

SELF-SPLICING AND ENZYMATIC ACTIVITY OF AN INTERVENING SEQUENCE RNA FROM *TETRAHYMENA*

Nobel Lecture, December 8, 1989

by

THOMAS R. CECHE

Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215, USA

A living cell requires thousands of different chemical reactions to utilize energy, move, grow, respond to external stimuli and reproduce itself. While these reactions take place spontaneously, they rarely proceed at a rate fast enough for life. Enzymes, biological catalysts found in all cells, greatly accelerate the rates of these chemical reactions and impart on them extraordinary specificity.

In 1926, James B. Sumner crystallized the enzyme urease and found that it was a protein. Skeptics argued that the enzymatic activity might reside in a trace component of the preparation rather than in the protein (Haldane, 1930), and it took another decade for the generality of Sumner's finding to be established. As more and more examples of protein enzymes were found, it began to appear that biological catalysis would be exclusively the realm of proteins. In 1981 and 1982, my research group and I found a case in which RNA, a form of genetic material, was able to cleave and rejoin its own nucleotide linkages. This self-splicing RNA provided the first example of a catalytic active site formed of ribonucleic acid.

This lecture gives a personal view of the events that led to our realization of RNA self-splicing and the catalytic potential of RNA. It provides yet another illustration of the circuitous path by which scientific inquiry often proceeds. The decision to expend so many words describing the early experiments means that much of our current knowledge about the system will not be mentioned. For a more comprehensive view of the mechanism and structure of the *Tetrahymena* self-splicing RNA and RNA catalysis in general, the reader is directed to a number of recent reviews (Cech & Bass, 1986; Cech, 1987, 1988a, 1990; Burke, 1988; Altman, 1989). Possible medical and pharmaceutical implications of RNA catalysis have also been described recently (Cech, 1988b).

Why Tetrahymena?

In the pre-recombinant DNA era of the early 1970's, much of the research on the structure and function of eukaryotic chromosomes utilized entire

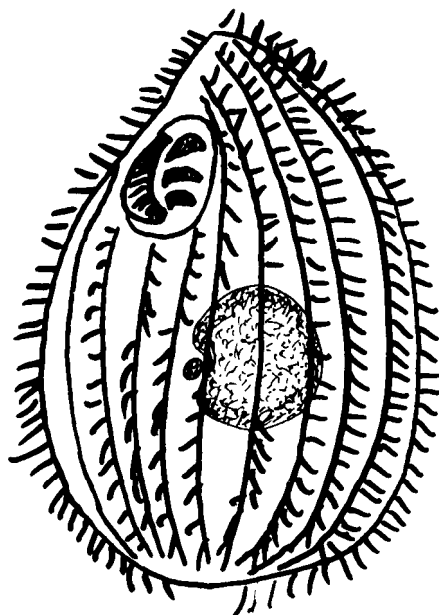


Figure 1. *Tetrahymena thermophila*, showing the transcriptionally active macronucleus and the germ-line micronucleus.

genomes as experimental systems. My own research with John Hearst in Berkeley and with Mary Lou Pardue at M.I.T. concerned the organization of DNA sequences and chromosomal proteins in the mouse genome. During my stay at M.I.T., I began to be dissatisfied with this global approach and became interested in the prospect of being able to dissect the structure and expression of some particular gene. Thus, when I set up my own laboratory in Boulder in 1978, I turned my attention entirely to the rDNA (gene for the large ribosomal RNAs) of the ciliated protozoan, *Tetrahymena* (Figure 1).

Unlike most nuclear genes, which are embedded in giant chromosomes, the genes for rRNA in *Tetrahymena* are located on small DNA molecules in the nucleoli; they are extrachromosomal (Engberg et al., 1974; Gall, 1974). Furthermore, in the transcriptionally active macronucleus the gene is amplified to a level of $\approx 10,000$ copies (Yao et al., 1974). These properties made it possible to purify a significant amount of the rDNA. The ability to purify the gene was not in itself a major attraction, because by this time the availability of recombinant DNA techniques ensured that no gene would long escape isolation and sequence analysis. Rather the attractive feature was the prospect of being able to isolate the gene complete with its associated structural proteins and proteins that regulated transcription (the synthesis of an RNA copy by RNA polymerase).

One feature of the *Tetrahymena* rDNA that was of only peripheral interest to me at that time was the presence of an intervening sequence (IVS) or intron, which interrupted the rRNA-coding sequences of the rDNA of some strains of *Tetrahymena pigmentosa* (Wild & Gall, 1979). In the course of mapping the RNA-coding regions of the rDNA of *Tetrahymena thermophila*,

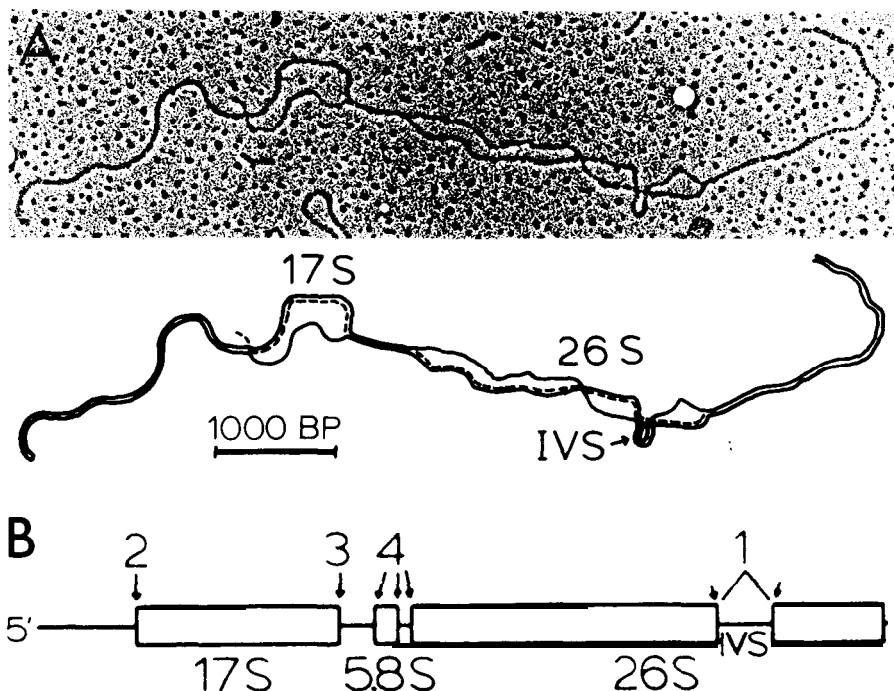


Figure 2. (A) Visualization of the RNA-coding portions of the *T. thermophila* rDNA by electron microscopy. The mature, processed rRNA was hybridized to the DNA under R-loop conditions. In the interpretation, each solid line indicates a single strand of DNA and each dashed line a strand of RNA. (Reproduced from Cech & Rio, 1979) (B) The pre-rRNA, thin lines representing portions that are removed during processing and open boxes representing the mature rRNA sequences. Numbered arrows indicate the usual order of RNA processing events: 1, splicing; 2 - 4, endonucleolytic cleavages.

