CHAPTER

18 Regulatory RNAs

OUTLINE

WE DISCUSSED IN THE PREVIOUS TWO CHAPTERS how transcription is regulated in prokaryotes and eukaryotes. We learned that this control is achieved using regulatory proteins—typically, sequence-specific DNA-binding proteins that either activate or repress transcription of nearby genes. The mechanistic details of gene regulation have been studied since François Jacob and Jacques Monod proposed their model of repression almost 50 years ago (Box 16-2). At that time, they could not say whether the *trans* factors (repressors) were proteins or RNA. It transpired that in the cases they studied (and indeed most other cases), the regulators were proteins that worked by binding the operator sites on DNA. But in their original paper they suggested that the regulators could just as easily be RNA molecules—indeed, they favored that possibility.

The idea that RNA molecules might be regulators was largely forgotten as more and more protein regulators were found in both prokaryotes and eukaryotes. But in recent years, there has been an explosion in the study of RNA regulators, particularly in eukaryotes, that operate at the level of transcription and especially translation. This new field emerged from two sources: the discovery of microRNAs, first reported in the early 1990s, and then the discovery of the phenomenon known as RNA interference in the late 1990s. Before we describe these forms of regulation—how they work and the applications they afford researchers—we consider cases of RNA-mediated gene regulation first described in bacteria.

REGULATION BY RNAs IN BACTERIA

Small RNAs have been recognized in prokaryotes for many years. Some are involved in regulating the replication of plasmids, and others are involved in regulating gene expression (see the discussion of Tn*10* in Chapter 11). Of the latter group, some of these RNAs control transcription—the 6S RNA of *Escherichia coli,* for example. This RNA binds to the σ^{70} subunit of RNA polymerase and down-regulates transcription from many σ^{70} promoters. The 6S RNA accumulates at high levels in stationary phase (the growth phase bacteria enter as nutrients become depleted and the cells stop dividing; see Chapter 22). In stationary phase, an alternative σ factor, σ^{S}, is made. This σ com-

petes with σ^{70} for core polymerase and directs the enzyme to promoters expressing genes for the multiple stress responses needed to survive stationary phase. By down-regulating transcription from σ^{70} promoters, 6S RNA helps this shift in expression to σ^{S} promoters.

In recent years, attention has focused on another group of small RNA molecules in bacteria that regulate translation and mRNA degradation. Interest in these small RNAs has been heightened by their similarity to RNAs that regulate gene expression in eukaryotes—the small interfering and microRNAs we discuss in the second half of this chapter. The bacterial small RNAs (called **sRNAs**) are larger (80–110 nucleotides) than those regulatory RNAs from eukaryotes (which range from 21 to 30 nucleotides). And they are not generally formed by processing of larger dsRNA precursors (as those eukaryotic RNA regulators are); instead, they are encoded in their final form by small genes. Many of these genes have been identified by bioinformatics, with close to 100 sRNAs being uncovered in *E. coli*. Of these, about a dozen have been characterized. Most sRNAs work by base pairing with complementary sequences within target mRNAs and directing destruction of the mRNA, inhibiting its translation or even in some cases *stimulating* translation.

Binding of an sRNA to its target mRNA is in most cases aided by the bacterial protein Hfq. This RNA chaperone is needed because the complementarity between the the sRNAs and their target mRNAs is typically imperfect and short, and so their interaction is weak. Hfq facilitates base pairing; also, by binding the sRNAs even before they are paired with their targets, Hfq increases the stability of these regulators.

A well-studied sRNA from *E. coli* is the 81-nucleotide RybB RNA. This sRNA binds several target mRNAs and triggers their destruction because the double-strand stretch of heteroduplex formed upon pairing is recognized as a substrate by the nuclease RNase E. Most of the mRNAs targeted by RybB encode iron storage proteins. Free iron is required by the cell under certain circumstances, but high levels are toxic. RybB regulates the levels of free iron by controlling the levels of iron storage proteins. RybB is expressed from a promoter recognized by a special σ factor called σ^{E} (like σ^{S}, a stress response σ factor). Expression of the gene encoding σ^{E} is itself regulated by RybB, and so this sRNA is part of an autonegative regulatory loop for σ^{E}.

The stationary-phase σ factor σ^{S}, mentioned above, is encoded by the *rpoS* gene of *E. coli*. Translation of *rpoS* mRNA is stimulated by two sRNAs: DsrA and RprA. Activation is achieved by a switch in alternative RNA base pairing: the small RNAs bind to a region of the mRNA that otherwise would pair with the ribosome-binding site, inhibiting translation. The *rpoS* gene is also acted on negatively by another small RNA, OxyS. Figure 18-1 shows these two mechanisms.

We shall return to *trans*-acting regulatory RNAs in the second half of this chapter where we consider their role in regulating gene expression in eukaryotes. But before turning to that topic, we consider other examples in bacteria of gene regulation mediated through alternative RNA pairing that operate in *cis*—that is, RNA regulatory elements that control expression of the genes within whose mRNAs they reside. The most striking examples are the so-called **riboswitches** that control metabolic operons and **attenuation** in biosynthetic operons. The *trp* genes of *E. coli* are the classic example of the latter mechanism and are where RNA-mediated regulation was discovered (we shall describe this case in detail in Box 18-1).

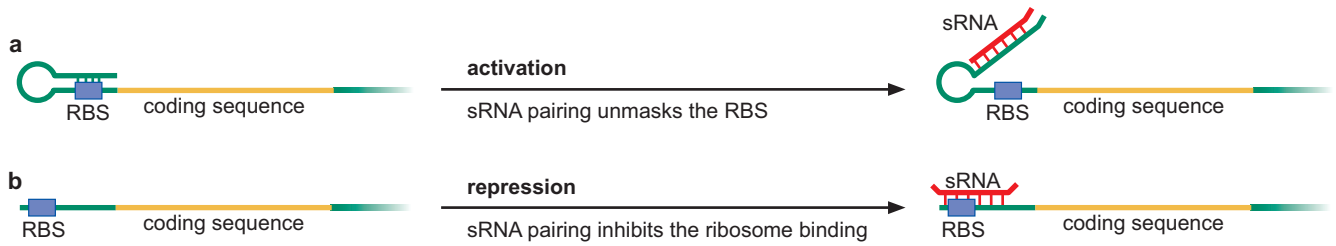


FIGURE 18-1 Activation and repression of translation by sRNAs. RBS indicates the ribosome-binding site. When this is occluded by base pairing with another RNA molecule—as in part b—or another region of the same RNA molecule—as in part a—translation is inhibited. (Adapted, with permission, from Gottesman S. et al. 2006. *Cold Spring Harbor Symp. Quant. Biol.* 71: 1–11, Fig. 1. © Cold Spring Harbor Laboratory Press.)

Riboswitches Reside within the Transcripts of Genes Whose Expression They Control through Changes in Secondary Structure

Riboswitches control gene expression in response to changes in the concentrations of small molecules. These regulatory elements are typically found within the 5′-untranslated regions of the genes they control, and they can regulate expression at the level of transcription or translation. They do this through changes in RNA secondary structure, as we shall see.

Each riboswitch is made up of two components: the **aptamer** and the **expression platform** (Fig. 18-2). The aptamer binds the small-molecule ligand and, in response, undergoes a conformational change, which in turn causes a change in the secondary structure of the adjoining expression platform. These conformational changes alter expression of the associated gene by either terminating transcription or inhibiting the initiation of translation. Both mechanisms are illustrated in the example shown in Figure 18-3, and which we now describe.

Riboswitches are, not surprisingly, typically found upstream of genes involved in the synthesis of the metabolite ligand recognized by the riboswitch in question. For example, in *Bacillus subtilis*, many genes involved in the use of the amino acid methionine have a 200-nucleotide-long untranslated leader RNA that acts as a SAM (*S*-adenosylmethionine)-sensing riboswitch. RNA polymerase initiates transcription at the promoter and transcribes through this leader region before entering the coding sequence of the downstream genes. Once transcribed into RNA, the leader region can adopt alternative structures through alternative patterns of intramolecular base pairing

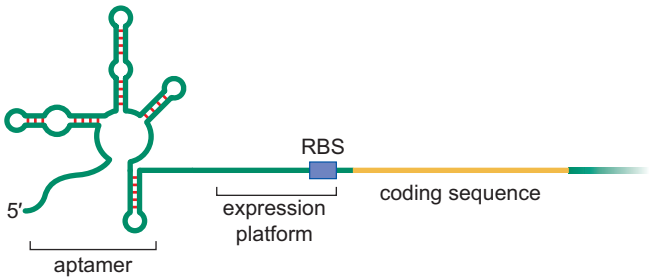


FIGURE 18-2 Organization of riboswitch RNAs. As described in the text, the aptamer binds the controlling metabolite, causing changes in the structure of the adjoining expression platform. The aptamers identified to date vary in size from 70 to 200 nucleotides; the expression platforms vary more in both size and character.

