RNA CATALYSIS AND THE ORIGIN OF LIFE

NORMAN R. PACE and TERRY L. MARSH
Department of Biology, Indiana University, Bloomington, IN 47405, U.S.A.

(Received 26 August, 1985)

Abstract. Until the discovery of catalytic RNAs, first the self-splicing intron in *Tetrahymena* and then the bacterial RNAse P, cellular enzymes had always seemed to be protein in nature. The recognition that RNA can catalytically make and break phosphodiester bonds simplifies some of the assumptions required of a rudimentary self-replicating entity. Available information on the chemistry of RNA-catalyzed reactions is reviewed, with particular attention to self-splicing introns and tRNA processing by RNase P. An explicit model for a self-replicating RNA is described. The model postulates a nucleotide binding/polymerization site in the RNA, and takes advantage of intrinsic fluidity in RNA higher order structure to dissociate parent and progeny complementary strands.

1. Introduction

Discoveries in nucleic acid biochemistry in recent years have fundamentally changed the way we may think of the origins of life on the earth. These discoveries have come on two fronts. One is the finding that RNA can catalyze the making and breaking of phosphodiester bonds, the linkages forming the nucleic acids. The second is less a momentous discovery than a realization, still growing, of the architectural intricacy of polynucleotides and the potential fluidity of their structures.

The discovery of catalysis by RNA was completely unexpected. It had long been presumed that only proteins could provide catalytic functions, although there was no reason for this presumption beyond the time-honored observation that enzyme activities, when purified, had always proven to be proteins. The recognition that RNA can manipulate phosphodiester bonds changes our outlook on the origins of life because it simplifies the assumptions required of a rudimentary, self-replicating entity. Prior to this, it had to be presumed that the first self-replicating entity required more or less the same information transfer process as modern cells: DNA → RNA → protein.

It was imagined that the genetic information – the 'genotype' – resided in nucleic acid, DNA, or RNA. However, catalytic replication of the genotype would require conversion of that information into 'phenotype' – something capable of carrying out the catalysis; by implication, this had to be a protein. One could simplify the process to some extent by presuming that RNA arose first, thus eliminating the transcription step, but there seemed no straightforward way to escape the requirement for a rudimentary translation apparatus, an ensemble of primitive ribosomes and tRNAs (for discussion see Woese, 1967; Orgel, 1968). Although one might invoke a translation apparatus that arose independently of the genotype in the prebiotic environment (Crick, 1968), the complexity of such assumption is unsavory. Moreover, no selective advantage would be conveyed upon the genotype. With the recognition of RNA catalysis, however,

Origins of Life **16** (1985) 97-116.

matters are conceptually simplified because we now know that the genotype can be the phenotype. The requirement for the primitive translation apparatus, to provide replication function, is postponed. The genotype, at least in principle, can replicate itself.

At least two examples of RNA-mediated catalysis have come to light. The first discovered was the 'self-splicing' reaction which occurs during the maturation of ribosomal RNA (rRNA) in *Tetrahymena*. The second was RNase P, which acts during the maturation of tRNA. Neither of these reactions is yet understood thoroughly in terms of mechanistic detail. They differ in the nature of the reaction course, but may prove similar in essence.

2. A 'Self-Splicing' RNA from *Tetrahymena*

Most RNA molecules are not functional immediately following transcription; rather, they require a variety of post-transcriptional 'processing' modifications (for review see Abelson, 1979). These modifications include the removal of precursor-specific sequences from the newly formed transcripts, as well as chemical alterations (e.g., methylation) of base and sugar components. Precursor-specific segments, mostly of unknown function, usually occur at the ends of the RNAs and must be removed by highly specific nucleases. Sometimes, however, the precursor-specific segments occur within the functional sequence. These internal precursor segments are termed 'intervening sequences' (IVS), or 'introns'; the adjacent, functional segments are 'exons' (expressed sequences). Maturation of an intron-containing RNA requires that the precursor chain be cleaved on both sides of the intron, and the exon boundaries then 'spliced' together to restore the continuous RNA chain.

The gene defining the 26S ribosomal RNA of *Tetrahymena thermophila* contains an intron, 413 nucleotides in length. It must be removed during the processing of the rRNA since the intron transcript is present in the precursor, but not in the mature, rRNA (Cech and Rio, 1979).

Cech and his colleagues, who had worked out the processing pathway for the 26S rRNA in *T. thermophila*, were interested in purifying the enzyme(s) responsible for the intron excision and subsequent ligation of the mature sequences. They had shown that isolated nuclei would carry out *in vitro* transcription, incorporating ribonucleoside triphosphates into the 26S rRNA precursor containing the IVS. Further incubation of the nuclei resulted in splicing of the rRNA and excision of an RNA fragment which, by sequencing, could be identified with the intron from the precursor rRNA (Zaug and Cech, 1980). Cech and co-workers then purified, from isolated nuclei, labeled precursur rRNA for use as an assay substrate for putative enzymes. The purified pre-rRNA indeed underwent intron excision when added back to nuclear extracts. More dramatically, however, intron excision also occurred in controls with no added extracts (Cech *et al.*, 1981). The only requirements for the reaction were appropriate mono- and divalent cations, and guanosine or a guanosine-containing 5′-nucleotide (GMP, GDP, or GTP).

To prove rigorously that intron self-excision is an intrinsic property of the RNA, Kruger *et al.* (1982) cloned the segment of rDNA containing the intron into an *E. coli* plasmid, then an *in vitro* transcription product was generated, using *E. coli* RNA polymerase. Upon deproteinization and incubation under the appropriate conditions, the *in vitro* transcript underwent the same IVS excision as seen with the precursor rRNA isolated from *Tetrahymena* nuclei. Since no *Tetrahymena* proteins were included in the transcription or processing reactions, the intron excision could not have been due to some sort of maturation protein associated with the pre-rRNA. Cech and his colleagues termed the pre-rRNA complex a 'ribozyme', to distinguish it from conventional protein enzymes. Subsequent work established that the guanosine triggers an intramolecular rearrangement which results in the excision of the intron and ligation of the mature domains of the rRNA (for review see Cech, 1985). The overall course of the reaction is outlined in Figure 1.

As seen in the figure, the guanosine co-substrate initiates cleavage of the precursor RNA chain at the intron-exon boundary by adding to the 5′ end of the intron. The requirement for guanosine or its nucleotides is quite specific; none of the other, usual

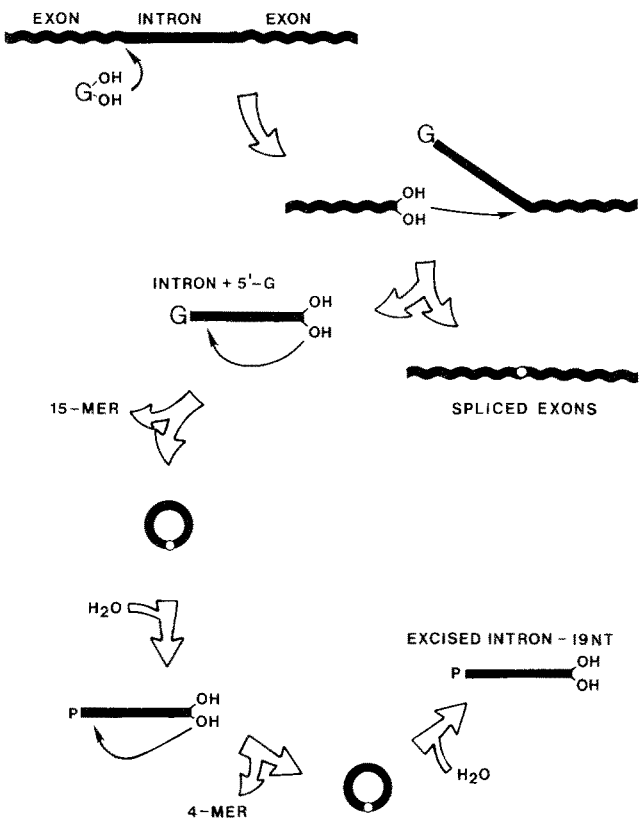


Fig. 1. The reaction course of the *Tetrahymena* self-splicing rRNA intron. See text for discussion.

nucleosides or nucleotides containing adenine, uracil, or cytosine will promote the reaction. Next, the newly exposed exon end is joined to the other exon, with concomitant excision of the intron. Overall, the excision and ligation reactions of biological relevance may be viewed as a series of transesterifications. Two phosphodiester bonds are made (the G addition and the splicing) and two are broken (the intron boundaries), so there is no net change in covalent bond energy during the reaction course.

