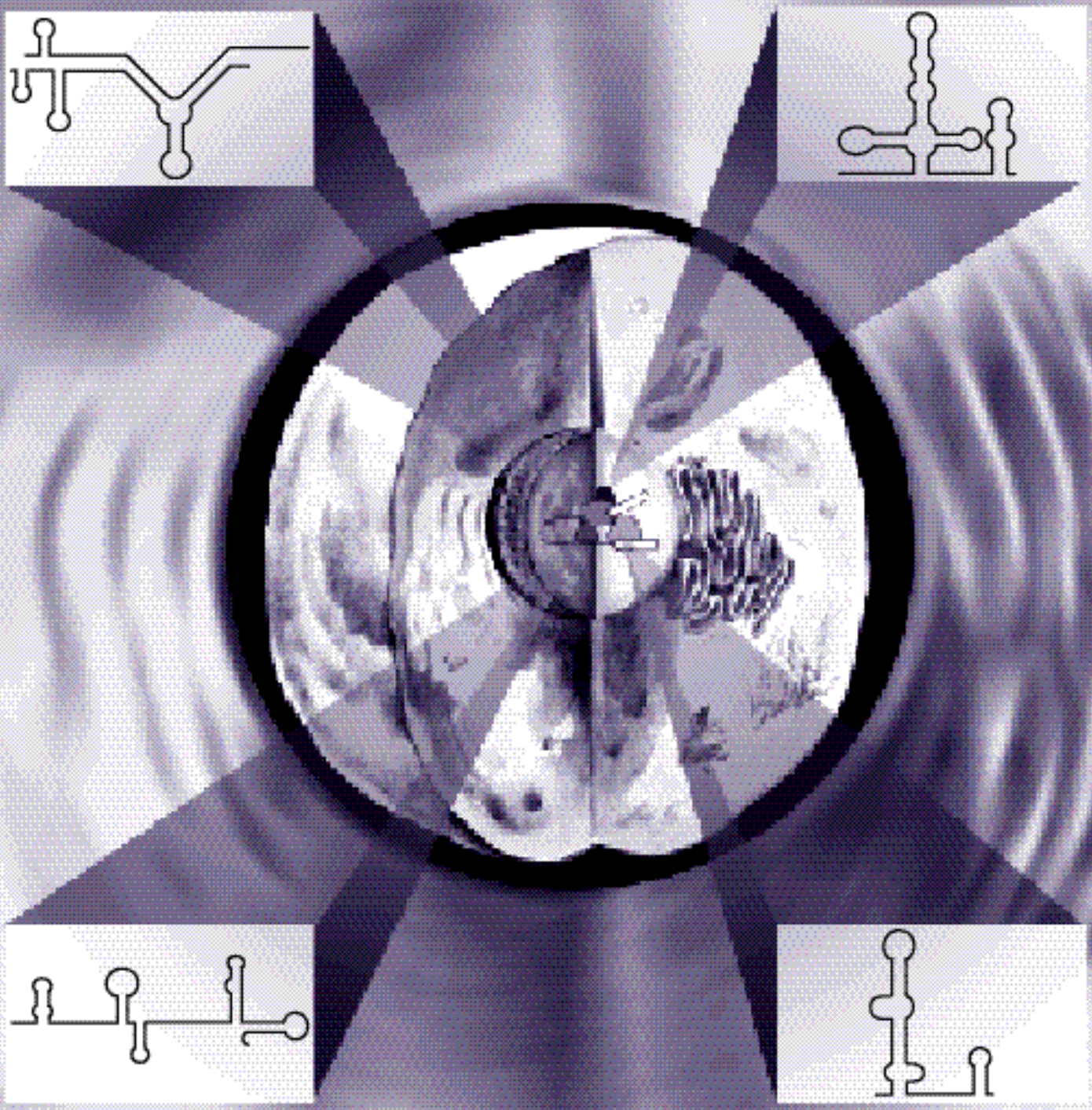
Bringing RNA into View

RNA and Its Roles in Biology



BSCS

Bringing RNA into View: RNA and Its Roles in Biology

Biological Sciences Curriculum Study (BSCS)
Pikes Peak Research Park
5415 Mark Dabling Blvd.
Colorado Springs, Colorado 80918-3842

Authors

John F. Atkins, *Eccles Institute of Human Genetics, University of Utah*
Andrew Ellington, *Indiana University*
B. Ellen Friedman, *Consultant*
Raymond F. Gesteland, *Eccles Institute of Human Genetics, University of Utah*
Harry F. Noller, *University of California—Santa Cruz*
Stephen M. Pasquale, *BSCS*
Richard D. Storey, *The Colorado College*
Olke C. Uhlenbeck, *University of Colorado, Boulder*
Alan M. Weiner, *Yale University*

This material is based on work supported by the National Science Foundation under grant No. 9652921. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the granting agency.

BSCS Development Team
Project Leadership
Joseph D. McInerney, *Principal Investigator*
Stephen M. Pasquale, *Project Director*

Project Coordination
Rose Johnson, *Production Assistant*
Dee Miller, *Production Assistant*
Barbara C. Resch, *Editor*

Project Support
Linda Ward, *Senior Executive Assistant*
Kevin Andrews, *Graphic Artist*
Lisa M. Chilberg, *Graphic Designer*

BSCS Administrative Staff
Timothy H. Goldsmith, *Chair, Board of Directors*
Rodger W. Bybee, *Executive Director*
Susan Loucks-Horsley, *Associate Executive Director*
Marcia Mitchell, *Financial Officer*

Field-Test Teachers
David Allard, *Texarkana College & Texas A&M University-Texarkana, Texas*
Gail Carmack, *The Science Academy of Austin at LBJ High School, Texas*
Alix Darden, *The Citadel, South Carolina*
Denise M. Foley, *Chapman University, California*
David A. McCullough, *Wartburg College, Iowa*
John Ohlsson, *University of Colorado, Boulder*
LeLeng P. To, *Goucher College, Maryland*

Design and layout by Angela Greenwalt

Illustration Credits
Figure 1.4a: © A. Barrington Brown, Science Source/Photo Researchers; Figure 1.4b: Courtesy of Harry F. Noller; Figure 1.4c: *Genes and Genomes* by Maxine Singer and Paul Berg, published by University Science Books, Sausolito, CA; Figure 2.3: John T. Ohlsson, University of Colorado, Boulder; Figure 3.6: stopwatch © 1999–2000 www.barrysclipart.com; Figures 3.9, T3.1, and T3.2: Gerald F. Joyce and Kathleen E. McGinness, The Scripps Research Institute; Figure 4.1b: Noller, H.F. (1993). On the origin of the ribosome. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (p. 141). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Figures 8 and 1.6 are adapted from Gesteland, R.F., & Atkins, J.F. (1993). *The RNA world*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Evaluation Form for
Bringing RNA into View: RNA and Its Roles in Biology

Your feedback is important. After you have used the module, please take a few minutes and return this form to BSCS, Attn: RNA, 5415 Mark Dabling Blvd., Colorado Springs, CO 80918-3842.

1. Please evaluate the *Faculty Background* by marking this form and providing written comments or suggestions on a separate sheet.

Sections Used	not helpful				very helpful
Overview	1	2	3	4	5
Background on RNA	1	2	3	4	5

2. Please evaluate the activities by marking this form and providing written comments or suggestions on a separate sheet. Rate the activities for their effectiveness at teaching concepts of RNA and its role in biology.

Activity 1: RNA Structure: Tapes to Shapes	1	2	3	4	5
Activity 2: RNA Catalysis	1	2	3	4	5
Activity 3: RNA and Evolution	1	2	3	4	5
Activity 4: RNA Evolution in Health and Disease	1	2	3	4	5