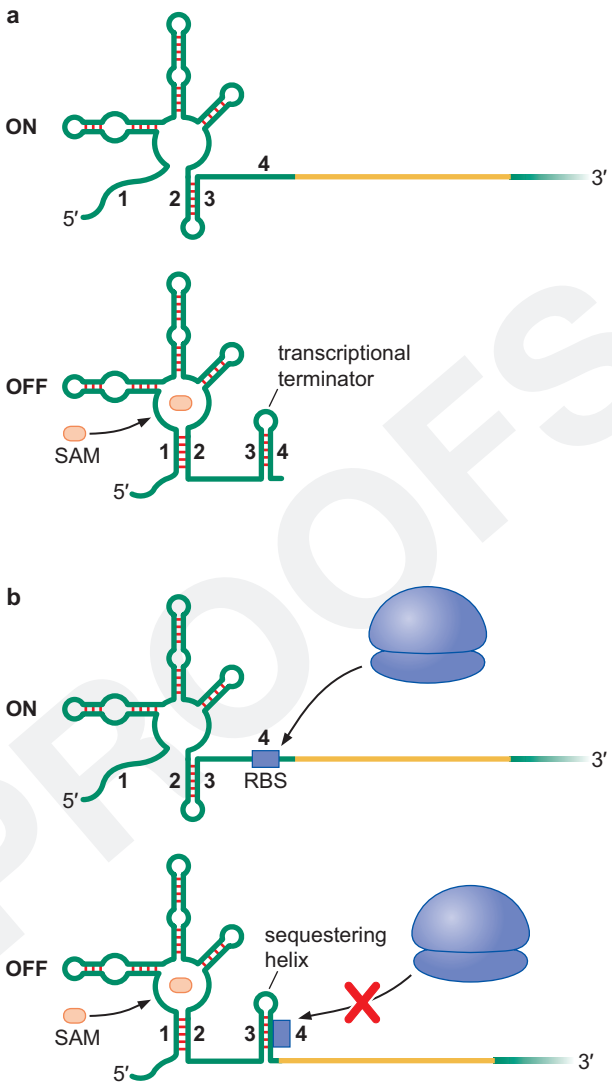


FIGURE 18-3 Riboswitches regulate transcription termination or translation initiation. Shown are two examples of a SAM-sensing riboswitch, in one case (a) regulating transcription termination, in the other (b) translation initiation. Numbers 1–4 indicate different sequence elements within the RNA upstream of the coding region (which is shown in yellow). In the absence of SAM, regions 2 and 3 form a stem loop; in the presence of SAM, regions 1 and 2 form a stem loop, and regions 3 and 4 do likewise. The consequence of that change in secondary structure controls transcription or translation as shown. (a) A stem loop of region 3 and 4 produces a transcriptional terminator, which triggers RNA polymerase to terminate transcription immediately after transcribing those regions and before entering the downstream coding region. The stem loop in this case is followed by a stretch of Us in the mRNA, another feature of transcriptional terminator (Fig. 12-12). (b) The stem loop formed by regions 3 and 4 inhibits translation initiation by sequestering the ribosome binding site, as shown.

(Fig. 18-3a). One arrangement includes a stem-loop transcriptional terminator (see Chapter 12). SAM—the ligand for this riboswitch—binds to the aptamer and stabilizes the secondary structure that includes this transcriptional terminator (as shown in the bottom part of Fig. 18-3a). Under these circumstances, transcription is terminated before polymerase has a chance to transcribe the downstream protein-coding segment of the gene. This form of transcriptional regulation is also called **attenuation**. In another case—at another gene—a SAM-sensing riboswitch can work by regulating translation. In that case, as shown in Figure 18-3b, the alternative secondary structure stabilized by SAM binding to the aptamer includes a stem loop that, although not a transcriptional terminator, does include the ribosome-binding site (RBS; within region 4). This conformational change sequesters the RBS and blocks ribosomes from initiating translation. This form of translation inhibition is thus essentially identical to that described for *trans*-acting sRNAs above (Fig. 18-1). The details of the changes in RNA secondary structure induced by SAM binding to a riboswitch are shown in Figure 18-4.

Many riboswitches have been identified and these respond to a range of different metabolites. Ligands include lysine and other amino

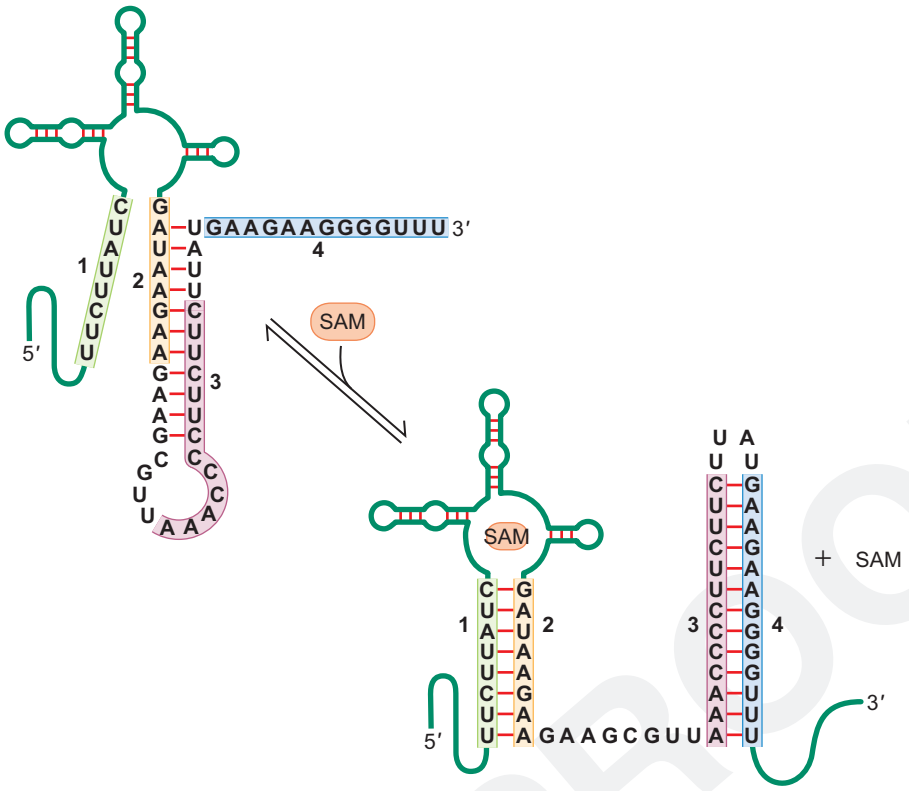


FIGURE 18-4 Changes in secondary structure of a SAM-sensing riboswitch. The sequences of regions 1–4 (described in Fig. 18-3) are here shown in detail and color-coded. The base pairing found in the two alternative secondary structures—that is, with and without SAM bound—are shown. (Adapted from Winkler W.C. et al. 2003. *Nat. Struct. Biol.* 10: 701–707, Fig. 5b. © Macmillan.)

acids, vitamin B12, coenzyme thiamine pyrophosphate (TPP), flavin mononucleotide (FMN), and guanine (Fig. 18-5).

Another kind of riboswitch responds to uncharged tRNAs, rather than to small-molecule ligands. Thus, certain genes, notably genes for aminoacyl-tRNA synthetases (see Chapter 14), are controlled by attenuation mediated by a 200–300-nucleotide-long, untranslated, leader RNA that directly and specifically interacts with the cognate, uncharged tRNA for the synthetase; the charged form of the tRNA does not fit in the binding pocket provided by the RNA secondary structure. Binding of uncharged tRNA stabilizes the leader RNA in its antitermination structure so that transcription into the adjacent synthetase gene can proceed. Specificity is achieved in part by a "codon–anticodon" interaction between the tRNA and the leader RNA. Because uncharged (but not charged) tRNA can bind to the leader, transcriptional readthrough is only stimulated when the cognate amino acid is in short supply and the level of uncharged tRNA in the cell rises.

Although most prevalent in bacteria, riboswitches are found in other organisms as well, including archaea, fungi, and plants. In some cases in these higher organisms, riboswitches are even involved in controlling alternative splicing (Chapter 13). Thus, for example, in a recent case described in the fungus *Neurospora crassa*, three TPP aptamers were identified, two of which inhibited, and the third stimulated, expression of genes through regulation of RNA splicing.