The intron retains its catalytic activity after excision, so it must be the seat of the catalytic mechanism (Zaug *et al.*, 1984). As further seen in Figure 1, the 3′ end of the excised intron attacks itself near its 5′ end, releasing a 15-nucleotide segment and forming a circular RNA, which again can open at the point of cyclization, and again close, releasing another four residues from the intron. This final, circular product remains labile at the site of cyclization and is subject to hydrolytic opening at that point. The hydrolysis may be directed by the enzymatic activity of the excised intron, since it results in a 5′-phosphorylated, linear form of the intron. Conventional chemical hydrolysis would yield the 2′- or 3′-phosphorylated products via a 2′,3′ cyclic phosphodiester intermediate. The only known RNA hydrolytic reactions with 5′-phosphorylated products are catalyzed by enzymes.

The catalytic activity of the intron has several roles which might have been thought the province of proteins only. The intron specifically binds the guanosine co-substrate which initiates the transesterifcation cascade; it arranges the two intron-exon boundaries in a geometry favorable for excision; and, it catalyzes the excision-ligation reactions and those undergone by the intron subsequent to excision. It is noteworthy that the intron carries out two types of chemical reactions: the several transesterification events which are readily imagined as occurring at the same active site, and hydrolysis (circle opening). It remains to be seen whether the transesterification and hydrolytic reactions are in essence equivalent, involving the same catalytic elements.

The specificity and catalytic properties of any enzyme, protein or RNA, lie in the complex geometry established between substrate and catalyst upon the association of the two. RNAs now are known to have intricate and precise secondary and tertiary structures. The notion that RNA can provide a catalytic binding site for a low molecular weight substrate, the guanosine, seemed novel, although specific binding sites for metals and intercalating agents had been established. Bass and Cech (1984) have characterized the nature of the co-substrate binding to some extent, analyzing the effects of co-substrate analogs on kinetic parameters. The rate of intron excision is dependent upon the concentration of the co-substrate in a manner common with protein enzymes; the Michaelis-Menten rate law is followed. At high concentrations of guanosine, the rate of intron excision is dependent upon the concentration of the precursor rRNA in a first-order fashion, so the excision must be intramolecular; excised introns do not catalyze other excisions.

The Km of the catalytic intron for guanosine is ca. 0.02 mM, reflecting a fairly high affinity. By contrast, the concentration of substrate required by the protein, RNase A, for half-maximal velocity is ca. 100-fold higher. The RNA catalysis is much slower than

the protein catalysis, however. The kcat for intron excision is only ca. 0.5 per minute, 1000-fold slower than hydrolysis by RNase A. It still is nuclear whether, in the cell, protein factors accelerate intron excision. Indeed, the rate of excision in isolated nuclei is ca. 50-fold higher than *in vitro*, with purified RNA (Kruger *et al.*, 1982). This higher excision rate in the nucleus need not be due to enzymatic action by a protein, however. Ribosomal proteins, for instance, might mold the mature domains of the pre-rRNA into a more favorable geometry for reaction. In essence, however, any involvement of auxiliary proteins is moot: the RNA is the catalyst.

The studies of guanosine analog participation in the reaction indicated that the pyrimidine moiety of the guanine base is the most important in binding. Alterations of the exocyclic O—6 or N—2, or the ring N—1 positions, diminished effectiveness in binding to the intron-containing pre-rRNA, but with little influence on maximum velocity (kcat). Any alterations of the 2′- or 3′-hydroxyl groups abolished reaction (and binding). With this information, Bass and Cech (1984) offered the schematic active site shown in Figure 2. The co-substrate is positioned, adjacent to the bond to be cleaved,

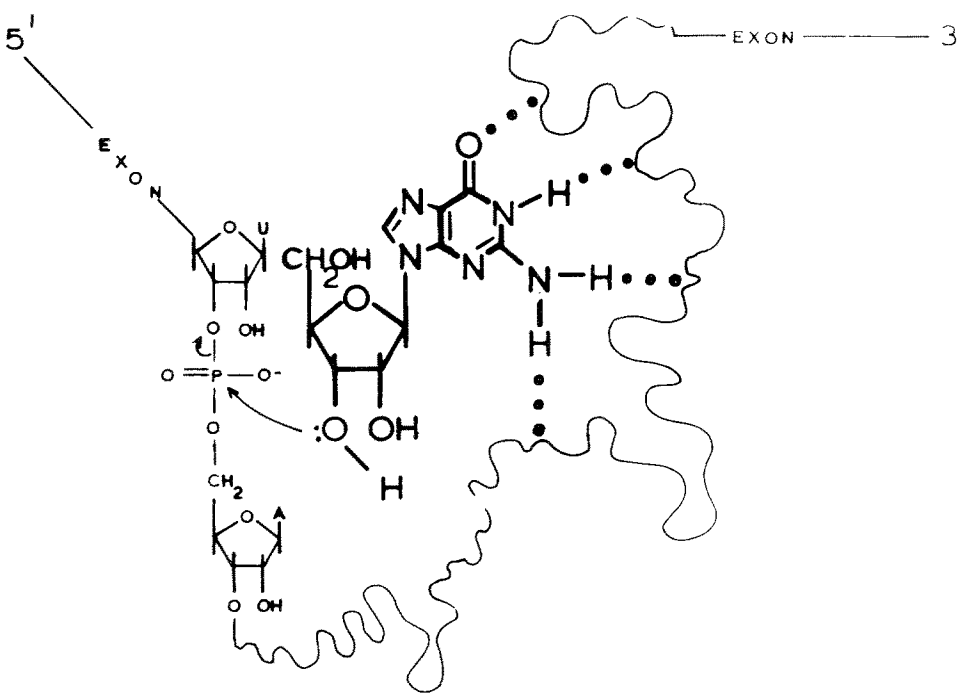


Fig. 2. A schematic view of guanosine activation by the self-splicing intron. (Reprinted from Bass and Cech, 1984 (with permission).) See text for discussion.

by specific contacts in the intron. Abstraction of the 3′ hydroxyl proton could drive the attack on the phosphodiester bond at the intron-exon boundary. If the geometry were then appropriate, the new 3′-OH could attack the next intron-exon linkage, as diagrammed in Figure 1. Most likely the reactive group is the ribose hydroxyl, not its deprotonated form, an alkoxide, a potent nucleophile. This follows from the observation that the guanosine-dependent, self-splicing reaction is largely independent of pH over its active range. A greater abundance of alkoxide ions, and hence higher reaction rate at higher pH, would be expected if free alkoxide were involved.

