9

Introns and the RNA World

Walter Gilbert and Sandro J. de Souza
Department of Molecular and Cellular Biology
The Biological Laboratories
Harvard University
Cambridge, Massachusetts 02138

The RNA World is a hypothesis about the origin of life based on the view that the most critical event is the emergence of a self-replicating molecule, a molecule that can both copy itself and mutate and, hence, evolve to more efficient copying (Gilbert 1986). Evolution works on variation and selection, and selection is always measured in terms of more efficient multiplication, the ability to make more of the entity in question. The concept of an RNA World is a way of answering the basic problem of what was the molecular biology involved at the beginning of life. Our understanding of the molecular basis of biology today is in terms of a genetic material, commonly DNA, translated through an apparatus involving RNA and the mechanism of protein synthesis to specify the positions of 20 amino acids in protein enzymes. That picture of life, in which the genetic material is of one chemical kind, DNA, made up of four bases, a second chemical, RNA, is used for structural and transfer purposes, and the enzymatic activities in the cell are a third chemical kind made up of 20 ingredient amino acids, creates a complex paradox in trying to formulate how life could have begun. This paradox was resolved by two realizations. One was that RNA is likely to be more primary than DNA, but the picture of an RNA–protein world, in which RNA is the genetic material specifying the positions of amino acids in proteins, still left one with a complex problem of beginnings. The second realization, however, was that there was no intrinsic reason that enzymatic activity must be limited to proteins. The discovery of the first two RNA enzymes showed that RNA molecules could carry out the phosphodiester bond transfers needed for RNA synthesis (Kruger et al. 1982; Guerrier-Takada et al. 1983).

Why RNA rather than DNA? The current biochemistry of these molecules suggests that RNA was antecedent to DNA. First, the synthesis of the deoxynucleotides is not primary, but secondary, to the synthe-

sis of the ribonucleotides. The biochemical processes in all cells today create ribonucleotide precursors, and then at the ribonucleotide diphosphate level, convert the sugars into the deoxy form using ribonucleotide reductases. This produces three of the deoxyribonucleotide precursors directly, but the fourth is produced as deoxyUDP and only later is the uracil methylated to produce the thymine of DNA. Second, the mechanism of DNA synthesis is completely dependent on previous RNA synthesis. The synthesis of the lagging strand of DNA, made in short pieces which are then connected, has each such piece initiated by an RNA primer which is then elongated into DNA and finally removed before the DNA strands are ligated. In general, DNA-copying enzymes cannot initiate new strands de novo but must elongate some preexisting primer, usually RNA. RNA-synthesizing enzymes, on the other hand, can initiate with a ribonucleotide triphosphate. Furthermore, the ends of linear DNA chromosomes are constructed by a telomerase function, which uses an RNA template to extend the 3′ end of the DNA chain. One last observation along these lines is that the RNA-synthesizing enzymes seem to be more primitive than that for DNA, in that they are less efficient and less rapid: The rate of RNA synthesis is about 50 bases per second, whereas DNA synthesis runs ten times faster, about 500–1,000 bases per second. For all of these reasons, RNA appears to be a biochemically primitive molecule that could have served as a precursor to a later DNA involvement. Thus, one would be led back from our current DNA–RNA–protein world to contemplate an RNA–protein world in which RNA would serve as the genetic material as well as the messengers translated by ribosomes into protein enzymes.

But which then came first? The classic chicken-and-egg problem, which needs a complicated protein machinery involving 20 amino acids in order to synthesize the enzymes necessary to synthesize new copies of the genetic material, which in turn dictates the structure of the protein-assembling machinery and the enzymes, would be simplified if either one or the other chemical entity served as the first structures. The suggestion that life begins with protein molecules creates a pattern of chemical reactions but provides no mechanism for genetic inheritance, since there is no form of protein–protein self-copying, replication, and mutation. The key aspect of evolution is the ability of molecules to grow; i.e., to replicate themselves, but in a form that embodies variation, mutation, and thus can provide the novel patterns on which natural selection will operate to improve the replication in a changing environment.