Don Rio and I found that this species also harbored an IVS in its rDNA (Figure 2; Cech & Rio, 1979; independently described by Din et al., 1979). Although intervening sequences had been discovered only two years before, by Phil Sharp's lab at M.I.T. and a group from Cold Spring Harbor Laboratory, there were already a large number of examples. Thus, the finding of another IVS was hardly cause for us to be distracted from our plan to investigate the proteins that regulated rDNA transcription.

RNA Splicing in Vitro

The first step towards biochemical dissection of the transcriptional process was to see if rRNA synthesis would proceed in a crude cell-free system. We isolated nuclei from *T. thermophila* (the nuclei provided both the RNA polymerase and the ribosomal chromatin templates) and incubated them with the nucleoside triphosphates and salts necessary for transcription. We also included α -amanitin, a mushroom toxin known to inhibit the polymerases that transcribe mRNA, tRNA and other small RNAs, which enabled us to focus on synthesis of the large rRNA.

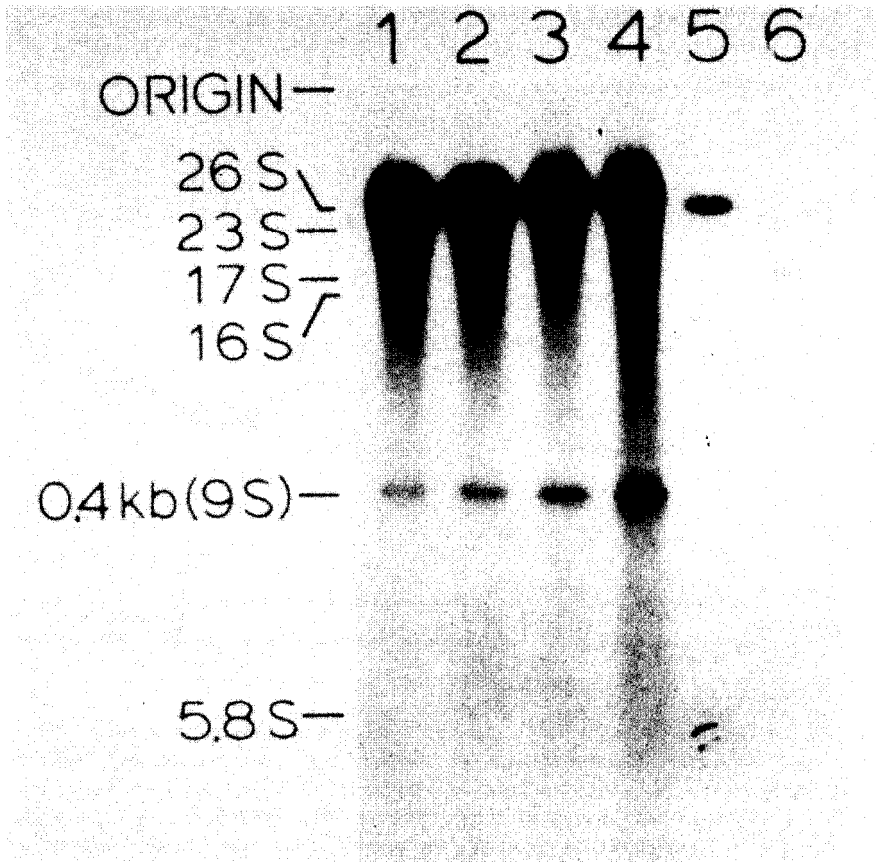


Figure 3. Transcription and splicing of pre-rRNA *in vitro* in isolated *T. thermophila* nuclei. (Lanes 1 - 4) RNA produced by incubation of nuclei at 30°C for times ranging from 5 - 60 min. The 0.4 kilobase species is the excised IVS RNA. (Lanes 5 and 6) Purified *Tetrahymena* 26S and 17S rRNAs, serving as molecular weight standards. RNA was analyzed by polyacrylamide gel electrophoresis. (Reproduced from Zaug & Cech, 1980; copyright by Cell Press)

When the products of these *in vitro* transcription reactions were separated by gel electrophoresis, they were found to consist of a somewhat heterogeneous distribution of high molecular weight RNA ≥ 26 S, the size expected for full-length pre-rRNA (Figure 3). In addition, there was a discrete low molecular weight product (≈ 9 S). The small RNA accumulated post-transcriptionally (Zaug & Cech, 1980). Thus, it seemed likely to be one of several short regions of the pre-rRNA that was cut out and ultimately discarded during the maturation process. These candidate regions included an external transcribed spacer at the 5' end, the internal transcribed spacers flanking the 5.8 S rRNA, and the IVS (Figure 2B).

Driven more by curiosity than by any conviction that the results would be of central importance to our research goals, I encouraged Art Zaug to identify the sequences encoding the small RNA. He confirmed that the

small RNA was encoded by the rRNA gene, and then mapped it to the intervening sequence (Zaug & Cech, 1980; see also Carin et al., 1980).

This was a finding of considerable excitement: the intervening sequence, synthesized as part of the pre-rRNA in our *in vitro* transcription reactions, was also being cleanly excised from the pre-rRNA *in vitro*. Despite a great deal of interest in the mechanism of RNA splicing (Damell, 1978; Abelson, 1979; Crick, 1979; Lerner et al., 1980), in only one other case - pre-tRNA in yeast - had RNA splicing been confirmed to occur *in vitro* (Knapp et al., 1979; Peebles et al., 1979). It seemed reasonable that rRNA splicing in *Tetrahymena* might follow a quite different path than tRNA splicing in yeast, so that detailed study of both systems would be justified.

Furthermore, in each *Tetrahymena* cell there were 10,000 identical genes each pumping out unspliced pre-rRNA at the rate of one copy per gene per sec. I reasoned that the nuclei might contain an unusually high concentration of the splicing enzyme to accomplish so much reaction, which could facilitate isolation of the first splicing enzyme. Little did we guess that the splicing "enzyme" would not exist in the traditional sense, but that something much more interesting lay in wait for us.

Self-splicing Unrecognized

Our strategy for purifying the splicing enzyme was conventional. We would find conditions in which pre-rRNA transcription occurred but splicing was inhibited, and purify the accumulated pre-rRNA to use as a substrate. We would then treat the pre-rRNA with extracts of *Tetrahymena* nuclei, which we already knew contained splicing activity, and would use gel electrophoresis to monitor the splicing reaction. Finally, we would obtain ever purer subfractions of the nuclear extract that retained activity, eventually isolating the splicing enzyme.

Isolation of unspliced pre-rRNA substrate proved to be straightforward (Cech et al., 1981). Art Zaug purified this RNA by standard SDS-phenol extraction and used it in a series of RNA splicing reactions. One set of test tubes contained substrate RNA and nuclear extract dissolved in the same solution of simple salts and nucleotides that had been conducive to RNA transcription and splicing in our earlier experiments with intact nuclei. Another tube, containing the same components except with the nuclear extract omitted, was to serve as the "splicing minus" control.

The very first attempt was successful. The RNA in the tubes containing the nuclear extract gave rise to the small band of RNA characteristic of the IVS. Surprisingly, however, the control RNA incubated only in salts and nucleotides produced the same amount of IVS. "Well, Art, this looks very encouraging, except you must have made some mistake making up the control sample." Yet several careful repetitions of the experiment gave the same result: release of the IVS occurred independent of the addition of nuclear extract, and therefore apparently independent of any enzyme. I became concerned that we weren't observing RNA splicing at all; perhaps the RNA we had been calling "precursor" had already been spliced *in vivo*,

and what we were observing *in vitro* was some disaggregation or release of the IVS from already spliced RNA. Clearly the reaction would be worthy of further pursuit only if we could show that a chemical transformation was occurring *in vitro*, and that it was the same cutting and rejoining that occurred in the living cell. We would have to teach ourselves some RNA chemistry.