One obviously important aspect of the self-splicing intron must be its higher order structure, that is its secondary and tertiary foldings. Although the nucleotide sequence of the rRNA intron has been known for some time (Kan and Gall, 1982), there is no way to predict, with confidence, even the secondary structure of polynucleotides. On the basis of free energy calculations and experiments using chemicals and enzymes which probe single-strand or duplex regions of RNA, Cech and co-workers have derived a credible model for the secondary structure of the *Tetrahymena* intron (Cech *et al.*, 1983; Inoue and Cech, 1985). The proposed structure offers no clues to the catalytic mechanism, but it provides a focus for experiments seeking the catalytic site by mutational analysis. A genetic approach to the splicing mechanism is now possible since the DNA which defines the intron has been cloned into the β-galactosidase gene in a bacterial plasmid. The intron retains the capacity to excise itself from the lac mRNA, and the excision is required to restore the proper translational reading frame for β-galactosidase production. Thus, mutations in the intron which damage the self-splicing mechanism confer a lac^- phenotype (Price and Cech, 1985).

3. How Widespread is Self-Splicing?

The generality of intron self-splicing remains to be established. Certainly not all introns can catalyze their own excision. There are several examples of introns in tRNAs and in mRNAs which require auxiliary enzymatic action for their removal from the functional RNA domains. Nonetheless, further examples of self-splicing are accumulating. The *Tetrahymena* rRNA intron has certain structural features in common with several other known mitochondrial and nuclear introns (Michel and Dujon, 1983; Waring and Davies, 1984). These therefore have been dubbed 'class I' introns. One of these, an intron in the *Neurospora* mitochondrial cytochrome b gene, has been shown to undergo self-splicing *in vitro*, in a manner analogous to the *Tetrahymena* rRNA intron (Garriga and Lambowitz, 1984). In the cell, however the cytochrome b intron excision clearly requires a nuclear gene product; mutants in the nuclear gene result in the accumulation of the mRNA precursor.

Introns other than class I types also can undergo self-excision *in vitro*. This occurs in at least some cases by a different mechanism than that employed by the *Tetrahymena* rRNA intron. Peebles, Perlman, and their colleagues have shown self-excision of a 'class II' intron from a precursor of the *Saccharomyces* mitochondrial cytochrome C oxidase subunit I mRNA (Peebles *et al.*, 1985). Class II introns are categorized by short

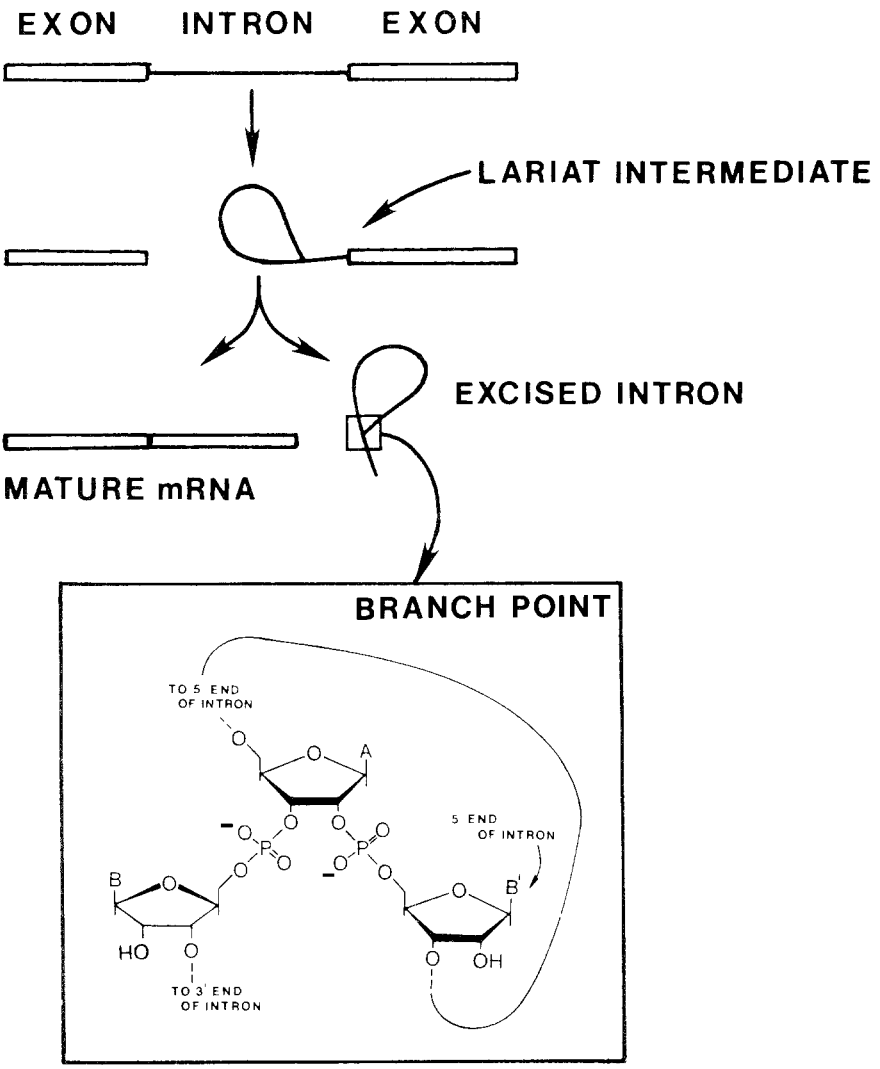


Fig. 3. Splicing by a 'lariat' intermediate. See text for discussion. The inset shows the sructure of the branch point.

sequence commonalities and are distinct from class I introns. In this case, as outlined in Figure 3, rupture of the phosphodiester band at the 5′ side of the intron is coupled with the formation of a novel, 2′,5′ linkage between the 5′ end of the intron and an intron sequence near, but not immediately at, its 3′ boundary. Subsequent condensation of the exon boundaries results in the release of a 'lariat' structure. RNA lariats first were identified among the reaction products which accumulate during mammalian precursor mRNA splicing in nuclear extracts *in vitro* (for review see Padgett *et al.*, 1985). Although these latter precursors have not yet been observed to undergo self-splicing, the potential seems clear. The capacity of some introns for self-excision *in vitro* possibly

reflects that all introns are primed by evolution to undergo facile removal, and that sometimes the facility is such that spontaneous excision may occur, albeit slowly.

4. What Do Introns Do for Cells?

Although a few introns are known to code for gene products important to their excision, the roles for most are unknown. Certainly the self-splicing intron in the *Tetrahymena thermophila* rRNA gene does nothing significant for the cell. Other, closely-related species of *Tetrahymena* do not have the intron in their 26S rRNA genes (Wild and Gall, 1979). Perhaps the *Tetrahymena* rRNA intron is a parasite, its replication apparatus the excision mechanism. If this were the case, we might expect to be able to trace the path of the excised intron back into the host genome.

Viewing the self-splicing intron as a 'selfish gene' (Doolittle and Sapienza, 1980) poses no conceptual problem, as it is becoming evident that DNA copies of RNA sequences – 'reverse transcripts' – can be inserted into the genome and thus propagated. Evidence that this has happened in the past is found in many 'pseudogenes' in eukaryotic cells. These are DNA sequences which closely resemble portions of active genes, but which precisely lack intron sequences present in the active gene, and as well lack some coding sequences, so cannot be active genes (for review see Baltimore, 1985). The best explanation for pseudogenes is that they are derived ftom functional mRNAs by reverse transcription. Since the mRNAs had undergone splicing before the synthesis of DNA copies inserted into the genome, the resulting genome copies, the pseudogenes, would precisely lack introns. Probably there are many opportunities for spurious reverse transcription of cellular RNAs. Many RNA viruses contain reverse transcriptase, and the enzyme sometimes is used during transposon migration (Boehe *et al.*, 1985).

5. RNase P - An RNA Enzyme

The *Tetrahymena* rRNA self-splicing is an intramolecular rearrangement triggered by the guanosine co-substrate. Although the intron can be forced *in vitro* to undergo intermolecular condensations (Sullivan and Cech, 1985; Zaug and Cech, 1985), it generally does not turn over: that is, affect more than one substrate molecule. The recognition of an RNA enzyme that turns over came in another RNA processing system, RNase P, which is involved in tRNA maturation.

