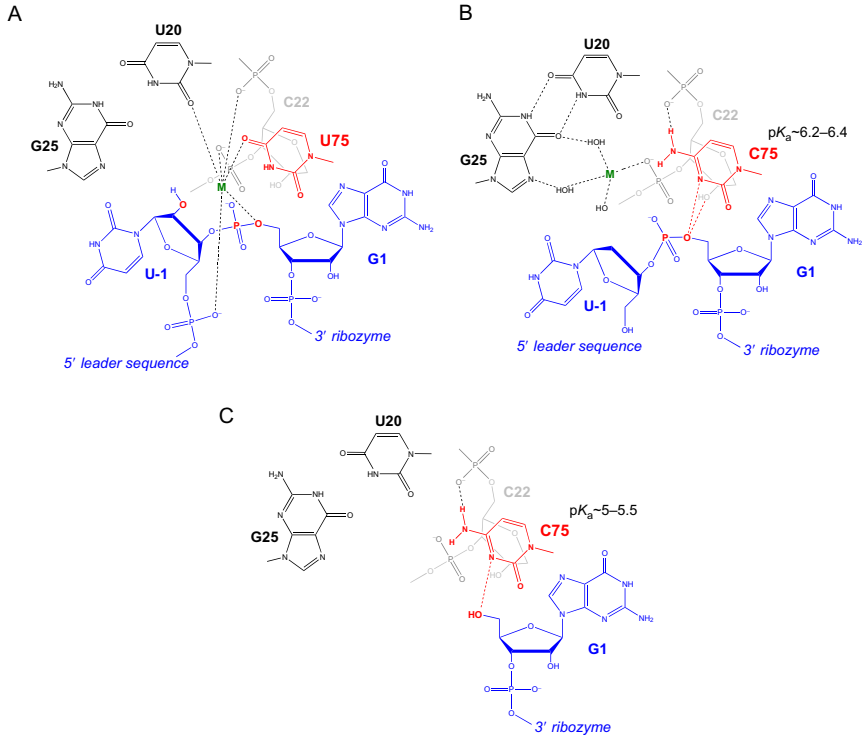

Figure 4.3 Active-site structures of the genomic HDV ribozyme. (A) Model of the C75U variant in precursor form.[31] (B) Model of the wild-type C75 with inhibited substrate strand lacking the U-1 2′ hydroxyl nucleophile.[33] The pK_a of C75 was measured using Raman crystallography.72. (C) Active-site groups in the product form of the ribozyme with the pK_a of C75 measured by NMR.[33,73] The substrate strand is colored blue, the active-site metal ion is green, and the catalytic nucleobase is red. The nucleophile, phosphorus of the scissile phosphate, and the 5′ oxygen of the leaving group are also colored red.

to that used for structure determination by X-ray crystallography, affording a rare combination of spectroscopic and diffraction data on nearly the same molecule.[36] When the protonation of the C75 nucleobase in the self-cleaved ribozyme was measured using solution NMR, the pK_a was estimated to be higher than that for free or base-paired cytosine, but not as high as the apparent pK_a of the reaction.[76] The authors concluded that while the active site of the product caused a shift in C75 pK_a, protonation of the nucleobase at neutral pH must be associated with an earlier step in the mechanism. Attempts to measure the pK_a by NMR in several inhibited forms of the

ribozyme failed due to spectral heterogeneity. This result led to the conclusion that a protonated C75 is likely associated with the phosphorane intermediate (or transition state) of the reaction, thus supporting a general acid or an electrostatic role for the nucleobase. Similar results were obtained by measuring the fluorescence of aminopurine placed next to the active-site cytosine.[77]

4.2. Metal ion requirements

A variety of divalent metal ions accelerate the ribozyme cleavage reaction. Self-scission is observed upon incubation with Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Mn^{2+}, Co^{2+}, and to a lesser extent Cd^{2+} and Ni^{2+} ions, but not in the presence of cobalt(III) hexammine, an exchange-inert structural analog of hexahydrated divalent metal ions.[2,10,12,27,35,51,66,78,79] Self-scission is also observed at high concentrations of monovalent ions, but the rate constants are more than two orders of magnitude lower than in the presence of divalent metal ions.[46,73] These observations suggest that divalent metal ions promote catalysis, but the lack of metal ion preference coupled with cobalt hexamine inhibition suggests that a hydrated form of divalent metal ion is required for efficient self-cleavage. Sulfur substitution of bridging and nonbridging oxygens on the scissile phosphate does not switch the metal ion preference toward thiophilic metal ions, suggesting that these positions are not directly contacted by the metal ions acting as Lewis acids.[80] These observations cannot rule out inner sphere coordination by a metal ion to other active-site groups. However, such coordination would require the active site to be able to accommodate multiple different metal ions.

Metal ion accommodation by the active site is further supported by experiments showing that changing the scissile phosphodiester configuration from 3′–5′ to 2′–5′ resulted in a change of preference from Ca^{2+} to Mg^{2+}.[66] While the 3′–5′ phosphodiester is cleaved somewhat faster in Ca^{2+}, the 2′–5′ linkage is cleaved two orders of magnitude slower in Mg^{2+}, and in the presence of Ca^{2+} the rate drops to background levels. Interestingly, the apparent pK_a values of the reaction in Mg^{2+} are independent of the type of phosphodiester undergoing scission.

Analysis of the divalent metal ion dependence in the presence of 1 M NaCl revealed that the genomic HDV ribozyme binds a structural Mg^{2+} with higher affinity than the active-site Mg^{2+}.[72] The active-site metal ion appears to act on the 2′ nucleophile because the apparent K_D of Mg^{2+} does not change between C76 and C76U variants when the 5′ leaving group is

activated.[67] A recent crystal structure of an inhibited precursor of the genomic ribozyme provided electron density at sufficiently high resolution so that the hydration of an active-site Mg^{2+} ion could be modeled. The structure shows that the metal ion does form inner sphere contacts, although not to the oxygens on the scissile phosphate, but rather to the phosphate between C22 and U23.[36] This structure was determined using a 2′ deoxy form of the −1 nucleotide, which together with the scissile phosphate is disordered in the crystal. Therefore, the ligands of the active-site metal ion could not be fully described. A previous crystal structure of a precursor ribozyme contained a ribose at the −1 position, but an active-site mutation caused rearrangement in the network of hydrogen bonds around the scissile phosphate.[35] Metal ions observed in the active site included Sr^{2+}, which made inner sphere contacts with the phosphates of C22 and U23, the base of U20, and with the 5′ oxygen of G1. Tl^{+}, a K^{+} isostere, was observed making an inner sphere contact with both the 2′ hydroxyl of the −1 nucleotide and the scissile phosphate.[81] The active site contained two other Tl^{+} ions, which were displaced by a single $Co(NH_3)_6^{3+}$ when it was soaked into the crystals. In the crystal structure of the ribozyme product, the active site appears to accommodate two different weakly bound metal ions at slightly distinct locations.[82] All together, the data support a model for the active site capable of binding several different metal ions that most likely activate the 2′ nucleophile by promoting its deprotonation.

Crystal structures of the precursor forms of the genomic HDV ribozyme show a poorly ordered leader sequence (the 5′ product), with only the sugar–phosphate backbone electron density observable in the most ordered case.[35,36,81] It is clear that the RNA strand has to make a sharp turn at the cleavage site to accommodate the leader, which makes no base-specific contacts with the ribozyme. This observation correlates well with the lack of sequence conservation upstream of the cleavage site among HDV-like ribozymes, although the identity of the −1 nucleotide affects the metal ion preference of the active site, suggesting that the base may affect the precise orientation of the scissile phosphate in the active site.[83] Terbium(III) footprinting of the genomic ribozyme revealed that the antigenomic ribozyme undergoes a conformational change between the precursor and products states and that the identity of the −1 nucleotide affects the metal ion accessibility of the groups lining the active site, suggesting that the active-site conformation is influenced by the 5′ sequence.[77,84,85] The models derived from the crystal structures show that the ribozyme does not undergo significant conformational change during the reaction. In solution, however, a FRET-based assay measured a significant change in the overall shape

of the molecule before and after the reaction.[77,86,87] When the ribozyme is labeled at the top of P2 and bottom of P4 helices, an inhibited precursor ribozyme showed compaction upon Mg^{2+} binding, whereas the product form showed extension of the molecule (approximately 50 and approximately 66 Å, respectively). These results suggest that the presence of the leader sequence in the precursor ribozyme causes bending in the structure such that the distal ends of P2 and P4 are closer before the reaction and further apart in the product form. Similar data were obtained for a *trans*-cleaving version of the ribozyme.[88] Interestingly, mutation of the active-site cytosine affects the overall shape of the molecule. Mg^{2+}-dependent conformational changes differed for the C75 and C75U variants, where the wild-type ribozyme showed shorter P2–P4 end-to-end distance in both precursor and product states.[87]

5. RIBOZYMES OUTSIDE OF THE HEPATITIS DELTA VIRUS

5.1. The *CPEB3* ribozyme

The first HDV-like ribozyme isolated outside of the virus was found embedded within the second intron of the mammalian gene encoding cytoplasmic polyadenylation element binding protein 3 (*CPEB3*).[2] Although the mammalian *CPEB3* ribozymes are identical to the HDV molecule in terms of secondary structure, they are quite divergent at the sequence level. Identification of the *CPEB3* ribozyme was thus not possible with sequence-based approaches. Instead, a novel *in vitro* selection technique was employed to isolate self-cleaving molecules from their genomic sources.

5.2. Selection scheme for *CPEB3* ribozyme identification

In a typical *in vitro* selection, a large pool of either synthetic or genomic sequences containing known regions flanking a tract of random nucleotides is evolved *in vitro* to contain a very small subset of the starting sequences that exhibit a desired trait.[89–91] This trait can be any selectable molecular phenotype, from ligand binding to enzymatic catalysis, so long as there exists a mechanism to distinguish between molecules with and without the chosen activity.

For selections of ligand-binding RNA or DNA sequences, discrimination between active and inactive sequences can be achieved simply by flowing the pool through a column containing the target ligand. Sequences unable to bind the target are washed through, whereas those capable of binding can be competitively eluted from the column.[92] This activity-enriched pool is then either directly polymerase chain reaction (PCR) amplified via

the known flanking regions in the case of DNA, or reverse transcribed and then PCR amplified for RNA molecules, and the process is repeated until a satisfactory level of enrichment is attained.

In order to isolate sequence-independent self-cleaving molecules using an *in vitro* selection, a fundamentally different approach is required to amplify active sequences. Because self-cleaved RNAs are covalently bisected, isolated sequences can no longer be directly amplified for use in subsequent rounds of selection unless the cleavage site is known *a priori*. This dilemma can be overcome in one of two ways. First, the RNA library can be circularized following transcription but prior to its incubation in media that support self-cleavage. After self-scission, the RNA will exist as an intact, but linear, molecule that can be enzymatically ligated and then amplified for use in subsequent rounds of the selection.[93,94]

Alternatively, the DNA can be circularized, and transcription of such a template enables a rolling-circle mechanism to yield concatemeric RNA molecules that each contains multiple copies of a single pool RNA sequence.[2,95–97] In all but monomer length transcripts, both flanking primers and promoter regions will be present in the proper orientations for productive reverse transcription and amplification to regenerate the DNA pool for subsequent rounds of selection.

The human *CPEB3* ribozyme was isolated using the rolling-circle transcription technique on a circular pool constructed from digested human genomic DNA (Fig. 4.4).[2] These 150 nucleotide genomic DNA fragments were ligated to primers, amplified using biotinylated forward primers and 5′ phosphorylated reverse primers, immobilized on a streptavidin column, and subsequently denatured to elute the phosphorylated reverse strand. The single-stranded sequences were splint-ligated, and the splint oligonucleotide was extended by DNA polymerase to yield nicked, double-stranded DNA (dsDNA) circles for use in rolling-circle transcriptions.

To select for self-cleaving ribozymes, the dsDNA pool was transcribed in a rolling-circle fashion to yield the multimeric RNA. These concatemers were gel-purified, incubated for 1 h in a buffer conducive to self-cleavage (10 mM Mg^{2+}), and the bands corresponding to self-cleaved dimeric RNA species were gel-purified again. Reverse transcription and PCR amplification with the biotinylated and phosphorylated primers restored the DNA pool for subsequent rounds of the experiment. After several cycles of selection, robust self-cleaving sequences came to dominate the population, among them the HDV-like fold found in the *CPEB3* gene.

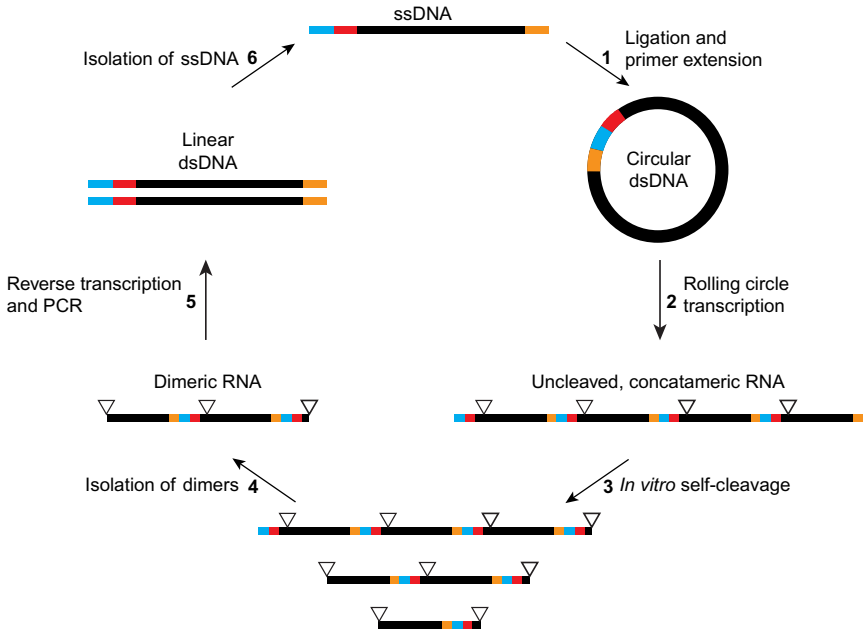

Figure 4.4 Selection scheme for the *in vitro* isolation of self-cleaving ribozymes. A single-stranded DNA pool containing a T7 promoter region (blue), constant primer regions (red and orange), and a stretch of random nucleotides (black) is (1) ligated and primer extended, followed by (2) rolling-circle transcription to yield concatemeric RNA molecules. These RNAs are then (3) incubated in buffer containing Mg^{2+}, wherein active sequences will self-cleave from concatemers to monomers (cleavage sites indicated by triangles). Dimers, which will contain both the forward and reverse primers in proper orientation, are then (4) isolated and (5) subjected to reverse transcription and PCR. The DNA is then (6) denatured to yield the full-length single-stranded DNA for use in subsequent rounds of selection.