3. What are the major strengths of this module?

4. What are the major weaknesses of this module?

5. Please rate the overall effectiveness of this module: not effective very effective

   1 2 3 4 5

6. Please provide a description of the classes in which you used this module: (circle response)

   College: 2 year 4 year High school: grade 9 10 11 12

   freshman sophomore junior senior Level of class: basic honors 2nd year

   How many students used the module? _____ How many students per class? _____

   Ethnicity (approximate % of minorities): _____

   Description of school: _____

   College: liberal arts science High school: urban suburban rural

7. Have you used BSCS materials before? ❒ yes ❒ no

8. Please provide your name and contact information below:

   Name_____

   School_____

   Mailing address_____ ❒ home ❒ work

   _____

   _____

   Phone_____ ❒ home ❒ work

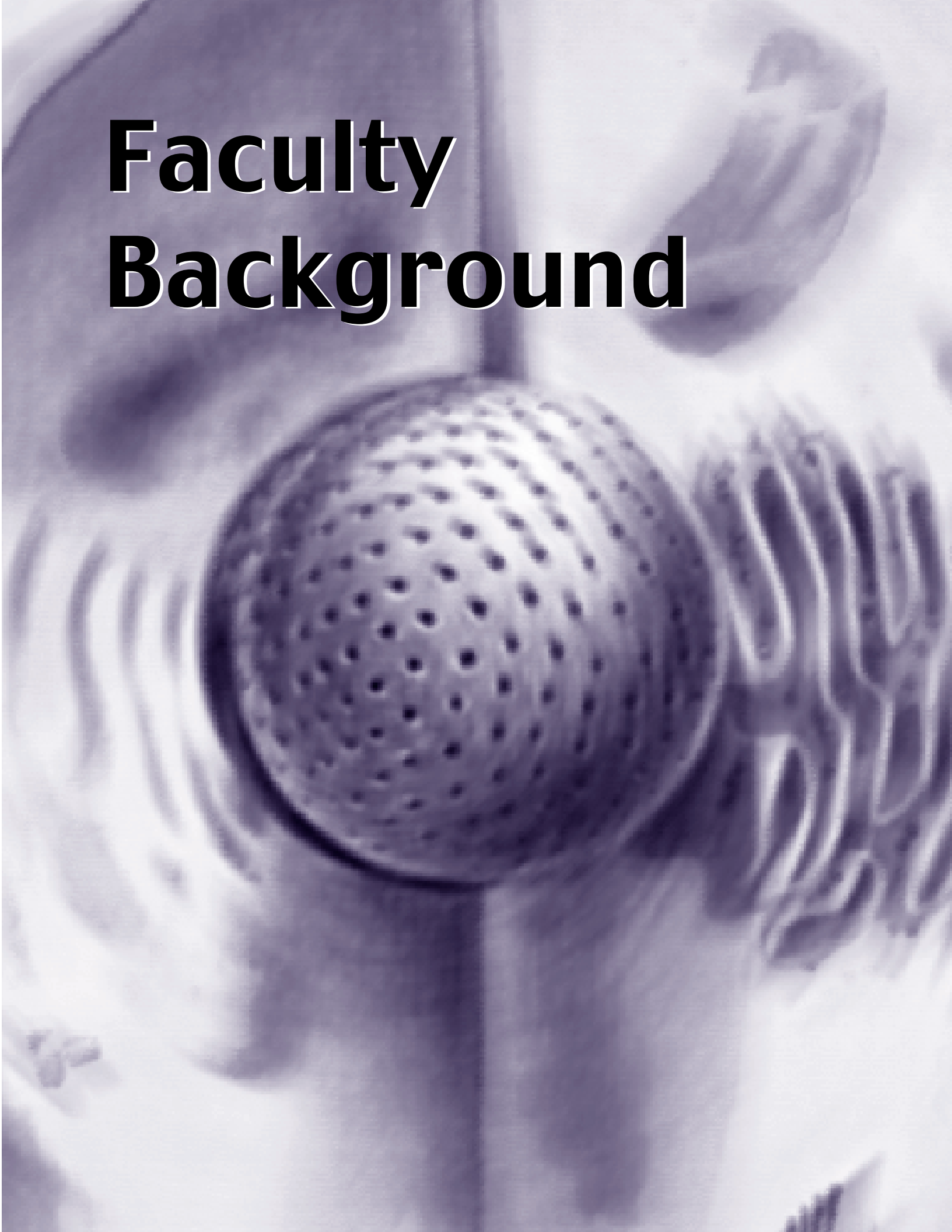
   FAX_____ ❒ home ❒ work

   E-mail address_____

   Was the address on your mailing label correct? ❒ yes ❒ no

Contents

Faculty Background



Overview

Introduction

The impetus for developing *Bringing RNA into View* for college biology classes is the recent and rapid growth in knowledge of the structures and diverse functions of RNA molecules. First described in 1947 as a cellular constituent involved in protein synthesis, RNA has since been shown to play several other essential roles in gene expression, including genome maintenance, processing and editing of primary transcripts, and localization of proteins within the cell. New RNAs continue to be discovered performing unexpected tasks in the cell.

Perhaps more than any other single discovery about nucleic acids since Watson and Crick's elucidation of the DNA double helix, the finding that RNA (and DNA) has catalytic ability has expanded our view of these molecules' potential, as evolutionary progenitors and contemporary biochemical players. The notion that an RNA world phase occurred early in prebiotic evolution (deriving both its information encoding and catalytic functions from RNA molecules) has inspired a burst of research into the molecular origins of life and the biochemical potential of nucleic acids.

Particularly revealing of the impressive biochemical potential of these molecules are recent studies expanding on the 30-year-old observation of Darwinian selection and evolution of nucleic acids in the test tube. Technological advances of the last 15 years have greatly increased the power and versatility of such *in vitro* experiments to create and select talented nucleic acid molecules, some of which have begun to reveal RNA's potential for true self-replication, for synthesizing its own monomer building blocks, and for practical use in combating viral and microbial infections as well as genetic disorders.

BSCS selected the topic of this module, RNA and its role in biology, as a key area for synthesizing these important new research findings into the fundamental concepts of college-level biology. In addition, this topic offers a useful opportunity to shift students from focusing on isolated facts to approaching biology conceptually; in short, the module helps students think about biological processes.

To develop this module, biologists at the Biological Sciences Curriculum Study in Colorado Springs and the Eccles Institute of Human Genetics at the University of Utah in Salt Lake City worked with an external advisory committee of scientists and educators, plus a variety of college faculty who conducted field tests of the module. This process identified the following major concepts for the module:

- Nucleic acids (DNA and particularly RNA) have two major functions: as informational molecules and as biochemical catalysts.
- The sequence of monomers in RNA dictates its three-dimensional structure and, consequently, its function.
- Molecules can be subject to natural selection.
- Evolution requires an iterative process of molecular replication, mutation, and selection.
- Modern roles for RNA suggest its major role in the origin of life.
- Studies of the origin of life and its evolution work with substantial data, such as modern observation of naturally occurring ribozymes (RNA catalysts) and *in vitro* molecular selection experiments.

- Modern roles for RNA include serving as a genome to a variety of viruses and smaller viroids.
- Most RNA-based viruses and viroids are significant pathogens of humans, animals, and agriculturally important crop plants, and as such they have a major social and economic impact.

How to Use the Module

The module provides background materials for faculty and a set of four educational activities for students. *Background on RNA* (in this section) provides you with an update on RNA research in a form that is accessible to busy faculty. This material is for your own use. It may extend your knowledge and thus be helpful for teaching the activities, but it is not essential to teaching the activities. The activities are inquiry-based explorations that offer you an alternative to lecture and stimulate student interest and responsibility for learning.

Figure 1 shows the layout of the materials. The module features four core classroom and laboratory activities that explore RNA's structure and function, RNA catalysis, RNA replication and evolution, and RNA's role in health and disease.

Notice that each activity appears in two sections, *Annotated Faculty Pages* and *Student Pages*. The *Student Pages* consist of introductory text and For Your Information essays that provide context and elaboration for the activity. Detailed protocols are provided for students, and Challenge Questions stimulate student thought and synthesis of ideas. The *Student Pages* and the *Copymasters* and *Templates* may be photocopied for classroom use.

The *Annotated Faculty Pages* contain the student text (**bold type**) plus annotations for the faculty (regular type). The annotations contain suggestions and hints for teaching each activity, answers to Challenge Questions, optional extension exercises, and reagent preparation instructions.

A summary of the activities is provided in Figure 2. Time and resources may not permit you to teach all four of the activities. However, we recommend that you teach Activity 1 (*RNA Structure: Tapes to Shapes*) before the other activities, to ensure that students understand the structural basis for the RNA functions explored in the later activities.

As your students proceed through the module, we recommend that you encourage them to ask questions, seek outside resources, and be aware of the way in which science attempts to understand natural processes. For example, call students' attention to citations in essays so that they begin to appreciate the significance of discussions that are based on primary scientific data rather than hearsay.

Figure 1 The module at a glance. *Bringing RNA into View* consists of three main components (*Faculty Background*, *Annotated Faculty Pages*, and *Student Pages*) as well as support materials.

Faculty Background
- Overview
- Background on RNA
- References

Annotated Faculty Pages
- Instructions and background for Activities 1–4

Copymasters and Templates
- Handouts and masters for Activities 1–4

Student Pages
- Student materials for Activities 1–4

Figure 2 Summary of the student activities.

Activity 1
RNA Structure: Tapes to Shapes
Students apply rules of base pairing and folding to construct physical models of RNAsequences. They use their models to explore structure-function relationships and the effects of mutation.

Activity 2
RNA Catalysis
Students explore catalytic RNA in the laboratory using a self-splicing group I intron. Students apply the techniques of *in vitro* transcription, RNA isolation, and acrylamide electrophoresis to study the kinetics of the splicing reaction.

Activity 3
RNA and Evolution
Students explore the replication of a catalytic RNA in the laboratory using a continuous *in vitro* system.

Activity 4
RNA Evolution in Health and Disease
Students explore the continuing evolution of RNAin the context of the emerging resistance of bacteria and viruses to therapeutic agents.

Background on RNA

The Path to the RNA World Hypothesis

The classic formulation of the flow of genetic information during gene expression holds that DNA is copied both to itself and to RNA, and RNA is then decoded to synthesize protein. This view was formalized by Francis Crick in 1968 as the central dogma of molecular biology (Figure 3).

Several assumptions are inherent in this traditional view: Information flow is unidirectional, as expressed by the one-way arrows; only DNA is a template that can be replicated; nucleic acid coding information must ultimately be translated to protein form if working catalysts are to result; and DNA, as the ultimate molecule that encodes genetic information, likely preceded RNA during the formation and subsequent evolution of biomolecules.