Transfer RNAs are the most subject of any RNAs to post-transcriptional modifications, commonly requiring a dozen or more different reactions (for review see Abelson, 1979). The tRNA precursors generally contain precursor-specific segments which must be removed from both termini during maturation, and they sometimes contain introns as well. Additionally, many of the tRNA nucleotides are modified (for instance, by methylation, thiolation, etc.), and sometimes extra residues are added to the 5′ and 3′ ends following the removal of precursor segments.

The enzyme responsible for removing 5′ precursor sequences, thereby generating the

mature 5′ termini of tRNAs, is RNase P. This enzyme, discovered in *Escherichia coli* in 1969, was the first RNA processing enzyme analyzed *in vitro* (for review, see Altman *et al.*, 1980). It evidently acts on all or most tRNA precursors, which means that the enzyme is capable of handling ca. 50 different substrates, the number of different tRNAs. This is evidenced by the fact that temperature-sensitive mutants in the enzyme accumulate precursors for all or most tRNA types at the restrictive temperature. Thus, RNase P seemed exceptionally versatile, since the nucleotide sequences of the many tRNAs differ substantially. As the secondary and tertiary structures of the tRNAs were elucidated, however, it became clear that the many tRNA sequences could present the same general aspect to the enzyme because they fold into a common, 'cloverleaf' secondary structure which, in turn, adopts an L-shaped tertiary configuration. By utilizing contacts with the folded, mature domains of the tRNA precursors, RNase P could recognize the many tRNAs by their commonalities, as arranged in space by their common foldings. The precursor tRNAs present to the enzyme a macromolecular surface, not a simple, linear, nucleotide sequence. This is borne out by studies of the structures of a number of tRNA mutants with reduced susceptibility to RNase P; bases involved in the secondary and tertiary structure foldings of the tRNAs are perturbed (Mazzara and McClain, 1980).

The *E. coli* RNase P proved a problematic enzyme to work with, resisting the efforts of several groups at purifying it using the conventional methods of column chromatography, salt precipitations, etc. (Altman *et al.*, 1980; Guthrie and Atchison, 1980). The reason for the difficulties became evident when Altman and his colleagues found that the enzyme consists of two components, both required for activity. Many of the attempted purification steps separated the two components, abolishing the ability to cleave isotopically labeled tRNA precursors. Most interesting, however, was the finding that one of the components was RNA, the other protein (Stark *et al.*, 1978). This was shown by treatment of the protein-RNA complex with a non-specific RNase, followed by removal or inactivation of the nuclease, resulting in destruction of RNase P activity. Moreover, the density of the active RNase P, as assessed by CsCl buoyant density centrifugation, was about $1.7\,g\,ml^{-1}$, consistent with a ribonucleoprotein complex. Pure protein would buoy at ca. $1.2\,g\,ml^{-1}$ and pure RNA at ca. $2\,g\,ml^{-1}$ in the CsCl gradients. The ribonucleoprotein nature of RNase P proved not limited to *E. coli*, but to be widely distributed. *Bacillus subtilis*, another bacterium, possesses an RNase P with protein and RNA elements analogous to those of *E. coli* (Gardiner and Pace, 1980; Gardiner *et al.*, 1985). The general features of the characterized bacterial RNases P are summarized in Figure 4. Additionally, RNA elements have been implicated in RNase P-like activities identified in extracts from *Schizosaccharomyces pombe*, (Kline *et al.*, 1981) and human cells and mitochondria (Doersen *et al.*, 1985), to list a few sources, and there are genetic indications for the occurrence in *Saccharomyces cerevisiae* mitochondria (Miller and Martin, 1983). None of these latter enzymes has yet received detailed characterization, however.

Although *in vivo* the RNase P undoubtedly functions as a ribonucleoprotein particle, at non-physiologically high salt concentrations only the RNase P RNA is required for

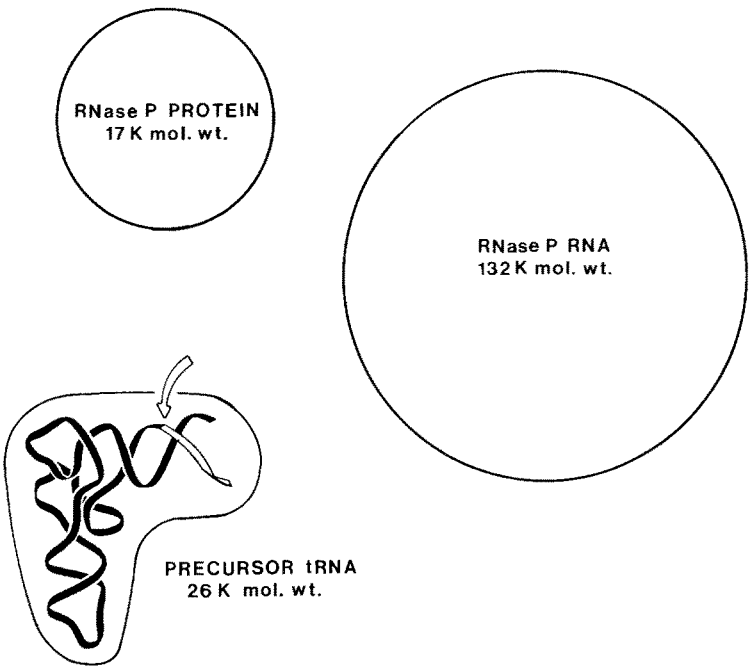


Fig. 4. RNase P holoenzyme components and substrate. The sizes of tRNA and the RNase P protein and RNA moieties are listed. The diagrams indicate the cross-sectional sizes of the molecules (correcting for partial specific volume), if globular in the case of the RNase P elements. The arrow indicates the point of cleavage in the tRNA precursor by RNase P.

accurate processing of tRNA precursors *in vitro* (Guerrier-Takada *et al.*, 1983). Proof that the RNA is the enzyme is rigorous: Guerrier-Takada and Altman (1984) have shown that an *in vitro* transcript of a cloned, RNase P RNA gene is active.

The salt concentrations which alleviate the requirement for the protein moiety of RNase P are extraordinarily high. The *B. subtilis* RNA-alone reaction requires ca. 2 M NH_4Cl and 200 mM $MgCl_2$ for maximal activity, 10-fold higher mono- and divalent cation concentrations than required by the holoenzyme. The *E. coli* reaction is more modest in its requirements, functioning at 100 mM NH_4Cl and 60 mM $MgCl_2$. The requirement for high salt concentrations is likely due in good part to a requirement for electrostatic screening. Both substrate and enzyme RNAs are highly negatively charged – one full negative charge per nucleotide phosphate. Without high cation concentrations, the resultant electrostatic repulsion between enzyme and substrate would prevent contact between the two. Thus, at least part of the role of the RNase P protein may be to provide an analogous cation screen. Indeed, the RNase P proteins of both *E. coli* and *B. subtilis* are strongly basic. The protein, if globular, is quite small relative to the RNA (Figure 4), however, so it likely makes contact with only part of the RNase P RNA.

The RNase P protein may have a role in addition to service as an electrostatic shield. In the cell, besides cleaving tRNA precursors, RNase P is responsible for the

maturation of at least one other RNA, '4.5S' RNA, a stable RNA of unknown function in *E. coli* (Bothwell *et al.*, 1976). Cleavage of the precursor of 4.5S RNA *in vitro* requires the holoenzyme; neither RNA nor protein alone are effective (Guerrier-Takada *et al.*, 1983). So, the protein may play some part in the specificity of the reaction. Moreover, the *E. coli* RNase P RNA seems incapable of acting alone on tRNA precursors which lack the ubiquitous sequence CCA at the 3′ terminus of the tRNA mature domain; the holoenzyme is required (Guerrier-Takada *et al.*, 1984). In contrast, the *B. subtilis* RNase P RNA seems to act equally effectively on precursors containing or lacking that sequence, so it is not an inherent requirement for catalysis by RNA. Rather, any difference between the specificities of the two catalytic RNAs likely reflects the genetic backgrounds of the organisms. In *E. coli*, the CCA termini of all the tRNAs are encoded in the DNA, whereas many of the *B. subtilis* tRNA genes lack the CCA (Vold, 1985). In those cases the sequence is added post-transcriptionally.