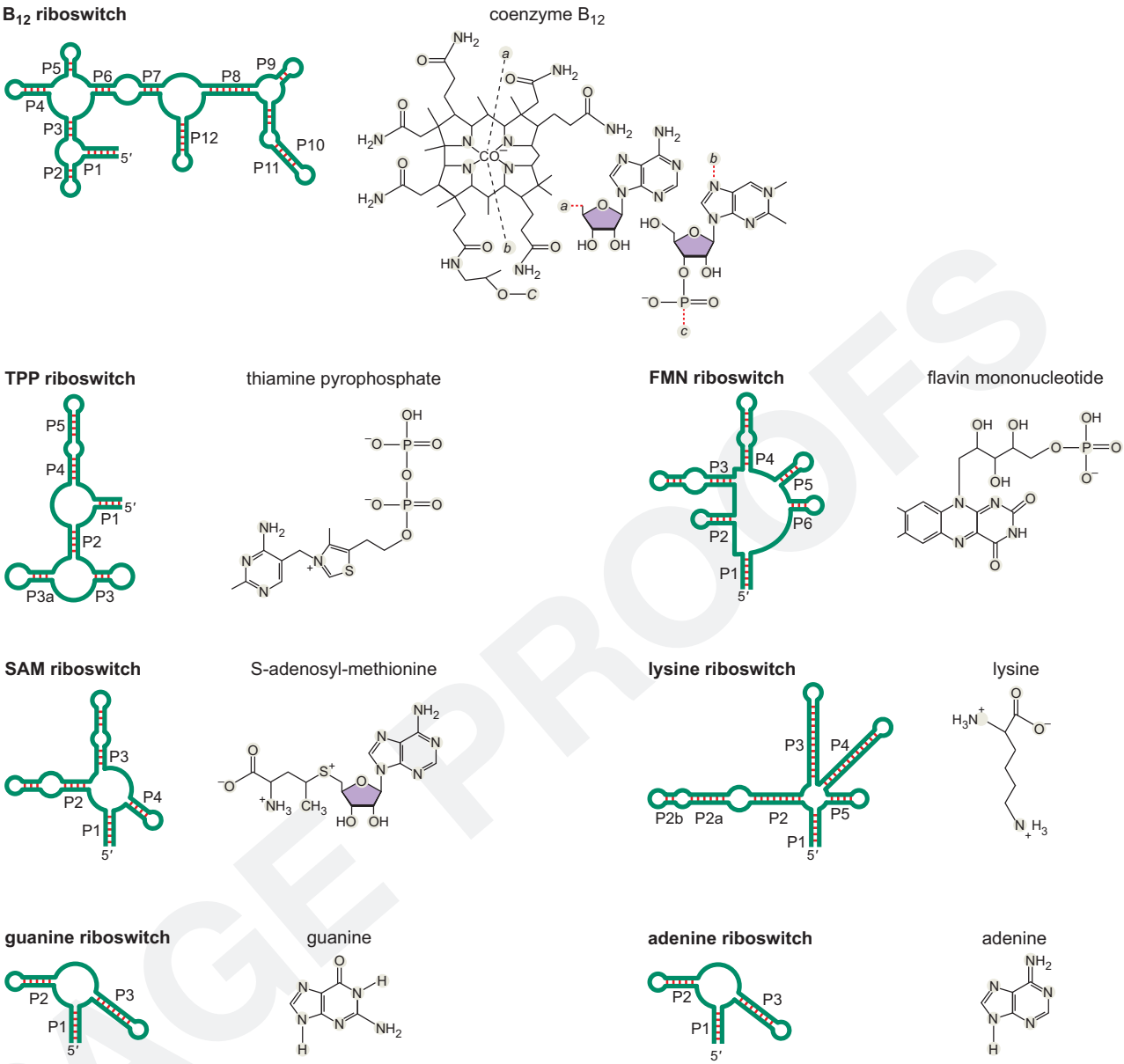


FIGURE 18-5 Riboswitches respond to a range of metabolites. The secondary structure of seven riboswitches and the metabolites they sense are shown here. (Adapted, with permission, from Mandal M. et al. 2003. *Cell* 113: 577–586, Fig. 7A. © Elsevier.)

Attenuation mediated by alternative RNA secondary structures was first discovered in studies of the tryptophan operon of *Escherichia coli*, as described in Box 18-1, Amino Acid Biosynthetic Operons Are Controlled by Attenuation. The *trp* operon contains genes responsible for the biosynthesis of tryptophan. Their expression is controlled in response to the level of tryptophan in the cell as measured by the availability of charged $tRNA^{trp}$. This example turned out to be one of several similar cases—attenuation linked to amino acid availability is a common control mechanism for amino acid biosynthetic operons. Understanding how the *trp* operon was controlled was the first time that alternative RNA secondary structures were shown to regulate gene expression in any organism.

ADVANCED CONCEPTS

BOX 18-1 Amino Acid Biosynthetic Operons Are Controlled by Attenuation

In *E. coli*, the five contiguous *trp* genes encode enzymes that synthesize the amino acid tryptophan. These genes are expressed efficiently only when tryptophan is limiting (Box 18-1 Fig. 1). The genes are controlled by a repressor, just as the *lac* genes are, although in this case, it is the *absence* of its ligand (tryptophan) that relieves repression.

Even after RNA polymerase has initiated a *trp* mRNA molecule, however, it does not always complete the full transcript. As with riboswitches, the decision to make a complete transcript is controlled by attenuation; in this case, most transcripts are terminated prematurely, before they include even the first *trp* gene (*trpE*). But attenuation is overcome if tryptophan levels are low in the cell; when tryptophan is limiting, polymerase does not terminate and instead transcribes all of the *trp* genes. Whether or not attenuation occurs depends on the ability of RNAs to form alternative secondary structures, just as it did with the riboswitches. In this case, however, the choice between alternative structures formed by the leader RNA is not controlled by binding of ligand directly to that RNA; instead, the choice of alternatives relies on the coupling of transcription and translation in bacteria.

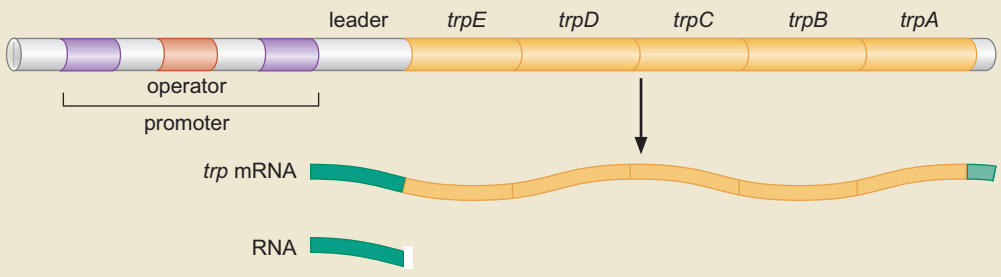
The sequence of the 5′ end of *trp* operon mRNA includes a 161-nucleotide leader sequence upstream of the first codon of *trpE* (Box 18-1 Fig. 2). Near the end of this leader sequence, and before *trpE*, is a transcription terminator, composed of a characteristic hairpin loop in the RNA (made from sequences in regions 3 and 4 of Box 18-1 Fig. 2), followed by eight uridine residues (see Fig. 12-12). Transcription usually stops after this terminator (and, we might have thought, should always stop), yielding a leader RNA 139 nucleotides long. This is the RNA product seen in the presence of high levels of tryptophan.

Three features of the leader sequence allow the terminator to be passed by RNA polymerase when the cellular concentration of tryptophan is low. First, there is a second hairpin (besides the terminator hairpin) that can form between regions 1 and 2 of the leader (see Box 18-1 Fig. 2). Second, region 2 also is complementary to region 3; thus, yet another hairpin consisting of regions 2 and 3 can form, and when it does, it prevents the terminator hairpin (3, 4) from forming. Third, the leader RNA contains an open reading frame encoding a short "leader" peptide of 14 amino acids, and this open reading frame is preceded by a strong ribosome-binding site (see Box 18-1 Fig. 2).

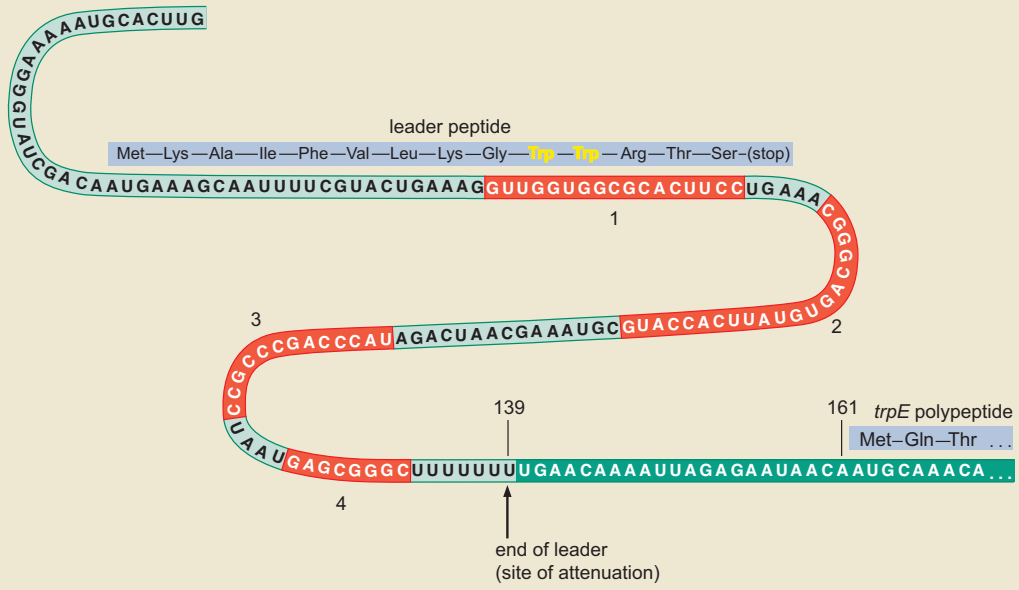
The sequence encoding the leader peptide has a striking feature: two tryptophan codons in a row. When tryptophan is scarce, there is very little charged tryptophan tRNA available, and the ribosome stalls when it reaches the two tryptophan codons. Under these circumstances, RNA around the tryptophan codons is within the ribosome and cannot be part of a hairpin loop. The consequence of this is shown in Box 18-1 Figure 3.