The ribonuclease P activity (Guerrier-Takada et al. 1983) and the self-splicing intron from *Tetrahymena* (Kruger et al. 1982) showed that RNA

molecules could cleave and join other RNA molecules; this is a sufficient enzymology to establish that an RNA molecule should be able to catalyze RNA-dependent RNA synthesis. Furthermore, one should expect that RNA would be capable of catalyzing a whole range of reactions. Broadly speaking, an enzyme is any structure that can bind to, and hence stabilize, the transition state of a chemical process. In that view of enzymology, the issue is one of the shape and complexity of a binding site rather than the chemical nature of the binding site; the issue is simply whether one can find hydrogen bonds and hydrophobic surfaces to construct binding pockets. RNA can serve these functions. In another view of enzymology, the nature of the enzyme is to bring into apposition with the substrate the cofactors and other groups that will participate in the chemical reaction. Again RNA molecules could, in principle, bring charges or metal ions to bear on the substrate.

In fact, today's view of the ability of RNA to catalyze phosphodiester bond formation and cleavage is that the RNA binds two essential magnesium ions which are used to handle the phosphate intermediates in exactly the same way as the protein RNA polymerases carry out those same processes using metal ions (Steitz and Steitz 1993; Steitz 1998).

Although protein enzymes frequently use cofactors, White (1976) pointed out that a large number of these cofactors are related to RNA moieties and look like residual pieces of RNA enzymes held in a protein framework to catalyze chemical reactions.

THE RNA WORLD

The minimal enzymology that RNA can do, phosphodiester bond cleavage and transfer, is enough enzymology to show that it would be possible for the first self-replicating molecule to be RNA, in the sense that in principle an RNA enzyme could copy other RNA molecules, including copies of itself. Although no such ribozyme has yet been created in the laboratory, Bartel's group has come extremely close by constructing a ribozyme that is capable of adding up to six bases in a template-directed fashion using RNA triphosphate precursors (Ekland and Bartel 1996). This is the key activity for an RNA-copying enzyme. The issue remains of finding a way to do the copying so that the two strands do not become inevitably hydrogen-bonded, and finding a way for the enzyme to accept an arbitrary RNA molecule. These are more mechanical problems than chemical problems; the critical demonstration is that an RNA molecule can do the synthetic chemistry. Thus, the RNA World contemplates a self-replicating RNA molecule, arising in a puddle containing all the RNA precursors,

catalyzing the formation of more molecules like itself, and in so doing, leading to mutant molecules and ever-better replication.

After this beginning, the RNA World picture considers a far more extensive use of RNA as enzymes: ribozymes that catalyze the synthesis of all the precursors needed to synthesize RNA; ribozymes to construct charge-neutralizing polyamines; and ribozymes to construct lipids. The full expression of the RNA World conceives of RNA-based organisms with RNA genetic material and RNA enzymes contained in lipid vesicles growing and multiplying. To have natural selection work to develop better RNA enzymes, one must effectively construct organisms, which can multiply and outgrow each other, in which a bounding membrane connects the mutated RNA genes with the better and more effective enzymatic ribozymes that are their products.

These views of the emergence of self-copying molecules imagine that one begins with pools filled with concentrated solutions of all the biochemical precursors. The appearance of a self-copying RNA would catalyze the formation of many molecules like itself in a pool. One could contemplate essentially pools of liquid as the first "organisms," each pool developing a selected "best" replicator; but to go much farther, one must have a way of enfolding the genetic material and the copying function in some boundary coat, some boundary membrane. This is needed if the genes and gene products are to be linked, as they must be if natural selection is to be able to identify a gene that makes a "better" product. Initially these membranes do not have to be impermeable to small molecules. They need only be able to hold together the macromolecular genes and gene products, because we have assumed a high concentration of precursors outside as well as inside these primordial cells. Only later, as one develops ribozymes to make precursors, does one need tighter membranes, with pores to let specific chemicals through and pumps to create and control gradients. Even at the beginning, however, a problem is posed by the negative charges on RNA and on lipids. Some supply of charge-neutralizing molecules, polyamines or oligo-lysines, would be necessary to permit RNA to be wrapped in lipid membranes and to aid in the wrapping (Jay and Gilbert 1987).