This straightforward view of gene expression has given way to a more detailed understanding of information flow in biology. Molecular genetic research, particularly during the last 15 years, provides exciting new insights that reveal the original formulation of the central dogma to be incomplete. Some of the assumptions listed above have been shown to be limiting. For example, the study of viruses having an RNA genome revealed that RNA, like DNA, can serve as a primary information-encoding molecule. Also, the discovery in 1970 of reverse transcriptase, an RNA virus-encoded protein catalyst that copies RNA-based information into DNA form, revealed that information flow in the biological world is in fact a two-way street between DNA and RNA.

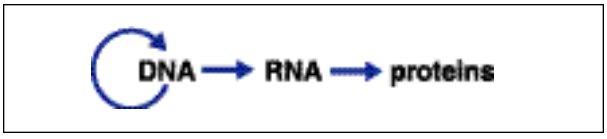


Figure 3 Central dogma of molecular biology.

Watson and Crick appreciated DNA's potential to be a template for its own replication as soon as they solved the double helical structure of the molecule. We now realize that RNA molecules likewise can play template roles for their own replication. In the process known as *template-directed nucleic acid replication*, an RNA sequence can serve as an informative scaffold onto which complementary bases align to produce a complementary strand (Figure 4a). Many viruses infecting plants and animals, such as plant viroids and polio virus, employ a two-step replication strategy in which their single-stranded RNA genome, acting as a template, is initially transcribed by a replicase enzyme to yield a complementary RNA strand. This molecule in turn serves as a template whose transcription regenerates copies of the original genome strand, thereby effecting replication (Figure 4b).

The discovery of RNA's informational and template functions expanded our ideas of the early evolution of information storage and expression mechanisms. However, the most dramatic and unexpected discovery to influence these ideas came early in the 1980s, when scientists in two different laboratories independently discovered that RNA has catalytic ability. Thomas Cech and his group at the University of Colorado were studying the splicing of large ribosomal RNA (rRNA) precursors in the protozoon *Tetrahymena*. These scientists serendipitously observed that the RNA precursor spontaneously

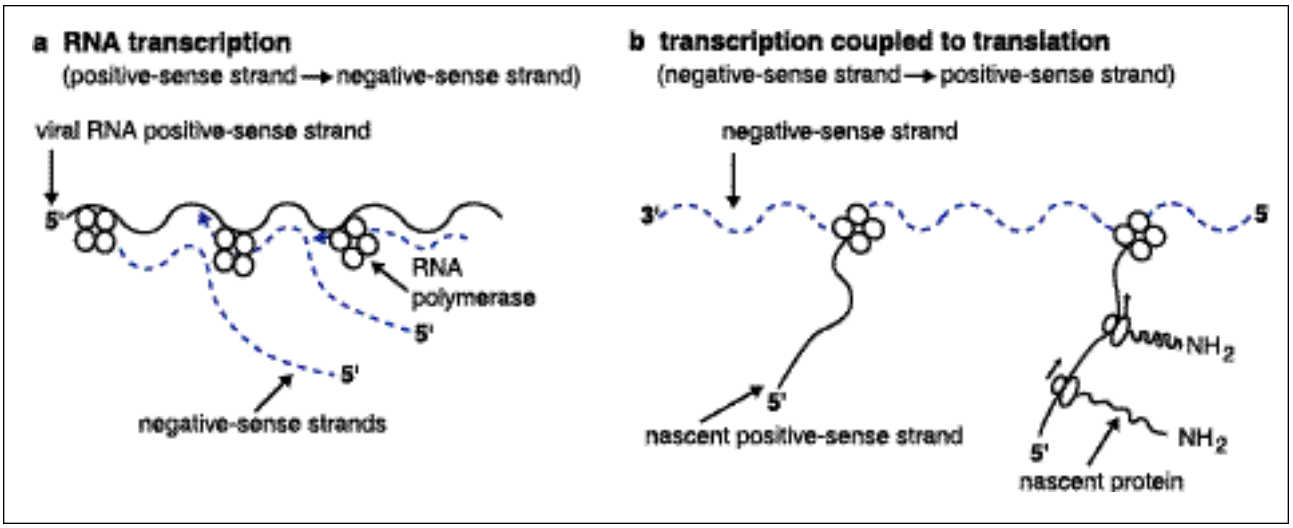


Figure 4 Viral RNA replication. The replication cycle of viruses typically consists of two steps. In (a), the positive-sense RNA genome is transcribed into many copies of a complementary negative-sense strand by RNA polymerase. In (b), the newly formed complementary strands become templates for synthesizing many copies of the original positive-sense genome. The new genomic strands may be translated as they are being synthesized.

changed size, becoming smaller after incubating in a protein-free buffer solution containing only $Mg2^{+}$ ions. Realizing how unusual this result was, they pursued it in earnest. In a series of brilliant experiments, Cech and his coworkers proved that this RNA, the "group I intron," has the inherent ability to catalyze its own excision from the RNA precursor. This RNA intron catalyzes its own self-splicing without the aid of any protein.

Meanwhile, in the lab of Sidney Altman at Yale University, researchers were continuing a long series of biochemical experiments to characterize an enzyme activity in *Escherichia coli*, called *RNase P*, that trimmed the 5′-end of a transfer RNA (tRNA) precursor. After exhaustive purification, the active enzyme was found to contain both a protein component *and* an RNA molecule. The conventional prejudice was that the protein must act as the "enzyme." But any effort to remove the RNA component eliminated the catalytic ability. Again, pursuit of an unexpected finding and diligent experimentation showed that the RNA molecule itself was sufficient to catalyze the trimming reaction. The RNA, in this case, was a catalyst acting not on itself but on another RNA.

Here were two different RNA molecules that catalyzed biochemical reactions, just like proteins. Nobel Prizes followed for both Cech and Altman—a new era had begun. Since this discovery, seven naturally occurring classes of ribozyme have been recognized, and hundreds of specific examples have been identified in a wide variety of organisms (Figure 5). The true prevalence of ribozymes in contemporary cells is unclear, but it is quite likely that new examples will be discovered.

Interestingly, the independent discovery of RNA catalysis by Cech and Altman occurred in research fields that were relative backwaters at the time. In the early 1980s, the hot research fever was in the first thrust of exploiting the new molecular biology technologies (for example, restriction enzymes, cloning, and sequencing) to explore emerging topics such as split genes, tumor viruses, and oncogenes. The unlikely discovery of catalytic RNAs in this setting is testimony to the importance in science of observing and attending to unexpected and unusual results—and being willing to pursue them despite prevailing fashions.

The RNA World Hypothesis

In addition to destroying the orthodox assumption that only proteins function as biological catalysts, the discovery of RNA's chemical versatility led to a dramatic change in how scientists view the likely sequence of molecular events during early evolution. Taken together, RNA's informational, template, and catalytic abilities led to the hypothesis that

Figure 5 Naturally occurring ribozymes.

Class	Size	Reaction	Source
group I intron	large: 413 NT in *Tetrahymena thermophilia*	intron excision	eukaryotes, eubacteria, and viruses
group II intron	large: 887 NT in yeast mitochondria	intron excision	eukaryotic organelles and eubacteria
RNase P	large: 350–410 NT	hydrolytic endoribonuclease	RNA subunit of eubacterial RNase P
hammerhead	small: 31–42 NT (enzyme strand can be 16 NT)	RNA cleavage	viral satellite RNA in plants, viroids, and newt satellite DNA
hairpin	small: 50 NT (minimum sequence)	RNA cleavage	(-) strand satellite RNA of tobacco ringspot virus
hepatitis Delta virus (HDV)	84 NT (required)	RNA cleavage	HDV
Neurospora VS RNA	881 NT (164 sufficient)	RNA cleavage	*Neurospora* mitochondria

RNA evolved, before the appearance of DNA or protein, in an *RNA world*. During this proposed phase of evolution, RNA is assumed to have provided both the coding and the catalytic abilities necessary and sufficient to initiate biological evolution. Specifically, if RNA's catalytic abilities during this time extended to its own *self-replication*, then molecular evolution automatically would have started as randomly variant RNAs were naturally selected on the basis of evermore efficient replication ability. This property of unaided self-replication, although an essential assumption of the RNA world hypothesis, remains to be demonstrated. No nucleic acid sequence possessing RNA replicase catalytic activity has yet been found in nature, but researchers have taken the first steps toward creating such a self-replicator RNA molecule *in vitro*. In a later section, we examine powerful new *in vitro* technologies for directing the evolution of nucleic acid molecules toward this and many other functions, and we discuss their implications for research in both evolutionary biology and biomedical science.

Figure 6 Modified central dogma.

The coding, template, and catalytic abilities of RNA also led to the current, expanded formulation of the central dogma, one that incorporates the RNA world hypothesis and places information flow in an evolutionary, historical context (Figure 6).

The discovery of ribozymes clearly was the catalyst for the current intense interest in the RNA world hypothesis. But the idea that RNA perhaps was the first genetic molecule is not new: Francis Crick and Leslie Orgel first proposed the possibility in 1968. Indeed, today's RNA world hypothesis builds upon a long history of research findings related to the origin of life and molecular evolution (Figure 7).