The Mystery of the Extra G

To verify the accuracy of RNA splicing in nuclei, we decided to determine the nucleotide sequence of the IVS RNA product. (It was not obvious that this would be particularly illuminating, since we had already shown that the IVS product came from the IVS region of the rDNA and that within the error limits of our measurements it was the correct size to account for the entire IVS.) When Art Zaugg labeled the RNA on its 5' end with ^{32}P and subjected it to sequencing reactions, he determined a sequence 5'-GAAAUAGNAA... (where N represents an unidentified nucleotide). The DNA sequence across the exon-IVS junction had been reported earlier for *T. pigmentosa* by Wild and Sommer (1980), and was being determined for *T. thermophila* by Nancy Kan in Joe Gall's laboratory. The *T. thermophila* DNA sequence predicted that the IVS RNA would begin with 5'-AAAUAGCAA. . . .

Thus, our RNA sequence was a perfect match to the DNA sequence except for the extra G residue on its 5' end. Art Zaugg meticulously checked and rechecked the identity of this terminal nucleotide using a variety of enzyme treatments and chromatography systems until there was no doubt: the IVS RNA began with an ordinary guanosine residue, linked to the next nucleotide by a standard 3' - 5' phosphodiester bond such as that produced by RNA polymerase. Clearly the Gall lab, known for the high quality of their science, must have made an error. We telephoned them, advising them that they had determined most of the sequence correctly but had apparently missed one G right at the 5' end of the IVS. Much to our surprise, they defended every nucleotide of their sequence: no ambiguity in the DNA sequence, at least in that region, and no chance of a G at the 5' splice site.

At about the same time, I was working to define the minimum components necessary for the release of the IVS from pre-rRNA. The original experiments had been done in a "transcription cocktail" that included the four nucleoside triphosphates, building blocks for RNA synthesis. I found that removal of three of the NTPs had no effect on the reaction, but the fourth, GTP, was required in micromolar concentration. In addition, IVS release required MgCl_2 and was stimulated by certain salts such as $(\text{NH}_4)_2\text{SO}_4$.

Was it a coincidence that GTP, the nucleotide required for IVS release in our simple *in vitro* system, was also the nucleotide that was found unexpectedly at the 5' end of the excised IVS? Or might there be a causal relationship between the two observations? The obvious hypothesis was that GTP was required so that it could be added to the 5' end of the IVS during splicing.

The test was simple: mix “P-labeled GTP with unlabeled pre-rRNA, and look for labeling of the IVS RNA concomitant with its excision. The experiment was the strangest I had ever performed. On the one hand, its success was a straightforward prediction from our existing knowledge of the system. On the other hand, it seemed incredibly naive and unrealistic to expect that simple mixing of a nucleotide with phenol-extracted, proteinase-treated RNA could possibly result in formation of a covalent bond. I certainly didn't want to be embarrassed in front of my graduate students and colleagues by the failure of such an experiment, so I did it very quietly.

The next day I ran the gel, exposed it for autoradiography, and developed the X-ray film. In the sample containing ^{32}P -GTP plus MgCl_2 there was a bright signal of radioactivity at the position of the IVS RNA. In a sample containing ^{32}P -GTP but no MgCl_2 , and in a sample in which ^{32}P -ATP had been substituted for the ^{32}P -GTP, there was no labeled IVS.

Over the next weeks, we confirmed several major features of the reaction. GTP addition was stoichiometric, one GTP per IVS RNA. The GTP was added precisely to the 5' end of the IVS by a normal 3' - 5' phosphodiester bond. Finally, the triphosphate was unnecessary; GMP and even guanosine were active. The last of these observations eliminated one otherwise very reasonable hypothesis, that the GTP was providing an energy source for ligation much as ATP is used by phage T4 RNA ligase. Had that been the case, forms such as guanosine and GMP which are missing the phosphoanhydride moiety would have been inactive.

It took only a few moments of thought to devise a simple splicing pathway that integrated our new information about the reaction. Addition of guanosine to the phosphorus atom at the 5' splice site must be occurring by a transfer of phosphate esters, or transesterification reaction (Figure 4A). Such a reaction would free the 5' exon, leaving a new 3' hydroxyl group at its 3' end. The simplest way to proceed from such proposed intermediates to the final observed products was to invoke a second transesterification reaction: attack of the 3' hydroxyl of the 5' exon at the 3' splice site. Thus, a single active site capable of promoting transesterification could be responsible for the entire splicing reaction.

The model shown in Figure 4A has undergone little change since we first described it in 1981 and drew a more explicit version in 1982 (Cech et al., 1981; Zaug & Cech, 1982). It has been a good predictor of a great many other IVS-catalyzed reactions since then (e.g., Zaug et al., 1983; Tabak et al., 1987). Furthermore, the model has been strongly bolstered by the isolation and characterization of the proposed intermediates (Inoue et al., 1986), by the demonstration of the reversibility of the reactions (Sullivan & Cech, 1985; Woodson & Cech, 1989), and by determination of the stereochemical course (McSwiggen & Cech, 1989; Rajagopol et al., 1989). An atomic-level model of transesterification occurring by an $\text{S}_{\text{N}}2(\text{P})$ mechanism is presented in Figure 4B.