5.3. Structural and enzymatic characteristics of the *CPEB3* ribozyme

A genomic selection for self-cleaving ribozymes is capable of revealing all self-cleaving sequences in a genome active under the selection conditions, but it does not provide any information as to the type of motif that is responsible for the reaction. To achieve the latter objective, extensive characterization of the identified sequences is required. Consequently, the secondary structure is only known for the motif residing in the *CPEB3* gene, although four sequences capable of self-scission were found in the human genome.[2]

The *CPEB3* ribozyme possesses several of the hallmark features associated with HDV ribozymes. *CPEB3* is able to catalyze its self-scission in a wide range of divalent metal ions including Mg^{2+}, Mn^{2+}, Ca^{2+}, and

Co^{2+}.[2] However, no reactivity was observed in the presence of monovalent cations, even at high concentrations (3 M Li^{+}), or in the presence of cobalt hexamine. Folding of the *CPEB3* ribozyme into an HDV-like secondary structure is further supported by data from phosphorothioate interference mapping. In HDV, several key interactions are made between the nonbridging oxygen atoms and the nucleotides essential to catalysis.[34–36] When the *CPEB3* ribozyme sequence is fit to an HDV-like secondary structure, the phosphorothioate substitutions that disrupt its scission correspond to the same interference positions in the HDV ribozyme.[2,65,80] The *CPEB3* sequence also exhibits a flat pH profile around pH 7, as is observed in both HDV ribozymes under similar conditions.[2,45]

Sequence-based searches using the human *CPEB3* sequence revealed a remarkable conservation of this motif across mammalian *CPEB3* genes (Fig. 4.5).[2] Corresponding motifs were not observed among analogs of the *CPEB3* gene in nonmammalian eukaryotic organisms.

Although the mammalian *CPEB3* ribozymes share a high degree of sequence conservation, the variations present have profound effects on the reaction. All mammals other than humans contain a C–G base-pair at the top of the P1 helix, whereas humans have a C · A mismatch at this location (Fig. 4.5). This might at least in part explain the slower cleavage rate observed for the human *CPEB3* ribozyme. The ribozyme found in *Canis familiaris* is most likely inactive because it contains a cytosine in lieu of the requisite guanosine at the cleavage site. This isolate also contains a single base-pair mismatch in the P2 helix and a U–G wobble pair at the top of P3. In the bovine ribozymes, a mutation at the base of the P3 helix transforms a G–C base pair into an A · C mismatch that would be expected to destabilize the overall structure and significantly decrease the self-cleavage rate. Several other mutations are observed among other mammalian *CPEB3* ribozymes; however, most of them occur in regions that are known to be nonessential for catalysis or that covary in support of the secondary structure (Fig. 4.5).

Despite sharing a common secondary structure, the catalytic rates between the human *CPEB3* and viral HDV ribozymes are highly divergent. HDV ribozyme catalyzes its self-scission with a $t_{1/2}$ approximately 3.5 min in 2 mM Mg^{2+} at 37 °C, whereas the human *CPEB3* ribozyme has a $t_{1/2}$ on the order of 1 h in 5 mM Mg^{2+} at 37 °C.[2,51] This is at least partially attributable to the relatively weak P1.1 found in the *CPEB3* sequences. In all isolates, this helix consists of a single G–C base pair and a U · U mismatch. Mutating this U · U mismatch into a canonical U–A base pair increases the catalytic rate of the ribozyme by an order of magnitude.[2]

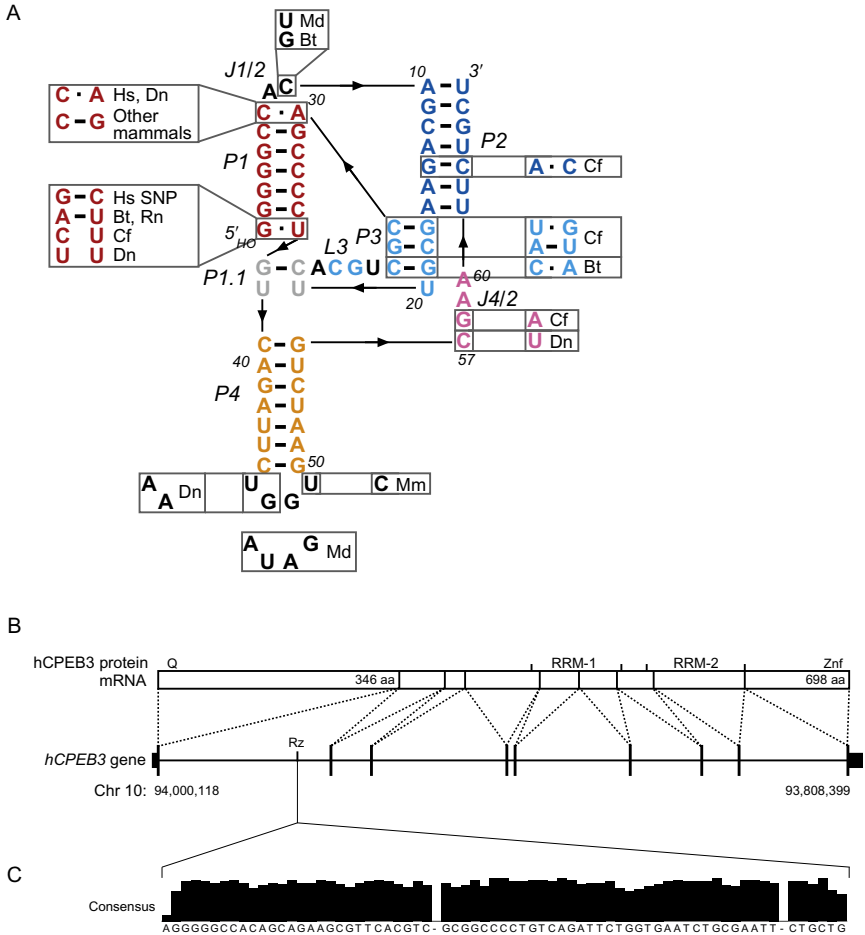

Figure 4.5 (A) Secondary structure of the human *CPEB3* ribozyme. SNPs and mutations of the ribozyme sequence in other mammals are indicated in boxes. Bt, cow; Cf, dog; Dn, armadillo; Hs, human; Md, opossum; Mm, mouse. (B) Genomic location of the human *CPEB3* ribozyme. (C) Consensus alignment of the *CPEB3* ribozyme found in various mammalian species.

Several studies have indicated that the catalytic rate of the HDV ribozyme can be modified upon the inclusion of 5′ and 3′ flanking sequences.[50,55,56,59,60] For instance, destabilization of the formation of alternate P1 and P3 helices by mutation of G11 to C and removal of the terminal U from the L3 region of genomic HDV enhances the catalytic rate by over three orders of magnitude.[60] The catalytic rates of the *CPEB3* ribozymes appear to be affected in a similar manner. Recently, the human version of the ribozyme was found

to cleave approximately 250-fold faster simply by mutating a single cytosine residue located −2 from the cleavage site to an A.[98] This mutation disrupts the formation of an alternate P1/P1.1 helix that extends the 3′ side of P1 helix through the P1.1 region, disrupting the nested double pseudoknot essential for activity. Although the roles that flanking regions play in ribozyme folding have currently been investigated only in the primate *CPEB3* motif, it is likely that similar upstream or downstream modifiers exist in other mammalian sequences that either speed up or slow down the cleavage reaction.

5.4. Expression and biological roles of the *CPEB3* ribozyme

The cytoplasmic polyadenylation element binding proteins are important for translation regulation, and *CPEB3* is specifically responsible for binding a novel translation-regulating signal that consists of a stem-loop with a U-rich bulge.[99,100] Expression of *CPEB3* occurs in a wide range of tissues and is increased in mouse hippocampus following kainate-induced seizure.[101] The *CPEB3* protein has been shown to repress translation of mRNAs containing its binding motif, perhaps because it interacts with features other than the conserved structures and proteins essential for *CPEB*-induced polyadenylation and thus prevents the mRNA from being polyadenylated and translated.[100]

The *CPEB3* ribozyme is located between the first and second coding exons of the *CPEB3* gene. This places it in the second intron of the gene because the 5′ UTR of the *CPEB3* mRNA is also spliced. Depending on the species, the ribozyme motif appears 10–25 kb upstream of the second coding exon. An RNA polymerase molecule moving at 2000–4000 nt/min would take 3–13 min to reach the second exon.[102,103] Although the initially reported *in vitro* cleavage rate (1 h^{-1} in 2 mM Mg^{2+} at 37 °C) indicates that the ribozyme-catalyzed self-scission reaction would not interfere with cotranscriptional splicing, recent data suggest that the *in vivo* reaction is much faster and could affect pre-mRNA stability before tagging or splicing of the second exon.[2,98] Indeed, a number of expressed sequence tags (ESTs) have been identified from *Mus musculus* (DT917655.1, DT914013.1, BY750066.1) and opossum (EC326704.1) that contain 5′ termini identical to the *in vitro* cleavage site of the ribozyme, providing strong evidence that the ribozyme is expressed and self-cleaves *in vivo*. The fact that these transcripts can be detected at all suggests that the ribozyme stabilizes the 5′ end of the transcript, protecting it from 5′ to 3′ exonucleases following the cleavage event. Furthermore, rapid amplification of cDNA ends (5′ RACE) experiments identified self-cleaved ribozymes in human testis and placenta, as well

as mouse testis, spleen, and brain tissues.[2] Although uncorroborated, it is likely that the ribozyme undergoes self-cleavage in human brain isolates as *CPEB3* is expressed at the RNA level in brain and likely plays a role in synaptic plasticity.[101,104] Ribozyme expression was also observed in mouse ovaries, kidneys, lungs, skeletal muscle, and bone marrow.[105]

Reverse transcription followed by quantitative polymerase chain reaction (RT-qPCR) performed on murine tissues revealed the cleavage state of the ribozyme to be tissue specific.[105] For example, in skeletal muscle *CPEB3* mRNA expression is high; however, the intron must be efficiently processed prior to ribozyme cleavage because no ribozyme-containing mRNAs are detected.[105] Alternatively, self-cleaved forms of the ribozyme are readily detected in ovarian and kidney tissues. Together, these data allude to the existence of an unknown regulatory mechanism for controlling both ribozyme-catalyzed self-scission and *CPEB3* expression.

Although the effects of the ribozyme self-cleavage are largely unknown, the *CPEB3* ribozyme appears to have an effect on episodic memory in humans.[106] A subset of human ribozyme isolates contain a single-nucleotide polymorphism (SNP) that changes the G–U wobble pair at the cleavage site to a G–C Watson–Crick pair. This mutation has been demonstrated to increase the overall rate of catalysis by approximately threefold.[2] Individuals homozygous for this SNP (rs11186856) experience decreased delayed episodic memory and specifically encounter difficulty recalling words and instances that have a high emotional valence (e.g., joy, poverty).[106] These findings suggest that the ribozyme may have a role in the neurobiology of emotional memory.

5.5. Bioinformatic identification of ribozymes

The *CPEB3* ribozyme is the first instance of an HDV-like self-cleaving RNA found outside of the namesake organism.[2,3] Other ribozyme motifs, specifically hammerhead ribozymes, have been readily detected in numerous organisms following their initial identification.[1,5,6,107–110] This contrast is likely due to several factors. First, the hammerhead motif is much less informationally complex than that of the HDV ribozymes. The genomic HDV molecule has an information content of approximately 50–55 bits, whereas a typical hammerhead sequence carries only 44 bits of information. This difference in complexity is reflected in current sequence data, where the hammerhead-like motifs are detected two orders of magnitude more frequently than HDV-like motifs when using comparable structure-based

search software on viral genomes.[3,111] With the advent of inexpensive DNA sequencing, however, searchable genomic space has expanded enough that the intricate HDV secondary structure can be detected in numerous loci.

The exponential increase in computer processing power has also facilitated the identification of complex RNA secondary structures on a broad scale. In a functional RNA molecule, it is typically the structure rather than the primary sequence that is conserved. In principle, a helix can form between any two nucleotides capable of base pairing, and the helical regions primarily dictate the overall structure that an RNA will assume. Consequently, two ribozymes with identical secondary structures can be unrecognizable on a sequence level. This is particularly striking for HDV ribozymes, where approximately 70 nucleotides are required to form the motif, but only 6 are invariant between the two structures (Fig. 4.2). Although sequence-based searches have been highly refined over the past few decades and have met with widespread success in protein-coding gene identification, their utility is thus limited in identifying catalytic RNAs. Furthermore, catalytic RNAs, unlike their protein counterparts, are not delimited by conserved translation initiation, termination, and splicing sites that can aid in the identification of protein-coding regions with similar function but different primary amino acid sequences. However, proteins with similar domain arrangements will often catalyze similar reactions, as do RNA molecules that share comparable folding patterns. Thus, a structure-based approach is more applicable to identify ribozymes.