Next, we discuss RNA shapes and the growing appreciation of their complexity and contribution to the diverse functions of RNA.

RNA Structure: Tapes to Shapes

A key principle of molecular structure is that shape determines function. For biological polymers composed of multiple subunits, the fundamental determinant of shape is the linear information represented in the sequence of individual subunits; the molecule is in effect an information tape. A specific three-dimensional shape emerges in the molecule as thermodynamically favored physical

Figure 7 Milestones in the evolution of the RNA world hypothesis.

1800s	The idea of spontaneous generation, belief in the ongoing creation of living organisms from nonliving materials, persists.
1828	Friedrich Wöhler synthesizes urea in the laboratory, eliminating "vital-force" as an agent in the synthesis of organic chemicals.
1859	Charles Darwin publishes *On the Origin of Species,* in which he proposes a theory of biological evolution based on the mechanism of natural selection.
1864	Louis Pasteur experimentally disproves ongoing spontaneous generation by showing that when liquids are boiled in order to kill any microorganisms present, and are subsequently kept sterile, no organisms appear in them.
1924	Alexander Oparin employs geological evidence in proposing that the early earth had a reducing atmosphere lacking oxygen and that the first single-celled organisms might have arisen from simple organic molecules present in this early atmosphere: in effect a restricted version of spontaneous generation.
1929	Biochemist J.B.S. Haldane proposes that life might have arisen on earth when Oparin's early atmosphere was subjected to energy in the form ultraviolet radiation and heat from the cooling earth.
1953	Graduate student Stanley Miller provides experimental support for the Oparin-Haldane hypothesis by mixing gases of the "primitive atmosphere" in a glass reaction vessel and subjecting them to electric current for one week; amino acids are formed *de novo.*
	Biologist James Watson and physicist Francis Crick publish their findings on the structure of DNA.
1961	Marshall Nirenberg and his colleagues begin their five-year project of cracking the genetic code by discovering that a messenger RNA made up entirely of the base uracil can be translated into a peptide made up entirely of the amino acid phenylalanine.
1962	Watson and Crick share a Nobel Prize for their work on DNA structure.
1967	Sol Spiegelman demonstrates the replication and evolution of RNAmolecules in the test tube.
1968	Francis Crick and Leslie Orgel propose that the first information molecule was RNA; Crick advances the central dogma of molecular biology.
1970	David Baltimore and Howard Temin independently discover reverse transcription of viral RNAgenomes into DNA.
1972	Harry Noller proposes a role for ribosomal RNAin the translation of messenger RNA into protein.
1982–83	Thomas Cech and Sidney Altman independently discover the first examples of catalytic RNAmolecules: ribozymes.
1986	Walter Gilbert coins the term "RNAworld" to describe the hypothesized time during which RNAwas the primary informational and catalytic molecule.
	Kary Mullis develops the polymerase chain reaction (PCR) technology that allows rapid copying of DNA and RNA sequences *in vitro* and enables large-scale laboratory studies of molecular evolution.
1989	Cech and Altman share a Nobel Prize for their discovery of catalytic RNA.
	Gerald Joyce develops the technique of *in vitro* amplification and selection of RNA(that is, directed evolution) using the PCR technique.
1992	Noller presents evidence for the catalytic involvement of the 23S rRNA in peptide bond formation.
1993	Mullis receives a Nobel Prize for his development of the polymerase chain reaction.
	Joyce further develops *in vitro* RNA amplification and evolution experimental procedures.
1995	Jack Szostak's laboratory takes the first steps toward the *in vitro* selection of a self-replicating RNA molecule.
1998	David Bartel and Peter Unrau use *in vitro* selection to demonstrate that RNAcan catalyze the formation of individual nucleotides.

interactions, typically noncovalent, occur between compatible subunits located at a distance from one another. Indeed, this important structure-function principle provides the entire rationale for evolution's invention of a genetic information-coding strategy almost 4 billion years ago.

The importance of linear information for molecular shape and function was first appreciated for proteins, the first biopolymers to be sequenced. Their covalently linked amino acid subunits are now known to interact further through a variety of weaker noncovalent associations; these include van der Waals forces, hydrogen bonds, ionic bonds, and hydrophilic interactions. These noncovalent interactions of amino acids, both locally and with more remote neighbors through folding of the molecule, give rise to higher-order protein shapes. More basic, local shape elements, such as hydrogen bond-stabilized alpha-helices and beta-sheets, can interact in a variety of ways within the folded protein. These higher-order, three-dimensional interactions are typically stabilized by hydrophobic and ionic forces to create a vast array of specific protein shapes. We recognize a variety of functional sites in proteins that result from their shapes: enzymatic active sites, binding pockets, regulatory sites, and domains for protein-protein interaction.

The science community's recognition of diversity of shape among nucleic acids developed more slowly, however. DNA's extended double helix, the first nucleic acid structure to be revealed, gave no hint of more complicated shapes. Only later, when the base sequence and three-dimensional structure of transfer RNA (tRNA) was worked out, was the ability of nucleic acids to adopt complex shapes confirmed.

The most important determinant of folding and shape in single-stranded nucleic acids, both RNA and DNA, is complementary base pairing via hydrogen bonding, according to the base-pair rules first established by Watson and Crick. For RNA, pairing of A with U and of G with C is the primary basis for folding. Even though each RNA molecule in a cell normally consists of a single continuous strand, RNA molecules frequently contain linear runs of bases that are complementary to other runs located elsewhere in the molecule. This allows the molecule to fold back and form double-stranded regions within itself.

These double-stranded regions of the molecule adopt a helical configuration similar to that found in double-stranded DNA. (A subtle but characteristic difference, however, is that helical regions in RNA adopt the A-form geometry, whereas those in DNA are most often B-form.) The helical regions alternate with more flexible single-stranded regions. Activity 1 demonstrates that even an RNA molecule of modest length can fold in several possible ways by bringing together different, more or less complementary regions. The degree of match and resulting thermodynamic stability of one structure over an alternative determines which form predominates in the cell.

Until recently, it was difficult to determine the folded structures of RNAs, and only a few were known. Textbooks typically show tRNA as a folded structure,

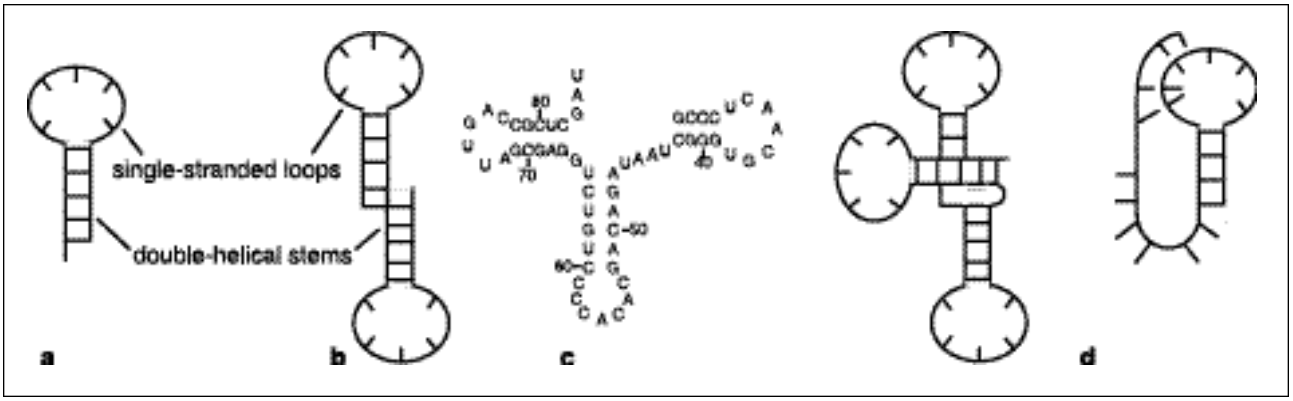


Figure 8 Some examples of RNA shapes and structural elements. Molecules consisting of one (a), two (b), or three (c) stem-loop elements. (d) A pseudoknot configuration.

whereas messenger RNA (mRNA) is shown as a linear thread lining up to be decoded. In fact, mRNA is known to fold into complex structures, and it is only "ironed out" by the passing ribosome assembly during translation. Recent advances in X-ray crystallography and nuclear magnetic resonance spectroscopy have opened up the investigation of RNA structure, and the number of resolved structures is increasing rapidly. Structural and sequence databases are beginning to reveal common motifs in RNA. Most such motifs are formed by conventional A-U, G-C base pairings, although occasional non-Watson-Crick pairings, such as G-G and G-A, can form by using alternate hydrogen-bond-forming sites. (For Activities 1 and 2, pairings other than A-U and G-C are and should be ignored.)