The requirement for high salt concentrations may be explained in part by a requirement for screening electrostatic repulsion between substrate and enzyme RNAs, but all results are not wholly consistent with this interpretation. For instance, if electrostatic screening were the only role for the high cation concentrations, then it would be expected that cations with smaller ionic radii, because they bind more tightly to nucleic acids than do larger ions, would stimulate the RNA catalyst at concentrations lower than those required for larger ions. In fact, the converse is true (Gardiner *et al.*, 1985). The small ions, Na^+ and Li^+, are almost ineffective in promoting the reaction as compared to the larger ones, K^+, NH_4^+, Cs^+, and Rb^+. This suggests that electrostatic shielding is not the only consideration in activating the RNA enzyme. Other factors, such as ion packing geometry or hydration phenomena, may be at work, as well.

Some data suggest that conformational transition, or at least conformational fluidity, may be important in the RNase P reaction. For instance, whereas high concentrations of NH_4Cl promote the RNA-alone reaction, $(NH_4)_2SO_4$ does not, and inclusion of high SO_4^{-2} concentrations in otherwise optimal reactions quenches activity. Relative to Cl^-, SO_4^{-2} has been shown to solidify macromolecular (protein) secondary and tertiary structures, as evidenced by an ability to stabilize them against denaturation (Von Hipple and Schleich, 1969). If conformational adjustments of either the substrate or enzyme RNAs were required during the reaction, then the inhibitory action of the SO_4^{-2} might be expected. Moreover, inclusion of mildly denaturing solvents, such as methanol, dimethyl sulfoxide, or ethylene glycol, in reactions with unsatisfactorily low salt concentrations, stimulates RNA catalysis to maximal activity, again suggesting some role for conformational mobility. The possible significance of structural fluidity to the RNase P reaction is discussed further below.

RNase P generates 5′-phosphate and 3′hydroxyl termini, as does the *Tetrahymena* self-splicing mechanism and most specific processing nucleases. The RNase P reaction course is different from that of self-splicing, however. As discussed, the self-splicing reaction is a series of transesterifications in which a ribose 3′—OH is thought to be the attacking nucleophile on the phosphodiester chain. In the case of RNase P, there is no

involvement of terminal ribose hydroxyl groups. Treatment of both substrate and enzyme RNAs with periodate, which oxidizes the ribose diol to a dialdehyde, does not diminish their reactivity (Marsh and Pace, 1985). Thus, the RNase P reaction seems hydrolytic. Consistent with this is the observation that the reaction rate increases with increasing pH, to ca. pH 9, where activity is lost because of denaturation of the RNA. It is not completely clear that the attacking nucleophile is hydroxide ion, however, since the reaction rate dependence on pH is less than first order.

Although different from the self-splicing transesterifications, RNase P and the *Tetrahymena* intron may prove to have intrinsic similarities, since the intron also undergoes hydrolytic opening of the cyclic form. Evidently this reaction involves free hydroxide ion; the reaction rate shows a first order dependence on pH over the active range (Zaug *et al.*, 1984). It is not yet known whether the intron 'active site' for the splicing mechanism also is used in circle opening or, instead, the labile bond is 'strained' in some manner. If the intron active site can manipulate both ribose hydroxyl groups and hydroxide ion, then it and RNase P could use the same general mechanism of action.

Protein hydrolases of RNA are well studied. The reactions carried out by these enzymes are commonly considered to be catalyzed by amino acid functional groups which donate or capture protons as needed to drive hydrolysis. An analogous proton exchange scheme, not entirely speculative, is shown for the RNase P reaction in Figure 5. Nucleic acid groups, of course, must be the proton sinks, but such exchange reactions are well known in the nucleic acids. Both bases and phosphodiester chain elements can serve in proton transfer. A few examples, in comparison with histidine, a common amino acid functional group, are diagrammed in Figure 6. Figure 6C certainly does not pertain to the RNase P reaction (above), but an internucleotide phosphate or a base exocyclic group, for example, could promote an equivalent alkoxide formation in an internal ribose (Figure 6D).

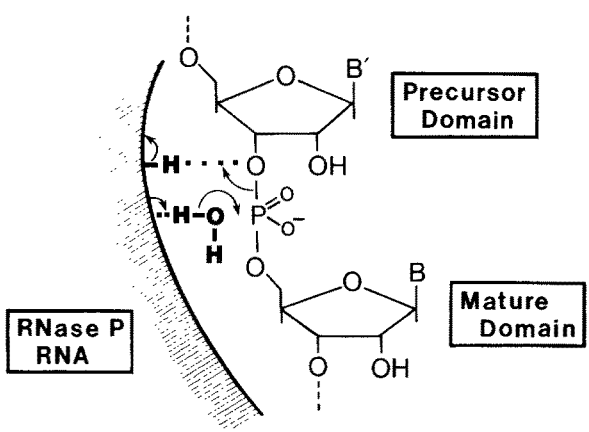
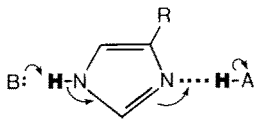


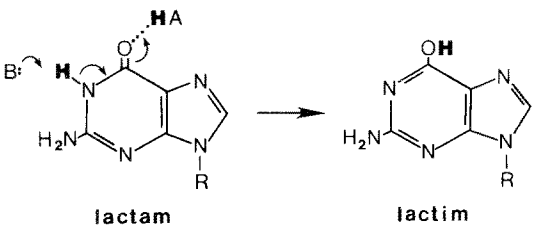
Fig. 5. A general scheme for hydrolysis of tRNA precursors by RNase P. See text for discussion.

Efforts to define the higher order structure of the RNase P RNA have not yet been fully successful. As pointed out above, we cannot yet predict, with good confidence,

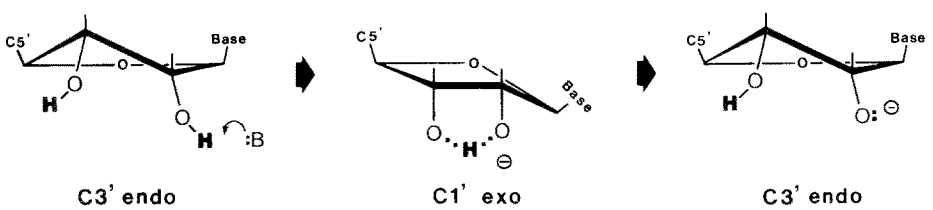
A. Histidine Proton Transfer



B. Base Tautomerization



C. Ribose As Proton Reservoir



D. Base Tautomer Activation Of Ribose 2'-OH

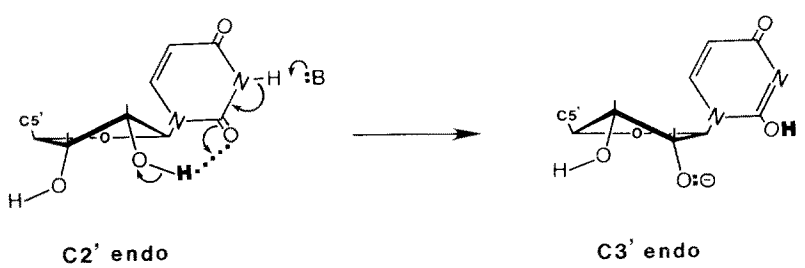


Fig. 6. Possible proton exchange mechanisms in: (A) histidine imidazole; (B) guanine; (C) 3'-terminal ribose; (D) internal ribose.

