Perhaps the greatest of all mysteries in the life sciences is the origin of life itself. How did life arise from organic and inorganic matter more than 3.5 billion years ago? Did it in fact arise on Earth? If not, how and where did it arise? Investigations into the origin of life have faced a conceptual challenge. Yes, it is possible for simple molecules of life, such as amino acids and nucleobases, to arise in the laboratory under conditions that are thought to resemble those that existed early in the history of Earth. This was demonstrated in classic experiments carried out in the 1950s by Stanley Miller and Harold Urey at the University of Chicago. Also, amino acids and nucleobases have been found in certain carbonaceous meteorites, such as the famous Murchison meteorite that fell to Earth in 1969. However, and as we have seen, replication of the carrier for genetic information (DNA or RNA) into daughter molecules and its transcription into mRNA require dedicated proteins. Indeed, we live in a Protein World in which the processes of living systems are carried out by molecular machines composed wholly or partly of proteins. Conversely, proteins are not carriers of genetic information and cannot duplicate themselves. Hence the following conundrum: If, in the beginning, life began with proteins, how was the sequence of amino acids in those proteins specified in the absence of nucleic acids? Conversely, if in the beginning life arose from nucleic acids, then how did those nucleic acids duplicate themselves and become decoded into sequences of amino acids in the absence of proteins?

The discovery that RNA can be both an enzyme and an information carrier offered a solution to this conceptual problem and led to the hypothesis that
life arose from an RNA World. Of course, we cannot go back in time and see if life evolved from RNA molecules, and there is no fossil record (with one possible exception as we will come to) that we can interrogate. What we might be able to do, however, is re-create in a test tube life or at least aspects of living systems, such as the ability to replicate and undergo Darwinian selection, and demonstrate that life could have evolved, in principle, from RNA molecules.

Ribozymes are enzymatic RNAs

We begin by asking what an enzyme is. As we have discussed, enzymes catalyze chemical reactions, and they are usually proteins. Enzymes mediate catalysis by facilitating the formation of the transition state, thereby lowering the activation energy of the reaction. Enzymes have an active site that binds substrates via intermolecular interactions (e.g., ionic interactions, hydrogen bonds, van der Waals interactions, etc.). Another important characteristic of enzymes is that they have a turnover number that is greater than one, meaning that a single enzyme molecule can catalyze the transformation of more than one set of reactants. In other words, enzymes are reusable; they are not consumed during the reaction.

The first enzyme to be purified and recognized as being a protein was urease, which catalyzes the hydrolysis of urea into carbon dioxide and ammonia. James Sumner made this seminal discovery in 1926, and over the subsequent years it led to the dogma that all enzymes are proteins. Only relatively recently has it been recognized that RNA molecules can be enzymes too. Enzymes composed of RNA are known as ribozymes. We will consider examples of RNA molecules that catalyze a variety of reactions, including the cleavage and formation of phosphodiester bonds and the formation of peptide bonds.

Some introns are self-splicing

The first clue that RNA molecules can facilitate a chemical reaction came from the discovery that certain introns are self-splicing. As you will recall from Chapter 10, the splicing of introns from RNA occurs via two transesterification reactions that are catalyzed by an RNA-protein complex called the spliceosome. The splicing of some introns, however, is not mediated by the spliceosome. Instead, these introns trigger their own removal. In self-splicing, the branch point adenosine is replaced with an exogenous guanosine nucleotide as a nucleophile that attacks the 5’ splice junction (Figure 1). That is, the nucleophile is a free nucleotide rather than a nucleotide that is part of the polynucleotide. The self-splicing intron RNA folds into a particular three-dimensional shape that creates a binding pocket to which the exogenous guanosine can bind. The 3’ hydroxyl of the guanosine is then used as a nucleophile for the first transesterification reaction, which frees the 3’ end of the upstream sequence and produces a linear intron. The second transesterification reaction proceeds just as it does in the spliceosome, with the 3’ hydroxyl of the freed upstream sequence acting as a nucleophile to attack the phosphodiester backbone at the 3’ splice junction.
Strictly speaking, self-splicing introns are not enzymes. Once self-splicing has taken place, the pre-RNA molecule has been replaced by mature RNA lacking the intron and the free, newly released intron. Thus, each intron can only splice itself out once.

**RNase P RNA is a *bona fide* ribozyme**

RNase P is an enzyme that catalyzes a step in the maturation of tRNA molecules. Like mRNAs, tRNAs are encoded by dedicated genes in the genome. The initial product of the transcription of such genes is a pre-tRNA molecule that contains an extension (or leader) at its 5’ end. RNase P removes this leader by hydrolytic cleavage, converting pre-tRNA into mature tRNA.

RNase P had been known for some time and was known to be a hetero-complex consisting of an RNA subunit and a protein subunit (Figure 2). The big surprise was the discovery that the protein subunit of the enzyme is largely dispensable for the reaction and that the RNA component is sufficient to carry out repeated rounds of cleavage of pre-tRNA substrate molecules.
The protein subunit enhances the affinity of the complex for the substrate and the rate of the reaction but is not itself needed for catalysis. Thus, the RNA subunit is a *bona fide* ribozyme, representing the first unambiguous demonstration that an RNA molecule can be an enzyme.