RNA's ability to fold in complex ways causes it to resemble proteins by having secondary and tertiary levels of structure. The determinants for folding these two molecules are quite different, however. Most proteins inside the cell are stabilized in a folded state with hydrophobic amino acid residues sequestered on the inside of the molecule, away from water, and by hydrophilic residues on the outside surface, exposed to the aqueous environment. Formal ionic interactions between amino acid side-chains of opposite charge also contribute to protein stability. Like RNA, proteins have many hydrogen bond interactions that help determine their shape. However, amino acids are not complementary in the manner of nucleotide bases, and thus there are no one-to-one pairing rules for amino acids. This fact makes it difficult to predict protein shapes from amino acid sequence alone. In the case of RNA, the accelerating accumulation of new sequence and crystallographic data makes the prospect look brighter for eventually predicting RNA three-dimensional shape from its sequence alone.

The folded structure of RNA is stabilized primarily by helical regions that form within the molecule based on Watson-Crick base pairing. Such helical regions differ in their degree of thermodynamic stability, depending on the number and nature of base pairs engaged in hydrogen bonding. Base stacking (the interaction between the electron clouds of the planar, cyclic bases when positioned on top of each other like a stack of dinner plates) also contributes to the stability of folded nucleic acids. Other more recently recognized contributors to folding and shape stability include the so-called ribose zipper, a juxtaposition of the minor grooves of two helical regions that is stabilized by hydrogen bonding of the 2′ OH group of ribose; the use of single-stranded loop regions that pair with "receptor" sequences elsewhere in the molecule; and the use of ions such as Mg^{++} to shield the uniform negative charge of the phosphate backbone. We next briefly describe some of the important structural motifs currently known to occur in folded RNAs, with emphasis on their known or proposed functions.

DOUBLE-STRANDED HELICES

With more than 50 percent of its bases in double-stranded form, a typical RNA contains a great deal of secondary structure. As mentioned previously, RNA helices are in the A-form whereas those in DNA are B-form. This difference in helix geometry creates significant differences in surface geography between RNA and DNA. Specifically, the major groove in RNA is quite deep and narrow, and the minor groove is shallow and wide, just the reverse of the B-form DNA helix. This difference in surface topography, along with the specific base sequence of the helix, determines the recognition and binding of other molecules to helical nucleic acids. For example, the shallow, wide, minor groove of RNA appears to be more accessible to protein side-chains and to present more hydrogen bonding opportunities than the major groove. Despite the different shapes of the two grooves, examples of protein binding to RNA appear to involve both to different extent. Because the binding of proteins to helical RNA and DNA can induce local bending and "melting" of base pairs, the limitations of helix groove size can be overcome to some extent. Indeed, there is some evidence that the single-stranded regions of the RNA, instead of the helices, may be more important as actual contact and recognition sites; the helical regions in this case serve to properly orient the single-stranded regions for presentation. Transfer RNA's single-stranded anticodon loop, which recognizes both mRNA and the synthetase enzyme that aminoacylates the tRNA, is a notable example of single-strand recognition ability.

HAIRPIN LOOPS

Like the anticodon loop, many important single-stranded recognition regions in RNA arise as part of a structure called a *hairpin loop*. The loop is created when a single strand of RNA bends back on itself to

form a double-stranded region. This creates a double-stranded stem and a single-stranded loop that caps the helix (see Figure 8). The number and size of hairpin loops vary among different RNA types; for example, the three loops in tRNA have 7–8 bases each, whereas the *Tetrahymena* self-splicing group I intron has six larger loops. So-called tetraloops, which have four unpaired bases atop their helical stem, are common in ribosomal RNAs (rRNAs); one such tetraloop in the 23S rRNA molecule appears to be the ribosome binding site of the toxic proteins ricin and sarin. In the *Tetrahymena* self-splicing intron, a tetraloop, along with its conserved receptor site elsewhere in the molecule, facilitates self-folding of the intron into the proper three-dimensional shape (Cate et al. 1996). Stem-loop structures are also found within the catalytic site of most ribozymes; for example, plant viruslike agents known as viroids contain self-cleaving RNA genomes whose catalytic site adopts the so-called hammerhead structure made up of three stem-loops. Smaller loops, known as *bulges*, are formed within a helical stem rather than at its end; they result when opposed bases are mispaired, causing them to pucker out from the helix.

There are many examples in which different aspects of gene expression, from mRNA transcription and translation to mRNA degradation, employ hairpin loops as control elements. Single-stranded loops may in general be preferred sites for the interaction of RNAs with regulatory molecules. For example, RNA is generally less susceptible to degradation by RNases when it contains a high proportion of hairpin secondary structure, and this can affect RNA's half-life in the cell. A well-studied example of structure controlling RNA degradation is the transferrin receptor mRNA. This mRNA encodes the cell surface receptor responsible for binding the plasma iron transport molecule, transferrin. A feedback mechanism responsive to plasma iron level increases the density of these transferrin receptors on the cell surface when iron levels are low, thereby enabling the cell to more efficiently scavenge transferrin-iron complexes from plasma. Operation of this system at the RNA level involves a regulatory protein, produced when iron levels are low, that binds to the transferrin receptor mRNA at a stem-loop structure near its 3′-end. The bound regulatory protein stabilizes the stem-loop, making the RNA less susceptible to degradation and thus able to be reused to make more copies of the receptor protein.

BASE TRIPLES

Unlike the two-way interaction within a standard pair of bases, a base triple is an interaction between three bases. Base triples are formed when a single-stranded region of an RNA nestles into the major or minor groove of a double-helical segment of the molecule; hydrogen-bonded triplets such as G-C-A can result. These triple-stranded regions help stabilize tertiary, three-dimensional structure and may be essential for certain RNA functions. For example, they have been found at the proposed catalytic regions in ribozymes such as the group I intron of *Tetrahymena*.

PSEUDOKNOTS

Pseudoknots represent a higher order, tertiary level of structure found in RNA. A pseudoknot results when some of the bases in an otherwise single-stranded loop pair with bases located outside that loop. This kind of interaction can potentially form a variety of distinct topologies, but all pseudoknots have two loops and two helical stems, usually with the stems sharing a common axis (see Figure 8). The multiple stems and loops of this RNA conformation provide more complex sites for interaction with proteins. In the 20 years since they were first identified, pseudoknots have been implicated in several examples of the regulation of gene expression. For example, pseudoknot structures in certain mRNAs of retroviruses, bacteria, and yeast appear to stimulate a gene-regulatory phenomenon called *ribosomal frame shifting*. In this process, pseudoknots in the mRNA, along with other specific base sequences, cause the ribosome to stall and slip during translation. The result is a change of reading frame that allows more than one protein sequence to be synthesized from a given mRNA. This is a clear example of an RNA structure that increases the compactness and efficiency of genetic information storage. Another gene regulation example, in bacteriophage T4, involves the binding of a transcription regulating protein to a pseudoknot in the promotor region of the gene 32 mRNA.

Pseudoknot structures also are found at the 3′-end of genomic RNAs of certain bacterial and plant viruses. Most intriguingly, these pseudoknots resemble the stem-and-loop configuration of tRNAs.

The exact function of these tRNA-like ends is unsettled, but they appear to be required for replicating of the virus RNA genome. Similar structures are found in a variety of other RNAs, such as the short molecules used to prime the reverse transcription of retroviral RNA genomes to cDNA copies, and the RNA transcripts made from certain fungal plasmids. Also, the chromosome-capping telomeres at the ends of eukaryotic chromosomes contain a TGG-rich sequence that is potentially able to base pair with the CCA sequence at the ends of these tRNA-like molecules. Weiner and Maizels (1987) have proposed the so-called genomic tag hypothesis, which posits that the tRNA-like sequence evolved early and functioned as a recognition tag that identified certain RNAs as genomes and somehow facilitated their replication. According to this view, the RNAs of similar shape found in today's viruses and cells can be considered molecular fossils—vestiges of a much earlier, RNA-dominated world.

These examples of RNA structure illustrate the general structure-function principle stated earlier: Polymer shape is determined by the linear order of subunits and their interactions within the molecule; three-dimensional shapes essential for the molecule's biological function emerge as a result. The structural motifs discussed in this section give only a hint of RNA's potential for shape diversity. Ongoing structural research employing a range of modern molecular techniques, such as RNA base sequence determination, site-directed mutagenesis, comparative analysis of sequences from different organisms, and advanced methods of X-ray crystallography, will no doubt reveal even more variety. That prospect is made all the more likely by the new RNAs that are being discovered in unexpected locations—and carrying out surprising functions—in the cell. In the next section, we discuss some of these newly recognized RNA functions.

The Diversity of RNA Function

Until recently, most biologists were aware of three types of cellular RNA: the standard trinity of mRNA, tRNA, and rRNA. Research largely within the last 20 years makes it clear that many more RNAs exist. They are found in diverse cellular locations, from the nucleus outward, carrying out a variety of key functions related to gene expression and metabolism. Some of the newly recognized RNAs are highly conserved between species as different as yeast and humans. And some, such as those bearing tRNA-like regions, appear to be related to familiar RNAs. The question of how these diverse and widely distributed RNAs arose and became so thoroughly integrated into cellular economy is central to the RNA world hypothesis.

Everything we know about evolution suggests that it is a conservative process that builds upon existing information to create novel structures and functions. That being so, if the RNA world hypothesis is correct and RNA was indeed the first information-encoding and catalytic entity, then we can predict that vestiges of those early RNA structures and functions, molecular fossils, should be present in contemporary organisms. Indeed, the diversity of RNAs and ribonucleotides in contemporary cells and their widespread involvement in key cellular functions provides much support for this hypothesis (Figure 9).