RNA secondary structure from nucleotide sequence. The best approach toward defining RNA secondary structure is a phylogenetic one, seeking canonical (Watson-Crick) base pairing schemes which are common to different – but homologous – nucleotide sequences from different organisms. If a possible base pairing scheme is not present in homologous sequences from different organisms, it likely does not exist.

Only two RNase P RNAs which are sufficiently different to be useful for structure comparisons have been sequenced, those of *E. coli* (377 nucleotides; Reed *et al.*, 1982; Sakamoto *et al.*, 1983) and of *B. subtilis* (396 nucleotides; unpublished). A third sequence, from *Salmonella typhimurium*, has been determined, but it is too close to the *E. coli* sequence to provide much phylogenetic counterpoint (Baer and Altman, 1985). The *E. coli* and *B. subtilis* RNase P RNAs certainly are homologous overall; heterologous reconstitutions, using the *E. coli* and *B. subtilis* RNAs and proteins, yield holoenzyme active in the low salt, protein-dependent reaction (Guerrier-Takada *et al.*, 1983). Yet the nucleotide sequences of the two RNAs are dramatically different. There are only a few small stretches of sequence conserved between the RNase P RNAs of the two organisms. Over most of the lengths of the RNAs the sequences are so disparate

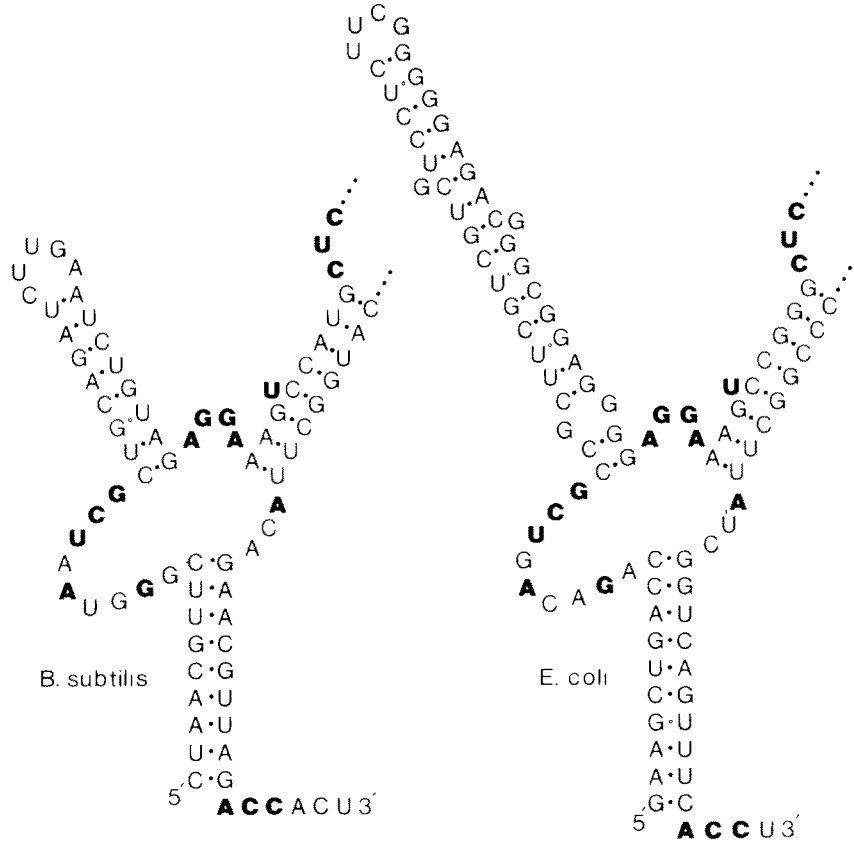


Fig. 7. Homologous secondary structures in the *B. subtilis* and *E. coli* RNase P RNAs. Conserved bases in putative single-stranded regions are presented in bold type.

that they cannot be credibly aligned. The alignment process is critical to comparative structure analysis because it defines the sequence blocks to be compared for pairing possibilities. Such disparity in the *E. coli* and *B. subtilis* RNase P RNA sequences was completely unexpected, based upon similar analyses of the ribosomal RNAs, which are more highly conserved in structure. Sequence determinations of the RNase P RNAs from other organisms will be required before a fully satisfactory secondary structure emerges.

Only ca. 30 % of the structures of the RNase P RNAs of *E. coli* and *B. subtilis* can be folded in a manner consistent with both sequences. In those ordered regions, however, the reason for the unexpected divergence of the two sequences becomes clearer. Differences in the two RNAs are due less to point mutations than to insertions and deletions of sequence blocks. An example of this is shown in Figure 7. In this region of the molecule the *E. coli* RNA has an extended helical domain; elsewhere, *B. subtilis* has segments not in the *E. coli* version. This type of structural variation in helical domains suggests that the RNase P activity of the RNA may be only one of several functions. For instance, although there is no evidence for such, the RNase P RNA might serve as the integrating element for a multi-enzyme complex for processing tRNAs, in an RNA-protein complex structurally analogous to the ribosome.

6. Why Is RNase P RNA?

The question, of course, arises: Why is the catalytic component of RNase P an RNA molecule? We know that proteins can specifically recognize RNAs without another RNA to serve as adaptor. Indeed, the other known, specific, processing nucleases (for example those involved in rRNA formation) are proteins (for review see Pace, 1984). RNase P has one unique aspect, however – it must handle many different substrates, the 50, or so, different tRNAs produced by cells. The different tRNAs have the same general form, the L-shaped tertiary structure, but they differ in their minor details, a consequence of different sequences, variable loop sizes, varying precursor-specific sequences, etc. The RNase P must accommodate all of these, possibly by molding – 'induced fit' – to the substrates. Perhaps RNase P activity is embodied in RNA because of a requirement for conformational fluidity, which may be important in the reaction (above). Proteins offer a greater wealth of chemically functional groups than do polynucleotides, but RNA structure may be intrinsically more mobile than that of proteins.

7. Are There Other RNA Enzymes?

Although the intron self-splicing reactions and RNase P are the first well-confirmed examples of RNA enzymes, others undoubtedly will emerge. Several putative, RNA-containing enzymes have been reported, the evidence consisting of RNA content in enzyme preparations and/or sensitivity of enzyme activity to RNase treatment. Some examples are the *E. coli* photoreactivating enzyme (Snapka and Sutherland, 1980), a

potato o-diphenol oxidase (Balasingam and Ferdinand, 1970), and a δ-aminolevulinic acid synthetase from *Chlamydomonas* (Huang *et al.*, 1984). However, all of these findings remain to be confirmed and extended, and in no case is RNA invoked as the active element. Another possible example of RNA self-cleavage has been reported by Apirion and his colleagues (Watson *et al.*, 1984). They find that incubation of a precursor of 'species I' RNA, a phage T4 RNA from infected *E. coli*, under the appropriate conditions, results in fragmentation at two specific sites. The reaction is potentiated by non-ionic detergents (e.g., Brij 58 or Nonidet P40), does not require a divalent cation, and yields 3′-phosphate and 5′-hydroxyl groups. Further confirmation of the RNA-dependent character of this reaction is necessary, however, as it bears the characteristics of trace contamination with an RNase A-like activity. Bonds cleaved in this reaction are CpA linkages, which are particularly sensitive to RNase A, and the reaction is inhibited by low concentrations of Na dodecylsulfate, a potent denaturant of proteins. In contrast, the *Tetrahymena* self-splicing reactions and RNase P are refractory to high levels of Na dodecylsulfate. On the other hand, the species I precursor may indeed contain phosphodiester bonds which are sufficiently strained that they undergo hydroxide-dependent hydrolysis even at physiological pH. However, since the cleavage points are in a precursor segment and do not yield a mature terminus, any utility to the cell is questionable.