**The large RNA of the ribosome is a ribozyme that catalyzes peptide bond formation**

The self-splicing intron and the RNase P ribozyme act on phosphorus centers, mediating transesterification reactions in one case and the hydrolysis of a phosphodiester bond in the other. One example of a ribozyme that acts on a carbon center is known, and it is found at the heart of the ribosome.

As discussed in chapter 11, the peptide bond-forming enzyme of the ribosome, the peptidyl transferase, catalyzes the transfer of the growing polypeptide chain from the peptidyl-tRNA in the P-site to the aminoacyl-tRNA in the A-site, creating a peptide bond. The ribosome consists of protein subunits and rRNA molecules. The large rRNA of the large subunit of the ribosome, it turn out, is responsible for the peptidyl transferase reaction. Early evidence for this finding came from biochemical experiments in which the peptidyl transferase reaction was carried out with a preparation of the large subunit that had been treated in such a way as to largely strip it of its protein components. Further evidence came later from X-ray crystallographic studies of the large subunit showing that the catalytic center for peptide bond formation is lined by the large rRNA with little or no access to the ribosomal proteins, which largely decorate the outer portions of the large subunit.

This was a stunning realization, because it meant that the most fundamental reaction in the Protein World is catalyzed by an RNA molecule. If contemporary life now represented by the Protein World arose from an
earlier RNA World, then conceivably the large rRNA ribozyme is a molecular fossil from this earlier age. Perhaps many of the chemical reactions of living systems were once catalyzed by ribozymes that over the course of evolution were replaced with protein enzymes. If so, self-splicing RNA, RNase P, and the peptidyl transferase are relics from the origin of life.

What is the catalytic mechanism of peptide bond formation? The answer to this question remains uncertain but evidence favors the following two-part hypothesis. First, the rRNA functions by aligning (properly positioning) the peptidyl-tRNA and aminoacyl-tRNA substrates so as to favor nucleophilic attack by the free amino group of the aminoacyl-tRNA on the carbonyl carbon of the peptidyl-tRNA. Second, and interestingly, the substrate itself, the aminoacyl-tRNA, is involved in the reaction via a proton shuttle. In this shuttle, the lone pair of electrons from the nitrogen nucleophile attacks the carbonyl carbon, as shown in Figure 3. As this happens, the bond between the carbonyl carbon and the 3’ oxygen breaks, and the electrons from that bond are used to capture a proton from the 2’ hydroxyl group of the tRNA. As the bond between the 2’ oxygen and the proton breaks, the electrons from the bond are used to capture a proton from the nitrogen atom. A new lone pair is formed on the nitrogen as the bond is broken between the nitrogen and the proton. Thus, the 2’ hydroxyl group acts as a catalyst by donating and then regaining a proton to speed up the rate of the reaction (i.e., it lowers the activation energy). Proton shuttling stabilizes the transition state, lowers the activation energy, and speeds up the reaction.

Evolving a ribozyme that is capable of self-replication

If the RNA World hypothesis is correct, then an important missing link in the chain of events from prebiotic chemistry to the emergence of a self-propagating, living system is an RNA molecule that is capable of replicating itself. Even if such a self-replicating RNA existed, we have little chance of finding it in the fossil record, and no such ribozyme is known in contemporary life forms. Therefore, the only way to test the RNA World hypothesis is to attempt to create a self-replicating RNA in the laboratory as a proof of principle. Although no such RNA replicase has been created so far, some striking progress has been made in creating RNAs that are capable of template-directed RNA synthesis by a two-stage artificial evolution strategy as we now explain.
The basic idea behind the strategy is to select for RNA molecules that have polymerizing activity from a pool of RNAs with randomized sequences. The first stage starts with RNA molecules consisting of two domains. The 5’ domains of these starting RNAs have random sequences, whereas the 3’ domain, the primer, is a fixed sequence (Figure 4A). The strategy takes advantage of the enormous diversity of a library of random RNA sequences \(4^n\) where \(n\) is the length of the 5’ domain), meaning that within the vast pool of RNAs is likely to be one or more molecules that are capable of carrying out phosphodiester bond formation. The pool of potential replicases is incubated with nucleoside triphosphates as substrates and a template RNA that contains a sequence complementary to the primer, representing the second, 3’ domain. After the incubation, RNAs that have grown longer by addition of one or more nucleotides at their 3’ terminus are selected by gel electrophoresis (Figure 4B). In other words, RNAs from the pool that exhibit polymerizing activity would be able to extend their 3’ terminus and would be identified by their increase in size.