Most of the newly recognized RNAs share a common functional theme: gene expression. Unexpectedly, the roles of these molecules extend to aspects of gene regulation beyond simple coding and template functions. They play key roles at several levels, from participation in DNA replication and maintenance to regulation of transcription, RNA processing and editing, translation of mRNAs, protein localization within the cell, and modification of protein function. Because some of these RNAs are known to be catalysts, we next highlight some ribozymes and their functions.

RIBOZYMES

Seven categories of naturally occurring ribozymes have been found in the 15 years since their discovery (see Figure 5). They occur in a wide range of organisms, including viruses, bacteria, fungi, and plants; however, the prevalence of ribozymes in today's biological world is presently unknown. The relatively few catalytic roles apparently still left for RNAs may represent evolutionary vestiges, former roles having been taken over through natural selection by the later appearing but structurally more versatile proteins. Indeed, if RNAs were the first biological catalysts, as proposed by the RNA world hypothesis, they were far from the most efficient: Proteins catalyze thousands to millions of times faster. The kinetic data of ribozymes reflects this sluggishness; ribozymes have quite low Km values,

Figure 9 Lines of evidence supporting the RNA world hypothesis.

- RNA is informational *and* catalytic *in vivo*; no other biomolecule has both properties.
- The nucleotide sequences of RNAs common to all organisms (for example, rRNAs) are highly conserved (similar) among the many different species studied, suggesting that RNA was a key molecule present early in evolution.
- RNA or ribonucleotides are involved in most critical cellular functions in all three domains of life:
  - Adenosine triphosphate (ATP) is a universal energy carrier.
  - Universal metabolic pathways employ adenine nucleotide coenzymes (NADH, NADPH, FAD, CoA).
  - Protein synthesis employs mRNAs, rRNAs, and tRNAs.
  - rRNA by itself can catalyze peptide bond formation.
  - DNA synthesis requires the prior conversion of ribonucleotides to their deoxy form.
  - The ribonucleotide uracil, found only in RNA, is the precursor for DNA's thymine.
  - RNA is the primer for DNA replication.
  - Ribonucleotide derivatives function as key signaling molecules in the cell (for example, cAMP, ATP).
- RNAs function as primers in DNAreplication and in reverse transcription of retroviral genomes.
- tRNA-like molecules are involved in nontranslational (nonprogrammed) polymerizations (for example, cell wall synthesis, polypeptide antibiotic synthesis).
- A tRNA-like molecule may have given rise to the RNA component of telomerase, the enzyme that maintains the ends of chromosomes.
- Enzymatic processing of mRNAs involves other small RNAs (snRNPs, RNase P).
- Protein sorting into the endoplasmic reticulum of all eukaryotes involves RNA(SRP-RNA).
- Ribonucleotides are used to activate and carry sugars during polysaccharide synthesis.

indicating very high affinity for their substrates and easy saturation. These kinetics seem well suited to ribozyme function, however; unlike protein enzymes, they typically catalyze only one reaction cycle (for example, their own removal from a larger molecule). Only two catalytic reaction mechanisms are known for naturally occurring ribozymes, transesterification and hydrolysis, and both employ OH groups as nucleophiles for cleavage of the RNA phospohodiester backbone. Notable examples are the coupled endonuclease-ligase reactions involved in splicing, and the coupled endonuclease-phosphatase phosphotransfer reactions that remove a substrate 3′ P and transfer it to the ribozyme.

Among the best-studied catalytic RNAs are the group I and group II introns, which autocatalyze their own excision from a larger precursor RNA and ligate the flanking exons. These introns are encoded in the genomes of a wide assortment of organisms, including bacteria, fungi, and plants. In eukaryotes, they are more commonly found in organelle genomes, such as fungal and plant mitochondria, and plant chloroplasts.

Group I and group II introns are rich in stem-loop secondary structure, particularly in their catalytic regions, and their higher order tertiary shape (at least in the case of the group I introns) appears to be stabilized by divalent cations like Mg^{++}. Most group I introns do not appear to require the aid (catalytic or otherwise) of proteins to self-splice. In contrast, the splicing of group II introns appears to benefit, both *in vitro* and *in vivo*, from the aid of maturase proteins that presumably help stabilize the correct tertiary structure of the intron required for self-splicing. Interestingly, at least some of these maturase proteins are encoded within the sequence of the intron itself; the intron carries coding "software" needed to make "hardware" that facilitates its catalytic role. Likewise, certain mobile introns capable of transposing to new genomic locations encode proteins, such as endonuclease and reverse transcriptase, that facilitate their movement. Enzymatic and coding abilities such as these are a far cry from the view held until quite recently that introns are nonfunctional genetic baggage or "junk DNA."

Unlike the autocatalytic processing of group I and group II introns, the processing of most eukaryotic mRNA precursors does not involve self-splicing introns. These introns instead require for their removal the actions of complex RNA-protein assemblies in the nucleus, known as spliceosomes. Spliceosomes are in effect macromolecular "splicing machines" and are reminiscent of ribosomes in being RNA-protein assemblies. The relative roles of the RNAs and proteins of spliceosomes, that is, which are catalysts and which are the structural elements, are not yet known. However, the catalytic

mechanism of splicing by spliceosomes resembles that of a group II intron, leading to speculation that the several spliceosomal RNAs may be an "intron-in-pieces": fragmented descendants of a once self-splicing intron.

RNA EDITING

Another level of gene expression regulation that involves the functional RNAs is the process of RNA editing. During editing, particular bases in the precursor RNA are added, deleted, or changed following transcription. The resulting edited RNA has a base sequence different from that encoded in the DNA and transcribed RNA precursor. Most examples of RNA editing involve eukaryotic mRNA precursors, and when these are edited the mature mRNA specifies a different protein product than that encoded in the gene. RNA editing thus increases the number and diversity of products from a given gene. As with gene regulatory processes like alternative splicing and *trans*-splicing, RNA editing effectively increases the information coding capacity of a genome. Although the detailed chemical and catalytic mechanisms of editing are not yet known, one version of the process employs small RNAs known as *guide RNAs*. The base sequences of the guide RNAs enable them to pair with the pre-edited target RNA and specify the bases to be added, deleted, or altered by complementary pairing.

Two additional examples illustrate the variety of RNA effects on gene expression. During the translation process of protein synthesis, a peptidyl transferase catalytic activity covalently links incoming amino acids to the growing polypeptide. This catalytic activity is known to reside in the large ribosomal subunit, which in *E. coli* consists of some 31 proteins and 2 distinct RNA molecules. Which of these ribosome components catalyzes peptide bond formation has long been a mystery. The conventional assumption was that it must be one of the proteins. Recent results, however, strongly suggest that the 23S rRNA component is the catalyst. Noller et al. (1992) extracted more than 95 percent of the proteins from the large subunit, leaving the 23S rRNA, and the peptidyl transferase activity, intact; treatments that damage the RNA eliminate the activity. More recently, Nitta et al. (1998) demonstrated that cloned segments of the 23S rRNA never exposed to ribosomal proteins could be reconstituted in the test tube and could catalyze peptide bond formation. The emerging picture is that of an RNA catalyst in the lead role of a reaction essential for life, with ribosomal proteins playing structural supporting roles. This view is consistent with the notion that complex ribosomes evolved when functions once carried out solely by RNAs were improved upon by the addition of proteins.

RNAs also play a role in the cellular localization of newly made proteins. In eukaryotes, a complex translocation machinery enables nascent polypeptides bearing an amino-terminal "signal sequence" to cross into the lumen of the endoplasmic reticulum (ER). A key element of this translocation machinery is the so-called signal recognition particle (SRP), a cytoplasmic RNA-protein assembly consisting of a core 7S RNA, to which six different protein components bind. The signal recognition particle binds to the free-floating ribosome-nascent peptide complex. It then carries the complex to the surface of the ER membrane, where the SRP binds to a receptor. The SRP in effect acts as a bridge to tether the ribosome-peptide to the surface of the ER, and thus facilitates movement of the newly made polypeptide into the lumen. The 7S RNA is necessary for SRP function and may be the component that binds directly to the ER membrane receptor protein.

RNA and Evolution: Molecular Evolution in a Test Tube

A basic assumption of the RNA world hypothesis is that early in the evolution of life, one or a few molecules came to dominate the pool of RNAs that had been randomly generated by nonbiological processes. Assuming that RNA's appearance preceded that of catalytic proteins, the RNA must have been able to replicate itself in order for evolution at the molecular level to get under way. The demonstrated catalytic ability of RNA makes the possibility of self-replication much more plausible. Even though no RNA capable of catalyzing its own replication has yet been found in nature, recent experiments in directed evolution (molecular evolution in a test tube) have demonstrated that RNA does indeed have this potential.