8. Fluidity in Nucleic Acid Structure

Our grasp of polynucleotide structure, although still rudimentary, has improved enormously in recent years. At the time our concepts of protein secondary and tertiary structure were taking form, the nucleic acids were considered essentially information tapes, chains of nucleotides with little native structure beyond the double helix. It is now abundantly clear that polynucleotide higher order structure is at least as complex as that of proteins (for review see Saenger, 1984). RNA, particularly, is capable of intricate, stable, tertiary structure. In tRNA, for instance, virtually every nucleotide is engaged by hydrogen bonding or other contact with another nucleotide. DNA seems more restricted in folding possibilities because it lacks the 2′-hydroxyl group, important in many tertiary contacts in RNA.

RNA structure can be viewed as short helices, formed by pairing of complementary sequences, which are arranged in precise topology by tertiary contacts between bases, sugars, and phosphates. The foundations of the duplex structures are the Watson-Crick complementary base pairs, A · U and G · C, but there is much evidence that non-canonical base pairs, G · U, G · A, A · C, etc., are no less important. Moreover, nucleic acids need not be base-paired to form helices; even 'single' strand' structures form surprisingly stable, helical stacks because of interactions between neighboring bases (Saenger, 1984).

The folding of an RNA molecule is not always unique. Sometimes the local structure can alternate between different pairings of complementary sequences. If two possible, mutually exclusive, complementary sequence pairings do not differ in their energetics,

the conformational transition between the two states probably is easily achieved. This is because complementary base pairs 'breathe' freely, so base-paired segments can easily be imagined to dissociate and reform in alternate pairings, if tuned by evolutionary pressure to do so. Such formation of alternate, complementary sequence pairings is thought to function in the cell, although data regarding this are sparse. One well-known example is the control of transcription by attenuation (for review see Crawford and Stauffer, 1980). In this case the occupancy of particular nascent mRNA sequences by a ribosome prevents or permits the formation of an RNA hairpin important for termination of transcription; thus, the RNA polymerase can sense active translation and terminate or complete the mRNA as needed by the cell.

Another biological mechanism which depends upon conformational fluidity of RNA secondary structure operates during the replication of bacteriophage single-strand RNA genomes. As a template RNA strand is copied into a complement by the RNA replicase, the new strand is displaced from base pairing interactions with the template by formation of local, intramolecular pairings. The conformational reordering prevents formation of an extended RNA–RNA duplex, known to be nonproductive during replication (Kramer and Mills, 1981).

The functional importance of conformational transitions in nucleic acids is only now being realized, but likely they will prove to be the key to many cellular events.

9. A Self-Replicating Entity

We do not yet know the structural details of any RNA catalytic site. However, the self-splicing intron and RNase P both testify that RNA can bind other RNAs and nucleotides, and break or make phosphodiester bonds. It is well established that appropriately 'activated' nucleotides, for instance the nucleotide 5′-imidazolides of Orgel and his colleagues, can spontaneously polymerize in a complementary, template-dependent fashion (Inoue and Orgel, 1983; Schwarz and Orgel, 1985). It is an easy leap of imagination that analogous condensations of prebiotic, activated nucleotides could be driven by an RNA catalytic site. Coupled with the intrinsic conformational fluidity of polynucleotide foldings, this mechanism could propagate a self-replicating entity. One scenario is diagrammed in Figure 8.

Consider the structure in Figure 8A to be a primitive, self-replicating molecule. It is an RNA chain, with an indicated catalytic site capable of binding nucleotides and catalyzing their addition to a growing chain on a complementary template. Near the other end of the RNA strand is a complement of the active site. The catalytic site (and its complement) need not be – indeed probably would not be – a specific nucleotide sequence span. Rather, as with modern proteins, the catalytic property likely would reside in the three dimensional juxtaposition of particular residues. The nucleotide sequence between the active site and its complement is such that no extremely stable duplex hairpins form; rather, many alternative structures are possible. As these alternate structures form and dissociate, plausibly in sequential fashion, the catalytic site would migrate along its own length (Figure 8B). In the presence of activated

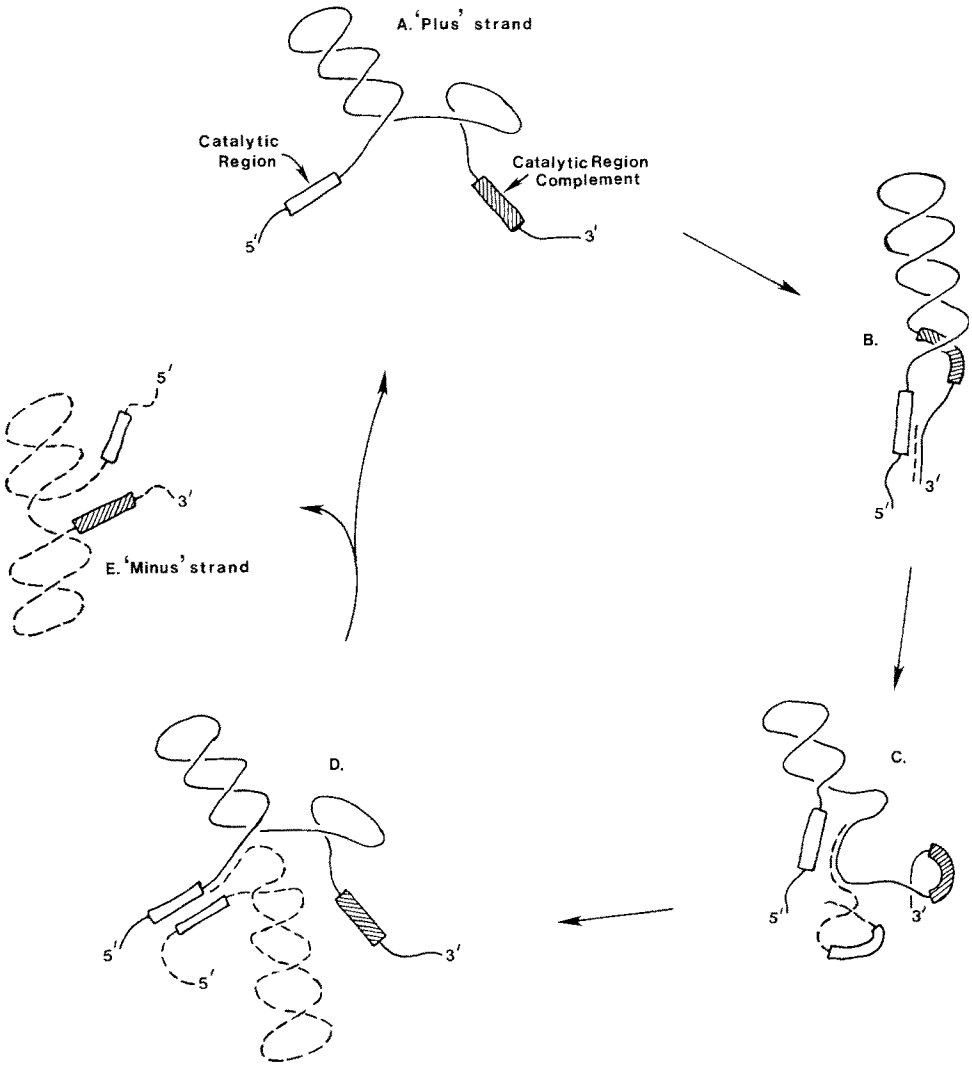


Fig. 8. An autocatalytic replication cycle. See text for discussion.

nucleotides, the active site could participate in the complementary alignment of the nucleotides on itself and catalyze their joining to form a new, complementary strand. As the new strand grows, it could displace itself from the template strand by the same mechanism that operates during phage RNA replication (above), by the formation of intramolecular hairpins (Figure 8C). Indeed, part of the selective pressure operating on the self-replicating entity would be the maintainance of intramolecular structure capable of displacing the complements from one another.