Structure-based searches for folded RNA molecules were first developed in the late 1980s. However, early programs required user-defined helices, and as such they could not be used to identify unknown base-pairing regions from a random chain of nucleotides.[112–114] The RNAMOT program overcame many of the limitations associated with previous motif-searching software by first allowing a helix to form from any string of nucleotides and subsequently analyzing downstream portions for complementarity.[115] RNAMOT was used as a scaffold to develop additional structure-based search programs, and it and its successors have been repeatedly used to identify novel tRNAs, group I self-splicing introns, hammerhead ribozymes, and ligand-binding RNAs at new genomic loci.[108,111,116,117] Comparable searches for HDV-like ribozymes were not performed until after the first crystal structure was solved and the motif was identified outside of the virus in the mammalian *CPEB3* gene.

The combined effects of a complex secondary structure, limited searchable genomic space, and a wealth of more readily detectable RNA motifs have slowed the identification of novel HDV ribozymes. Nevertheless,

HDV-like ribozymes have now been identified in a plethora of organisms ranging from mammals, insects, fish, and other deuterostomes to plants, protists, bacteria, and other viruses.

5.6. Structure-based searches reveal additional HDV ribozymes outside of mammals

The presence of a conserved HDV ribozyme in mammalian genomes suggested that the motif might be found in other organisms. Although *in vitro* genomic selection could be used to identify the ribozyme in alternative organisms, the nature of the experiment is prohibitive to high-throughput searches for a single RNA motif. Rather, a bioinformatics methodology can be employed that is capable of analyzing a multitude of genomes rapidly for a particular catalytic structure.

Bioinformatic searches generally take two forms: sequence- or structure-based. As mentioned previously, the former has been employed extensively in mapping organisms' proteomes, while the latter has achieved great success in identifying noncoding RNAs. In a structure-based search, the identity of nucleotide or amino acid at a particular location is not as important as the interactions in which the residue can participate, often with distant portions of the molecule. Whereas a DNA or RNA sequence-searching algorithm needs only to match the identity of a particular sequence point-by-point against the user-specified input, a structure-searching algorithm must first determine whether a particular structure element can form, and then whether that element can form within the folding space required to produce the entire motif.

Base pairs dominate the secondary structures of RNA and DNA, and these interactions should form the basis of any structure-based search. However, since a helix can form between any base-pairing nucleotides, and the 3′ sequence is not defined until the 5′ sequence has been specified, an iterative search method is required. A number of such iterative search tools exist including the RNABOB package, which was used to discover numerous HDV-like ribozymes.[118]

RNABOB is an extension of the earlier structure-based search program RNAMOT that permits the user to describe any RNA structure through an intuitive "descriptor" design.[115] Helices and single-stranded regions are independently defined in length, composition, and number of mismatches, and then they are ordered in relation to each other as they would appear in the secondary structure (Fig. 4.6). For helices, the user can define either a specific sequence or allow for any nucleotide, in which case the software will

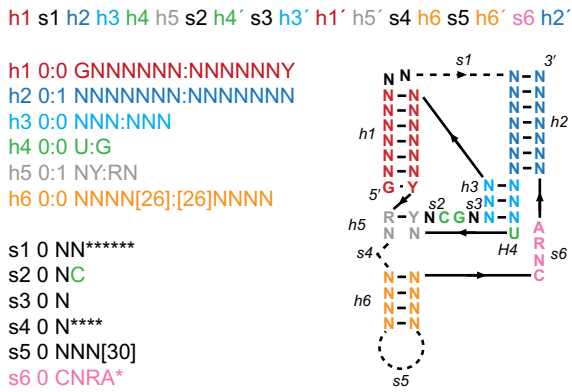

Figure 4.6 A typical RNABOB descriptor. Each motif that forms the secondary structure is individually defined on the left, and their relation to one another is specified in the single line appearing at the top of the figure. In the descriptor, an "h" designation corresponds to a helical region, and an "s" designation corresponds to a single-stranded portion of the molecule. When defining a helix, a number to the right of the colon indicates the number of allowed mismatches to that helix, whereas a number to the left or following an "s" designation defines the number of allowed variations to a defined sequence. Asterisks or bracketed numbers in the descriptor specify the maximum number of additional nucleotides a region can contain (e.g., s1 is at least two nucleotides but can contain as many as eight nucleotides).

search for regions of complementarity between any portions of the genome that are spaced appropriately to produce the gross structure of the motif.[119]

Following a genomic search, RNABOB outputs a file that contains all sequences capable of assuming the defined structure with each region defined in the descriptor individually delimited. This allows for easy parsing into spreadsheets, where the data can be further organized according to a particular region, such as the highly conserved core of the HDV ribozyme. Since optimal descriptors often allow for various mutations and insertions that independently may not disrupt proper folding but collaboratively lead to inaccurate structures, organizing the data in such a way allows suboptimal sequences to be quickly eliminated from the pool of candidate motifs.

5.7. HDV-like ribozymes found using RNABOB

A genome-wide search for sequences capable of assuming an HDV-like structure using the RNABOB software revealed a plethora of *in vitro* and *in vivo* active ribozymes in a wide range of organisms. The first HDV-like ribozymes identified through structure-based searches included several

unique families from the genomes of the mosquito *Anopheles gambiae* and the purple sea urchin *Stronglyocentrotus purpuratus*. Active motifs were also identified in the lamprey eel *Petromyzon marinus*, the lancelet *Branchiostoma floridae*, and two nematodes *Caenorhabditis japonica* and *Pristionchus pacificus*. Outside of eukaryotes, additional HDV-like ribozymes were found in the human gut commensal bacterium *Faecalibacterium prausnitzii*, and the insect *Chilo* Iridescent Virus type 6 (CIV).[3] These newly defined HDV-like motifs subsequently served as starting points for additional sequence and structure-based searches for HDV ribozymes.

5.8. The *A. gambiae* ribozymes

A number of self-cleaving sequences have been detected in the African malaria mosquito *A. gambiae*. When fitted to an HDV secondary structure model, the ribozymes can be differentiated into two classes, drz-Agam-1 and drz-Agam-2 (Fig. 4.7). Isolates of the drz-Agam-1 family are approximately the same length and structure as the HDV ribozymes, whereas the drz-Agam-2 family is farther removed from the canonical HDV secondary structure.

The J4/2 region of the drz-Agam-1 family of ribozymes is of particular interest due to the presence of a second cytosine adjacent to the catalytic cytosine. Self-scission is eliminated upon mutation of both of these residues to uridines; however, C to U substitution of just the upstream cytosine yields a construct that exhibits only residual activity (C.T. Webb, personal communication). This suggests flexibility in the HDV core, where both cytosines are potentially positioned to act in catalysis. However, the detailed mechanism of self-scission for this variant ribozyme has not been studied.

The drz-Agam-2 family differs from the drz-Agam-1 sequences in that it possesses a much larger P4/L4 domain and a large insertion in the J1/2 region of the secondary structure. The J1/2 regions of the genomic and antigenomic HDV ribozymes also vary in length, although not to the extent observed between the mosquito families. Among the drz-Agam-2 sequences, the J1/2 peripheral domain is approximately 50 nt in length, representing approximately one-third of the total sequence space for the structure. Computational modeling predicts that the J1/2 insertion forms extensive base-pairing interactions leading to the creation of a sixth helix in the secondary structure. The formation of the sixth helix likely enhances

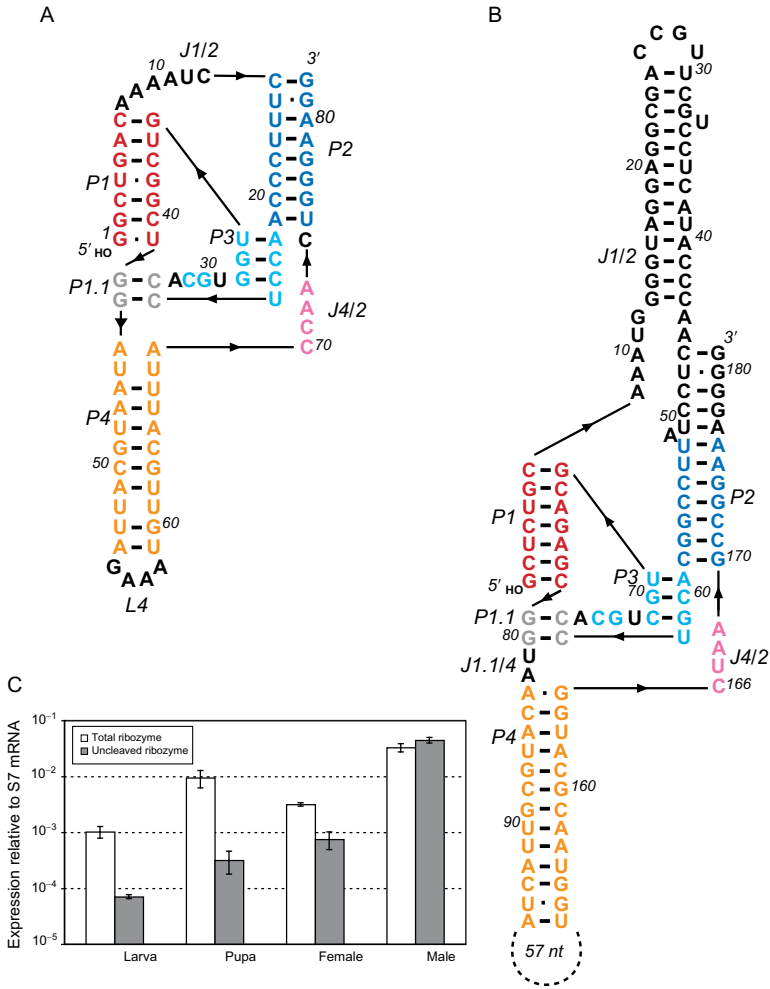

Figure 4.7 Secondary structures for ribozymes from the (A) drz-Agam-1 and (B) drz-Agam-2 families, corresponding to the drz-Agam-1-1 and drz-Agam-2-1 ribozymes, respectively. (C) *In vivo* expression data obtained by qPCR for the drz-Agam-1-1 ribozyme. Expression for other drz-Agam ribozymes follows a similar pattern.

overall folding of the motif, as such sequences exhibit rapid cleavage with low sensitivity to Mg^{2+} and temperature.[3]

The *in vivo* expression of the drz-Agam ribozymes was confirmed through 5′ RACE experiments and RT-qPCR. Expression was found to vary between sexes and life stages of the insect, with the highest expression

observed in adult male mosquitoes. Despite fast *in vitro* self-scission, the ribozymes found in adult male mosquitoes appear in their intact, uncleaved forms. However, self-cleaved forms of the ribozyme are readily detected in both pupal and larval stages of the insect and show lower overall expression relative to the adult, indicating that both transcription and self-scission events are regulated *in vivo*.

Unlike expression, which can be regulated on transcriptional or post-transcriptional levels, regulation of self-scission is likely due to the direct alteration of ribozyme structure to prevent formation of the catalytic core. Such ribozyme disruption might be achieved by binding an antisense transcript, protein, or small molecule effector, or by forming an alternative secondary structure with a distal element of the same transcript. Alternative folding, however, would require formation of a competing structure in a sex- and development-specific manner that would again likely involve nucleic acid, protein, or other effectors.

The genomic distribution of the drz-Agam ribozymes is generally not conserved, as the sequences map to introns and intergenic regions of the genome. However, BLAT analysis of the drz-Agam-2 sequences placed them at the 5′ end of the non-long terminal repeat (non-LTR) retransposable element (RTE).[3,120] This retrotransposon was first identified in *Caenorhabditis elegans*, and encodes an apurinic–apyrimidinic (AP) endonuclease and a reverse transcriptase protein.[121] In a typical non-LTR retrotransposition event, the 5′ end of the motif is often truncated due to inefficient or incomplete retrotransposition followed by insertion into the host genome. Thus, it is somewhat surprising to find the highly complex drz-Agam-2-2 sequence in its entirety forming the 5′ terminus of the element. That this sequence appears in its full-length, active form suggests that the ribozyme-catalyzed self-cleavage event is necessary for the viability of the RTE. Recently, HDV-like ribozymes have been found at the beginning of additional classes of transposable elements, supporting a model in which self-scission by the ribozyme is responsible for the cotranscriptional processing of the elements.[4]

5.9. R2 ribozymes

The best-understood class of retrotransposons found to be associated with HDV-like ribozymes is the R2 family.[122] This class of transposons inserts site-specifically into the 28S rDNA of most insect species. It encodes an endonuclease and reverse transcriptase protein for use in the

retrotransposition cycle. The mechanism of transposition begins with cotranscriptional processing of the element from the surrounding rRNA followed by endonuclease-catalyzed cleavage at the insertion site of the host DNA. Target-primed reverse transcription and second-strand synthesis reconstruct the retrotransposon DNA for insertion at the new locus.[123,124]

Until recently, the factor responsible for the cotranscriptional processing of the element was unknown, even though the processing was discovered over 25 years ago.[125] Recently, it has been confirmed that this event is due to the action of a self-cleaving ribozyme of an HDV-like motif (Fig. 4.8).[4] Rapid cleavage of the R2 sequence from the surrounding transcript was observed *in vitro*, indicating that the event is autocatalytic in nature rather than dependent on additional cellular factors. Subsequent work performed on truncated versions of the R2 transposon demonstrated that an approximately 200 nt region at the start of the sequence was essential to this process. When this region was fitted to various models of ribozyme secondary structures, it was found to readily assume a drz-Agam-2-like structure. This similarity between the R2 and drz-Agam-2 family of sequences also renders the R2 ribozymes readily identifiable in sequence- and structure-based searches using the drz-Agam-2 core.[126]

Sequence analysis of the ribozyme-containing regions of R2 elements revealed that areas of conservation among isolates map to the conserved core portions of the ribozyme such as the L3 and J4/2 segments, whereas divergent sequences are found in the variable portions of the ribozymes, and covariation is present between nucleotides that would form the base-paired helices in the folded RNA structure. Thus, a consensus sequence can be constructed in which the P3, P1.1, L3, and J4/2 portions of the ribozymes are identical, the J1/2 region varies in both length and composition, and the P1 and P4 helices contain mutations that maintain base pairing (Fig 4.8). In R2 ribozymes isolated from *D. melanogaster*, *D. sechellia*, *D. simulans*, and *D. ananassae*, the P2 helix is invariant, although several nucleotides are covaried in isolates from *D. falleni* and *D. pseudoobscura*.