Sol Spiegelman and his group first demonstrated evolution and selection of RNA molecules in a test tube in the 1960s. Spiegelman's *in vitro* experiments

started with a mixed population of RNA genomes from the bacteriophage Qβ, which were transferred serially to a series of tubes supplying only ribonucleotides and replicase (the RNA polymerase that copies the viruses' RNA genome inside host cells). By limiting the time available for replication in each tube, the experiments imposed speed of RNA replication as the selection criterion. Several cycles of *in vitro* replication and transfer led to variant RNAs that could be copied at greatly increased speed. By altering the physical or chemical selection conditions, populations of RNA adapted to the imposed conditions came to predominate in the mixture.

The modern era of *in vitro selection* (also referred to as directed evolution) began in 1989–90 and was made possible by technical advances in RNA and DNA synthesis (for example, automated oligonucleotide synthesis), nucleic acid amplification (for example, polymerase chain reaction), and selection methods (for example, affinity chromatography). Starting with a large synthetic population (pool) of randomly varying RNA or DNA molecules, the goal of *in vitro* selection experiments is to amplify those variants that are able to meet some experimentally imposed selection criterion, such as the ability to bind to a particular target molecule or catalyze a particular chemical reaction. After 10 to12 cycles of selection and amplification with mutation, molecules well adapted to the selection criterion predominate in the pool (Figure 10).

A parallel can be drawn between the variation, selection, and amplification aspects of these *in vitro* experiments and natural, biological evolution. However, the action of selection in these experiments is more direct than in biological evolution: In organisms, molecular structure-function is selected somewhat indirectly through a complex organismal phenotype; with *in vitro* methods the molecule's base sequence, which directly determines its function, *is* the selected phenotype. It might be said that the generation of highly functional molecules from a random assemblage of sequences is akin to a tornado assembling a 747 from random parts, creationist views notwithstanding.

The RNA and DNA molecules found by this powerful experimental approach tell us a lot about the potential range of nucleic acid functions and their evolutionary potential. In fewer than 10 years, the method has uncovered many new synthetic ribozymes, supplementing the seven known natural classes. Significantly, one of these synthetic RNAs can copy an RNA template, forming short complementary strands. Further laboratory refinements of this molecule may yield a bona fide RNA-dependent RNA polymerase, a key ribozyme activity required for self-replication and assumed in RNA world ideas about the origin of life.

In vitro evolution experiments selecting for *DNA* molecules have shown that single-stranded DNA, like RNA, has catalytic potential. Perhaps the distinction

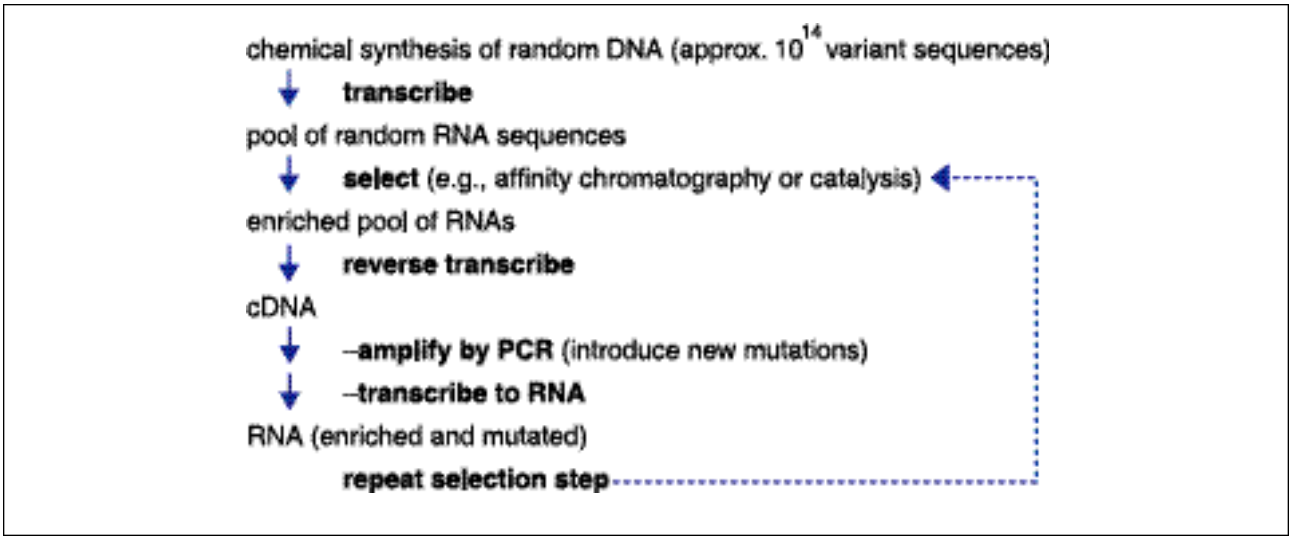


Figure 10 The strategy of directed evolution *in vitro*. Starting with a large, randomly synthesized population of nucleic acid, repeated cycles of selection followed by amplification can yield a particular sequence well suited to the selection criterion.

between RNA and DNA is not as great as we have come to believe. Recall that the only chemical difference is one extra oxygen atom in RNA (in the 2′ OH group of ribose). Also, both RNA and single-stranded DNA can fold into complex shapes. If there really is little difference between these two closely related molecules, why is DNA today the storage form of most genetic information, and why is RNA the active form performing a variety of cellular tasks? The current thinking is that DNA appeared later in the RNA world, most likely as an RNA derivative, and DNA's succession to the role of information repository may be related to its greater chemical stability (due to its fully double-stranded configuration and lack of a 2′ OH nucleophile). Or perhaps DNA proved to have too small a range of catalytic ability to out-compete RNA in a functional role. It remains an intriguing question whether life could have evolved equivalently if the earliest catalytic events had been DNA-based and RNA had appeared later as the storage form.

RNA, Health, and Disease

Many viruses have RNA genomes. Some are important human pathogens: polio, HIV, flu, and measles, among others. Numerous examples are also found among plant viruses, several of which have major economic impacts in agriculture and forestry. A common feature of all RNA viruses is a high level of variation among their genome sequence, the result of a high mutation rate (10^{-3}–10^{-4} per base pair per replication). The frequency with which mutations arise in a population is determined in part by how often the organism reproduces. Viruses replicate millions of times each day, so random mutations are constantly arising. Another factor affecting the frequency of mutation is the fidelity of the replication process itself. The replicase enzyme that copies RNA genomes occasionally makes random errors, inserting the incorrect monomer (for example, A opposite G, or U opposite C). Indeed, the polymerases that copy DNA also make errors, but cells have evolved a molecular quality-control, proofreading mechanism that can correct mistakes most of the time, keeping the DNA mutation rate low (less than 10^{-7}). RNA lacks such a correction mechanism, a condition that may be related to its more limited role as a genome in nature. Although populations of RNA viruses have greater variation in genome sequence than DNA viruses, variation is essential to the adaptation and evolution of all viruses, indeed, all life forms.

Because of this high mutation rate, the many progeny viruses made in an infected cell constitute a population of variants, with typically one to three changes in each viral genome. The genome sequence of a population of RNA viruses thus is not unique, but rather is a population of variant "quasispecies." Another source of genetic variation in viruses is their ability to exchange whole blocks of genes among progeny by the process of recombination. Most of the genome alterations are deleterious, and those viruses will not survive. Other genome changes, however, can confer advantages for different aspects of viral behavior: faster replication, increased virulence, or decreased sensitivity to antiviral drugs.

In the case of HIV, it has been estimated that each of its 10,000 RNA bases is mutated more than 10,000 times each day in an infected person (Coffin, 1995). HIV's particularly rapid mutation rate and the chronic nature of the infection make it possible to detect the emergence within an individual patient of new mutant strains during the course of the disease. The virus thus can evolve within the microcosm of a single human host. In similar ways, new strains of cold and flu viruses continue to emerge and plague humankind.

With both RNA and DNA viruses, the war between the hosts and the viruses that infect them is sophisticated, subtle, and evolving: The host inactivates the virus or limits its propagation, the virus subverts the cell to make more virus. The battle shifts back and forth with move and countermove. A critical move by the host is to mount an immune response against the virus. Unfortunately, the RNA virus's ability to rapidly generate genetic diversity produces an ever-moving target for the immune system, as well as for developers of vaccines and antiviral drugs. In the case of HIV, the virus has evolved an additional powerful strategy to ensure its persistence in the host, that of attacking and disabling the immune system itself.

Some viruses subvert cellular defenses by employing RNA molecules as weapons. The cellular interferon response is one example. When infected with virus, cells typically react by secreting the potent signaling molecule interferon. This protein protects neighboring cells from becoming victims by temporarily shutting down their protein-synthetic machinery. This clever cellular defense strategy

denies the virus access to the one source of new proteins needed to complete its life cycle. Adenovirus, however, has evolved a more clever way to undermine this defense: Specialized viral RNA molecules (for example, VAI RNA) effectively block interferon's action and prevent the shutdown of cellular protein synthesis. The cell has no option but to contribute slavishly to the production of viral progeny.

Viruses can likewise defeat medical therapies. Although vaccines have been dramatically successful in curtailing some viral diseases, such as polio, they are less effective at combating the rapidly evolving RNA viruses, such as rhinovirus (common cold), influenza, and HIV. For these, the current approach is to develop antiviral drugs, chemical agents that either block entry of the virus into host cells or, once inside, block key steps in the viral reproduction cycle. Current treatment of HIV, for example, employs drugs such as Acyclovir, AZT, and protease inhibitors. By targeting different aspects of viral reproduction, these drugs are proving effective at slowing the course of HIV infection.