A ribosome caught at the tryptophan codons (part b) masks region 1, leaving region 2 free to pair with region 3; thus, the terminator hairpin (formed by regions 3 and 4) cannot be made, and transcription is not attenuated. If, on the other hand, there is enough tryptophan (and therefore enough charged Trp tRNA) for the ribosome to proceed through the tryptophan codons, the ribosome blocks sequence 2 by the time RNA containing regions 3 and 4 has been made. Thus, the terminator forms, attenuating transcription, and the *trp* genes are not transcribed.

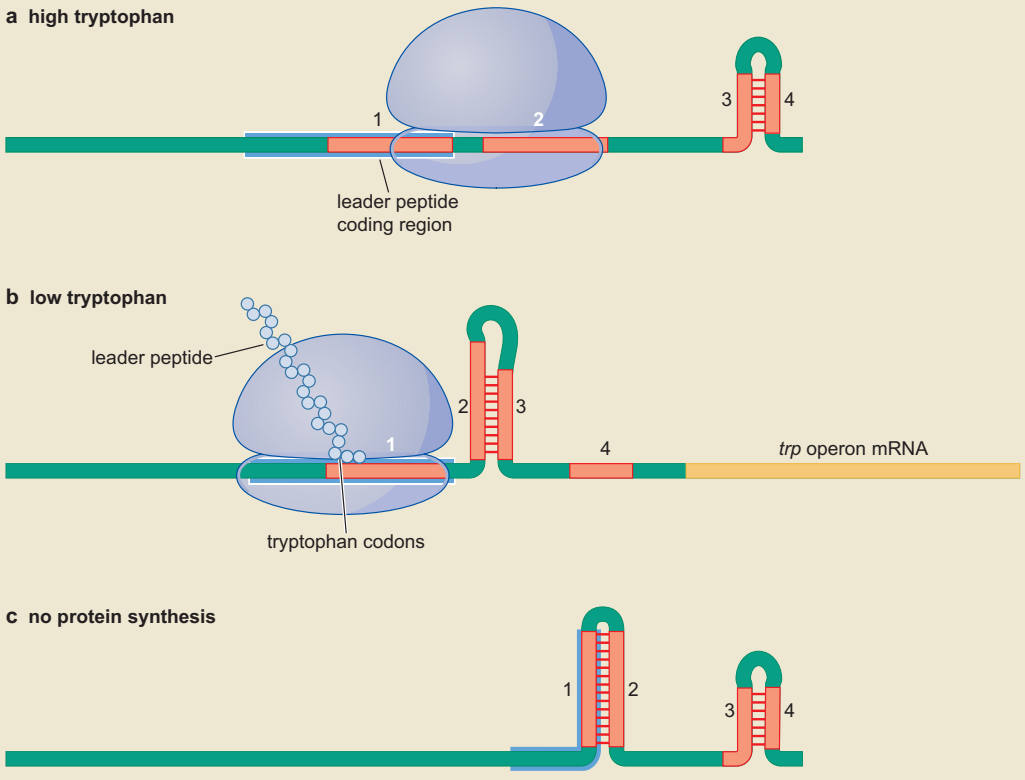
The *trp* operon is controlled by repression and attenuation, providing a two-stage response to progressively more stringent tryptophan starvation. But attenuation alone can provide robust regulation: other amino acid operons such as *leu* and *his* rely entirely on attenuation for their control. In the case of the leucine operon, its leader peptide has four adjacent leucine codons, and the histidine operon leader peptide has seven histidine codons in a row.



BOX 18-1 FIGURE 1 The *trp* operon. The tryptophan operon of *E. coli,* showing the relationship of the leader (see text) to the structural genes that code for the Trp enzymes. The gene products are anthranilate synthetase (product of *trpE*), phosphoribosyl anthranilate transferase (*trpD*), phosphoribosyl anthranilate isomerase-indole glycerol phosphate synthetase (*trpC*), tryptophan synthetase β (*trpB*), and tryptophan synthetase α (*trpA*).



BOX 18-1 FIGURE 2 ***trp* operator leader RNA.** Features of the nucleotide sequence of the *trp* operon leader RNA.



BOX 18-1 FIGURE 3 **Transcription termination at the *trp* attenuator.** The figure shows how transcription termination at the *trp* operon attenuator is controlled by the availability of tryptophan. The blue box shows the leader peptide coding region. (a) Conditions of high tryptophan: sequence 3 can pair with sequence 4 to form the transcription termination hairpin. (b) Conditions of low tryptophan: the ribosome stalls at adjacent tryptophan codons, leaving sequence 2 free to pair with sequence 3, thereby preventing formation of the 3-4 termination hairpin. (c) No protein synthesis: if no ribosome begins translation of the leader peptide AUG, the hairpin forms by pairing of sequences 1 and 2, preventing formation of the 2-3 hairpin, and allowing formation of the hairpin at sequences 3-4. The Trp enzymes are not expressed.

RNA INTERFERENCE IS A MAJOR REGULATORY MECHANISM IN EUKARYOTES

We have seen how RNA molecules can regulate expression of genes in prokaryotes. We have (in Chapter 17) already seen examples of regulatory RNA elements in the transcripts of some eukaryotic genes, but these function through binding regulatory proteins (the HIV TAT protein was one such case). It is, however, now apparent that RNAs have a more widespread and unanticipated role in gene regulation in eukaryotes. These new modes of regulation are described in this section.

Several types of very short RNAs repress—or silence—expression of genes with homology to those short RNAs. This silencing, called **RNA interference** (**RNAi**), manifests in different ways: sometimes by inhibiting translation of the mRNA, in other cases through destruction of the mRNA, and in yet others by transcriptional silencing of the promoter that directs expression of that mRNA. As we describe below, these short RNAs are generated by special enzymes from longer double-stranded RNAs (dsRNAs) of various origins.

It is still unclear how widespread the action of regulatory RNAs will turn out to be, and the detailed mechanism used to silence the target genes in any given case is still emerging. But it is already clear that the roles of these RNAs range from developmental regulation (well studied in the worm *Caenorhabditis elegans* and the plant *Arabidopsis*; see Chapter 22) to mechanisms that protect organisms against viral infection. Furthermore, RNAi has been adapted for use in several organisms as a powerful experimental tool, providing an easy means of "turning off" expression of any specific gene.

Short RNAs That Silence Genes Are Produced from a Variety of Sources and Direct the Silencing of Genes in Three Different Ways

Before describing aspects of the production and function of these short silencing RNAs in more detail, we first provide an overview of how this type of silencing works (illustrated in Fig. 18-6).

The small RNAs have different names depending on their origin. Those made artificially or produced in vivo from dsRNA precursors are typically called **small interfering RNAs** (**siRNAs**). Another group of regulatory RNAs are the **microRNAs** (**miRNAs**). The latter are derived from precursor RNAs which are encoded by genes expressed in cells where those miRNAs have specific regulatory functions.

Both siRNAs and miRNAs are generated from longer RNA molecules by the enzyme **Dicer**, an RNase-III-like enzyme (Fig. 5-24) that recognizes and digests longer dsRNA or the stem-loop structures formed by miRNA precursors (see below). The siRNA and miRNA products are typically 21–23 nucleotides long; their production is shown as the first step in Figure 18-6.

These small RNAs inhibit expression of homologous target genes in three ways: they trigger destruction of the mRNA encoded by the target gene, they inhibit translation of the mRNA, or they induce chromatin modifications within the target gene and thereby silence its transcription. Remarkably, whichever route is used in any given case, much of the same machinery is required. This machinery includes a complex called the **RNA-induced silencing complex** (**RISC**). A RISC

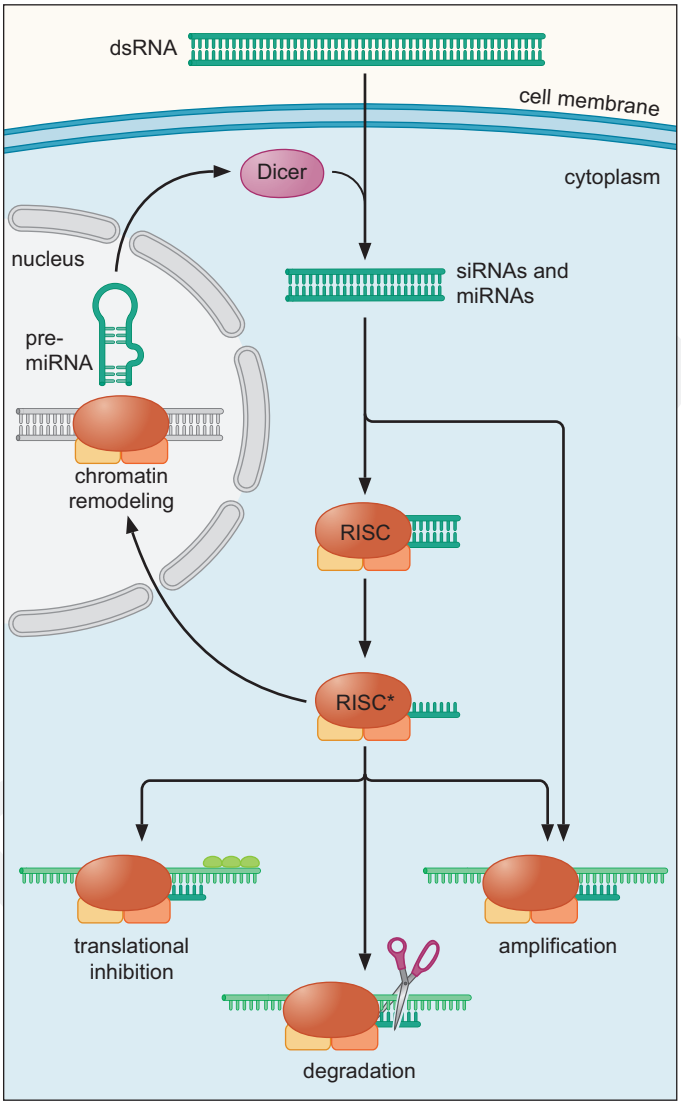


FIGURE 18-6 **RNAi silencing.** RNAi switches off the expression of a gene when dsRNA molecules that have homology to that gene are introduced, or made, in the cell. This effect involves processing of the dsRNA to make siRNAs and miRNAs by the enzyme Dicer. Another enzyme involved only in the case of miRNAs—Drosha—is not shown here, but is described later. The siRNAs and miRNAs direct a complex called RISC (RNA-induced silencing complex) to repress genes in three ways. It attacks and digests mRNA that has homology with the siRNA; it interferes with translation of those mRNAs; or it directs chromatin-modifying enzymes to the promoters that direct expression of those mRNAs. (Adapted, with permission, from Hannon G.J. 2002. *Nature* 418: 244–251, Fig. 5. © Macmillan.)