The low sequence diversity in the P2 and P3 helices among most R2 ribozymes is somewhat surprising. For the P3 helix, this may be attributable to its proximity to the active site. In all families of HDV-like ribozymes, including the viral sequences, the nucleotide composition of this helix consists of an A–U base pair followed by two C–G base pairs, or simply three C–G base pairs.[2,3,34] In the genomic HDV ribozyme, only the cytosine of the middle C–G base pair (C18–G29) forms any contacts with the remainder of the structure, forming a hydrogen bond between the C18 O2 position

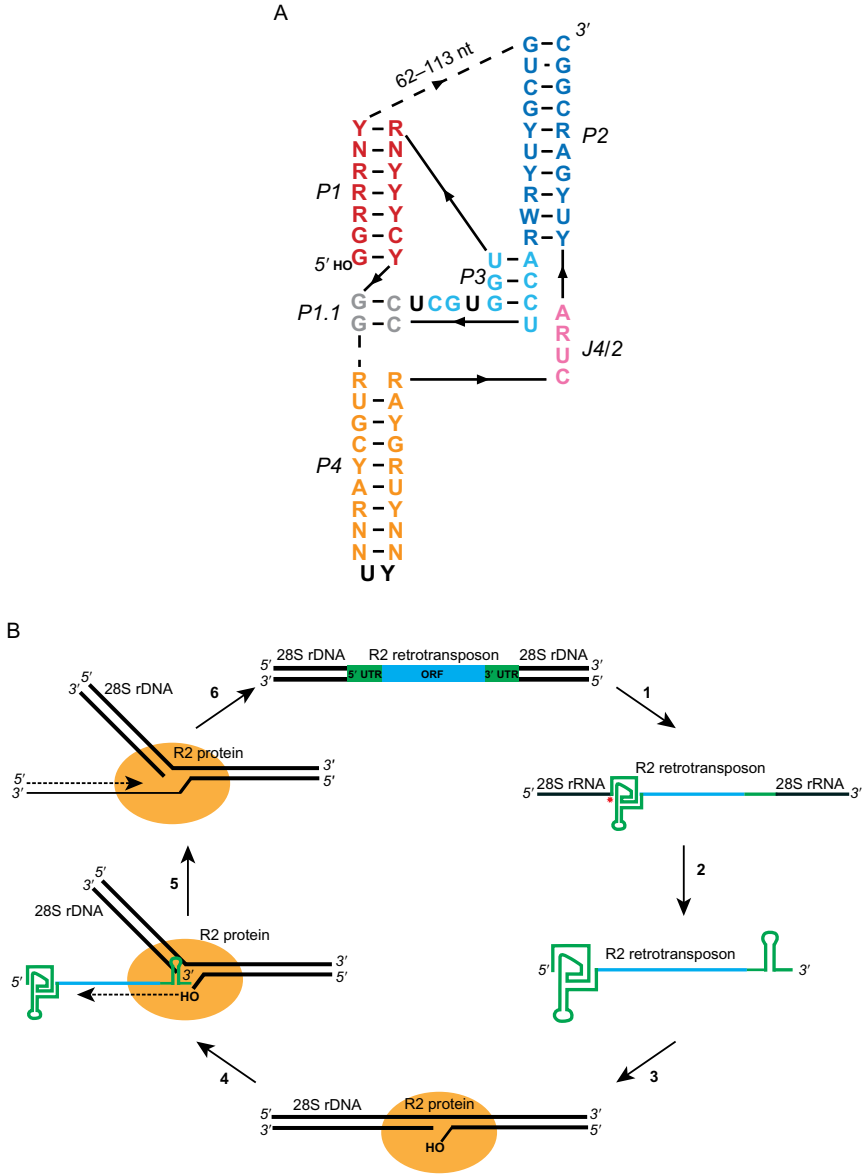

Figure 4.8 (A) The consensus sequence and secondary structure for the *Drosophila* R2 ribozymes with IUPAC nucleotide abbreviations. (B) The R2 retrotransposition event. R2 DNA is (1) cotranscribed from the 28S ribosomal locus (rDNA) to yield the ribozyme-containing R2 RNA embedded in the 28S rRNA. The R2 retrotransposon is then (2) excised from the surrounding transcript at the 5′ junction via the action of the HDV-like ribozyme that will cleave the RNA at the position marked with a red star and at the 3′ junction by unknown factors. Following terminal processing, the R2 RNA is (3) translated

(Continued)

and the 2′ OH of the conserved A78 residue in J4/2.[34–36] *In vitro s*election experiments have shown that variants that maintain base pairing in the P3 are tolerated, although wild-type sequences appear with much higher frequency.[44] Formation of a properly base-paired P3 is essential for catalysis; therefore, the observed conservation may instead reflect the rarity of two compensatory mutations occurring in this region in a single replicative event.[41]

Unlike the P3 helix, the P2 helix varies greatly in both nucleotide composition and length among the HDV-like motifs. In the R2 ribozymes, this region is potentially longer than the six to eight base-pair norm, suggesting, in conjunction with its high level of conservation, that it has an important role in catalysis or retrotransposition. Comparable conservation for this motif is not found among isolates from the drz-Agam-2 family.[3] The high levels of *in vivo* expression of the *Drosophila* R2 ribozymes provide ample opportunity for random mutations to arise, but that none of these mutations has accumulated in the P2 helix suggests an additional role for this motif.

The observed conservation of P2 in the R2 ribozyme may be linked to a role in translation. Since mRNA generated cotranscriptionally from a larger transcript via the actions of a self-cleaving ribozyme would not possess the 5′ methyl-guanosine cap that ensures stability of the message and promotes translation, other methods must be employed to achieve protection and begin translation. Resistance to degradation is represented at least in part by the unusual high stability of the HDV ribozyme structure, as folding and self-cleavage are observed even in high concentrations of denaturants for the viral sequences and *CPEB3*.[2,52,127] Translation initiation, however, can only be reconciled if the ribozyme is acting as an internal ribosome entry site (IRES). There exist several classes of IRESes with the underlying features that they form complex secondary structures, contain a recognizable codon, and are capable of starting translation in the absence of some or all of the canonical translation initiation factors.[128]

Recent evidence suggests that the retrotransposon ribozymes are indeed capable of acting as IRESes.[126] Constructs containing either wild-type or inactive C/U mutant ribozymes upstream of a luciferase reporter gene

Figure 4.8—Cont'd to produce the R2 protein, which nicks the target DNA and yields a free, 3' OH. The free 3' OH is used in (4) target-primed reverse transcription to create the R2 cDNA, which forms the template for (5) second-strand synthesis of the R2 DNA. The full-length R2 retrotransposon is then (6) reinserted into the genome to yield an active retrotransposon at a new 28S rDNA locus.

demonstrated levels of activity comparable to a known HCV IRES in both *in vitro* and *in vivo* translation assays. The amount of luciferase produced from the ribozyme constructs was markedly higher than the levels obtained from the HCV IRES *in vitro*, with the catalytically inactive mutant ribozymes showing the highest levels of expression. This difference was much less pronounced during *in vivo* experiments using transfected S2 cells, perhaps due to insect-specific translation factors missing from the rabbit reticulocyte lysate used for *in vitro* studies. Regardless, expression was found to vary among ribozymes, with the *D. simulans* sequences producing the highest amounts of luciferase.

In R2 ribozymes, the J1.1/4 and P4 junction contains the consensus sequence "RUG" that would code for either a canonical methionine as the first amino acid, or a valine residue. The lack of conservation observed at this presumed translation start site indicates that if this location is indeed responsible for translation initiation, then it likely does so in a noncanonical manner that does not depend on an AUG start codon.

Within the J4/2 region of the catalytic core, a "URA" stop codon can be found in-frame with the upstream RUG codon, and the small open-reading frame would yield a seven amino acid peptide. This small open-reading frame could serve to localize and prime the remainder of the mRNA for translation by the ribosome, as little nucleotide conservation is found in the P4 region between isolates. However, if the URA codon does not completely halt translation, elongation would proceed through the 3′ side of the P2 helix and incorporate a highly conserved amino acid sequence. With the exception of the R2 ribozyme from *D. pseudoobscura*, the mutations found in the P2 helix are all silent with respect to amino acid sequence. These observations thus support a model in which the ribozyme serves to initiate translation in a noncanonical manner during which the amino acids coded for by the P2 helix form the *N*-terminus of the R2 protein of the retrotransposon (Fig. 4.8B).

Although the R2 ribozymes exhibit minimal sequence variation, their catalytic rates are quite divergent.[126] In the HDV and *CPEB3* ribozymes, the upstream sequences have been shown to markedly decrease self-cleavage activity if they are capable of extending the P1 helix by base pairing with nucleotides required to form the P1.1 helix, thereby disrupting active-site formation of the molecule.[55,56,59,98] An identical event is observed among the R2 ribozymes. As the number of alternate base pairs that the upstream sequence can form is increased from zero to three, the rate constants decrease from approximately 80 to 0.30 per hour, respectively.[126]

5.10. Additional RT-associated ribozymes

Structure-based searches for HDV-like ribozymes within the EST database, known RT-containing mRNAs, and genomic repeat regions as defined by RepeatMasker revealed many more ribozymes that map to various non-LTR retrotransposons.[126] In addition to the R2 ribozymes in *Drosophila*, the R2 elements of the termites *Kalotermes flavicollis* and *Reticulitermes lucifugus*,[129] the sea squirt *Ciona intestinalis*,[130] the black-legged tick *Ixodes scapularis* (GenBank™ accession ABJB010506112), the horseshoe shrimp *Triops cancriformis* (GenBank™ accession EU854578), and the Zebra finch *Taeniopygia guttata*[131] were all found to harbor HDV-like ribozymes in their 5′ terminal region. Other rDNA-associated retrotransposons with ribozymes include the R4 element[132] in the human intestinal roundworm *Ascaris lumbricoides*, the horse intestinal roundworm *Parascaris equorum*, and the R6Ag1 and R6Ag3 elements in *A. gambiae*.[133] With the exception of the *A. gambiae* elements, all of these ribozymes possess the extended J1/2 region observed in the *Drosophila* isolates. Since this region can adopt numerous conformations without abolishing self-cleavage, conservation of the extended J1/2 region is likely attributable to each ribozyme's function as an IRES. The JAM1 RTE[134] from the yellow fever mosquito *Aedes aegypti*, the SARTPx2[133,135] element from the Asian swallowtail butterfly *Papilio xuthus*, and the Baggins retrotransposons from *Drosophila* are also under the influence of HDV-like ribozymes. Both the JAM1 and SARTPx2 associated ribozymes, like the R6Ag1 and Ag3 ribozymes from *A. gambiae*, lack the extended J1/2 region that is a hallmark of other rDNA-associated retrotransposon ribozymes. These ribozymes do, however, contain P4 helices that are longer than those seen in the *Drosophila* R2 sequences, perhaps in part to compensate for the loss of stability associated with a shorter J1/2 motif.

Several ribozymes have been identified upstream of regions that display high similarity to RT-coding motifs in a number of organisms, including *A. aegypti*, the butterfly *Heliconius numata*, the flatworm *Schistosoma mansoni*, the sea slug *Aplysia californica*, the common sea urchin *Paracentrotus lividus*, the Indonesian coelacanth *Latimeria menadoesnsis*, the sea lamprey *Petromyzon marinus*, the little skate *Leucoraja erinacea*, and the painted turtle *Chrysemys picta*.

5.11. HDV-like ribozyme cores similar to drz-Agam-1

In addition to drz-Agam-like ribozymes found in retrotransposons, a number of other sequences have been identified in genomes and ESTs from

various organisms using sequence-based search methods. Although sequence-based searching is generally challenging for folded, functional RNA molecules, additional HDV-like motifs were found based on their highly similar P3 and L3 regions relative to the drz-Agam-1 family of sequences.[3] In insects, ribozymes were found in the pea aphid *Acyrothosiphon pisum*, a second mosquito, *Anopheles funestus* (EST CD578134), and in several ESTs from the central nervous system of the insect *Rhodinus prolixus*, a vector for Chagas disease parasite *Trypanosoma cruzi* (e.g., EST FG545964). HDV-like ribozymes with identical P3 and L3 regions were also found in several fungi and plants.