In the second stage, the polymerase chain reaction (PCR) is used to amplify the 5’ domain (the 5’ domain only) of the size-selected RNAs (Figure 4A). The PCR amplification is conducted in such a way (termed error-prone amplification) as to introduce new nucleotide substitutions (mutations) in the 5’ domain, thereby “evolving” the pool of RNAs. We then re-attach 3’ primers to the 3’ ends of these amplified and mutated 5’ domains. Thus, we have in effect created a new pool of starting RNAs that had shown some level of polymerizing activity and have now been further mutated. We incubate these RNAs with nucleotide substrates and template and once again select for RNAs that have grown in size relative to the other RNAs in the pool. This should enrich for RNAs that have enhanced polymerizing activity. We keep repeating the cycle of incubation, size selection, and error-prone amplification until we obtain an RNA replicase that is capable of robustly extending the primer. We also periodically verify that the selected RNAs carry out extension at the 3’ end in a template-dependent manner.

So far it has been possible to obtain an RNA replicase that is capable of extending the 3’ terminus by close to 100 nucleotides. Thus, artificial evolution experiments of this kind have established that RNA molecules are indeed capable of polymerizing nucleoside triphosphates in a template-directed manner.

Many fundamental questions remain unanswered in attempting to understand if and how life arose from RNA. Among these is the source of nucleoside triphosphate substrates for RNA synthesis in the prebiotic world. Another question is how polynucleotides arose that were long enough to exhibit some level of polymerizing activity as a stepping stone to a full replicase. A further question is how replicases would be capable of self-replication. It seems difficult to imagine that a replicase molecule could directly make a copy of itself. If not, then one RNA replicase would have to copy another. But then the product would be the complement (it would have the complementary sequence), which would not be expected to have replicase activity. So a replicase would have to make copies of its complement to generate additional replicase molecules.
Figure 4 Self-replicating ribozymes can be evolved in the laboratory from random sequences of RNA

(A) Shown is a selection scheme to evolve self-replicating ribozymes from random sequences of RNA. In this scheme a pool of random RNA sequences (green) are attached to a fixed primer sequence. This fixed primer sequence is complementary to a portion of an RNA template molecule (purple), which is added to the pool of random RNAs. RNA molecules with polymerizing activity become longer when incubated with template and NTPs via the addition of nucleotides to their 3’ ends (blue). Size selection is used to select for RNAs that have become longer. The active RNAs are amplified using PCR, mutagenized to increase diversity, and reattached to a primer sequence. This results in a new pool of RNAs that can be used for another cycle of selection. Processes described in the text as stage one are indicated with black arrows; stage two is indicated with a red arrow. (B) Shown is a cartoon representation of a gel electrophoresis experiment used to separate active replicases from RNAs that did not elongate. This results in a new pool of RNAs that can be used for another cycle of selection. Processes described in the text as stage one are indicated with black arrows; stage two is indicated with a red arrow. (B) Shown is a cartoon representation of a gel electrophoresis experiment used to separate active replicases from RNAs that did not elongate. This results in a new pool of RNAs that can be used for another cycle of selection. Processes described in the text as stage one are indicated with black arrows; stage two is indicated with a red arrow.
Yet one more challenging question is the nature of the Darwinian system that would have allowed for the selection of superior replicases. If perchance a mutant replicase arose that had superior replicase activity, it would have to compete with all the other replicase molecules as a substrate for its own replication. Being a superior replicase would not mean that it is also a superior template for other replicases. Therefore, the Darwinian system would have needed a means by which superior replicases could outcompete other replicases.

Of course, we are left with many other kinds of questions and challenges. For example, how did living systems become partitioned into cell-like compartments that were capable of growth and division? How did the transition occur to a Protein World that was capable of generating and utilizing chemical energy and creating the vast array of metabolites and small molecules upon which contemporary life is based? As you can see, and in contrast to the topics we considered in Chapters 1-12, RNA and the Origin of Life raises many more questions than it answers.

**Summary**

Not all enzymes are proteins. Some enzymes known as ribozymes are RNAs. Self-splicing by certain introns is mediated by the intron itself. Self-splicing is not, however, a catalytic process, as it has a turnover number of one. An example of a bona fide ribozyme is RNase P, which removes the leader RNA from pre-tRNA molecules by hydrolytic cleavage. RNase P is a heterocomplex of RNA and protein in which the RNA component is necessary and sufficient to catalyze the cleavage reaction and can act repeatedly on one pre-tRNA molecule after another. Another example of a ribozyme is peptidyl transferase. The ribosome is composed of rRNAs and proteins, but peptide bond formation is catalyzed by the large rRNA of the large subunit rather than by a ribosomal protein.

The discovery of ribozymes raised the hypothesis that the contemporary Protein World arose from an earlier RNA World in which the reactions of life were catalyzed by RNA molecules. If so, then the peptidyl transferase may be a bridge between the RNA World and the Protein World in that an RNA molecule catalyzes the formation of the most fundamental reaction in the protein world, peptide bond formation.

If life arose from an RNA World, then the carrier of genetic information, the primordial RNA genome, must have been capable of catalyzing its own replication. This would require that the primordial RNA genome was both a repository of genetic information and a ribozyme capable of self-replication. No self-replicating ribozyme is known, but artificial evolution has been used to select for RNA molecules that are capable of polymerizing nucleoside triphosphates.