Although we can imagine a plethora of ribozymes to do all necessary reactions, the issue arises of how random processes could yield RNA enzymes of appreciable size. This is the classic size paradox, emerging for RNA. An RNA enzyme might be 300–600 bases long in order to function. Even for 300 bases, that is 4^{300} or 10^{180} molecules. Clearly, one could not get such a molecule by a random process. One way around this problem is to hypothesize that the enzymatic activities are carried by rather short

RNA pieces. We think a more likely possibility is that the RNA World had an intron–exon structure. The RNA genetic material consisted of RNA exons held together by self-splicing RNA introns, either group I or group II introns, both of which functions are catalyzed by RNA molecules acting as ribozymes. In this picture, the RNA genetic material would have an extended, presumably a more linear, structure adapted for copying. After the gene is copied, the introns would splice themselves out to leave a set of RNA exons tied together so that they could fold up and become a ribozyme. This use of introns at the RNA level does three things. First, it solves the size paradox by making complex RNA functions out of shorter, simpler pieces, RNA exons 30–40 bases long. Second, it provides a way of distinguishing genetic material in a form that could be copied from RNA folded up to form enzymes, and thus removes a requirement that the ribozyme be in a form that can be directly copied. Third, it provides a mechanism for enhanced illegitimate recombination. There may be a background of recombination in an RNA-copying world, just by the RNA synthetase occasionally jumping from molecule to molecule during the replication and thus creating recombinant products, but this would be primarily homologous recombination. The critically useful process created by the intron–exon structure at the RNA level is essentially transposition of exons. Two introns surrounding an RNA exon can cut out across the exon and make a transposon that can then enter an intron in some other molecule. Figure 1 shows this process. This concept provides a way to shuffle exons at the RNA level to create novel RNA molecules. Finally, as recently discussed by Jeffares et al. (1998), recombination tends to minimize the "Eigen limit" problem; i.e., the limits in genome size imposed by high rates of replication error (see Eigen 1993).

This picture of a fully developed RNA World uses RNA genetic material, molecules about a few thousand (1–10,000) bases long with an intron–exon structure. The splicing out of the introns ties the exons together to make functional ribozymes, which make the precursors, the charge neutralizers, and the membranes. There would be many copies of the genes in each cell so that the division does not have to be extremely accurate.

Is the genetic material single- or double-stranded? If the genetic material were single-stranded, with the ribozyme polymerase copying a plus-strand into a minus-strand and a minus-strand into a plus-strand, the intron–exon structure might be unstable. If the RNA genetic material were double-stranded, the intron–exon structure would be stable, and upon transcribing to make a plus-strand, that strand can either be copied back to make a new double-stranded molecule or can splice out its introns to become a ribozyme. In this picture, one has as many transcription units
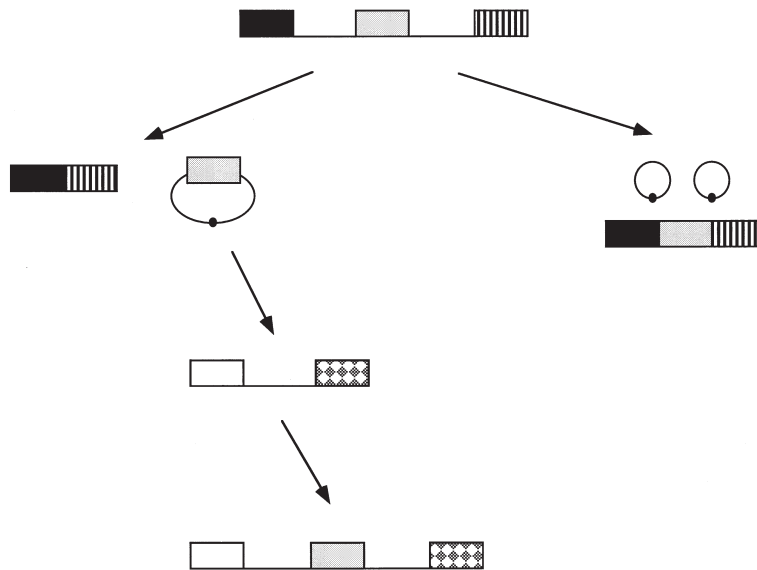
Figure 1   Schematic view of an intron–exon structure at the RNA level, spliced in two ways. The introns are shown as type I, although they could as easily be type II. Either both introns splice out to produce a ribozyme, or the extreme ends of the two introns splice together to carry an intervening exon to a new position in a novel gene.

as one has ribozymes, and about as many separate chromosomes as ribozymes.