In the fungus kingdom, the antifungal biocontrol fungus *Trichoderma atroviride* possesses a confirmed HDV-like ribozyme ($k_{obs} = 1.8\ h^{-1}$ in 10 mM Mg^{2+} at 37 °C), as does the mycosporadic fungus *Trichoderma virens*.[7] The red bread mold *Neurospora crassa* contains a ribozyme active *in vitro* and which appears self-cleaved at the 5′ end of an EST in the organism (EST GE968064).[7] Putative ribozymes are also found in the RNase H gene of the fungus *Ajellomyces capsultans* (EST XM_001543659), the filamentous fungi *Podospora anserine*, the invasive fungus *Neosartorya fischeri*, and the plant pathogenic fungi *Magnaportha grisae*, *Phaeosphaeria nodurm SN15*, and *Alternaria brassicicolla*.[7] Plant HDV-like ribozymes with identical cores are detected in ESTs from the artichoke *Cynara scolymus* (EST GE606767) and the sunflower *Helianthus annus* (EST DY925396).[3]

Ribozymes sharing the drz-Agam-1 core are also found in ESTs from the tapeworm *Moniezia expansa* (EST FE942699).[3] However, these ribozymes are likely inactive or under tight regulation *in vivo*, as the EST extends beyond the presumed cleavage site for the motif. An active HDV-like ribozyme is present in the free-living marine protist *Diplonema papillatum* (k_{obs} = approximately 100 h^{-1} in 10 mM Mg^{2+} at 37 °C), and resides between the splice leader for the 5S rRNA and 5S rRNA itself. Since the sequence exhibits robust *in vitro* activity, it is likely involved in the processing of noncoding RNAs in the organism. The pacific abalone *Haliotis discus* also contains an HDV-like ribozyme that is one of the smallest HDV-like sequences identified to date at just 58 nt.[7]

5.12. *Stronglyocentrotus purpuratus* ribozymes

Structure-based searching revealed multiple HDV-like ribozymes in the genome of the purple sea urchin *S. purpuratus* (Fig. 4.9). Based on the secondary structure, the ribozymes can be divided into four families of sequences, drz-Spur-1 to -4. This distinction is made largely based on the

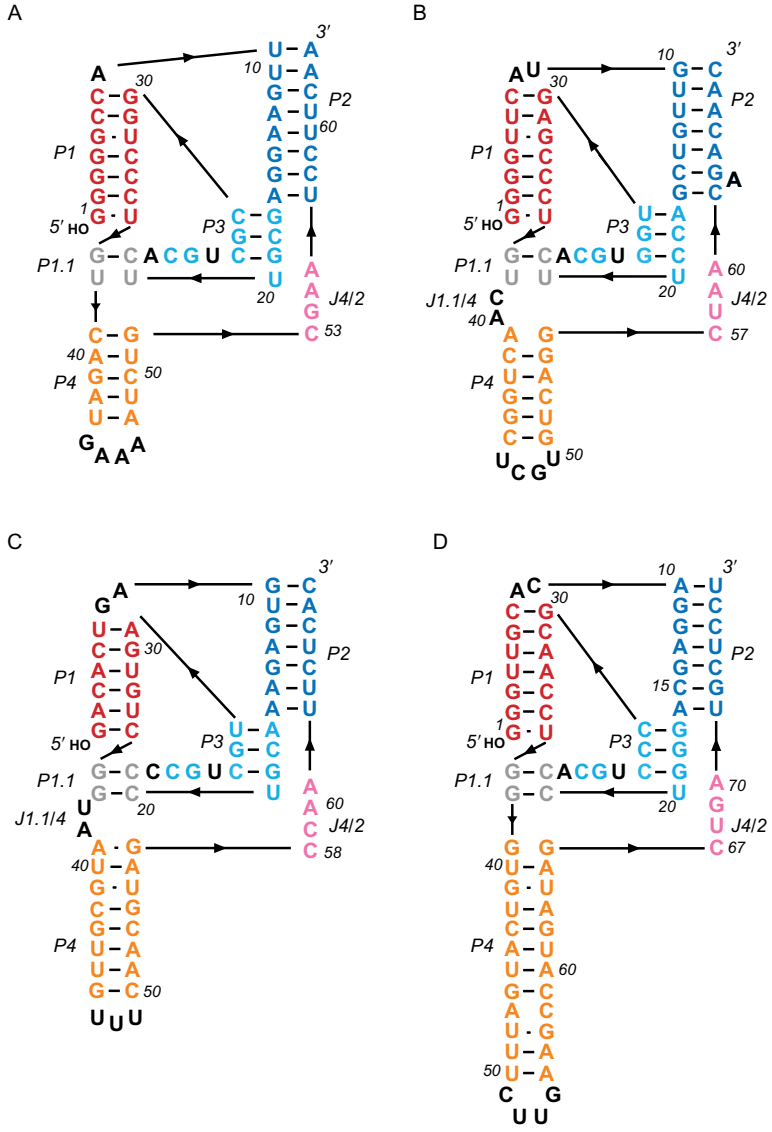

Figure 4.9 Secondary structures for the (A) drz-Spur-1, (B) drz-Spur-2, (C) drz-Spur-3, and (D) drz-Spur-4 ribozymes. Both drz-Spur-1 and drz-Spur-3 are found associated with predicted reverse transcriptase genes, suggesting a role in retrotransposition for these sequences.

length of the helices for the various families and the composition of the J4/2 region. A slightly longer P2 helix of 8 bp, a short, 5 bp P4 helix, no J1.1/4 region, and a J4/2 active site of CGAA composition characterize the drz-Spur-1 family. In the drz-Spur-2, -3, and -4 families of sequences, the P2 helix is 7 bp in length. Drz-Spur-2 is unique in that it also possesses a J1.1/4 segment and a J4/2 region of CUAA. The only other family of *S. purpuratus* ribozymes with a J1.1/4 region is the drz-Spur-3 family, which is the only known HDV-like ribozyme with a 6 bp P1 helix. The J4/2 region of this family is composed of CCAA. Much like in the drz-Agam-1 family of sequences, the second cytosine in this region also appears to function in catalysis, as self-cleavage activity is lost only if both cytosines are mutated. Finally, the drz-Spur-4 family is distinguished by its extended P4 helix and "CUGA" J4/2 region. In all of the drz-Spur families, multiple copies of the sequences are present throughout the sea urchin genome.

By far, the most common class of sea urchin ribozymes is the drz-Spur-1 family. BLAT and BLAST analysis of the drz-Spur-1 sequence identifies well over 30 full-length copies dispersed throughout the genome. One copy, which is transcribed *in vivo* (EST CD324081), is found in the first exon of a predicted non-LTR retrotransposase gene. Given the recently identified role of HDV-like ribozymes in the cotranscriptional processing of retrotransposons, it is likely that the drz-Spur-1 ribozyme plays a crucial part in the transposition event.[3] Association with such an element might also explain the much higher copy number of drz-Spur-1 ribozymes relative to the other drz-Spur families.

A number of the putative drz-Spur-1 ribozymes contain mutations that would render them inactive. These include active-site cytosine to uridine mutations, deletions on the 3′ side of the P1 helix, and mutations to the guanosine nucleotide that forms the 3′ side of the cleavage site. However, the latter mutation may not prove fatal to self-cleavage activity, as the drz-Spur-1 sequences contain a second guanosine adjacent to the cleavage site guanosine that could serve to shift the scission location one nucleotide downstream. This would yield a 6 nt P1 helix in the drz-Spur-3 family of sequences, which show robust self-cleavage despite the shortened P1 region ($k_{obs} = 12\ h^{-1}$ in 1 mM Mg^{2+} at 22 °C).

Unlike many of the other HDV-like ribozymes identified through secondary structure searches, the ribozymes from *S. purpuratus* frequently appear within or near genes or predicted genes. Three copies of drz-Spur-1 appear in the 22nd intron of a predicted homeobox gene, although one of these copies contains an active-site C53U mutation. Drz-Spur-2 and drz-Spur-4 both appear four times

in the genome, with one copy of drz-Spur-2 residing in the first intron of a predicted RNase PH gene, and one copy of drz-Spur-4 appearing antisense to the ninth intron of a predicted gene in the HECTc superfamily. Drz-Spur-3, which occurs five times in the genome, resides 68 nt downstream of a predicted protein-coding gene, within the 3rd and 13th introns of centromere binding and metabotropic glutamate receptor genes, respectively, and antisense to the 4th intron of a predicted reverse transcriptase gene. Expression data for the *S. purpuratus* ribozymes are limited, although in several instances spliced ESTs are found in close proximity to the motifs.

Sequence searches performed based on the drz-Spur-1 and -2 RNAs revealed additional ribozymes in the little skate *L. erinacea* and the common sea urchin, *P. lividus*. ESTs from early life stages of both of these organisms indicate that the RNA is transcribed *in vivo*. In *P. lividus*, the ribozymes appear to be self-cleaved in gastrula or pluteus stages of the organism (EST AM565620), despite having a single-nucleotide deletion of the residue adjacent to the catalytic cytosine in the J4/2 region. As previously noted, these sequences appear in close proximity to putative RTEs like those found in *A. gambiae*, where expression and self-cleavage state are also found to vary over life stages. In the little skate, sequences related to the drz-Spur-1 motif are likely inactive due to a mutation of the conserved guanosine residue in L3, whereas those based on the drz-Spur-2 motif are found in embryonic ESTs in apparently self-cleaved form (EST EE992768).[3]

5.13. *Branchiostoma floridae* ribozymes

Two families of HDV-like ribozymes have been identified in the lancelet *B. floridae*, drz-Bflo-1 and -2. The two families differ in several regions. The J4/2 of drz-Bflo-1 consists of CUAA, whereas that of drz-Bflo-2 consists of CGAA. The L3 of drz-Bflo-1 contains an extra pyrimidine adjacent to the 3′ side of the P3 helix similar to the genomic form of the HDV ribozyme.[3,34–36] Drz-Bflo-1 also possesses a 2 bp P1.1 followed by a J1.1/4 region and an extensively base-paired P4. The P1.1 helix of drz-Bflo-2 contains only a single base pair, and the ribozyme has a much shorter P4 helix. At only 63 nucleotides from the cleavage site to the last base pair in P2, drz-Bflo-2 is one of the smallest naturally occurring HDV-like ribozyme identified.

Both families of *B. floridae* ribozymes are found in multiple copies in the genome. Two copies of drz-Bflo-1 are present, with one residing in a peptidyl arginine deaminase gene (EST BW733175). Numerous copies of drz-Bflo-2 are present in both intergenic and intragenic regions, although

deletions to the L3 core and P3 helices in conjunction with mutations to the cleavage-site guanosine likely render all but a few of these sequences inactive. Nevertheless, ESTs are detected for many drz-Bflo-2 copies, indicating that they are transcribed *in vivo* (e.g., EST BW882616).

5.14. *Petromyzon marinus* ribozymes

The genome of the lamprey eel *P. marinus* contains an HDV-like ribozyme, drz-Pmar-1, that was identified using structure-based searches. Drz-Pmar-1, like the drz-Spur-1 and -2, drz-Bflo-2, and *CPEB3* sequences, contains only a single G–C base pair in the P1.1 helix. Several copies of the ribozyme are present in the genome, although the exact numbers are unknown due to incomplete genome assembly for the organism. One of these copies resides in a zinc finger gene, and it is expressed *in vivo* (EST FD713707). Several of the copies are likely inactive due to a poorly paired P1 helix or active-site C56U mutations. Interestingly, ribozymes sharing the drz-Pmar-1 sequence are readily detected in the genomes of the spiny dogfish shark *Squalus acanthias*, where the sequence is expressed (EST ES605876), and the coelacanth *Latimeria menadoensis*, where drz-Lmen-1 cleaves *in vitro* and shows strong temperature dependence.[7]

5.15. Nematode ribozymes

Nematodes represent one of the most diverse phylums of eukaryotic organisms on the planet, with estimates of species diversity ranging from 100,000 to 1,000,000.[136] It is, therefore, not surprising that HDV-like ribozymes are readily detected in this vast sequence space. Two nematodes, *C. japonica* and *P. pacificus*, possess ribozymes with confirmed *in vitro* activity (Fig. 4.10).[3] In both organisms, only a single family of HDV-like self-cleaving molecules is found, but each family is present many times throughout the genome. In *P. pacificus*, the copy number rivals that of the most abundant R2 ribozymes, and the copy number in the *C. japonica* genome is greater still.

There are approximately 120 copies of the drz-Cjap-1 family of ribozymes in the *C. japonica* genome. Because the *C. japonica* genome is not completely assembled, it is likely that some of these are overlapping reads from the sequencing of the organism. However, alignment of the copies reveals many to have differentiated upstream regions but conserved sequences downstream of the ribozymes (Fig. 4.10). Furthermore, none of the ribozymes appears in regions that map to known *C. japonica* genes, but an RTE-like RT-coding region does appear downstream of the

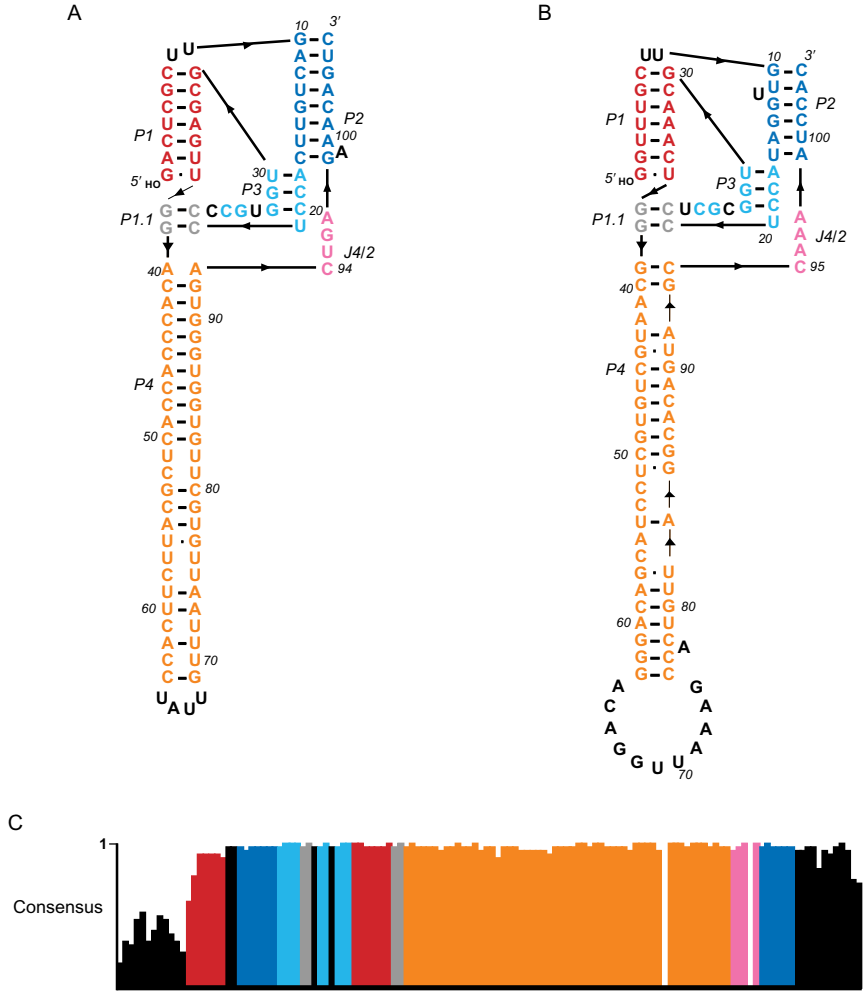

Figure 4.10 Secondary structures of (A) the drz-Ppac-1 and (B) drz-Cjap-1 ribozymes. (C) Alignment of 80 drz-Cjap-1 sequences. The coloring of the consensus alignment is identical to that seen in the secondary structure, with 10 nt upstream and downstream of the ribozyme also shown (black). Gaps in the consensus sequence correspond to positions at which a single copy of the ribozyme contains a nucleotide insertion.