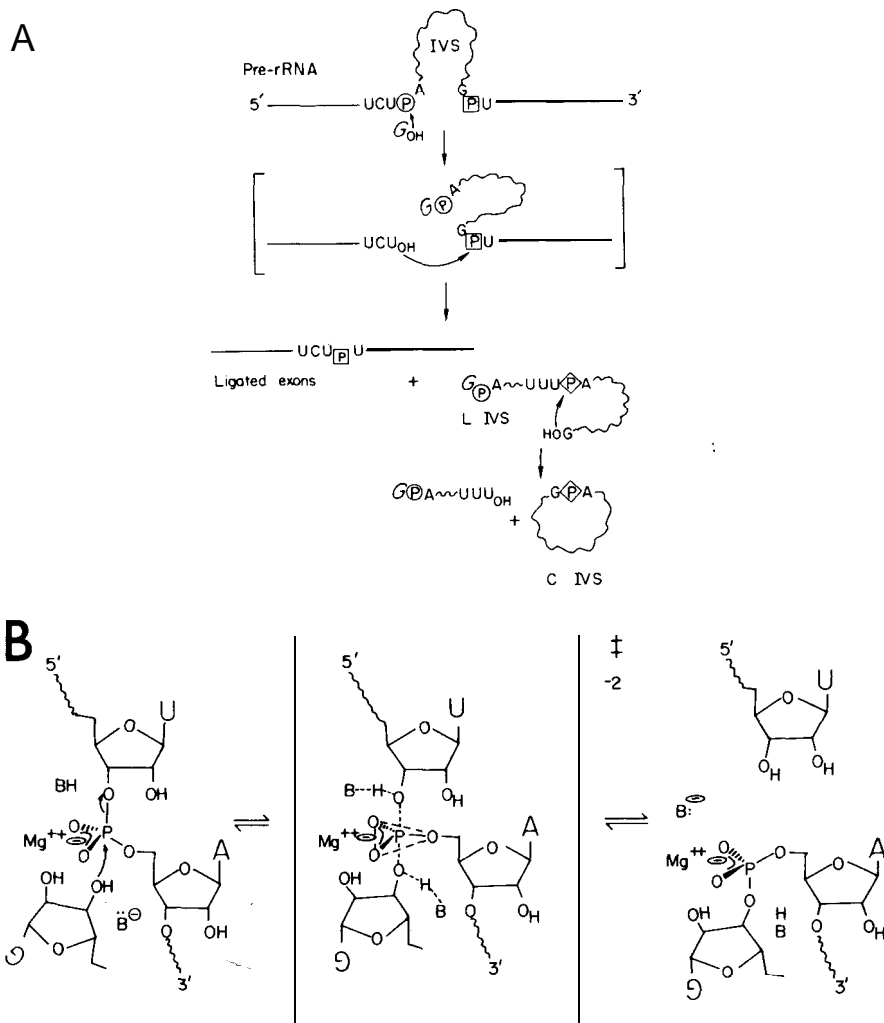


Figure 4. (A) Self-splicing of the *Tetrahymena* pre-t-RNA by consecutive transesterification reactions. Straight lines, exons (mature rRNA sequences); wavy line, IVS; circle, 5' splice-site phosphate; square, 3' splice-site phosphate; diamond, cyclization site phosphate. (B) Model for the initial step involving nucleophilic attack by the 3'-hydroxyl group of guanosine on the phosphorus atom at the 5' splice site. The hypothesis of an in-line, S_N2 (P) reaction with inversion of configuration around phosphorus was subsequently confirmed (McSwiggen & Cech, 1989). The hypotheses of acid-base catalysis and coordination of Mg²⁺ to the phosphate, enhancing the electrophilicity of the phosphorus atom and stabilizing the trigonal bipyramidal transition state, are untested. (Reproduced from Cech, 1987; copyright by the AAAS)

Self-splicing Recognized

The guanosine-addition reaction provided us with the proof we needed: specific RNA bond breakage and formation were occurring in our simple *in vitro* splicing reaction. Such a chemically difficult reaction between very unreactive molecules certainly had to be catalyzed. But what was the catalyst?



Figure 5. IVS RNA excision and cyclization activities are resistant to protease treatment. (Lane S) Pre-rRNA' purified by SDS-phenol extraction. This RNA was then treated exhaustively with proteases or (NT) not treated. Incubation with GTP was performed, (lanes -) in the absence of $MgCl_2$, or (lanes +) with 10 mM $MgCl_2$. Excision of the IVS RNA as a linear molecule (L) and subsequent conversion to a circular form (C) were undiminished by protease treatment. (Lane M) isolated linear and circular IVS RNAs. (Reproduced from Grabowski et al., 1983, by permission of Alan R. Liss, Inc.)

Our first hypothesis was that the splicing activity was a protein tightly bound (perhaps even covalently bonded) to the pre-rRNA isolated from *Tetrahymena* nuclei. This would have to be a very unusual protein-RNA complex to survive the multiple forms of abuse to which we had subjected it: boiling in the presence of the detergent SDS, SDS-phenol extraction at temperatures as high as 65°C and extensive treatment with several nonspecific proteases (Figure 5). That we took this hypothesis seriously provides an indication of how deeply we were steeped in the prevailing wisdom that only proteins were capable of highly efficient and specific biological catalysis.

In the same paper, we described an alternative hypothesis:

The resistance of the splicing activity to phenol extraction, SDS and proteases can also be interpreted in a more straightforward manner. The rRNA precursor might be able to undergo splicing without the participation of a protein enzyme. A portion of the RNA chain could be folded in such a way that it formed an active site or sites that bound the guanosine cofactor and catalyzed the various bond-cleavage and ligation events. If one of the RNA molecules produced in the reaction (for example, the free IVS) retained its activity and catalyzed additional splicing events, then it would be an example of an RNA enzyme. (Cech et al., 1981)

As we accumulated negative result upon negative result trying to identify a protein stuck to the pre-rRNA, the alternative “RNA only” hypothesis began to appear more and more attractive. But how could we obtain a positive result to prove it?

The best strategy we could devise was to synthesize the RNA in as artificial a manner as possible, so that it was never in contact with the *Tetrahymena* cells that up to now had been our sole source of the RNA. Complete chemical synthesis of an RNA the size of the pre-rRNA (or even the size of the IVS) was and still is beyond the scope of available technology. The next best approach was to synthesize RNA from a recombinant DNA template using purified RNA polymerase.

A bacterial plasmid (Figure 6) encoding the IVS and a portion of the flanking rRNA sequences, situated so as to allow transcription by *Escherichia coli* RNA polymerase, was already being constructed in the lab by Kelly Kruger. The original purpose was to facilitate synthesis of large quantities of pre-rRNA substrate for isolation of the splicing enzyme. The cloning

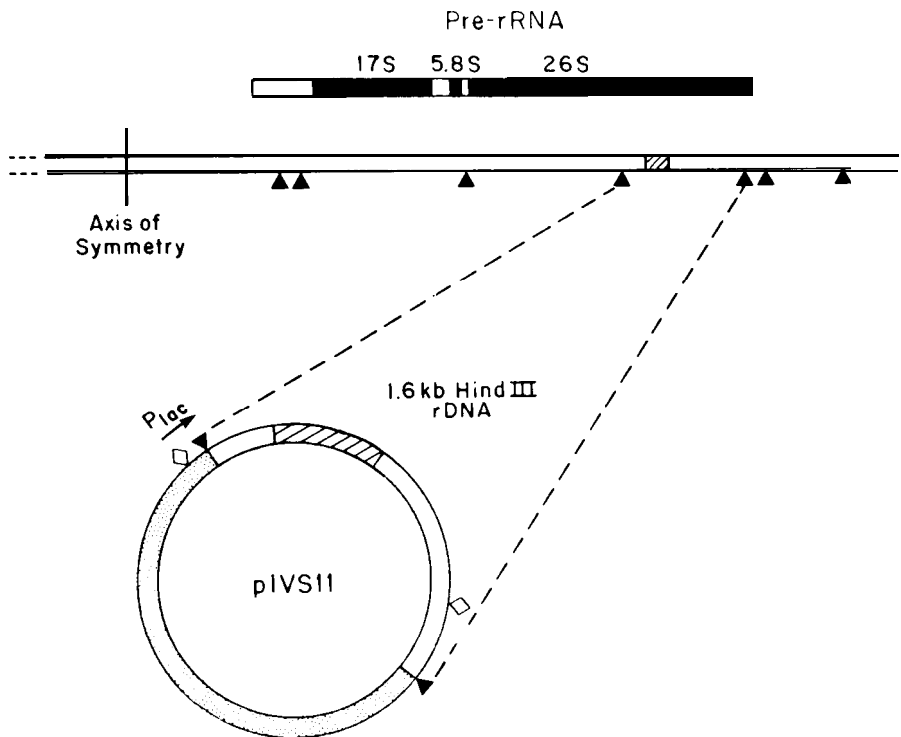


Figure 6. Plasmid constructed to enable synthesis of an artificial, shortened version of the pre-rRNA. (Top) Diagram of the natural pre-rRNA, with transcribed spacers shown as open boxes, mature rRNA sequences as solid boxes, and the IVS as a hatched box. (Middle) The *T. thermophila* rDNA. One half of a palindromic rDNA molecule is shown. (Bottom) Plasmid pIVS11 containing the 1.6 kilobase *Hind* III fragment of the rDNA inserted adjacent to P_{lac} , a promoter for transcription by *E. coli* RNA polymerase. Restriction endonuclease sites are *Hind* III (□) and *Eco* RI (▲). (Adapted from Kruger et al., 1982; copyright by Cell Press)

took longer than any of us had anticipated, such that by the time it was accomplished early in 1982 we were desperate to have the plasmid for a different purpose: to produce synthetic pre-rRNA for the self-splicing test.

