# RNA Can Be a Catalyst

## *The discovery that RNA can catalyze biochemical reactions revolutionizes ideas on biological catalysis and early evolution*

Everyone knows that biological catalysis is the sole province of proteins. Period. The discovery by Thomas Cech and his colleagues* at the University of Colorado of a biochemical reaction mediated by RNA in the absence of protein therefore comes as something of a surprise. And the importance of the finding reaches beyond the world of catalysis in that the notion of catalytic RNA immediately opens new perspectives on the basic components required in the early evolution of life. Harry Noller of the University of California at Santa Cruz calls Cech's work "one of the most exciting discoveries of the decade."

Cech and his colleagues happened upon their discovery "through a series of accidents." When Cech went to Colorado 5 years ago he chose to work with the ciliated protozoan *Tetrahymena thermophila*, an especially convenient organism with which to study certain aspects of eukaryotic molecular biology. He was interested in chromatin structure as it impinges on the regulation of gene activity. The gene selected was the superabundant ribosomal RNA gene. What Cech's team finished up studying was the organism's precursor ribosomal RNA from which an intervening sequence 413 nucleotides long is precisely excised and the coding regions rejoined, all in the total absence of protein. "We were forced to conclude that the catalytic activity was in the RNA molecule itself."

Unlike the rigid molecular structure typically formed by DNA, RNA is a much more flexible molecule with the potential to assemble itself into a diversity of conformations. In recent years there has been a quickening of interest in RNA as an active as opposed to solely structural molecule. For instance, ribosomal RNA is now thought by some to play more than a passive role in protein synthesis, particularly in the peptidyl transferase reaction. Sidney Altman of Yale University struggled for some years to establish ribonuclease P, a 30 to 1 combination of RNA to protein, as a legitimate enzyme. And the many ribonucleoprotein particles derived from small nuclear and cytoplasmic RNA's are expected by some to show activity in this respect too.

The convincing demonstration at the University of Colorado that RNA can indeed catalyze reactions is certain to give a fillip to this line of research. The difficulties to be faced are, however, enormous. While it is possible to calculate some aspects of secondary structure in RNA, albeit with wide margins for error, analysis of tertiary structure in large RNA molecules is at present not possible. It is principally in the tertiary structure that catalytic activity may lie.

Cech chose to work with the ribosomal RNA gene in *Tetrahymena* because it is extremely plentiful in that organism and because the organism provides a eukaryotic biochemical environment not available to a gene cloned into a bacterial host.

*Tetrahymena* has two nuclei: one large, the other small. The micronucleus packages the genetic material and is transcriptionally inactive. Its larger partner, which derives from the genetic nucleus in the organism's sexual phase, is polyploid and contains at least 45 copies of all the genes. The ribsomal RNA gene is even more numerous, there being about 20,000 copies of a head-to-head dimer, each of which exists as a mini-chromosome. This abundance of material compares favorably with what can be achieved with plasmids and phages with recombinant DNA technology. Moreover, most of the genes appear to be transcriptionally active so as to serve the tremendous demand for ribosomes during the organism's 2½ hour growth period between cell divisions.

> "We were forced to conclude that the catalytic activity was in the RNA molecule itself."

One of the first things Cech and his colleagues found was that the ribosomal RNA has an intervening sequence (IVS). Joseph Gall and Martha Wild at Yale University had earlier found an intervening sequence in the ribosomal RNA of some strains of *Tetrahymena pigmentosa*. The ribosomal RNA introns in *T. thermophila* and *T. pigmentosa* are virtually identical. The discovery of these introns was interesting in itself because, unlike the case in *Drosophila*, the intervening sequence is present in every copy of the gene. Moreover, the intron interrupts a coding sequence which is highly conserved through *Escherichia coli*, *Tetrahymena*, frogs, and humans. "For these reasons it was clear that the intervening sequence must be spliced out of the RNA," says Cech. "We said, this might be a good system in which to isolate an RNA splicing enzyme."

RNA processing, particularly messenger RNA splicing, is currently an intense and fashionable field of study. However, when Cech and his colleagues got their first information on potential splicing in *Tetrahymena*, that is early in 1979, they were intent on studying transcription. They therefore noted the presence of the intervening sequence and the absolute requirement for splicing, and continued.

The team developed an isolated nuclei transcription system that was not specifically designed to preserve other related reactions, including splicing. It became immediately obvious, nevertheless, that splicing was in fact occurring in this system. "We thought this meant that the splicing enzymes were tightly associated with the preribosomal RNA."