ribozyme. Such observations are consistent with a model in which the drz-Cjap-1 sequences perform a role in retrotransposition comparable to other non-LTR-associated ribozymes, likely acting as a 5′ cap for the RNA and helping to initiate translation.[126]

In *P. pacificus*, the drz-Ppac-1 family of sequences can be found approximately 30 times, with both the upstream and downstream regions highly conserved. Conservation in regions flanking drz-Ppac-1, in conjunction with the high copy number, suggests that the ribozymes represent a rare class of retrotransposons with site-specific insertion points.[122] Analysis of sequences downstream from the ribozymes indicates the presence of RTE-like proteins. However, all the drz-Ppac-1 ribozymes appear intergenically so that it is unclear what the element might target for site-specific insertion since. It is possible that the ribozymes are in fact not intergenic, but appear so due to insufficient genome annotation in *P. pacificus*. High-throughput transcriptome and proteome analysis for the organism was recently completed, and the forthcoming data will perhaps answer some of these questions.[137]

Two independent isolates of drz-Ppac-1 have been tested *in vitro*. Despite sharing almost identical nucleotide composition, the two ribozymes exhibit very different kinetic properties.[3] The drz-Ppac-1–1 sequence was found to self-cleave in 1 mM Mg^{2+} at room temperature ($k_{obs} = 1.2\ h^{-1}$). However, the second isolate, drz-Ppac-1-2, did not show any self-scission at room temperature, and only residual activity in 1 mM Mg^{2+} at 37 °C ($k_{obs} = 0.78\ h^{-1}$). Although flanking sequences are known to play a crucial role in catalytic rates of HDV-like ribozymes, they are likely not contributing to the observed rate difference between the *P. pacificus* ribozymes because they are nearly identical between the *in vitro* constructs.[56] The drz-Ppac-1 sequences diverge primarily in their P4 helices, and two base pairs also changed in the P1 helix between the two isolates. Although the P4 helix is not essential for catalysis, it is expected to promote proper folding of the ribozyme. Therefore, the difference in catalytic rates might reflect the strength of base-pairing interactions in this helix.

5.16. HDV-like ribozymes in bacteria and viruses

The drz-Frpa-1 and drz-CIV-1 from *F. prausnitzii* and CIV, respectively, are currently the only two HDV-like ribozymes identified outside of eukaryotes and the HDV genome. As the number of bacterial and viral strains and species is far greater than those of eukaryotes, it is likely that many more HDV-like ribozymes reside in these bastions of sequence space that must be searched as new genomes and populations of microorganisms become available. Alternatively, direct, metagenomic searches from environmental samples using HDV-like motif descriptors could be used to uncover additional HDV-like ribozymes.

The secondary structure of drz-Fpra-1 is somewhat unique in that it contains an A–U base pair at the start of the P1.1 helix (Fig. 4.11). In all other isolates of HDV-like ribozymes, the P1.1 helix begins with a G–C base pair. This could contribute to the slow rate of cleavage observed for the isolate

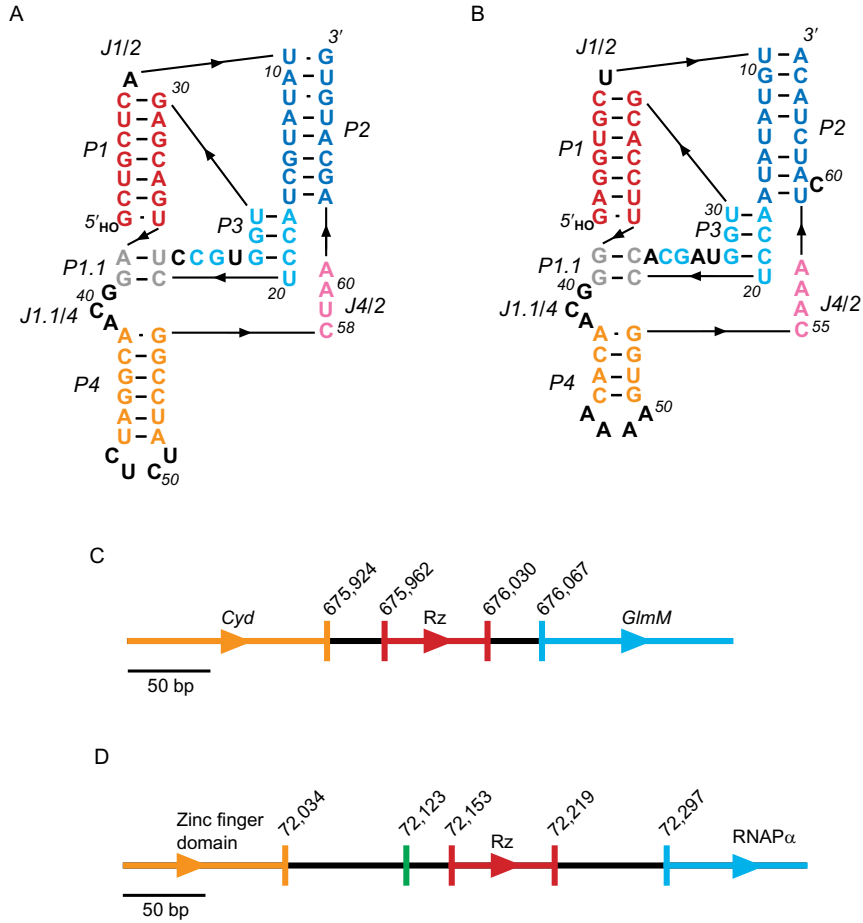

Figure 4.11 Secondary structures of the (A) *Faecalibacterium prausnitzii*, drz-Fpra-1, and (B) *Chilo* iridescent virus, drz-CIV-1, ribozymes. (C) Genomic location of the drz-Fpra-1 ribozyme. Red corresponds to the ribozyme, orange indicates the predicted *Cyd* coding region, and blue represents the predicted *GlmM* coding region. Arrows indicate direction of transcription, and genomic locations are listed above. (D) Genomic location of the drz-CIV-1 ribozyme. Red indicates the ribozyme, green corresponds to a TATA box, orange corresponds to a zinc finger protein-coding region, and blue indicates the viral RNA polymerase alpha subunit coding region.

(0.47 h^{-1} in 10 mM Mg^{2+} at 37 °C), as the A residue might be able to form an alternate Watson–Crick pair with the uridine residue in the leader sequence, extending the P1 helix and disrupting the formation of P1.1.

The self-cleavage activity of the *Faecalibacterium* drz-Fpra-1 ribozyme is very temperature dependent. At 37 °C, the temperature at which the bacteria live and grow, between 60% and 80% of the ribozyme, is found self-cleaved in a 24 h assay ($k_{obs} = 0.47\ h^{-1}$ in 10 mM Mg^{2+}). However, at room temperature, only 20% of the ribozyme population self-cleaves within 24 h. The cause of this phenomenon is unknown, but the fact that the ribozyme does not self-cleave to the same endpoints at different temperatures suggests that the sequence has been optimized to function under the conditions in which *F. prausnitzii* grows.

The genomic location of drz-Fpra-1 further supports a role for the ribozyme in the biochemistry of the bacterium. The ribozyme resides between two genes: downstream of cytochrome *d* ubiquinol oxidase (*Cyd*) and upstream of phosphoglucosamine mutase (*GlmM*) (Fig. 4.11). These two protein open-reading frames are separated by only 144 nt, indicating that they may be part of a bi-cistronic mRNA that is processed by the ribozyme posttranscriptionally. In the absence of additional cellular factors, the observed slow *in vitro* rate of scission implies that very few messages are likely to be cleaved *in vivo*. However, cotranscriptional self-scission has not been measured for the ribozyme, therefore *in vivo* cleavage rate and extent are difficult to estimate. Additional factors that accelerate the self-cleavage activity of drz-Fpra-1 may indeed be present, as both gene products are essential for metabolism inside the bacteria and would likely be upregulated following nutrient intake.[138,139] Interestingly, the bacterium contains a second copy of the *GlmM* gene, along with all of the upstream drz-Fpra-1 ribozyme except for the 5′ side of the P1 helix. This could be due to a retrotransposition event of the self-cleaved ribozyme-terminated *GlmM* mRNA, in which the 5′ end of the ribozyme was lost.

The drz-CIV-1 ribozyme is also found between two genes that reside in close proximity (Fig. 4.11). It maps 144 nt upstream of the viral DNA-dependent RNA polymerase large subunit start codon, and 120 nt downstream of the stop codon of a DNA-binding protein. Two competing pathways have been proposed for this ribozyme. In one, the ribozyme could serve to inactivate the translation of the large subunit of the polymerase by cleaving the 5′ cap of the mRNA following the initial production of the subunit during the *immediate-early* stage of viral infection.[140] Alternatively, drz-CIV-1 might serve to process the polymerase mRNA from the

surrounding transcript and promote cap-independent translation initiation of the message. The recent discovery of the R2 ribozymes and their proposed roles in translation initiation make the possibility of the latter task more likely, although the *in vitro* self-cleavage rate in physiological conditions suggests that the majority of the transcripts would be cleaved on a time scale too slow to promote expression of the protein.[2,3,140]

6. EVOLUTIONARY RELATIONSHIP AMONG HDV-LIKE RIBOZYMES

The widespread occurrence of HDV-like ribozymes raises questions as to the origins of these motifs. Did the HDV ribozyme motif originate in the distant past and remain relevant as genomes evolved, or are the many examples of this motif a product of convergent evolution? The observed variability among the *Drosophila* R2 ribozymes is consistent with a model in which a single HDV-like ribozyme arose early in the evolutionary development of the fruit fly and has been subjected to random mutations that maintain the crucial catalytic requirements for self-scission. The rate of background mutations in *Drosophila* has been estimated to be around 11 mutations/kb every 1 million years, so for the observed approximately 50 mutations to occur between the most dissimilar R2 ribozymes, the sequence would have had to originate over 25 million years ago.[141] Also, R2 ribozymes have been postulated to spread vertically since no evidence of horizontal gene transfer has been observed.[4,133]

The evolutionary relationship between the R2 ribozymes and HDV-like ribozymes in other organisms is more speculative. The HDV-like ribozymes containing long inserts in the J1/2 region are largely associated with transposable elements. It is therefore reasonable to assume that they share a common ancestral sequence that propagated via retrotransposition.

Whether these sequences are also evolutionarily related to HDV-like ribozymes that do not contain extended J1/2 regions remains to be determined. Currently, only *A. gambiae* mosquitoes have been identified with ribozymes containing both normal length and long J1/2 regions. Given the high informational content of HDV-like ribozymes, it is unlikely that the drz-Agam-1 and -2 families are examples of convergent evolution in a single genome. However, 100 bp insertions specific to a nonessential region of a functional RNA molecule are infrequent in occurrence, except if there is some benefit in either having or not having the insertion. As the

J1/2 insertion is common to retrotransposon-associated ribozymes, it likely represents a gain-of-function mutation that is necessary for the ribozyme's role in the retrotransposition cycle. If this is the case, the drz-Agam-2 family could be the product of selective acquisition of a stable structure upstream of a drz-Agam-1 ribozyme following a transposition event that brought a drz-Agam-1 ribozyme and an ancestral retrotransposon into close proximity. Such insertion would require the initial drz-Agam-1 5′ P1 strand to be lost through mutation or deletion and a new 5′ P1 strand to take its place. It seems plausible that such a sequence could appear by chance since the minimum length of the P1 helix in active ribozymes is only 6 bp.

The level of sequence variation observed between the HDV-like ribozyme families can be explained in one of two ways. Either HDV-like ribozymes arose very early and provided a function that was retained in subsequent replication and speciation events, slowly accruing mutations that led to the diversity one finds today, or the HDV ribozyme motif evolved independently several times as a need arose in a given genome for a self-cleaving RNA molecule.

The hypothesis that the observed HDV-like ribozyme diversity is the product of a single, ancestral self-cleaving HDV RNA is backed by the fact that HDV-like ribozymes have never been detected during *in vitro* selection experiments for RNAs capable of self-scission. During these experiments, the starting DNA pool can consist of 10^{16} unique sequences.[142–144] This is well beyond the sequence diversity found in any single genome, and yet no HDV-like ribozymes have been isolated *in vitro* despite their widespread occurrence *in vivo*. Based on the informational content of the HDV motif, a starting pool of $>10^{16}$ sequences would yield on average a single HDV-like self-cleaving motif. Furthermore, the existence of less complex RNA motifs capable of catalyzing the same self-scission reaction *in vivo* would further decrease the chances of the HDV ribozyme motif evolving independently on multiple occasions. Hammerhead ribozymes for instance, which are found widely dispersed in nature, have also been isolated independently several times *in vitro*.[5,6,110,142]

Alternatively, several lines of evidence point to convergent evolution for HDV-like ribozymes. It is unlikely that every instance of the motif evolved independently as certain underlying features can be found among ribozymes located in distantly related organisms. For instance, despite mutational data showing that the J4/2 region can consist of the consensus sequence "CNRA," which has eight possible permutations, only five variants to this region are observed *in vivo*.[3,42,4] If the motifs all evolved independently, all

combinations of sequences in the J4/2 region should appear in approximately equal distributions. That this is not the case points to common ancestral sequences that have been subjected to random mutations to unconserved portions to establish the variability seen today.