The plasmid was grown up in *E. coli*, carefully deproteinized, and incubated with purified *E. coli* RNA polymerase under conditions that we already knew were inhibitory for splicing. The polymerase was then destroyed and the RNA purified by gel electrophoresis under denaturing conditions. Upon addition of GTP, $MgCl_2$, and salt, the TVS RNA was released from this artificial, shortened pre-rRNA. The site at which GTP broke the RNA chain was exactly the position that served as the 5' splice site *in vivo*, providing some confidence that self-splicing was relevant to splicing as it occurred in the living cell.

We held a relatively subdued celebration in the lab. Between sips of champagne we compiled a list of possible general names for RNA molecules able to lower the activation energy for specific biochemical reactions. It was then that we coined the term "ribozyme," for a ribonucleic acid with enzyme-like properties.

RNA in Circles

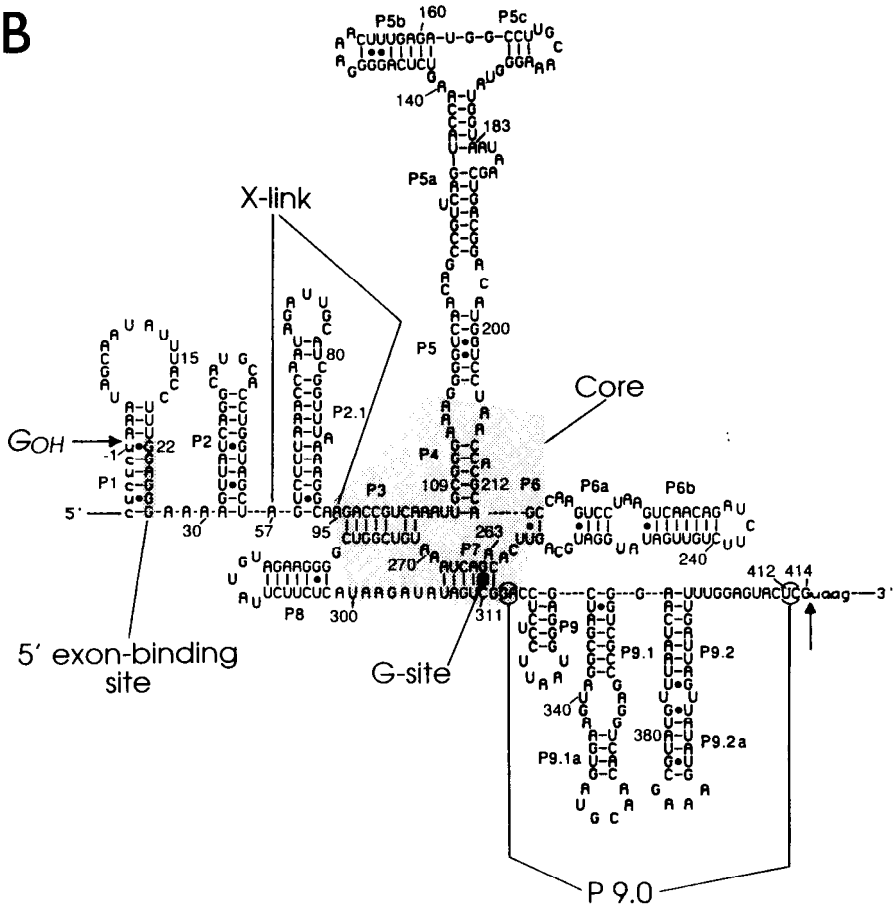
Concurrent with our studies of pre-rRNA splicing in isolated nuclei, we were pursuing a post-splicing phenomenon: the conversion of the excised intervening sequence RNA into a circular RNA molecule in isolated *Tetrahymena* nuclei (Grabowski et al., 1981). The circular form survived treatment with protease and various denaturants, which suggested that it was a covalently closed circle of RNA. As Paula Grabowski, then a graduate student in the laboratory, characterized the cyclization reaction, she found that it occurred with extensively deproteinized RNA in a simple buffered $MgCl_2$ solution. I dismissed her original observation of protein-free cyclization as being an artifact of incomplete denaturation of the linear RNA. Yet, as the experiments proved reproducible, I began to derive some solace in the knowledge that two researchers, Zaug and Grabowski, studying two different reactions, splicing and cyclization, were both finding activity in the absence of added protein. Having two strange results somehow made me more comfortable than just a single strange result. The two sets of observations came together when we found that the plasmid transcripts that underwent self-splicing produced IVS RNA that underwent self-cyclization (Kruiger et al., 1982).

The circular IVS RNA was not formed by end-to-end joining of the linear form. Instead, the 3' end of the linear IVS attacked an internal phosphorus atom near the 5' end of the molecule, clipping off a short oligonucleotide in the process (Figure 4A). Thus cyclization, like RNA splicing, occurred by transesterification (Zaug et al., 1983).

How Does the RNA Do It?

At first glance, RNA seemed ill suited to be a catalyst. With its four rather similar bases, RNA would appear greatly limited in its ability to form a

B



conserved sequence blocks interacted to form a common set of short base-paired regions, serving to fold the intervening sequences into the same fundamental secondary structure (Figure 7A). Michel called these the group I introns, to distinguish them from a second group of mitochondrial introns that shared a different structure (Michel & Dujon, 1983).

The *Tetrahymena* rRNA IVS contained the conserved sequence elements and secondary structures characteristic of the mitochondrial group I (Michel & Dujon, 1983; Waring et al., 1983). Furthermore, a very similar structure model of the *Tetrahymena* IVS was independently derived by another method, free energy minimization as constrained by experimentally determined sites of cleavage of the folded RNA by various nucleases (Cech et al., 1983). A current version of the secondary structure is shown in Figure 7B.

This convergence of two previously noninteracting sets of ideas was important in several respects. In terms of splicing mechanisms, it was now unlikely that the *Tetrahymena* intron would be unique; we could extend our knowledge of its mechanism by comparing and contrasting splicing of different members of the group. Second, a believable model of the second-

dary structure of the *Tetrahymena* IVS was now in hand, and one could begin to formulate structure-function relationships. Finally, the similarity between the *Tetrahymena* rRNA and fungal mitochondrial introns might be revealing their origin; perhaps they were transposable elements able to enter both nuclear and mitochondrial compartments (Cech et al., 1983).