Because the system appeared to offer an effective way of studying splicing, Cech wrote a grant proposal to NIH, entitled "Enzymic splicing of ribosomal RNA precursor." The proposal was funded with a high priority score.

A method of isolating and accumulating unspliced RNA was eventually developed and it should have allowed closer scrutiny of the splicing reaction. It did, but not with the expected result. Excision of the intervening sequence proceeded at comparable rates both in the presence and absence of nuclear extract, as long as certain salts and nucleotides were present. "I said, this can't be right. Art Zaug repeated it five times in five different ways, but each time with the same result. Even then we didn't

*Kelly Kruger, Paula J. Grabowski, Arthur J. Zaug, Julie Sands, Daniel E. Gottschling, Thomas R. Cech, *Cell* **31**, 147–157 (1982).


0036-8075/82/1126-0872$01.00/0 

believe it, so we put it aside thinking it must be an artifact."

One possible problem was that the preribosomal RNA that Cech and his colleagues isolated might have been an aggregate of the mature ribosomal RNA and the excised intervening sequence. If this were the case, events observed would not be real splicing but simply the dissociation of the two component molecules from the aggregate. So, from the beginning of 1980 to the summer of 1981 the Colorado team switched its attention from trying to understand the nature of the activity of events to simply untangling the events themselves.

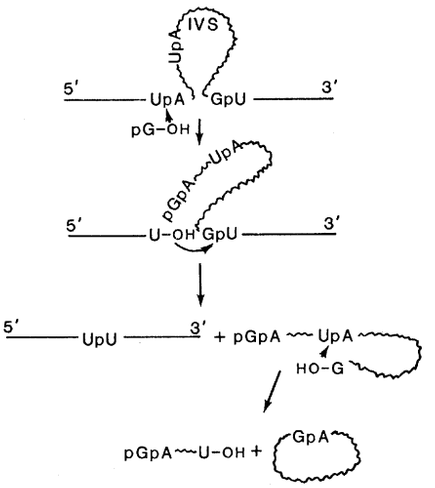
It turned out that there was no artifact. "We were able to see that covalent bonds were being broken and re-formed and that a guanosine cofactor was covalently added in a phosphodiester bond to the end of the intervening sequence. This isn't going to happen spontaneously. There is a real chemical change going on in the absence of protein enzymes."

The overall result of the reaction is the precise excision of the intervening sequence followed by the ligation of the 3′ and 5′ ends of the exon, and the subsequent cyclization of the intron. The details of the mechanism are intriguing, especially as it involves the repetition of a single chemical reaction, a *trans* esterification that results in cleavage and subsequent ligation, which acts first on the preribosomal RNA and then on the excised intron. In each case, the cleavage and cyclization events appear to proceed as concerted reactions.

The cleavage-ligation assault on both the preribosomal RNA and on the excised intervening sequence is initiated by a nucleophilic attack on a UA† link. In the first case, however, the nucleophile is a guanosine cofactor which becomes covalently linked to the 5′ end of the intervening sequence. In the second case a guanine residue at the 3′ end of the intron makes the attack. The outcome of this second reaction is the cyclization of the intervening sequence with the loss of a 15-nucleotide fragment from its 5′ end. Although somewhat disguised, both reactions are RNA recombinations.

The transesterification reactions involve essentially no change in free energy, partly because the reaction products contain the same number of ester bonds as the reactants. Nevertheless, an activation barrier must be overcome, and this is mediated by the RNA molecule.

Cech and his team have been able to establish that, for the first reaction, the RNA provides a specific binding site for




**Self-splicing RNA**

*A guanosine cofactor initiates the sequence of reactions by nucleophilic attack at a UpA link at the 5′ end of the intervening sequence. Excision of the IVS and ligation of the exons proceed as a concerted reaction. A second nucleophilic attack by the 3′ end of the IVS on a UpA link near the 5′ end initiates a second cleavage-ligation reaction, this time producing a small circular RNA and a 15-nucleotide fragment.*

the guanosine cofactor. The amino and keto groups of the guanine appear to form hydrogen bonds with yet-to-be-identified targets in the binding site. A good candidate for the binding target is of course a cytosine residue, and this possibility can be investigated by means of photoaffinity labeling.

Cech guesses that the guanosine nucleoside is bound in a very specific orientation with respect to the UA linkage. "The A would be about equal distance to the U as it is to the G," he comments. "That alone is not sufficient to make the link flip over. You need to overcome the activation energy." One way this could happen is if the UA linkage is stretched because of the RNA's conformational arrangement. "Tension on the oxygen-phosphate bond could weaken it so that it breaks. The bond might then re-form with the U or with the G." A high concentration of guanosine cofactor would drive the reaction to GA formation. "Another possibility is that the RNA withdraws electrons from the phosphate residue, thus initiating nucleophilic attack."