Furthermore, HDV-like ribozymes are found with and without a J1.1/4 region. Ribozymes without this region are quite heterogeneous in nucleotide composition at the top of the P4 helix, whereas those with a J1.1/4 region contain almost exclusively an A–G mispair at this location. In the crystal structure of the genomic HDV ribozyme, the A43 and G74 residues are base paired, and G74 forms additional tertiary contacts with other nucleotides in the J1.1/4 region. These interactions provide a number of stabilizing effects for the ribozyme. Therefore, it is unlikely that they would be lost during the evolution of the sequence. Rather, the presence or absence of a J1.1/4 region might provide a fine-tuning mechanism for the self-cleavage rate of the ribozyme following multiple, convergent paths of evolution to an HDV-like core. The P3, L3, and P1.1 core regions of the ribozymes are all very similar, and identical between closely related retrotransposon ribozymes regardless of the presence of a J1.1/4 region. This is indicative of the minimal sequence space that the core regions can occupy, whereas the remainder of the motif can arise from a diverse set of nucleotides so long as proper base pairing is maintained. Combined, these data allow for a model in which the HDV ribozyme motif evolved independently in several evolutionary lines and subsequently spread among closely related species.

Although the evolutionary path of HDV-like ribozymes remains to be determined, the widespread nature of this motif cannot be denied. The motif has been identified across all kingdoms of life except Archaea, and in a variety of biological roles including retrotransposition and genome processing, with putative roles in mRNA stability and translation regulation. Recently, hammerhead ribozymes have also been identified in a variety of species, including mammals.[5,6,110] These data indicate that the self-cleaving ribozymes, and likely ribozymes in general, are much more common in biological systems than previously expected.

REFERENCES

1. Symons RH. Plant pathogenic RNAs and RNA catalysis. *Nucleic Acids Res.* 1997;25:2683–2689.
2. Salehi-Ashtiani K, Luptak A, Litovchick A, Szostak JW. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science.* 2006;313:1788–1792.

3. Webb C-HT, Riccitelli NJ, Ruminski DJ, Luptak A. Widespread occurrence of self-cleaving ribozymes. *Science*. 2009;326:953.
4. Eickbush DG, Eickbush TH. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol Cell Biol*. 2010;30:3142–3150.
5. de la Pena M, Garcia-Robles I. Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. *RNA*. 2010;16:1943–1950.
6. de la Pena M, Garcia-Robles I. Intronic hammerhead ribozymes are ultraconserved in the human genome. *EMBO Rep*. 2010;11:711–716.
7. Webb CHT, Luptak A. HDV-like self-cleaving ribozymes. *RNA Biol*. 2011;8: 719–727.
8. Seehafer C, Kalweit A, Steger G, Graf S, Hammann C. From alpaca to zebrafish: hammerhead ribozymes wherever you look. *RNA*. 2011;17:21–26.
9. Perreault J, Weinberg Z, Roth A, et al. Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. *PLoS Comput Biol*. 2011;7:1–13.
10. Wu HN, Lin YJ, Lin FP, Makino S, Chang MF, Lai MM. Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proc Natl Acad Sci USA*. 1989;86:1831–1835.
11. Forster AC, Symons RH. Self-cleavage of virusoid RNA is performed by the proposed 55-nucleotide active site. *Cell*. 1987;50:9–16.
12. Kuo MY, Sharmeen L, Dinter-Gottlieb G, Taylor J. Characterization of self-cleaving RNA sequences on the genome and anti-genome of human hepatitis delta virus. *J Virol*. 1988;62:4439–4444.
13. Hampel A, Tritz R. RNA catalytic properties of the minimum (−) sTRSV sequence. *Biochemistry*. 1989;28:4929–4933.
14. Saville BJ, Collins RA. A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell*. 1990;61:685–696.
15. Rizzetto M, Canese MG, Arico S, et al. Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers. *Gut*. 1977;18:997–1003.
16. Rizzetto M, Hoyer B, Canese MG, Shih JW, Purcell RH, Gerin JL. Delta Agent: association of delta antigen with hepatitis B surface antigen and RNA in serum of delta-infected chimpanzees. *Proc Natl Acad Sci USA*. 1980;77:6124–6128.
17. Ponzetto A, Cote PJ, Popper H, et al. Transmission of the hepatitis-B virus-associated delta-agent to the eastern woodchuck. *Proc Natl Acad Sci USA*. 1984;81:2208–2212.
18. Taylor JM. Structure and replication of hepatitis delta virus RNA. *Hepatitis Delta Virus*. 2006;307:1–23.
19. Branch AD, Robertson HD. A Replication cycle for viroids and other small infectious RNAs. *Science*. 1984;223:450–455.
20. Chen PJ, Kalpana G, Goldberg J, et al. Structure and replication of the genome of the hepatitis delta-virus. *Proc Natl Acad Sci USA*. 1986;83:8774–8778.
21. Gudima S, Dingle K, Wu TT, Moraleda G, Taylor J. Characterization of the 5' ends for polyadenylated RNAs synthesized during the replication of hepatitis delta virus. *J Virol*. 1999;73:6533–6539.
22. Nie XC, Chang JH, Taylor JM. Alternative processing of hepatitis delta virus antigenomic RNA transcripts. *J Virol*. 2004;78:4517–4524.
23. Chao M, Hsieh SY, Taylor J. Role of 2 forms of hepatitis delta virus-antigen-evidence for a mechanism of self-limiting genome replication. *J Virol*. 1990;64:5066–5069.
24. Wong SK, Lazinski DW. Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proc Natl Acad Sci USA*. 2002;99:15118–15123.
25. Lazinski DW, Taylor JM. Expression of hepatitis delta virus RNA deletions: cis and trans requirements for self-cleavage, ligation, and RNA packaging. *J Virol*. 1994;68:2879–2888.

26. Modahl LE, Macnaughton TB, Zhu NL, Johnson DL, Lai MMC. RNA-dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases. *Mol Cell Biol*. 2000;20:6030–6039.
27. Sharmeen L, Kuo MY, Dinter-Gottlieb G, Taylor J. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J Virol*. 1988;62:2674–2679.
28. Wadkins TS, Perrotta AT, Ferre-D'Amare AR, Doudna JA, Been MD. A nested double pseudoknot is required for self-cleavage activity of both the genomic and antigenomic hepatitis delta virus ribozymes. *RNA*. 1999;5:720–727.
29. Reid CE, Lazinski DW. A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs. *Proc Natl Acad Sci USA*. 2000;97:424–429.
30. Englert M, Sheppard K, Gundllapalli S, Beier H, Soll D. Branchiostoma floridae has separate healing and sealing enzymes for 5'-phosphate RNA ligation. *Proc Natl Acad Sci USA*. 2010;107:16834–16839.
31. Englert M, Beier H. Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins. *Nucleic Acids Res*. 2005;33:388–399.
32. Schutz K, Hesselberth JR, Fields S. Capture and sequence analysis of RNAs with terminal 2', 3'-cyclic phosphates. *RNA*. 2010;16:621–631.
33. Chan CM, Zhou C, Huang RH. Reconstituting bacterial RNA repair and modification in vitro. *Science*. 2009;326:247.
34. Ferre-D'Amare AR, Zhou K, Doudna JA. Crystal structure of a hepatitis delta virus ribozyme. *Nature*. 1998;395:567–574.
35. Ke A, Zhou K, Ding F, Cate JH, Doudna JA. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature*. 2004;429:201–205.
36. Chen JH, Yajima R, Chadalavada DM, Chase E, Bevilacqua PC, Golden BL. A 1.9 A crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry*. 2010;49:6508–6518.
37. Been MD, Perrotta AT, Rosenstein SP. Secondary structure of the self-cleaving RNA of hepatitis delta virus: applications to catalytic RNA design. *Biochemistry*. 1992;31:11843–11852.
38. Ferre-D'Amare AR, Zhou K, Doudna JA. A general module for RNA crystallization. *J Mol Biol*. 1998;279:621–631.
39. Tanner NK, Schaff S, Thill G, Petit-Koskas E, Crain-Denoyelle AM, Westhof E. A three-dimensional model of hepatitis delta virus ribozyme based on biochemical and mutational analyses. *Curr Biol*. 1994;4:488–498.
40. Perrotta AT, Been MD. Core sequences and a cleavage site wobble pair required for HDV antigenomic ribozyme self-cleavage. *Nucleic Acids Res*. 1996;24:1314–1321.
41. Been MD, Wickham GS. Self-cleaving ribozymes of hepatitis delta virus RNA. *Eur J Biochem*. 1997;247:741–753.
42. Legiewicz M, Wichlacz A, Brzezicha B, Ciesiolka J. Antigenomic delta ribozyme variants with mutations in the catalytic core obtained by the in vitro selection method. *Nucleic Acids Res*. 2006;34:1270–1280.
43. Nishikawa F, Kawakami J, Chiba A, Shirai M, Kumar PK, Nishikawa S. Selection in vitro of trans-acting genomic human hepatitis delta virus (HDV) ribozymes. *Eur J Biochem*. 1996;237:712–718.
44. Nehdi A, Perreault JP. Unbiased in vitro selection reveals the unique character of the self-cleaving antigenomic HDV RNA sequence. *Nucleic Acids Res*. 2006;34:584–592.
45. Perrotta AT, Shih I, Been MD. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science*. 1999;286:123–126.

46. Nakano S, Chadalavada DM, Bevilacqua PC. General acid–base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science*. 2000;287:1493–1497.
47. Doherty EA, Doudna JA. Ribozyme structures and mechanisms. *Annu Rev Biochem*. 2000;69:597–615.
48. Rosenstein SP, Been MD. Hepatitis delta virus ribozymes fold to generate a solvent-inaccessible core with essential nucleotides near the cleavage site phosphate. *Biochemistry*. 1996;35:11403–11413.
49. Wadkins TS, Shih I, Perrotta AT, Been MD. A pH-sensitive RNA tertiary interaction affects self-cleavage activity of the HDV ribozymes in the absence of added divalent metal ion. *J Mol Biol*. 2001;305:1045–1055.
50. Perrotta AT, Been MD. A toggle duplex in hepatitis delta virus self-cleaving RNA that stabilizes an inactive and a salt-dependent pro-active ribozyme conformation. *J Mol Biol*. 1998;279:361–373.
51. Perrotta AT, Been MD. The self-cleaving domain from the genomic RNA of hepatitis delta virus: sequence requirements and the effects of denaturant. *Nucleic Acids Res*. 1990;18:6821–6827.
52. Smith JB, Dinter-Gottlieb G. Antigenomic Hepatitis delta virus ribozymes self-cleave in 18 M formamide. *Nucleic Acids Res*. 1991;19:1285–1289.
53. Duhamel J, Liu DM, Evilia C, et al. Secondary structure content of the HDV ribozyme in 95% formamide. *Nucleic Acids Res*. 1996;24:3911–3917.
54. Perrotta AT, Been MD. Cleavage of oligoribonucleotides by a ribozyme derived from the hepatitis delta virus RNA sequence. *Biochemistry*. 1992;31:16–21.
55. Nishikawa F, Roy M, Fauzi H, Nishikawa S. Detailed analysis of stem I and its 5' and 3' neighbor regions in the trans-acting HDV ribozyme. *Nucleic Acids Res*. 1999;27:403–410.
56. Chadalavada DM, Knudsen SM, Nakano S, Bevilacqua PC. A role for upstream RNA structure in facilitating the catalytic fold of the genomic hepatitis delta virus ribozyme. *J Mol Biol*. 1999;301:349–367.
57. Chadalavada DM, Senchak SE, Bevilacqua PC. The folding pathway of the genomic hepatitis delta virus ribozyme is dominated by slow folding of the pseudoknots. *J Mol Biol*. 2002;317:559–575.
58. Diegelman-Parente A, Bevilacqua PC. A mechanistic framework for co-transcriptional folding of the HDV genomic ribozyme in the presence of downstream sequence. *J Mol Biol*. 2002;324:1–16.
59. Jeong S, Sefcikova J, Tinsley RA, Rueda D, Walter NG. Trans-acting hepatitis delta virus ribozyme: catalytic core and global structure are dependent on the 5' substrate sequence. *Biochemistry*. 2003;42:7727–7740.
60. Brown TS, Chadalavada DM, Bevilacqua PC. Design of a highly reactive HDV ribozyme sequence uncovers facilitation of RNA folding by alternative pairings and physiological ionic strength. *J Mol Biol*. 2004;341:695–712.
61. Brown AL, Perrotta AT, Wadkins TS, Been MD. The poly(A) site sequence in HDV RNA alters both extent and rate of self-cleavage of the antigenomic ribozyme. *Nucleic Acids Res*. 2008;36:2990–3000.
62. Chadalavada DM, Cerrone-Szakal AL, Bevilacqua PC. Wild-type is the optimal sequence of the HDV ribozyme under cotranscriptional conditions. *RNA*. 2007;13:2189–2201.
63. Pereira MJ, Behera V, Walter NG. Nondenaturing purification of co-transcriptionally folded RNA avoids common folding heterogeneity. *PLoS ONE*. 2010;5:e12953.
64. Bravo C, Lescure F, Laugaa P, Fourrey JL, Favre A. Folding of the HDV antigenomic ribozyme pseudoknot structure deduced from long-range photocrosslinks. *Nucleic Acids Res*. 1996;24:1351–1359.