contains, in addition to the siRNA or miRNA, various proteins including a member of the **Argonaute** family.

The siRNA or miRNA must be denatured to give a **guide RNA**—the strand that gives the RISC specificity, as we shall see—and a **passenger RNA**, which usually gets discarded. The resulting complex, the mature RISC, is then directed to target RNAs containing sequences complementary to the guide RNA. These target RNAs are degraded or their translation is inhibited. Typically, the choice depends in part on how closely the guide RNA matches the target mRNA: if the sequences are highly complementary, the target is degraded; if the match is not as good (i.e., if there are several base-pairing mismatches), the response is more often inhibition of translation. In those cases where the target RNA is degraded, Argonaute is the catalytic subunit that carries out the initial mRNA cleavage; for this reason Argonaute is often called "**Slicer**" and mRNA cleavage is called "slicing."

A RISC can also be directed into the nucleus where it recruits other proteins that modify the chromatin around the promoter of the gene complementary to the guide RNA (shown on the left of Fig. 18-6). This

modification leads to silencing of transcription (Chapter 17). Establishing silencing in the centromeric regions of the yeast *Schizosaccharomyces pombe*, for example, requires the RNAi machinery. In this case, it is believed that regions of the centromere (see Chapter 7) are transcribed to produce RNAs that hybridize with other RNAs from the same region. The resulting dsRNAs are recognized by Dicer and cleaved to produce the siRNAs responsible for directing the RNAi machinery to the centromeres.

Another feature of RNAi silencing is worth noting—its extreme efficiency. Thus, very small amounts of dsRNA are often enough to induce a near complete shutdown of target gene expression. A factor adding to efficiency could be the action of an **RNA-dependent RNA polymerase** (RdRP), an additional enzyme required in many cases of RNAi including centromeric silencing in fission yeast. This polymerase can amplify the inhibitory signal: the RdRP generates dsRNA after recruitment to the mRNA by the original siRNA (shown on the right of Fig. 18-6). This feedback process generates large amounts of siRNA. RdRP has not yet been identified in mammalian cells, but high efficiency likely still results from the fact that slicing is catalytic—that is, each RISC can cleave several mRNAs.

Thus, although in the first section of this chapter we saw examples of small RNAs regulating gene expression in *trans* in bacteria, the mechanism of both production and action of such RNAs in eukaryotes is very different.

SYNTHESIS AND FUNCTION OF miRNA MOLECULES

miRNAs Have a Characteristic Structure That Assists in Identifying Them and Their Target Genes

As we have seen, miRNAs are one class of regulatory RNA that silences gene expression through the RNAi pathway. Indeed, in terms of their production and function, miRNAs are perhaps the best understood class. As mentioned above, miRNAs are encoded in the genome as segments of longer transcripts. Their characteristic structure helps identify them and predict the target genes they might regulate.

The functional form of an miRNA is typically about 21 or 22 nucleotides (it can vary from 19 to 25 nucleotides). These short RNAs are generated by two RNA cleavage reactions from a longer RNA transcript (called a pri-miRNA, for *pri*mary) that carries a hairpin-shaped secondary structure. The first cleavage liberates the stem loop, called the pre-miRNA; the second generates the mature miRNA from the pre-miRNA. One of the first identified, and best-characterized, miRNAs is *let-7*, which regulates development at the larval-to-adult transition in the worm *C. elegans* (see Chapter 22). The structure of the pre-miRNAs for *let-7* and some other naturally occurring miRNAs are shown in Figure 18-7.

It was thought initially that one "arm" of pre-miRNA stem-loop structure would be the regulatory miRNA. Recently, however, numerous examples have been identified where both "arms" of the structure give rise to functional miRNAs, each with its own set of target genes (in these cases the two miRNAs are labeled red and blue in Fig. 18-7). It now appears that having miRNAs produced from both arms is common. The pre-miRNAs can be encoded by any part of a transcript: that is, they

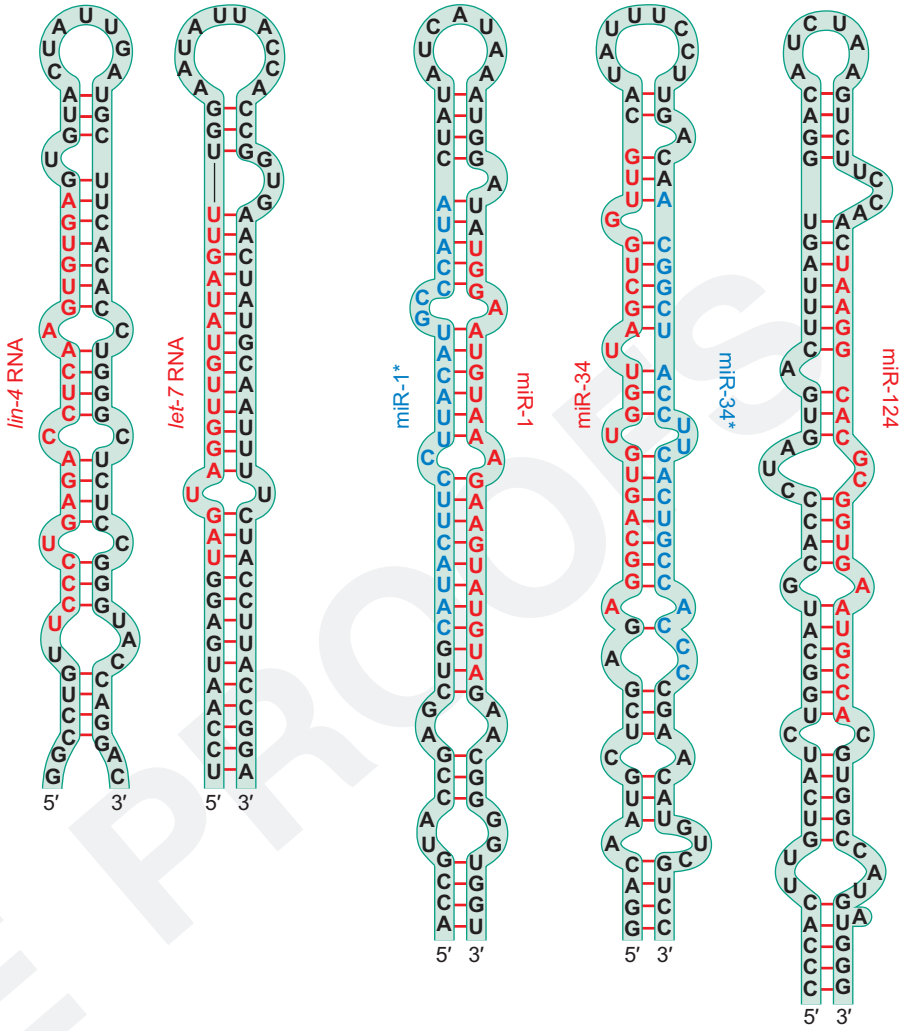


FIGURE 18-7 Structure of some pre-miRNAs prior to processing to generate the mature miRNAs. The sequences in red are miRNAs. In some cases, both "arms" of a stem loop can generate a functional miRNA. In such cases, the second miRNA is shown in blue—for example, miR-1 (red) and mi-R1* (blue), as well as with miR-34 (red) and miR34* (blue). The miRNAs shown are all from the worm. *lin-4* and *let-7* were identified genetically; those called miR are found by bioinformatics. (Modified, with permission, from Lim L.P. et al. 2003. *Genes Dev.* 17: 991, Fig. 6. © Cold Spring Harbor Laboratory Press.)

might fall within coding regions, within leader regions, or within introns (Fig. 18-8).