Although it is yet to be established, it seems possible that the cyclization reaction occurs at the same guanine active site as the one where the initial excision took place. The Colorado team has shown that the binding site for cyclization is within the intervening sequence but has not yet been able to exclude the possibility that the active site for excision might involve exon sequences.

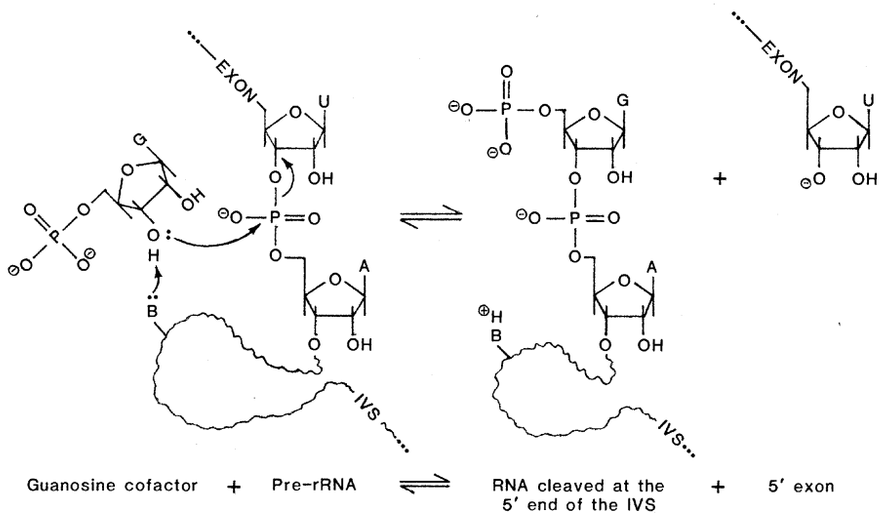
If the two sites were one and the same—that is, within the intervening sequence—there would have to be some means by which the cyclization step always comes second. Cech points out that the 15-nucleotide fragment can fold back on itself to form a marginally stable five base-pair mini-helix. "Maybe it's a dynamic structure," speculates Cech. "Maybe it is intact during the first step and then melts out exposing the second UA for the cyclization step." Tenuous but tantalizing support for the importance of the helix comes from the structure of the intron from a related organism, *Tetrahymena pigmentosa*, whose sequence was determined by Gall and Nancy Kan. There are two base changes in the 15-nucleotide fragment, one A to U and the other U to A: the integrity of the mini-helix is thus maintained. The role of this helix is speculative but testable, notes Cech. By making single base changes in model molecules it will be possible to investigate the effect on splicing efficiency.

Although the reactions described here proceed in the absence of protein and are apparently catalyzed by the RNA itself, they are rather slow in vitro. The half-life of the precursor molecule to the mature RNA plus linear intervening sequence stage is about 2 minutes. This compares with just 2 seconds in the living cell. Cech attributes the difference not to traditional enzymic activity in vivo but to the probable interaction between the RNA and structural proteins. "My intuition is that the real molecule that is doing this reaction in the cell is a ribonucleoprotein," he says. "Nascent RNA is bound up with ribosomal proteins, so you wouldn't expect RNA in the cell to have the same conformation as when it is naked in vitro. I think this will account for the difference in reaction rate." In any case, true enzymic acceleration of the reaction would normally be of the order of $10^5$ not $10^2$.

When a *Tetrahymena thermophila* organism is manufacturing ribosomal RNA it inevitably makes a vast amount of small cyclized intervening sequence as a by-product. The organism apparently has an effective disposal system because the RNA circles remain at a low steady state concentration. Cech suspects the circles have no particular function. About half the *Tetrahymena* species lack intervening sequences in their ribosomal RNA and therefore do not produce them.

RNA circles have been suggested to be potentially interesting molecules. They have been implicated as regulatory elements, and there is evidence that cyclized introns in yeast mitochondria might carry coding sequences. From a

†U, uracil; A, adenine; G, guanine; C, cytosine.




***Possible mechanism for initiating self-splicing RNA***

*RNA is in general stable in the presence of free nucleotides. The reaction scheme shown therefore does not normally occur. However, if an "active site" in the RNA were to bind the guanosine factor in a special orientation relative to the splice junction, and if other groups on the RNA (such as the B, a basic group) serve to aid the flow of electrons, the reaction will take place.*

mechanistic viewpoint it would be interesting to know if these circles are formed by a cleavage-ligation reaction. Is a short polynucleotide also formed with them? And is there a UA link or some equivalent being attacked by the end of the intron, which is a G in all these systems?