65. Prabhu NS, Dinter-Gottlieb G, Gottlieb PA. Single substitutions of phosphorothioates in the HDV ribozyme G73 define regions necessary for optimal self-cleaving activity. *Nucleic Acids Res*. 1997;25:5119–5124.
66. Shih IH, Been MD. Ribozyme cleavage of a 2,5-phosphodiester linkage: mechanism and a restricted divalent metal-ion requirement. *RNA*. 1999;5:1140–1148.
67. Das SR, Piccirilli JA. General acid catalysis by the hepatitis delta virus ribozyme. *Nat Chem Biol*. 2005;1:45–52.
68. Oyelere AK, Kardon JR, Strobel SA. pK(a) perturbation in genomic Hepatitis Delta Virus ribozyme catalysis evidenced by nucleotide analog interference mapping. *Biochemistry*. 2002;41:3667–3675.
69. Perrotta AT, Wadkins TS, Been MD. Chemical rescue, multiple ionizable groups, and general acid–base catalysis in the HDV genomic ribozyme. *RNA*. 2006;12:1282–1291.
70. Shih IH, Been MD. Involvement of a cytosine side chain in proton transfer in the rate-determining step of ribozyme self-cleavage. *Proc Natl Acad Sci USA*. 2001;98:1489–1494.
71. Cerrone-Szakal AL, Siegfried NA, Bevilacqua PC. Mechanistic characterization of the HDV genomic ribozyme: solvent isotope effects and proton inventories in the absence of divalent metal ions support C75 as the general acid. *J Am Chem Soc*. 2008;130:14504–14520.
72. Nakano S, Proctor DJ, Bevilacqua PC. Mechanistic characterization of the HDV genomic ribozyme: assessing the catalytic and structural contributions of divalent metal ions within a multichannel reaction mechanism. *Biochemistry*. 2001;40:12022–12038.
73. Perrotta AT, Been MD. HDV ribozyme activity in monovalent cations. *Biochemistry*. 2006;45:11357–11365.
74. Nakano S, Bevilacqua PC. Proton inventory of the genomic HDV ribozyme in Mg(2+)-containing solutions. *J Am Chem Soc*. 2001;123:11333–11334.
75. Gong B, Chen JH, Chase E, et al. Direct measurement of a pK(a) near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography. *J Am Chem Soc*. 2007;129:13335–13342.
76. Luptak A, Ferre-D'Amare AR, Zhou K, Zilm KW, Doudna JA. Direct pK(a) measurement of the active-site cytosine in a genomic hepatitis delta virus ribozyme. *J Am Chem Soc*. 2001;123:8447–8452.
77. Harris DA, Rueda D, Walter NG. Local conformational changes in the catalytic core of the trans-acting hepatitis delta virus ribozyme accompany catalysis. *Biochemistry*. 2002;41:12051–12061.
78. Nakano S, Cerrone AL, Bevilacqua PC. Mechanistic characterization of the HDV genomic ribozyme: classifying the catalytic and structural metal ion sites within a multichannel reaction mechanism. *Biochemistry*. 2003;42:2982–2994.
79. Suh YA, Kumar PK, Taira K, Nishikawa S. Self-cleavage activity of the genomic HDV ribozyme in the presence of various divalent metal ions. *Nucleic Acids Res*. 1993;21:3277–3280.
80. Jeoung YH, Kumar PK, Suh YA, Taira K, Nishikawa S. Identification of phosphate oxygens that are important for self-cleavage activity of the HDV ribozyme by phosphorothioate substitution interference analysis. *Nucleic Acids Res*. 1994;22:3722–3727.
81. Ke A, Ding F, Batchelor JD, Doudna JA. Structural roles of monovalent cations in the HDV ribozyme. *Structure*. 2007;15:281–287.
82. Ferre-D'Amare AR, Doudna JA. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J Mol Biol*. 2000;295:541–556.
83. Perrotta AT, Been MD. A single nucleotide linked to a switch in metal ion reactivity preference in the HDV ribozymes. *Biochemistry*. 2007;46:5124–5130.

84. Harris DA, Tinsley RA, Walter NG. Terbium-mediated footprinting probes a catalytic conformational switch in the antigenomic hepatitis delta virus ribozyme. *J Mol Biol*. 2004;341:389–403.
85. Sefcikova J, Krasovska MV, Sponer J, Walter NG. The genomic HDV ribozyme utilizes a previously unnoticed U-turn motif to accomplish fast site-specific catalysis. *Nucleic Acids Res*. 2007;35:1933–1946.
86. Pereira MJ, Harris DA, Rueda D, Walter NG. Reaction pathway of the trans-acting hepatitis delta virus ribozyme: a conformational change accompanies catalysis. *Biochemistry*. 2002;41:730–740.
87. Tinsley RA, Harris DA, Walter NG. Magnesium dependence of the amplified conformational switch in the trans-acting hepatitis delta virus ribozyme. *Biochemistry*. 2004;43:8935–8945.
88. Tinsley RA, Walter NG. Long-range impact of peripheral joining elements on structure and function of the hepatitis delta virus ribozyme. *Biol Chem*. 2007;388:705–715.
89. Joyce GF. Amplification, mutation and selection of catalytic RNA. *Gene*. 1989;82:83–87.
90. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. *Nature*. 1990;346:818–822.
91. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*. 1990;249:505–510.
92. Wilson C, Nix J, Szostak J. Functional requirements for specific ligand recognition by a biotin-binding RNA pseudoknot. *Biochemistry*. 1998;37:14410–14419.
93. Pan T, Uhlenbeck OC. In vitro selection of RNAs that catalyze self-cleavage reactions with Pb2+. *FASEB J*. 1992;6:A412.
94. Jayasena VK, Gold L. In vitro selection of self-cleaving RNAs with a low pH optimum. *Proc Natl Acad Sci USA*. 1997;94:10612–10617.
95. Daubendiek SL, Kool ET. Generation of catalytic RNAs by rolling transcription of synthetic DNA nanocircles. *Nat Biotechnol*. 1997;15:273–277.
96. Diegelman AM, Kool ET. Generation of circular RNAs and trans-cleaving catalytic RNAs by rolling transcription of circular DNA oligonucleotides encoding hairpin ribozymes. *Nucleic Acids Res*. 1998;26:3235–3241.
97. Diegelman AM, Daubendiek SL, Kool ET. Generation of RNA ladders by rolling circle transcription of small circular oligodeoxyribonucleotides. *Biotechniques*. 1998;25:754–758.
98. Chadalavada DM, Gratton EA, Bevilacqua PC. The human HDV-like CPEB3 ribozyme is intrinsically fast-reacting. *Biochemistry*. 2010;49:5321–5330.
99. Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev*. 1999;63:405–445.
100. Huang YS, Kan MC, Lin CL, Richter JD. CPEB3 and CPEB4 in neurons: analysis of RNA-binding specificity and translational control of AMPA receptor GluR2 mRNA. *EMBO J*. 2006;25:4865–4876.
101. Theis M, Si K, Kandel ER. Two previously undescribed members of the mouse CPEB family of genes and their inducible expression in the principal cell layers of the hippocampus. *Proc Natl Acad Sci USA*. 2003;100:9602–9607.
102. Ucker DS, Yamamoto KR. Early events in the stimulation of mammary tumor virus RNA synthesis by glucocorticoids. Novel assays of transcription rates. *J Biol Chem*. 1984;259:7416–7420.
103. Adelman K, La Porta A, Santangelo TJ, Lis JT, Roberts JW, Wang MD. Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc Natl Acad Sci USA*. 2002;99:13538–13543.

104. Alarcon JM, Hodgman R, Theis M, Huang YS, Kandel ER, Richter JD. Selective modulation of some forms of Schaffer collateral-CA1 synaptic plasticity in mice with a disruption of the CPEB-1 gene. *Learn Mem*. 2004;11:318–327.
105. Luptak A, Szostak JW. Mammalian self-cleaving ribozymes. In: Lilley DMJ, Eckstein F, eds. *Ribozymes and RNA Catalysis*. Cambridge, UK: Royal Society of Chemistry; 2007.
106. Vogler C, Spalek K, Aerni A, et al. CPEB3 is associated with human episodic memory. *Front Behav Neurosci*. 2009;3:4.
107. Epstein LM, Gall JG. Self-cleaving transcripts of satellite DNA from the newt. *Cell*. 1987;48:535–543.
108. Ferbeyre G, Smith JM, Cedergren R. Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol Cell Biol*. 1998;18:3880–3888.
109. Przybilski R, Graf S, Lescoute A, et al. Functional hammerhead ribozymes naturally encoded in the genome of Arabidopsis thaliana. *Plant Cell*. 2005;17:1877–1885.
110. Martick M, Horan LH, Noller HF, Scott WG. A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature*. 2008;454:899–902.
111. Ferbeyre G, Bourdeau V, Pageau M, Miramontes P, Cedergren R. Distribution of hammerhead and hammerhead-like RNA motifs through the GenBank. *Genome Res*. 2000;10:1011–1019.
112. Saurin W, Marliere P. Matching relational patterns in nucleic acid sequences. *Comput Appl Biosci*. 1987;3:115–120.
113. Shapiro BA. An algorithm for comparing multiple RNA secondary structures. *Comput Appl Biosci*. 1988;4:387–393.
114. Margalit H, Shapiro BA, Oppenheim AB, Maizel JV. Detection of common motifs in RNA secondary structures. *Nucleic Acids Res*. 1989;17:4829–4845.
115. Gautheret D, Major F, Cedergren R. Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA. *Comput Appl Biosci*. 1990;6:325–331.
116. Steinberg S, Gautheret D, Cedergren R. Fitting the structurally diverse animal mitochondrial tRNAs(Ser) to common three-dimensional constraints. *J Mol Biol*. 1994;236:982–989.
117. Bourdeau V, Ferbeyre G, Pageau M, Paquin B, Cedergren R. The distribution of RNA motifs in natural sequences. *Nucleic Acids Res*. 1999;27:4457–4467.
118. Eddy SR. RNABOB: a program to search for RNA secondary structure motifs in sequence databases. Unpublished.
119. Riccitelli NJ, Lupták A. Computational discovery of folded RNA domains in genomes and in vitro selected libraries. *Methods*. 2010;52:133–140.
120. Malik HS, Eickbush TH. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. *Mol Biol Evol*. 1998;15:1123–1134.
121. Youngman S, van Luenen HGAM, Plasterk RHA. Rte-1, a retrotransposon-like element in Caenorhabditis elegans. *FEBS Lett*. 1996;380:1–7.
122. Jakubczak JL, Burke WD, Eickbush TH. Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. *Proc Natl Acad Sci USA*. 1991;88:3295–3299.
123. Luan DD, Eickbush TH. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol Cell Biol*. 1995;15:3882–3891.
124. Kurzynska-Kokorniak A, Jamburuthugoda VK, Bibillo A, Eickbush TH. DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon. *J Mol Biol*. 2007;374:322–333.

125. Jamrich M, Miller Jr OL. The rare transcripts of interrupted rRNA genes in Drosophila melanogaster are processed or degraded during synthesis. *EMBO J*. 1984;3:1541–1545.
126. Ruminski DJ, Webb CHT, Riccitelli NJ, Luptak A. Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (HDV)-like self-cleaving ribozymes. *J Biol Chem*. 2011;286:41286–41295.
127. Rosenstein SP, Been MD. Self-cleavage of hepatitis delta virus genomic strand RNA is enhanced under partially denaturing conditions. *Biochemistry*. 1990;29: 8011–8016.
128. Kieft JS, Viral IRES. RNA structures and ribosome interactions. *Trends Biochem Sci*. 2008;33:274–283.
129. Ghesini S, Luchetti A, Marini M, Mantovani B. The non-LTR retrotransposon R2 in termites (Insecta, Isoptera): characterization and dynamics. *J Mol Evol*. 2011;72:296–305.
130. Kojima KK, Fujiwara H. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. *Mol Biol Evol*. 2004;21:207–217.
131. Kojima KK, Fujiwara H. Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol Biol Evol*. 2005;22:2157–2165.
132. Burke WD, Muller F, Eickbush TH. R4, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes. *Nucleic Acids Res*. 1995;23:4628–4634.
133. Kojima KK, Fujiwara H. Evolution of target specificity in R1 clade non-LTR retrotransposons. *Mol Biol Evol*. 2003;20:351–361.
134. Warren AM, Hughes MA, Crampton JM. Zebedee: a novel copia-Ty1 family of transposable elements in the genome of the medically important mosquito Aedes aegypti. *Mol Gen Genet*. 1997;254:505–513.
135. Kubo Y, Okazaki S, Anzai T, Fujiwara H. Structural and phylogenetic analysis of TRAS, telomeric repeat-specific non-LTR retrotransposon families in Lepidopteran insects. *Mol Biol Evol*. 2001;18:848–857.
136. Blumenthal T, Davis RE. Exploring nematode diversity. *Nat Genet*. 2004;36: 1246–1247.
137. Borchert N, Dieterich C, Krug K, et al. Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res*. 2010;20:837–846.
138. Mengin-Lecreulx D, van Heijenoort J. Characterization of the essential gene glmM encoding phosphoglucosamine mutase in *Escherichia coli*. *J Biol Chem*. 1996;271:32–39.
139. Garcia-Horsman JA, Barquera B, Rumbley J, Ma JX, Gennis RB. The superfamily of heme-copper respiratory oxidases. *J Bacteriol*. 1994;176:5587–5600.
140. Schnitzler P, Sonntag KC, Muller M, et al. Insect iridescent virus type 6 encodes a polypeptide related to the largest subunit of eukaryotic RNA polymerase II. *J Gen Virol*. 1994;75(Pt 7):1557–1567.
141. Tamura K, Subramanian S, Kumar S. Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. *Mol Biol Evol*. 2004;21:36–44.
142. Salehi-Ashtiani K, Szostak JW. In vitro evolution suggests multiple origins for the hammerhead ribozyme. *Nature*. 2001;414:82–84.
143. Tang J, Breaker RR. Structural diversity of self-cleaving ribozymes. *Proc Natl Acad Sci USA*. 2000;97:5784–5789.
144. Wilson DS, Szostak JW. In vitro selection of functional nucleic acids. *Annu Rev Biochem*. 1999;68:611–647.