THE RNA–PROTEIN WORLD

The concept that there is rather complete development of an RNA World accumulating in proto cells, with an RNA genetic material and RNA enzymology capable of making a variety of the biochemical structures needed, produces a simple way to develop an RNA–protein world. The reason for the simplicity is that one can introduce proteins first as short homo-oligo peptides. Their functions are to bind to RNA molecules to enhance the catalytic functions of the ribozymes. One would expect the process to begin with the ability to activate and charge a single amino acid onto an RNA molecule and to develop a precursor to the ribosome to catalyze the peptide bond formation. One must develop a way of encoding that process into mRNA in order to make short oligopeptides. One conjectures that the first amino acid used would be either lysine or arginine: lysine because of its chemical simplicity and codon simplicity; poly A, or

arginine, because of its much more extensive codon usage in the ultimate code. The first product would be short, positively charged oligopeptides to bind to RNA molecules to aid in charge neutralization made on a primitive RNA ribosome (Maizels and Weiner 1987). Products and structures that increase the fidelity of the proto-translational system would be positively selected (see discussion on proto-ribosomes in Jeffares et al. 1998). The ability of a ribozyme to activate an amino acid has recently been demonstrated (Illangasekare et al. 1995; see Chapter 7). A link between ribosomes and ribozymes is also suggested by the fact that group I introns have their function affected by kanamycin, gentamycin, and neomycin, anti-ribosome antibiotics (von Ahsen et al. 1991). A further link is the observation that an engineered ribozyme could hydrolyze an aminoacyl ester substrate (Piccirilli et al. 1992), and peptide bonds have been formed by selected ribozymes (Zhang and Cech 1997). Beginning with a single amino acid and a single transfer/activating RNA, one can then easily imagine mutant forms developing the ability to activate, encode, and transfer other amino acids, building up first to some 5 amino acids to carry out most of the protein functions and ultimately extending to the 20 amino acids (21 including seleno-cysteine) that are currently used.

Such a picture of the gradual development of the protein-synthetic machinery, for the transition from an RNA World to an RNA–protein world, does not require the big bang of a beginning for protein enzymes that spring into action completely functional. The first use of proteinaceous material is as oligopeptides to support the RNA enzymes. Gradually the protein chains become more complex, the support they can offer to the ribozymes becomes more extensive, and ultimately, protein molecules emerge that themselves carry the full enzymatic action. These can then replace the ribozyme functions for a number of biochemical processes. This view would suggest a unitary origin for the transfer RNAs arising from one common ancestor. At the RNA level, the activating ribozymes would have a unitary origin, but at the protein level the replacement of the ribozyme function need not be unitary. In fact, the protein-activating enzymes belong to two major families of dissimilar structure and differing chemical detail (Eriane et al. 1990).

As the mechanism of protein synthesis develops using RNA messages to encode proteins, the intron–exon structure at the RNA World level means that one can develop an intron–exon structure for the genetic information at the protein world level. In this hypothesis, the exons would encode small functional or folding elements of the ultimate proteins which could, through shuffling, create novel protein structures. The first protein enzymes were probably aggregates of short polypeptide chains, each

folding up as a small component of the final structure: These products as exons were linked ultimately into a single gene by introns and then moved into novel combinations by exon shuffling. Experimental evidence showing the plausibility of such aggregates exists for triosephosphate isomerase (Bertolaet and Knowles 1995) and tRNA synthetase (Shiba 1995). The end product of an RNA–protein world would be a cell with the ability to handle and to synthesize RNA and protein precursors, able to do a great deal of biochemistry. This cell would have a cell membrane and probably a cell wall to support the contents against osmotic pressure differentials. This cell would have pores and pumps to scavenge organic precursors from the environment. Such a cell has all the attributes of the last common ancestor except for DNA. The genetic material still is RNA, even with an intron–exon structure. The exon structure would match the one we have detected today in the intron positions in ancient conserved proteins (de Souza et al. 1996), in which we infer that there is a pattern of original exons which represent modules, compact units of protein structure, of the order of 21 Å, 28 Å, and 33 Å in diameter, corresponding roughly to oligopeptide units 15, 22, and 30 amino acids long. This exon–intron structure, we think, would have been fully developed at an RNA World level.