It is interesting that intervening sequences of both messenger RNA and ribosomal RNA all end with a G. This might be a coincidence or it might be important. The splicing mechanism worked out for *Tetrahymena thermophila* might not be just a special case. Except for the G, the splice sites in messenger and ribosomal RNA introns are, however, different. It remains possible that the consensus sequence for the splice sites in messenger RNA has to do with recognition of the sites, not with the mechanism of splicing.

The chance of there being a common splice mechanism within all ribosomal RNA's is, however, greater. The existence of a U just before the introns and a G at the end is noteworthy. And so too is a 16-nucleotide consensus sequence in nuclear and mitochondrial ribosomal RNA introns and in mitochondrial messenger RNA introns, a discovery made recently by John Burke and Uttam Rajbhandary at the Massachusetts Institute of Technology. "In our mechanism the G and the U play an important role," says Cech. "Both undergo nucleophilic attack and appear to be conserved. Maybe the conserved sequence of 16 nucleotides has something to do with a guanine binding site. This sequence would be an interesting target for site-specific mutagenesis."

Like all chemical reactions, those involved in splicing the ribosomal RNA in *Tetrahymena* are potentially reversible: the excised intervening sequence might be able to reinsert into an RNA molecule. This, in other words, would represent an RNA transposon. "Such a mechanism would be of interest in the origin of introns," remarks Cech. "Yes, this is an interesting idea," comments Phillip Sharp of the Massachusetts Institute of Technology.

Altman notes the similarity in size between the *Tetrahymena* ribosomal RNA intron and plant viroids, infectious RNA circles. "Prior to Cech's discovery you would say that viroids must interact with proteins in order to be infective," says Altman. "But now we can see it is possible that viroid RNA might do its own catalysis." Cech believes that the excised intervening sequence from *Tetrahymena* ribosomal RNA is a potentially dangerous species because of its reactivity, but that in this case cyclization renders it innocuous. But with the notion of catalytic RNA now demonstrated, the intellectual links between ancient splicing systems, RNA transposons and introns, and infectious RNA circles are easy to make.

Cech and his colleagues outline the enzyme-like properties of their ribosomal RNA in the following way: "(1) It lowers the activation energy for specific bond cleavage and formation events. (2) Its activity is dependent upon a precise structure. . . . (3) It has a specific binding site for the guanosine cofactor. (4) Two or more domains of the RNA form an active site or sites for the phosphoester transfer reactions." However, the RNA lacks one classic property of a catalyst: the catalytic activity is apparently lost after the chemical reaction. This is enough, some argue, to debar attaching the label "enzyme" to the RNA. Others contend that the molecule's ability to lower the activation energy for a specific reaction is sufficient for it to attain the status of an enzyme. In any case, the Colorado team has tentatively coined the term ribozyme, "a ribonucleic acid molecule that has the intrinsic ability to break and form covalent bonds."

It is a reasonable assumption that the splicing in *Tetrahymena* ribosomal RNA represents an ancient reaction mechanism, a vestige of a once common state of biochemical catalysis. The discovery has inevitably stirred tremendous excitement among those interested in the origin of life, particularly among those who favor RNA over DNA as the primordial genetic material. "If RNA can both store information and carry out catalytic reactions it is obviously a nicer molecule for the earliest genetic material," says Noller. "We are only now beginning to understand what RNA can do. When enzymes were discovered to be made up of chains of amino acids it was not obvious from the structure how they could catalyze the complex chemical reactions they so clearly did. We are in a similar position now with respect to RNA."

The catalytic potential of RNA immediately offers a way around the perennial chicken and egg problem of early evolution: proteins are required for making nucleic acids, and vice versa; worse, proteins are required for protein manufacture. "RNA recombination could have preceded protein synthesis," says Sharp. "Primitive ribosomes could have worked without proteins," adds Noller. "You don't need proteins to do protein translation."

Cech acknowledges the possibility that as RNA can make and break covalent bonds, it might indeed have the potential for self-replication. "But," he cautions, "it's a big jump to go from the *Tetrahymena* system in which there is an intramolecular recombination event to one in which you have an RNA molecule marching down another RNA making a copy of it." Despite the present gulf between observation and speculation, there are those who are prepared to make that jump. "Cech's discovery is raising our awareness of what RNA can do," says Noller. "You can be sure there is a great deal more to be discovered."—**ROGER LEWIN**