Waiting for Number Two

If RNA catalysis were of any general significance to biology, there would be additional examples. Throughout 1982 and most of 1983, none came forth. Yet there seemed to be some reasonable candidates. In the fall of 1983, we wrote an article in which we speculated:

Several enzymes, such as RNase P, 1,4- α -glucan branching enzyme and potato *o*-diphenol oxidase, have RNA components essential for their catalytic activities. The peptidyl transferase activity of ribosomes also requires an RNA-protein complex. It remains to be seen whether the RNA is directly involved in the active site of any of these ribonucleoprotein enzymes. (Bass and Cech, 1984)

The speculation about RNase P was already outdated when our paper was published in 1984, because by that time Guerrier-Takada et al. (1983) had announced that the RNA component was the catalytic subunit of that ribonucleoprotein enzyme. This was followed by the report early in 1984 that a synthetic *E. coli* RNase P RNA, transcribed *in vitro* from a recombinant DNA template, also had intrinsic catalytic activity; the possibility that catalysis was due to protein contamination was thereby eliminated (Guerrier-Takada and Altman, 1984). Similarly, the RNA subunit of *Bacillus subtilis* RNase P acted as an enzyme *in vitro* (Guerrier-Takada et al., 1983; Marsh & Pace, 1985). The RNase P discovery was very exciting to us. Not only did it provide a second example of an RNA molecule that lowered the activation energy for a specific biochemical reaction, but it was the first proven case of an RNA molecule that catalyzed a reaction without itself undergoing any net change. RNA catalysis was not restricted to the realm of intramolecular catalysis.

Within the next year, self-splicing of additional group I intervening sequences was reported. Garriga and Lambowitz (1984) found that the first IVS of the cytochrome *b* pre-mRNA from *Neurospora* mitochondria underwent self-splicing *in vitro*. Several self-splicing group I RNAs from yeast mitochondria were characterized by van der Horst and Tabak (1985). Most unexpectedly, self-splicing group I IVSs were found in three bacteriophage T4 mRNAs (Belfort et al., 1985; Ehrenman et al., 1986; Gott et al., 1986); RNA splicing took place in a prokaryote. In all cases, splicing occurred by the same G-addition pathway as splicing of the *Tetrahymena* nuclear rRNA IVS.

The mitochondrial group II intervening sequences have conserved sequences and secondary structures distinct from those of group I (Michel and Dujon, 1983). Peebles et al. (1986) and Van der Veen et al. (1986)

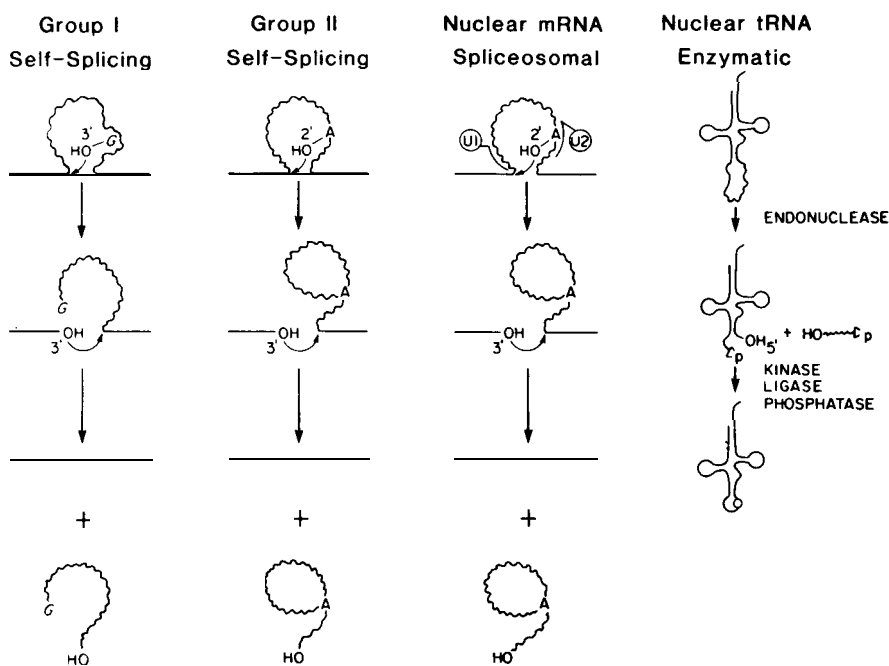


Figure 8. Mechanism of splicing of the four major groups of precursor RNAs. Wavy lines indicate IVSs, smooth lines indicate flanking exons. For nuclear mRNA splicing, many components assemble with the pre-mRNA to form the spliceosome; only the UI and U2 small nuclear ribonucleoproteins are shown here. (Adapted from Cech, 1986a; copyright by Cell Press)

discovered that pre-mRNA containing a group II intervening sequence was self-splicing *in vitro*. The reaction did not require guanosine, and occurred by formation of a branched “lariat” RNA. The proposed mechanism is shown in Figure 8.

The fundamental chemistry of group II RNA splicing appears to be the same as that of nuclear pre-mRNAs, which do not self-splice. Instead, nuclear mRNA splicing requires assembly of the substrate with a large complex of proteins and small nuclear ribonucleoproteins to form the spliceosome (Brody & Abelson, 1985; Frendewey & Keller, 1985; Grabowski et al., 1985). The mechanistic similarities have led to the speculation that nuclear mRNA splicing may also be RNA catalyzed, with much of the catalysis being provided in the form of the small nuclear RNAs (Kruger et al., 1982; Maniatis and Reed, 1987).

Enzymologists Outraged

Although our description of RNA self-splicing was shocking to many, it was quickly accepted by the scientific community. In contrast, our use of the words “catalysis” and “enzyme-like” to describe the phenomenon provoked some much more heated reactions.

Our reasons for emphasizing the relationship between RNA self-splicing and biological catalysis might be better appreciated in the context of the

Table 1. Defining characteristics of biological catalysis.

Characteristic	Self-splicing	L - 19 IVS RNA
1. Rate acceleration	✓	✓
2. Specificity	✓	✓
3. Catalyst regenerated	*	✓

* No, although the active site is preserved through the reactions. (From Cech, 1988c, by permission of Alan R. Liss Inc.)

definition given in Table 1. First, biological catalysts achieve rate accelerations of the order of 10^5 - to 10^{13} -fold, bringing the reactions into a time scale that is useful for living systems. Self-splicing clearly meets this criterion; although the quantitation of rate acceleration can be done only for the site-specific hydrolysis reaction promoted by the RNA active site, even this relatively slow side-reaction occurs 10^{10} -fold faster than the estimated uncatalyzed rate (Figure 9). Second, the extraordinary specificity of biological catalysts is evident in the self-splicing reaction; the molecule selects GTP as the attacking group for the first step of the reaction and is able to choose 2 of the 6000 nucleotides in the pre-rRNA as splice sites. On the other hand, the IVS RNA is clearly not regenerated in exactly the same form as it entered the reaction; after all, the purpose of self-splicing is to convert pre-rRNA to ligated exons plus excised IVS. Nevertheless, enzymologists do speak of intramolecular catalysis (Jencks, 1969; Bender et al., 1984; Fersht, 1985), and we thought that such a descriptor was particularly appropriate for self-splicing.