Two lines of argument further support this concept. The peptidyl transferase activity today still appears to reside in the ribosomal RNA: The basic mechanism of protein chain extension is an RNA-catalyzed one, not a protein-synthetic one (Noller et al. 1992). This is clearly an indication that an RNA enzymology preceded a protein enzymology. We described the intron–exon structure of the RNA World as being type I and type II introns. Even in the RNA World, the introns could be of type III, for which the catalysis is done not by *cis*-acting RNA structures, but by *trans*-acting ones. Today the RNA spliceosomes, the particles involved in the splicing of nuclear pre-messenger RNAs, are RNA–protein particles in which the small nuclear RNAs involved are most likely to carry out the details of the catalytic reactions (for review, see Sharp 1994). Both the ribosomes and the RNA-splicing mechanism look like remnants of an ancient RNA World.

THE DNA–RNA–PROTEIN WORLD

The genetic material of the RNA–protein world would have its genetic material in the form of RNA molecules about 10,000 bases long, RNA molecules the length of the molecules in RNA double-stranded or single-stranded viruses; the reason is that the error rate in copying RNA is likely

to be too high to permit the creation of extremely long molecules. One expects the error rate to be high because there is no proofreading mechanism, and thus the errors are likely to be on the order of 10^{-3} or 10^{-4}, about down to the tautomerism rate for the bases. Thus, at the RNA–protein world level, one expects a large number of short chromosomes, each encoding one or at most a few genes, and the assortment of these chromosomes into daughter cells probably taking place at random because the parent cell has large numbers of duplicates. The introduction of DNA solves these problems. DNA-based enzymology can develop a full-fledged error-correcting mechanism that ultimately drives the error rate down to 10^{-9}. This permits long chromosomes, an ordered development of a mitotic process, and a well-defined segregation of chromosomes into daughter organisms. The first enzymes involved would be the enzymes that create the DNA precursors from the RNA precursors, ribonucleotide reductases, and a reverse transcriptase function that can convert an RNA chain into a DNA chain. Ultimately then, we need DNA-directed RNA polymerases and DNA-directed DNA polymerases, as well as a telomerase. This picture suggests that the reverse transcriptase and the telomerase functions are of equal age and very old. Since there is an intron–exon structure for protein genes at the RNA–protein world level, this structure is simply copied into the DNA. This leads us to an intron–exon structure for the DNA genes, the exons still being primarily units of protein folding and function. Of course, at both the RNA and the DNA levels, simple exons can be fused together and reused as more complicated exons.

At both the RNA and DNA levels, tying together of the exons that correspond to the short polypeptides that might as an aggregate form enzymes increases their genetic linkage, so that the entire complement of polypeptides needed to generate some enzymatic activity can be passed in a simple form from parent to daughter. However, genetic linkage is not a required concept. Pieces of protein need not be genetically linked for the organism to survive, but if they are linked, it is easier to pass the function as a whole to the offspring and that pattern, we think, would be quite valuable in the very early stages of evolution. The intron–exon structure of a gene provides a certain amount of genetic linkage in that the exons are held together, but the separation of the exons along the genetic material provides an enhanced recombination rate, over evolutionary time, that could lead to better novel combinations, and to the creation, by illegitimate recombination, of entirely new structures.