The distinctive secondary structure of a primary transcript carrying an miRNA (pri-miRNA) has made it possible to predict their presence based on the calculated secondary-structure fold of the RNA sequence. Furthermore, in many cases, candidates for the regulated target genes can also be predicted, because silencing depends on sequence complementarity between the target and the mature miRNA. The base pairing between miRNA and target RNA is initiated by interactions of so-called "seed residues"—typically the sequence between bases 2 and 9 of the 22-nucleotide miRNA. This is the region of highest complementarity, and so it is the region most useful in identifying candidate target genes. Of course, establishing that an miRNA really exists requires that its

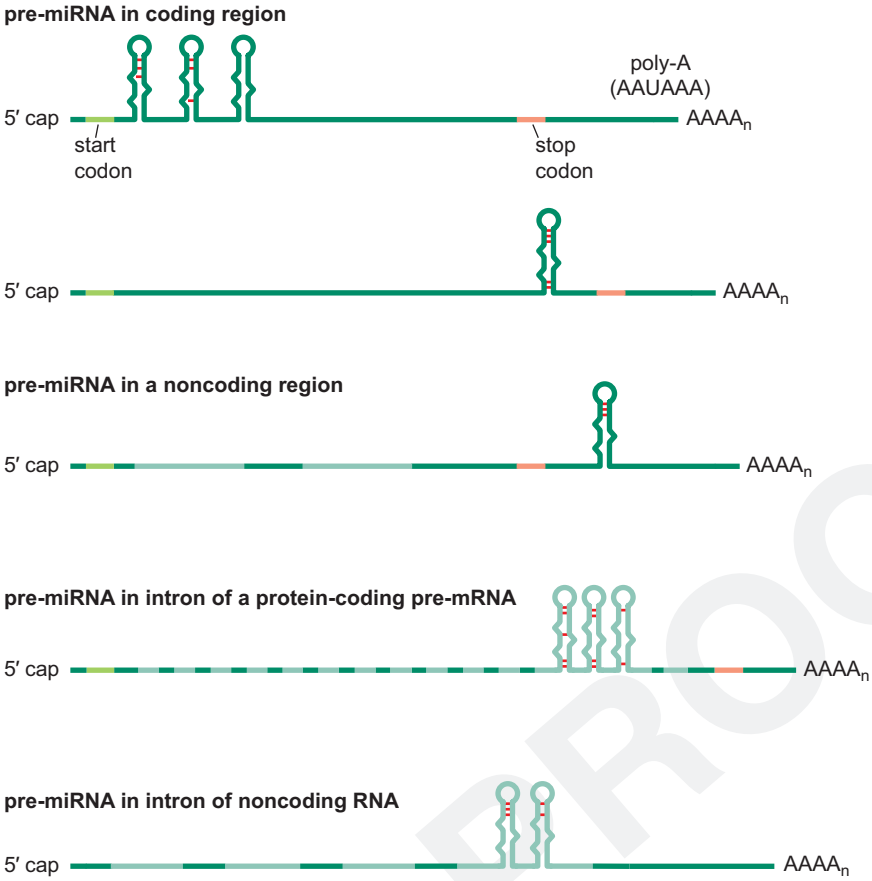


FIGURE 18-8 miRNAs are coded in both introns and exons in RNA. Intronic sequences are shown in light green.

presence be detected in cells (e.g., by northern blotting) and that gene expression from target mRNA be affected by its presence.

The two cleavage reactions required to generate the miRNA from these primary transcripts are mediated by two distinct RNases. One is Dicer, which we have already introduced, and is required for almost all cases of RNAi. The other, specifically required for miRNA processing, is **Drosha**. A characteristic of both these enzymes is that they recognize and cleave RNAs on the basis of the structure of their substrates rather than their specific sequence. We now turn our attention to how these enzymes work.

An Active miRNA Is Generated through a Two-Step Nucleolytic Processing

Two specialized RNA-cleaving enzymes are required to process the initial pri-miRNA transcript, containing the stem-loop structure, into the mature miRNA. The first enzyme is Drosha, a member of the RNase III family of enzymes. Drosha makes two cleavages that cut the stem-loop region of the RNA (pre-miRNA) out of the primary transcript RNA (pri-miRNA). This enzyme works together with an essential specificity subunit protein (called Pasha in some organisms and DGCR8 in others), and together these two proteins form an active **microprocessor complex**. The pre-miRNA generated by Drosha

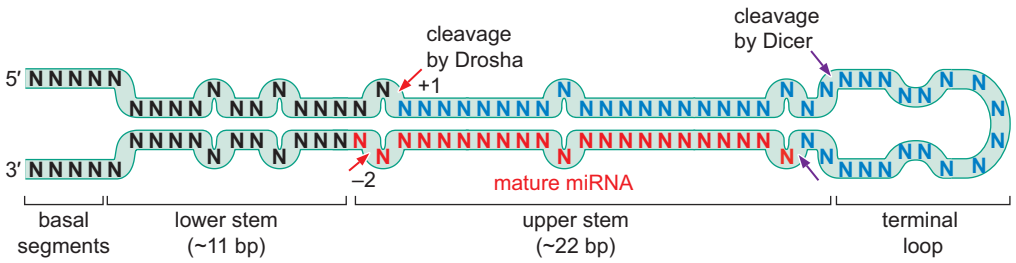


FIGURE 18-9 Overview of the structure of pri-RNA showing Dicer and Drosha cleavage sites. The region in red becomes the mature miRNA. Note that the basal segments must be single-stranded for proper recognition by the Drosha complex.

is usually approximately 65–70 nucleotides long. Drosha resides in the cell's nucleus, and the Drosha-catalyzed cleavage event occurs in this cellular compartment.

The base-paired stem in the pri-miRNA is typically about 33 bp in length (three helical turns of dsRNA) and contains only a few mismatches (Fig. 18-9). At the "top" of the stem is a loop of variable size (usually relativity large, about 10 nucleotides); the sequence of this loop region is not critical for the processing reactions. Importantly, for processing by Drosha, single-stranded RNA (ssRNA), lacking significant secondary structure, is needed flanking each side (5′ and 3′) of the stem-loop. It is the ssRNA–dsRNA junctions that are in large part responsible for determining the cleavage specificity of Drosha.

The stem region can be divided into two functional segments: an approximately 11-bp lower stem and an approximately 22-bp upper stem (Fig. 18-9). Drosha cleaves 11 bp away from the dsRNA–ssRNA junctions—that is, between the lower and upper stems in the pri-miRNA (Fig. 18-10). The two cleavages thus generate the approximately 65-nucleotide pre-miRNA composed of the 22 bp (two helical turns) of dsRNA and the top loop. The RNase III family enzymes are specific for dsRNA and cleave it in a manner that leaves a 2 nucleotide overhang on the 3′ ends of the dsRNA product. This 3′ overhang is important for recognition of that RNA molecule by the next enzyme in the pathway, Dicer.

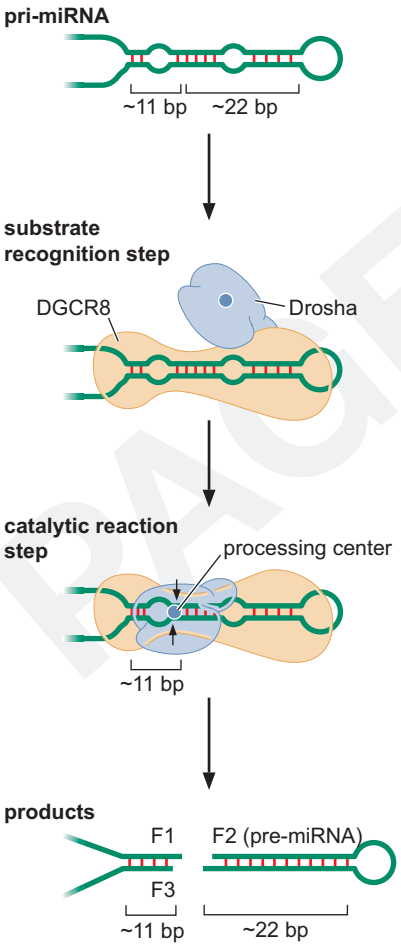


FIGURE 18-10 Recognition and cleavage of pri-miRNA by the microprocessor complex. Three fragments are generated by cleavage, labeled F1, F2 (the pre-miRNA), and F3.

Dicer Is the Second RNA-Cleaving Enzyme Involved in miRNA Production

The pre-miRNA liberated by Drosha is exported to the cytoplasm where the second RNA cleavage reaction, carried out by Dicer, takes place. As with Drosha, Dicer selects its cleavage sites using a measuring, rather than sequence-specific, mechanism. A high-resolution structure of Dicer provides insight into how this likely occurs.

Dicer is constructed of three modules (Fig. 5-23): two RNase III domains and a dsRNA-binding domain called the PAZ domain (named for three proteins that contain this domain: Piwi, Argonaute, and Zwille).