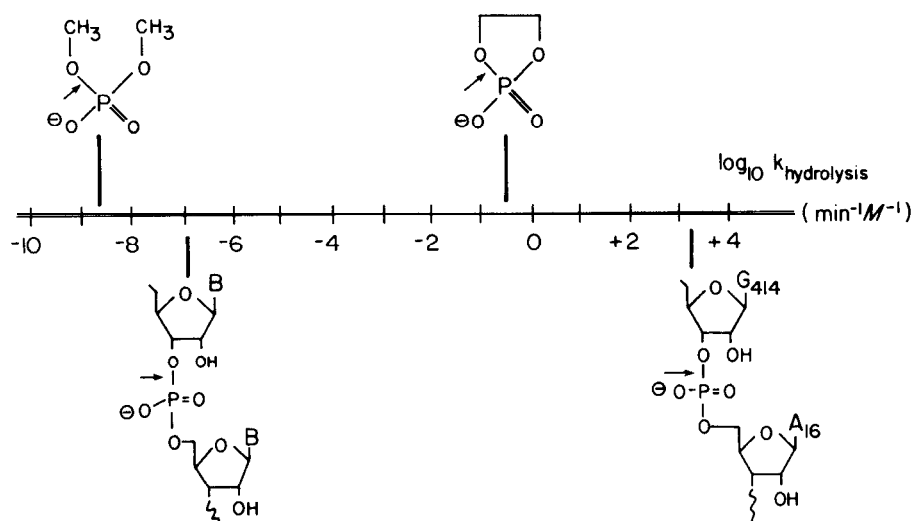


Figure 9. The IVS RNA has an extremely efficient catalytic center. Second-order rate constants (42°C) for the alkaline hydrolysis of phosphate diesters are displayed on a logarithmic scale. Arrows designate the P-O bond that undergoes cleavage. Above line, data for dimethyl phosphate and ethylene phosphate from Kumamoto et al. (1956) and Haake and Westheimer (1961). Below line, hydrolysis of RNA (left) uncatalyzed and (right) catalyzed by the catalytic center of the *Tetrahymena* IVS. (Reproduced from Cech, 1987; copyright by the AAAS)

The enzymologists who reviewed Bass & Cech (1984) were far from convinced. All three referees wrote thoughtful reviews, expressing considerable interest in the data but chastising us for our naivete about enzymes. For example:

Enzymes are true catalysts: they speed the rate of a reaction, but are themselves unchanged by the reaction. In the present instance, the ribozyme acts in a "one-shot" reaction and is permanently changed so that it can no longer cycle as does a true catalyst. The authors are well aware of this, but appear to ignore this key feature in making their comparisons.

How fundamental was this distinction between self-splicing and catalysis? Opinions would probably vary as much today as they did six years ago. Instead of engaging in a protracted argument, we decided to test experimentally whether the self-splicing RNA could be converted into a multiple-turnover catalyst. Arthur Zaugg made a slight alteration of the self-splicing IVS, removing its first 19 nucleotides. The resulting L - 19 IVS RNA met all three criteria of a biological catalyst (Table 1 and Zaugg & Cech, 1986a).

Shortened forms of the *Tetrahymena* IVS have multiple enzymatic activities. Depending on the substrates with which they are presented, these RNA enzymes can catalyze nucleotidyltransfer, phosphotransfer and hydrolysis reactions; nucleotidyltransfer can result in either endonucleolytic cleavage or polymerization of RNA (Zaugg & Cech, 1986a,b; Zaugg et al., 1986; Kay & Inoue, 1987; Been & Cech, 1986, 1988; Doudna & Szostak, 1989).

As an example, consider the reaction diagrammed in Figure 10. A form of the IVS RNA missing both its splice sites catalyzes the cleavage of other RNA molecules after sequences resembling that of the normal 5' exon, CUCU. The reaction is an intermolecular version of the first step of RNA self-splicing, in which RNA cleavage is accompanied by covalent joining of guanosine to the 5' end of the 3' cleavage product. Because the IVS RNA is

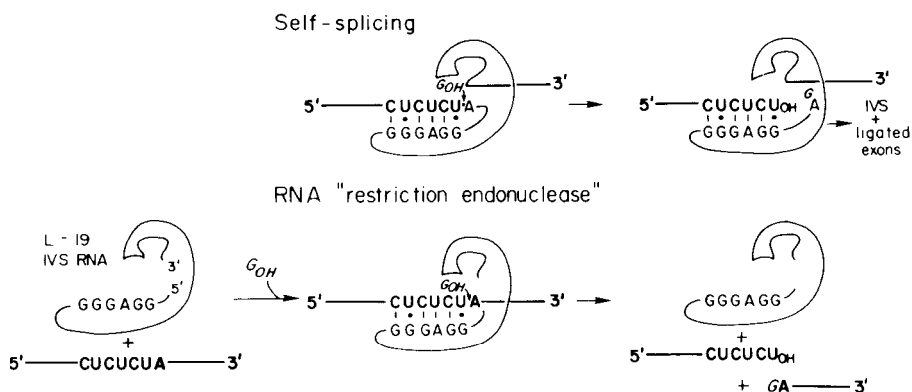


Figure 10. A shortened version of the *Tetrahymena* IVS RNA has enzymatic activity as an endonuclease. The mechanism is an intermolecular version of the first step of pre-rRNA self-splicing. Thin letters and lines, IVS sequences; bold letters and thick lines, exon sequences (top) or substrate RNA sequences (bottom); *G* in italics, free guanosine or GTP. (Reproduced by permission from Zaugg et al., 1986; copyright Macmillan Journals Limited)

unaltered by the reaction, it can sequentially bind and process a large number of substrate RNA molecules. The k_{cat}/K_m for cleavage of an oligoribonucleotide substrate containing a CCCUCU recognition sequence is $10^8 \text{ M}^{-1} \text{ min}^{-1}$ (Herschlag & Cech, 1990), well within the range of protein enzymes.

Site-directed mutagenesis of the 5' exon-binding site of the IVS redirects substrate specificity as predicted by the rules of Watson-Crick base-pairing (Zaug et al., 1986; Murphy & Cech, 1989). Thus, it has been possible to create a whole set of "RNA restriction endonucleases" that may be of use for the sequence-specific cleavage of RNA.

Converting a self-processing RNA into an RNA enzyme by physically separating its internal substrate from its catalytic center has been generally successful. The "hammerhead" and "hairpin" ribozymes both found in plant infectious agents (viroids or viral satellite RNAs), undergo self-cleavage leaving 2',3'-cyclic phosphate termini (Prody et al., 1986; Forster & Symons, 1987). Both have been converted into RNA enzymes (Uhlenbeck, 1987; Haseloff & Gerlach, 1988; Hampel & Tritz, 1989). A group II IVS can cleave RNA containing a 5' splice-site or ligated exon RNA in an intermolecular reaction (Jacquier & Rosbash, 1986; Jarrell et al., 1988), although in these cases multiple turnover was not demonstrated. Even the intramolecular lead-cleavage reaction of tRNA can be converted into an enzymatic system (Sampson et al., 1987). The ability of pieces of a structured RNA molecule to self-assemble, recreating an active unit, therefore appears to be a general property of the RNA biopolymer. In addition, if the fragment which undergoes reaction is secured to the rest of the molecule by a relatively weak interaction such as a few base-pairs, it will dissociate quickly enough to permit multiple turnovers.