We have discussed the origins of life focusing on the first self-replicating event. By self-replicating we mean the appearance of a molecule that can copy molecules like itself to produce more of its own kind

along with a notion that that copying could be occasionally inaccurate, hence able to introduce mutations that would permit a population of molecules varying about a norm to emerge. This variation in structure permits the emergence of molecules that copy more effectively under any specified conditions, and as those conditions change, the molecules can evolve to make more of themselves under the new conditions. We regard this as the crucial aspect of evolution: that it involves exponential replication, including both multiplication and variation, which can be worked upon by natural selection. Before the emergence of such a molecule, there was a period of prebiotic synthesis of material, which we have not discussed. That synthesis needs to include ways of forming the precursors for RNA synthesis, oligonucleotides or RNA triphosphates, or polyphosphates. The synthesis of these molecules could even be catalyzed by inorganic catalysts, such as clays or other material, which might produce high-molecular-weight precursors. But these issues of where the precursors are synthesized, or whether there is a background of inorganic or organic catalysts that can be used to produce complex molecules, to our mind are not the same as the question of where the evolution begins. The replication of a nucleic acid in the presence of a catalyst, even though the nucleic acid moiety under these conditions will modify and evolve (Spiegelman 1971; Eigen 1987), is not a full model of evolution because the multiplication or amplification is not self-sustaining. It is only the ability of a molecule effectively to copy itself, to make more of its own kind, that creates the exponential growth that is characteristic of life.

REFERENCES

Bertolaet B. and Knowles J. 1995. Complementation of fragments of triosephosphate isomerase defined by exon boundaries. *Biochemistry* **34:** 5736–5743.

de Souza S.J., Long M., Schoenbach L., Roy S.W., and Gilbert W. 1996. Intron positions correlate with module boundaries in ancient proteins. *Proc. Natl. Acad. Sci* **93:** 14632–14636.

Eigen, M. 1987. New concepts for dealing with the evolution of nucleic acids. *Cold Spring Harbor Symp. Quant. Biol.* **52:** 307–320.

———. 1993. The origin of genetic information: Viruses as a model. *Gene* **135:** 37–47.

Ekland E. and Bartel, D. 1996. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382:** 373–376.

Eriane G., Delarue M., Poch O., Gangloff J., and Moras, D. 1990. Partition of tRNA synthetase into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347:** 203–206.

Gilbert W. 1986. Origin of life: The RNA world. *Nature* **319:** 618.

Guerrier-Takada C., Gardiner K., Marsh T., Pace N., and Altman S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35:** 849–857.

Illangasekare M., Sanchez G., Nickles T., and Yarus M. 1995. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267:** 643–647.

Jay D. and Gilbert W. 1987. Basic protein enhances the incorporation of DNA into lipid vesicles: Model for the formation of primordial cells. *Proc. Natl. Acad. Sci.* **84:** 1978–1980.

Jeffares D., Poole A., and Penny D. 1998. Relics from the RNA world. *J. Mol. Evol.* **46:** 18–36.

Kruger K., Grabowski P., Zaug A., Sands J., Gottschling D., and Cech T. 1982. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena. Cell* **31:** 147–157.

Maizels N. and Weiner A. 1987. Peptide-specific ribosomes, genomic tags, and the origin of the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **52:** 743–749.

Noller H., Hoffarth V., and Zimniak L. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256:** 1416–1419.

Piccirilli J.A., McConnell T.S., Zaug A.J., Noller H.F., and Cech T.R. 1992. Aminoacyl esterase activity of the Tetrahymena ribozyme. *Science* **256:** 1420–1424.

Sharp P.A. 1994. Split genes and RNA splicing. *Cell* **77:** 805–815.

Shiba K. 1995. Dissection of an enzyme into two fragments of intron-exon boundaries. In *Tracing biological evolution in protein and gene structures.* (ed. M. Go and P. Schimmel), pp. 11–21. Elsevier Press, The Netherlands.

Spiegelman S. 1971. Extracellular strategies of a replicating RNA genome. *Ciba Found. Symp.* **1971:** 45–73.

Steitz T. 1998. A mechanism for all polymerases. *Nature* **391:** 231–232.

Steitz T. and Steitz J. 1993. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci.* **90:** 6498–6502.

von Ahsen U., Davies J., and Schroeder R. 1991. Antibiotic inhibition of group I ribozyme function. *Nature* **353:** 368–370.

White H. III 1976. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7:** 101–104.

Zhang B. and Cech T. 1997. Peptide bond formation by in vitro selected ribozymes. *Nature* **390:** 96–100.