Evolutionary considerations, however, dictate that the *way* in which these drugs are administered is a key factor in their long-term effectiveness. Administered singly, any given drug soon loses effectiveness as the rapidly mutating virus generates a population that is resistant. If the drug is replaced by another in an attempt to hit the remaining resistant population, that population will in turn give rise to a new population that is resistant to both drugs . . . and so it goes. The unintended but predictable outcome of this type of serial drug treatment is the eventual creation of a viral population that is resistant to several drugs. Such a multiply resistant virus could spread quickly in the host population. A much better strategy, one that makes rational use of evolutionary principles, is to use all three drugs in combination. The key to this approach is that each drug *independently* and *simultaneously* hits a different step in the virus's reproduction cycle. For a triply resistant viral population to emerge in the presence of all three drugs, three independent resistance mutations would have to occur together in a single founder *virus*, a much less probable occurrence.

The alternative to devising separate drug treatments for each different type of virus is to design a broad-spectrum antiviral agent that is effective against many viruses. Unfortunately, because viruses are dependent on host cell machinery (such as that for protein synthesis), any drug that interferes with a key synthetic step common to all viruses likely also would inhibit cellular function. The situation is different for bacteria, which have evolved enough differences at the molecular level from eukaryotic cells, including RNAs, that it is possible to develop drugs that specifically target bacteria. Agents such as penicillin, which impairs synthesis of the unique molecular outer layers of bacterial cells, were dramatically effective when first introduced. Also, the protein synthetic apparatus of bacteria is sufficiently different in molecular detail from that of eukaryotes that it is a good target for antibiotics. Some of these antibiotics inhibit bacterial protein synthesis by binding directly to RNA components of the bacterial ribosome. Structural differences between bacterial and nonbacterial rRNAs account for the specificity of the antibiotic in this case. Mutations that change the structure of the bacterial rRNAs, such that the antibiotics no longer bind, confer resistance on the pathogen. Once again, we see that the ongoing process of evolution, fueled by mutations such as these in the bacterial and viral populations, ultimately undermine the effectiveness of even the most powerful therapeutic agents in our arsenal. Activity 4 of the module explores RNA's role as a target for certain antibiotics and antiviral agents, as well as the acquisition of resistance.

Two novel strategies focusing on RNA with potential application against viral pathogens are currently being explored in the laboratory: *in vitro selection* and *antisense technology*. The former approach employs the *in vitro* selection strategy described above to select functional nucleic acid molecules that bind to and inactivate viral components. Ribozymes able to cleave DNA have been developed, as have deoxyribozymes able to cleave RNA. Looking ahead, it may one day be possible to engineer into the genome of affected cells a ribozyme sequence that has been selected *in vitro* for its ability to inactivate the RNA, DNA, or protein of a viral pathogen. The coding sequence for a ribozyme having, for example, anti-HIV activity might be incorporated into a population of the patient's immune cells. Subsequent expression of the ribozyme "bullet" in these cells and their descendants conceivably could have significant therapeutic effects against the virus.

Another potential application of *in vitro*-selected nucleic acids is as ligand-binding molecules; such molecules could be useful as biochemical reagents or as potential therapeutic agents. The term *aptamer* describes *in vitro*-selected RNAs or DNAs that have the ability to bind with specificity to another molecule (aptamers, unlike ribozymes, are not catalytic). Aptamer sequences with unique shapes have been developed to bind a wide range of ligands, including small organic dyes, coenzymes, amino acids, vitamins, and viral proteins. The binding specificity of aptamer RNAs is high enough that they can distinguish between ligands as similar as theophylline and caffeine, which differ by only a single methyl group. These designer nucleic acids promise to be useful additions to the growing list of functional biomolecules, including protein enzymes and antibodies, that already have been generated by *in vitro* techniques.

The second RNA-related approach to therapeutics, *antisense technology*, attempts to inactivate and neutralize unwanted mRNAs that derive from a mutated target gene. This approach takes advantage of the tendency of single-stranded RNAs to bind to sequences complementary to themselves. The strategy is to engineer into affected cells an antisense copy of the defective gene, that is, one in which the 5′-3′ orientation of the gene has been reversed relative to its promotor. Insertion of the inverted gene copy creates a situation in which RNA transcripts from the inverted gene, known as *antisense transcripts*, arise from what is normally its nontranscribed coding strand. The antisense transcripts are thus complementary to the sense transcripts from the defective gene, which arise from the opposite, template strand. Antisense and sense RNA molecules can bind and neutralize one another, effectively silencing the expression of the defective target gene, perhaps a provirus or an oncogene. Successful therapeutic applications of both the *in vitro* selection and antisense approaches await advances in gene therapy technologies needed to incorporate engineered sequences into the genome in a functional state.

References and Related Literature

Abbott, I., et al. (1977). Comparative ecology of Galápagos ground finches (*Geospiza* Gould): Evaluation of the importance of floristic diversity of interspecific competition. *Ecological Monographs*, *47*: 151–184.

Cairns-Smith, A.G. (1985). *Seven clues to the origin of life*. New York: Cambridge University Press.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R., & Doudna, J.A. (1996). RNA tertiary structure mediation by adenosine platforms. *Science, 273*: 1696–1699.

Cech, T.R. (1998). Structure and mechanism of the large catalytic RNAs: Group I and group II introns and ribonuclease P. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 239–269). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Cleland, A., Watson, H.G., Robertson, P., Ludlam, C.A., & Brown, A.J.L. (1996). Evolution of zidovudine resistance-associated genotypes in human immunodeficiency virus type 1-infected patients. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 269*: 6–18.

Coffin, J.M. (1995). HIV population dynamics *in vivo*: Implications for genetic variation, pathogenesis, and therapy. *Science, 267*: 483–489.

Ekland, E.H., Szotak, J.W., & Bartel, D.P. (1995). Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science, 269*: 364–370.

Eldredge, N. (1995). *Reinventing Darwin: The great debate of the high table of evolutionary theory.* New York: John Wiley & Sons, Inc.

Ellington, A.D., & Joyce, G.F. (1990). *In vitro* selection of RNA molecules that fold into specific ligand-binding ligands. *Nature, 346*: 818–822.

Gesteland, R.F., & Atkins, J.F. (Eds.). (1998). *The RNA world* (2nd ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Gilbert, W. (1986). The RNA world. *Nature, 319*: 618–620.

Howard Hughes Medical Institute Holiday Lecture on Science. (1995). *The Double Life of RNA* [videotape set]. (A four-tape series of lectures by Nobelist Thomas Cech, discussing his discovery of ribozymes and other important functions of RNA. Summaries of the lectures can be found online at http://www.hhmi.org/.)

Jenison, R.D., Gill, S.C., Pardi, A., & Polisky, B. (1994). High resolution molecular discrimination by RNA. *Science, 263*: 1425–1429.

Joyce, G.F. (1992, December). Directed molecular evolution. *Scientific American, 267*(6): 90–97.

Joyce, G.F., & Orgel, L.E. (1993). Prospects for understanding the origin of the RNA world. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 1–25). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Landweber, L.F., Simon, P.J., & Wagner, T.A. (1998). Ribozyme engineering and early evolution. *BioScience, 48*(2): 94–103.

Mills, D.R., Peterson, R.L., & Spiegelman, S. (1967). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proceedings of the National Academy of Sciences, 58*(1): 217–224.

Nitta, I., Kamada, Y., Noda, H., Ueda, T., & Watanabe, K. (1998). Reconstitution of peptide bond formation with *Escherichia coli* 23S ribosomal RNA domains. *Science*, *281*: 666–669.

Noller, H.F., Hoffarth, V., & Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction methods. *Science, 256*: 1416–1419.

Orgel, L.E. (1994, October). The origin of life on the earth. *Scientific American, 271*(4): 77–83.

Simons, R.W., & Grunberg-Manago, M. (Eds.). (1998). *RNA structure and function*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Smith, J.M., & Szathmary, E. (1995). *The major transitions in evolution*. Oxford: W.H. Freeman (Spektrum).

Szostak, J.W., & Ellington, A.D. (1993). *In vitro* selection of functional RNA sequences. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 511–533). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Unrau, P.J., & Bartel, D.P. (1998). RNA-catalyzed nucleotide synthesis. *Nature, 395*: 260–263.

Weiner, A.M., & Maizels, N. (1987). 3′ Terminal tRNA-like structures tag genomic RNA molecules for replication: Implications for the origin of protein synthesis. *Proceedings of the National Academy of Sciences, 84*: 7383–7387.

Wright, M.C., & Joyce, G.F. (1997). Continuous *in vitro* evolution of catalytic function. *Science*, *276*: 614–617.

Wyatt, J.R., & Tinoco, I. (1993). RNA structural elements and RNA function. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 465–496). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Zaug, A.J., McEvoy M.M., & Cech, T.R. (1993). Self-splicing group I intron from *Anabaena* pre-tRNA: Requirements for base-pairing of the exons in the anticodon stem. *Biochemistry, 32*: 7946–7953.