Origin of Life Fantasies

The discoveries of RNA self-splicing and the enzymatic activity of RNase P RNA rekindled earlier speculation concerning the possible role of RNA in the origin of life (Woese, 1967; Crick, 1968; Orgel, 1968). Contemporary cells depend on a complex interplay of nucleic acids and proteins, the former serving as informational molecules and the latter as the catalysts that replicate and express the information. Certainly the first self-reproducing biochemical system also had an absolute need for both informational and catalytic molecules. The dilemma was therefore *Which came first, the nucleic acid or the protein, the information or the function?* One solution would be the co-evolution of nucleic acids and proteins (Eigen, 1971). The finding that RNA can be a catalyst as well as an information-carrier lent plausibility to an alternative scenario: the first self-reproducing system could have consisted of RNA alone (Sharp, 1985; Pace & Marsh, 1985; Orgel, 1986).

Perhaps coincidentally or perhaps because of its ancestry, one of the reactions catalyzed by the *Tetrahymena* IVS RNA enzyme is a nucleotidyl transfer reaction with fundamental similarity to the reaction catalyzed by RNA replicases (Zaug & Cech, 1986a; Been & Cech, 1988). A specific model

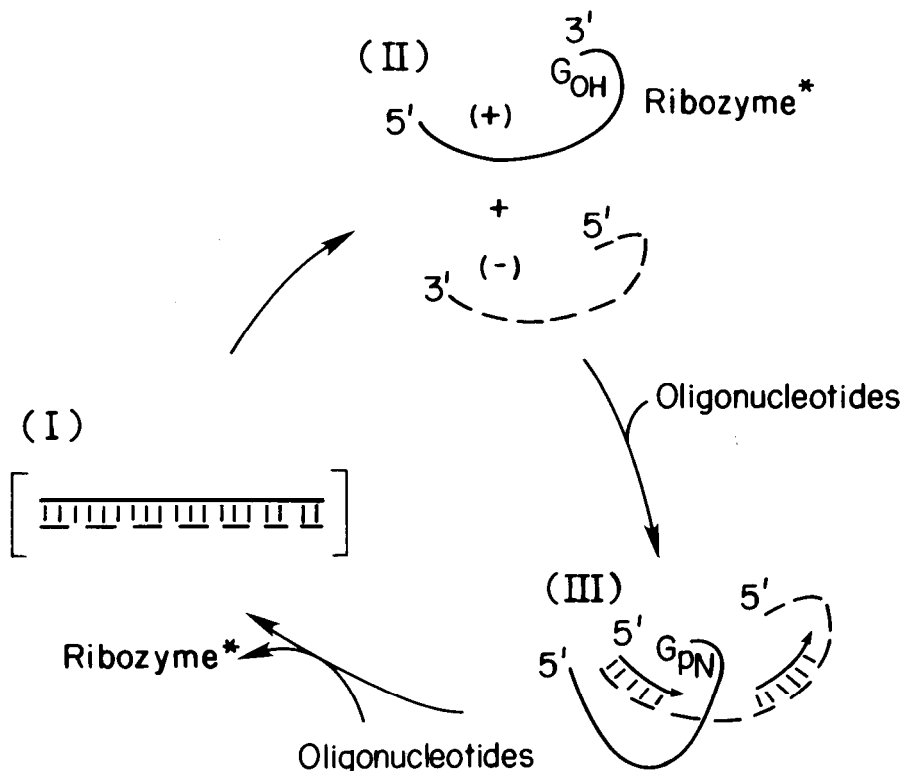


Figure 11. Hypothetical model for RNA self-replication involving an RNA catalyst (ribozyme*). Double-stranded RNA (I) undergoes strand separation to give ribozyme* ((+) strand) and the complementary (-) strand (II). The ribozyme* catalyzes synthesis of a new (+) strand, using the (-) strand as a template (III). The detailed mechanism is described by Cech (1986b). Completion of synthesis reforms the double-stranded RNA (I). A second cycle is needed to achieve replication of the starting material.

for RNA self-replication based on the properties of the catalytic center of the *Tetrahymena* IVS RNA is given in Figure 11. The RNA enzyme, ribozyme*, differs from the L - 19 IVS RNA in that it utilizes an external rather than an internal template (Cech, 1986b). Separation of the template region from the rest of the catalytic center of the RNA with retention of activity has recently been achieved by Doudna and Szostak (1989).

Now that we have examples of catalytic RNAs, it has been entertaining to look back at earlier speculations about the catalytic potential of RNA. As a representative example, consider the following:

... in the evolutionary scheme, folded nucleic acid structures were abandoned by nature in favour of proteins. Although nucleic acids may have performed many enzymatic tasks in primitive cells, this is much more efficiently done by proteins. . . . With nucleic acids the four bases are all of one structural type, though the existence of many modified bases points to an evolutionary proliferation giving much greater possibilities in form-

ing structures. Nevertheless, the use of bases cannot match the enormous flexibility provided by having twenty amino acids which fall into three or four different structural types. (Klug et al., 1974)

One implication is that nucleic acids, if they could have any such activity, would by nature be inferior catalysts. At one time I had a similar bias, but it was gradually dispelled as we quantitated the rate acceleration and specificity inherent to the catalytic center of the *Tetrahymena* ribozyme. The second conclusion, regarding the limited versatility of RNA catalysts, still strikes me as being correct. In all well established examples, the substrate for an RNA catalyst or ribonucleoprotein enzyme is RNA or DNA or, in the case of protein synthesis on ribosomes, the closely related aminoacyl-tRNA. To its credit, RNA can form a specific binding site for at least one amino acid (Yarus, 1988), and there is evidence that covalent linkage of the terminal protein to poliovirus RNA is at least in part RNA-catalyzed (Tobin et al., 1989). Nevertheless, it seems unlikely that RNA can match the enormous variety of binding sites that can be formed from amino acid side chains. The list of RNA catalysts is still growing quite rapidly. Yet it seems likely that if an entire list of biological catalysts is ever complete, it will include more proteins than RNA molecules.

ACKNOWLEDGEMENTS

I dedicate this lecture to my wife Carol and to my parents, Robert and Annette. I am indebted to my coworkers over the past decade, who performed the experiments and contributed many of the ideas about RNA catalysis described here. In addition to those whose work I have described, there are many who made important contributions that have gone unmentioned here but not at all unappreciated. Funding for our research has been provided principally by the National Institutes of Health and the American Cancer Society, and more recently by the Howard Hughes Medical Institute; I thank them for their uninterrupted and enthusiastic support. Finally, I have enjoyed the comradeship and helpful criticism of the RNA research community, both in Boulder and worldwide.

REFERENCES

- Abelson J. (1979) *Ann. Rev. Biochem.* **48**, 1035-1069.
Altman S. (1989) *Adv. Enzym.* **62**, 1-36.
Barford, E. T. & Cech, T. R. (1989) *Molec. Cell. Biol.* **9**, 3657-3666.
Bass, B. L. & Cech, T. R. (1984) *Nature* **308**, 820-826.
Bass, B. L. & Cech, T. R. (1986) *Biochemistry* **25**, 4473-4477.
Been, M. D. & Cech, T. R. (1986) *Cell* **47**, 207-216.
Been, M. D. & Cech, T. R. (1988) *Science* **239**, 1412-1416.
Belfort, M., Pedersen-Lane, J., West, D., Ehrenman, K., Maley, G., Chu, F. & Maley, F. (1985) *Cell* **41**, 375-382.
Bender, M. L., Bergeron, R. J. & Komiyama, M. (1984) *The Bioorganic Chemistry of Enzymatic Catalysis* (John Wiley, New York).