Figure 18-11a shows in cartoon form the organization of the Dicer protein and how it is believed to interact with a dsRNA molecule. In the right panel, Figure 18-11b, is the structure of Dicer, modeled with a substrate RNA. The overall structure of the protein is shaped like a hatchet. The PAZ domain is at the bottom of the handle, where it forms a binding pocket for the 3′ end of the dsRNA substrate. The

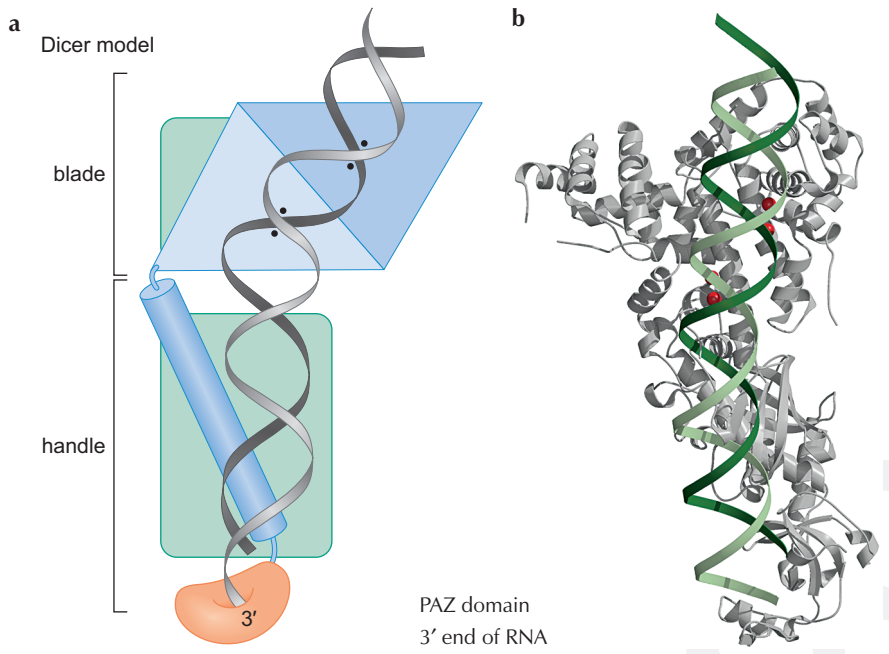


FIGURE 18-11 Dicer structure and organization. (a) The scheme shows Dicer organization. (b) Dicer structure modeled with dsRNA reveals how length is measured. The protein is shown in gray, with nuclease active sites indicated by the red spheres (and as black dots in part a). The RNA is in green. The structure shown contains only the RNase III and PAZ domains. The Dicer protein also contains ATPase and other domains. (b, MacRae I.J. et al. 2006. *Science* 311: 195–198. PDB Code: 2FFL; note that the RNA was modeled into the structure and was not part of the crystal structure.) Image prepared with MolScript, BobScript, and Raster 3-D.

handle of the hatchet is formed by a linker domain and contains a positively charged binding surface for the RNA molecule. The top "blade" region comprises the two RNase domains, arranged in a symmetrical dimer. Each RNase domain carries an active site and is responsible for cleaving one of the two strands of the substrate RNA. Thus, Dicer will act on any dsRNA, regardless of sequence, and will cleave this molecule 22 nucleotides from its end. The PAZ domain anchors the 3′ terminus of the substrate RNA to position the active sites of the enzyme approximately 22 nucleotides away in a ruler-like fashion (see Fig. 18-11). Indeed, the occurrence of differently sized PAZ domains correlates with the different sizes of Dicer products found in different organisms.

Incorporation of a Guide Strand RNA into RISC Makes the Mature Complex That Is Ready to Silence Gene Expression

The actions of Drosha and then Dicer generate the 21–25-nucleotide RNA molecule that will guide regulation of gene expression. The active form of the regulatory miRNA is the single-strand form—at this stage called the "guide" RNA—incorporated into a RISC protein complex. Within this complex, the RNA guide strand recruits RISC to a target RNA. It has been argued that the length of approximately 22 nucleotides is just long enough to uniquely specify a single target gene in the large genomes of complex eukaryotes using RNA–RNA base pairing.

RISC is a multicomponent complex that includes the guide miRNA. The guide RNA base-pairs with the target mRNA, and, as a consequence, promotes silencing of gene expression. The central component of this regulatory complex is a protein called Argonaute, which is in many cases an RNA-cleaving enzyme. The best understood mechanism of gene silencing is RISC-mediated cleavage—or slicing—of the target mRNA. However, many organisms have multiple members of the Argonaute protein family. For example, there are eight distinct Argonautes in humans, but not all of these Argonautes, when incorporated into a RISC complex, have slicer activity. RISCs containing other Argonautes must silence gene expression using non-slicer-dependent mechanisms, such as repression of translation.

Generation of the active RISC and slicing occur as follows. The short dsRNA generated by Dicer is incorporated into RISC. The dsRNA is denatured to provide the guide strand and the passenger strand. According to one model, the passenger strand is cleaved by Argonaute and then ejected from the complex; according to another, the passenger strand is removed without first being cleaved, likely by an RNA helicase. The resulting RISC—called mature RISC—with its single-stranded guide RNA is now ready to recognize and slice the target mRNA.

WEB STRUCTURAL TUTORIAL

As we saw with Dicer, the structure of an Argonaute protein provides a framework for understanding the mechanism of target RNA recognition and cleavage by RISC (Fig. 18-12; see also Structural Tutorial 18-1). Like Dicer, Argonaute has both a PAZ domain and an RNase domain. The PAZ domain specifically recognizes the 3′ end of the guide RNA. The bound guide RNA is base-paired to the target RNA, and the architecture of the complex is such that this binding positions the active site of the RNase

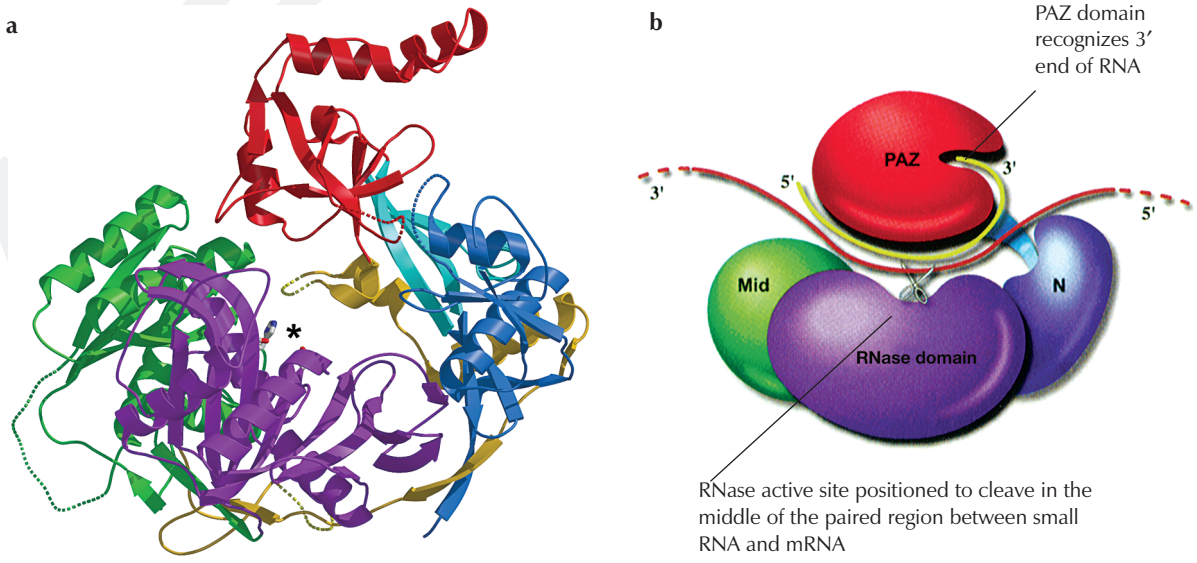


FIGURE 18-12 Argonaute structure, showing RNA-binding regions and an RNase H–like nuclease domain. (a) Crystal structure of Argonaute. The domains are colored as in part b, with the blue domain being the amino-terminal part of the protein, and the green domain in the middle. (b) Cartoon of the Argonaute domains. The arrow shows the RNase active site positioned to cleave in the middle of the paired region between small RNA and miRNA. (a, Song J.J. et al. 2004. *Science* 305: 1434–1437, Fig. 4C. PDB Code: Iu04. Image prepared with MolScript, BobScript, and Raster 3-D. b, Adapted, with permission, from Song J.J. et al. 2004. *Science* 305: 1434–1437, Fig. 4C. © AAAS.)

domain appropriately to cleave the target RNA strand. Cleavage occurs nearly in the middle of the guide RNA–target RNA duplex, between the 10th and 11th nucleotide from the 5′ end of the guide RNA.

As we have already mentioned, in some cases, mature RISC can inhibit translation rather than slicing mRNA. The mechanism of this translational repression is still under scrutiny, but it appears that miRNAs lead, in some cases, to the sequestration of mRNA in so-called processing bodies (P-bodies) within the cytoplasm where translation is repressed. Binding of miRNA can also destabilize the poly-A tail (see Chapter 12) of the target mRNA, disrupting translation initiation (Chapter 14).

siRNAs Are Regulatory RNAs Generated from Long Double-Stranded RNAs

As we have seen, the critical steps in synthesis of an active miRNA are Drosha cleavage, followed by Dicer cleavage, followed by incorporation of the guide RNA into the RISC complex. But only miRNAs are made from large hairpin precursors. In contrast, the precursor RNA for the siRNA pathway is a longer dsRNA. As a consequence of this different initial substrate, Drosha is not needed for the generation of siRNAs. Cleavage by Dicer is still required, however, and again generates a suitable 21–22-nucleotide RNA for incorporation into RISC. In plants, even miRNAs are generated by Dicer alone; it is not clear how they manage to forgo prior action of Drosha.