At some point the catalytic site must itself be copied (Figure 8D), a process which would disrupt the activity by the transient deposition of a complementary strand. However, if the polymerization process were to occur from one end of the template to

the other – presumably necessary to generate an intact, full-length product with reasonable frequency – then by the time the active site was copied, a new one would be available for action (Figure 8D). That new catalytic site would derive from its complement in the original template. Upon completion of the product strand, complete dissociation from the template would occur, again as during phage RNA replication. Both parent and progeny (Figure 8E) strands, although complementary, would contain catalytic sites (and their complements) and so be capable of propagation.

The scheme shown in Figure 8 is formally analogous to the 'tidal pool' scenario for nucleic acid replication, in which activated nucleotide monomers align and are spontaneously polymerized on a complementary template, the complements dissociating upon decrease in ionic strength or increase in temperature. The scheme goes beyond the tidal pool scenario, however. The randomness of nucleotide positionings and polymerization on the complementary template are replaced with directed placement and condensation by the catalytic site, and a discrete dissociation mechanism is considered. Thus, the imagined entity acquires *phenotype* – catalysis and structural fluidity – upon which selective pressure may bear.

Acknowledgement

The authors' studies are supported by NIH grant GM34527.

References

Abelson, J.: 1979, *Ann. Rev. Biochem.* **48**, 1035.
Altman, J., Bowman, E. J., Garber, R. L., Kole, R., Koski, R. A., and Stark, B. C.: 1980, in D. Soll, J. N. Abelson and P. R. Schimmel (eds.), *Transfer RNA: Biological Aspects*, Cold Spring Harbor Laboratory, pp. 71–97.
Baer, M. and Altman, S.: 1985, *Science* **228**, 999.
Balasingham, K. and Ferdinand, W.: 1970, *Biochem. J.* **118**, 15.
Baltimore, D.: 1985, *Cell* **40**, 481.
Bass, B. L. and Cech, T. R.: 1984, *Nature* **308**, 820.
Boeke, J. D., Garfinkle, D. J., Styles, C. A., and Fink, G. R.: 1985, *Cell* **40**, 491.
Bothwell, A. L. M., Garber, R. L., and Altman, S.: 1976, *J. Biol. Chem.* **251**, 7709.
Cech, T. R.: 1985, *Int. Rev. Cytol.* **93**, 3.
Cech, T. R. and Rio, D. C.: 1979, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5051.
Cech, T. R., Zaug, A. J., and Grabowski, P. J.: 1981, *Cell* **27**, 487.
Cech, T. R., Tanner, N. K., Tinoco, I., Weir, B. R., Zuker, M., and Perlman, P. S.: 1983, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3903.
Crawford, I. P. and Stauffer, G. V.: 1980, *Ann. Rev. Biochem.* **49**, 163.
Crick, F. H. C.: 1968, *J. Mol. Biol.* **38**, 367.
Doersen, C.-T., Guerrier-Takada, C., Altman, S., and Attardi, G.: 1985, *J. Biol. Chem.* **260**, 5942.
Doolittle, W. F. and Sapienza, C.: 1980, *Nature* **284**, 601.
Gardiner, K. and Pace, N. R.: 1980, *J. Biol. Chem.* **255**, 7507.
Gardiner, K. J., Marsh, T. L., and Pace, N. R.: 1985, *J. Biol. Chem.* **260**, 5415.
Garriga, G. and Lambowitz, A. M.: 1984, *Cell* **39**, 631.
Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S.: 1983, *Cell* **35**, 849.
Guerrier-Takada, C. and Altman, S.: 1984, *Science* **223**, 285.
Guerier-Takada, C., McClain, W., and Altman, S.: 1984, *Cell* **38**, 219.

Guthrie, and Atchison, R.: 1980, in D. Soll, J. N. Abelson and P. R. Schimmel (eds.), *Transfer RNA: Biological aspects*, Cold Spring Harbor Laboratory, pp. 83–97.
Huang, D.-D., Wang, W.-Y., Gough, S. P., and Kannangara, C. G.: 1984, *Science* **225**, 1482.
Inoue, T. and Cech, T. R.: 1985, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 648.
Inoue, T. and Orgel, L. E.: 1983, *Science* **219**, 859.
Kan., M. C. and Gall, J. G.: 1982, *Nucl. Acids Res.* **10**, 2809.
Kline, L., Nishikawa, S., and Soll, D.: 1981, *J. Biol. Chem.* **256**, 5058.
Kramer, F. R. and Mills, D. R.: 1981, *Nucl. Acids Res.* **9**, 5109.
Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R.: 1982, *Cell* **31**, 147.
Marsh, T. L. and Pace, N. R.: 1985, *Science* **229**, 79.
Mazzara, G. P. and McClain, W. H.: 1980, in D. Soll, J. N. Abelson and P. R. Schimmel (eds.), *Transfer RNA: Biological Aspects*, Cold Spring Harbor Laboratory, pp. 3–27.
Michel, F. and Dujon, B.: 1983, *EMBO J.* **2**, 33.
Miller, D. L. and Martin, N. C.: 1983, *Cell* **34**, 911.
Orgel, L. E.: 1968, *J. Mol. Biol.* **38**, 381.
Pace, N. R.: 1984, in D. Apirion (ed.), *Processing of RNA*, CRC Press, pp. 2–34.
Padgett, R. A., Grabowski, P. J., Konarska, M. M., and Sharp, P. A.: 1985, *Trends Biochem. Sci.* **10**, 154.
Peebles, C. L., Mecklenburg, K. L., Perlman, P. S., Tabor, J. H., and Cheng, H. L.: 1985, *Cell*, in press.
Price, J. V. and Cech, T. R.: 1985, *Science* **228**, 719.
Reed, R. E., Baer, M. F., Guerrier-Takada, C., Donis-Keller, H., and Altman, S.: 1982, *Cell* **30**, 627.
Saenger, W.: 1984, *Principles of Nucleic Acid Structure*, Springer Verlag, New York, 556 pp.
Sakamoto, K., Kimura, N., Nagawa, F., and Shimura, Y.: 1983, *Nucl. Acids Res.* **11**, 8237.
Schwarz, A. W. and Orgel, L. E.: 1985, *Science* **228**, 585.
Snapka, R. M. and Sutherland, B. M.: 1980, *Biochemistry* **19**, 4201.
Stark, B. C., Kole, R., Bowman, E. J., and Altman, S.: 1978, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3717.
Sullivan, F. X. and Cech, T. R.: 1985, *Cell* **42**, 639.
Vold, B. S.: 1985, *Microbiol. Rev.* **49**, 71.
Von Hipple, P. and Schleich, T.: 1969, *Accts. Chem. Res.* **2**, 257.
Waring, R. B. and Davies, R. W.: 1984, *Gene* **28**, 277.
Watson, N., Gurevitz, M., Ford, J., and Apirion, D.: 1984, *J. Mol. Biol.* **172**, 30.
Wild, M. A. and Gall, J. G.: 1979, *Cell* **16**, 565.
Woese, C. R.: 1967, *The Genetic Code*, Harper and Row, New York, 200 pp.
Zaug, A. J. and Cech, T. R.: 1980, *Cell* **19**, 331.
Zaug, A. J. and Cech, T. R.: 1985, *Science* **229**, 1060.
Zaug, A. J., Kent, J. R., and Cech, T. R.: 1984, *Science* **224**, 574.
Zaug, A. J., Kent, J. R., and Cech, T. R.: 1985, *Biochemistry* **24**, 6211.