A fascinating series of observations and experiments led to our current understanding of small regulatory RNAs in eukaryotes. These began in the late 1980s with seemingly mystifying results of attempts to overexpress pigment genes in petunias (to make them a deeper purple, but ending up with the flowers turning white). Next was the surprising discovery of regulatory genes from worms whose products turned out to be miRNAs, and then experiments showing that introducing dsRNAs into worms silenced complementary gene expression. This story is described in Box 18-2, History of miRNAs and RNAi.

Small RNAs Can Transcriptionally Silence Genes by Directing Chromatin Modification

We have now seen how miRNAs and siRNAs can silence genes by inhibiting the translation of target mRNAs or directing their destruction. Regulatory RNAs can also act at the level of transcription, switching off expression of target genes by directing histone modification of the promoter. This mechanism has been most extensively studied in centromeric silencing in the fission yeast *S. pombe*.

We noted in Chapter 17 that in yeast, genes placed in certain regions of the genome are typically silenced. In the case we described in detail in that chapter, genes placed near the telomeres in *Saccharomyces cerevisiae* were silenced. Genes in the mating-type locus of that yeast and of *S. pombe* are also silenced. In *S. pombe*, the centromeres are another silenced region of the genome. In both organisms, silencing involves histone modifications. But unlike cases of silencing in *S. cerevisiae*, which lacks the RNAi machinery, centromeric silencing in *S. pombe* requires that pathway.

The centromeres of *S. pombe* have a sequence organization more like that of higher eukaryotes (e.g., flies and humans) than that of *S. cerevisiae* (see Fig. 7-8). Each centromere has a central region, of

BOX 18-2 History of miRNAs and RNAi

In 1989, Richard Jorgensen, working at the biotech company Advanced Genetic Sciences in Oakland, California, was trying to make petunia plants with more deeply purple flowers than existing strains. The strategy seemed straightfoward: he would introduce into the plants an additional copy of the pigment gene (encoding chalcone synthase) under the control of a strong promoter. These plants would make more chalcone synthase and the flowers would be more purple. What he actually got were plants with varying degress of paler flowers, many that were sectored—with purple and white regions—and even some that were completely white (Box 18-2 Fig. 1).

Although disappointing, these results were intriguing. In attempting to understand what was going on, Jorgensen uncovered various features of the phenomenon, called **cosupression** (because expression of both the transgene and the endogenous gene are repressed). The greater the expression of the transgene, the lower the level of chalcone synthase; this was true whether increased expression resulted from multiple copies of the transgene or from use of stronger promoters driving the transgene. It was also noted that some plants had variegated patterns of pigmentation, and that different variegation patterns could be found in different flowers on the same plant. These patterns were sometimes inherited, but on other occasions apparently altered at random. These observations suggested to Jorgensen and others (particularly Marjori Matzke, who also was investigating this phenomenon) that they were dealing with an epigenetic phenomenon.

Other investigators were trying to make plants resistant to viral infection. One approach was to overexpress in plants a dominant-negative derivative of a common viral replication factor: this protein was expected to block replication of any infecting virus that used this common replicative mechanism. Although the dominant-negative viral product blocked replication of the potato virus from which it was derived, its specificity of action was surprisingly tightly restricted to that virus. It was also shown that the protein itself was not even needed—just the RNA.

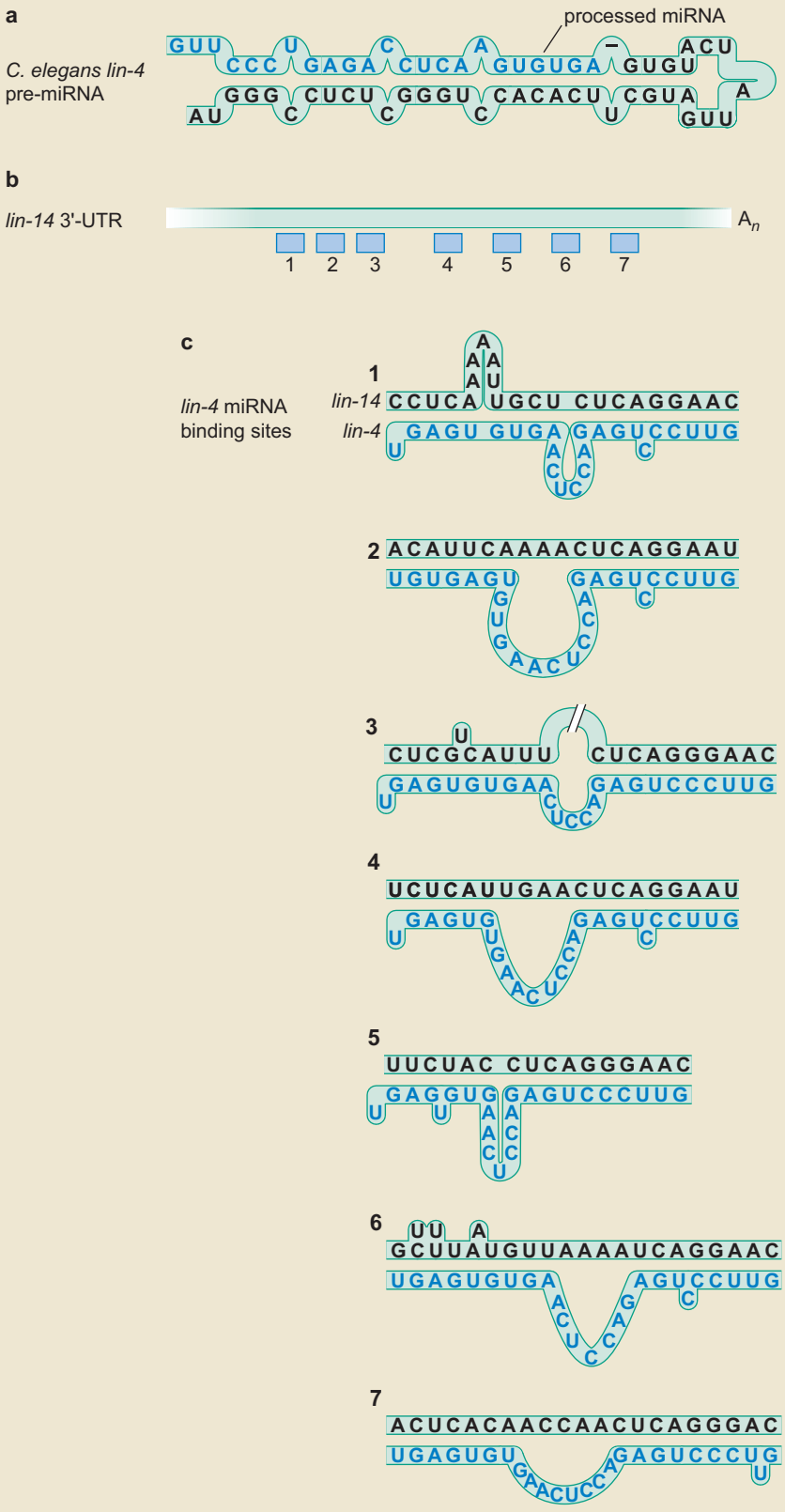
Meanwhile other researchers were using antisense RNA to knock down expression of the *par-1* gene in worms. Their intention was to prove that this gene was responsible for a particular developmental phenotype. Antisense RNA produced in vitro and injected into the developing worm induced the phenotype predicted for the loss of *par-1* expression. But it was found that *sense* RNA had the same effect. This was only included in the experiment as a negative control of course; it was not expected to have any effect on expression. RNAs unrelated to the *par-1* gene had no effect.

The explanation for this RNA-dependent gene repression was provided by Andrew Fire and Craig Mello in experiments that earned them the 2006 Nobel Prize in Physiology or Medicine. They demonstrated that it was in fact neither sense nor antisense RNA that silenced the gene—it was the dsRNA produced by a mix of the two. It turned out that the RNA preparations of sense or antisense were both contaminated with small amounts of the opposite strand, and it was the resulting double-strand population that caused silencing. When dsRNAs were deliberately prepared, they were shown to be very potent in eliminating expression of the target gene. Hence, the phenomenon of RNAi had been discovered, a finding published in 1998.

BOX 18-2 FIGURE 1 Petunia flower. An example of the effects of overexpressing the pigment gene chalcone synthase in what would otherwise be a completely purple petunia flower. (Courtesy of Richard A. Jorgensen, University of Arizona.)

Mechanistic insights came thick and fast from several labs. First, dsRNAs were shown to trigger degradation of homologous mRNAs in extracts from *Drosophila* cells, an assay that led to the identification of RISC. The identifcation of siRNAs—the species that directs RISC to the target genes—was reported in plants in 1999. Dicer, the nuclease that creates them, was described in 2001. And the final major component of the pathway, Slicer, was identified in 2005 when the crystal structure of Argonaute revealed the protein to be an RNase.

In addition to showing that Dicer was needed to generate the siRNAs, it was shown to also be needed for miRNA genes to function during worm development. The first miRNA and its target had been described in 1993, by Victor Ambros and Gary Ruvkun, respectively. At the time, this observation was seen as a neat but eccentric oddity; the *lin-4* gene encoded a small RNA that acted on a target gene, *lin-14*, by virtue of sequence complementarity between the miRNA and regions of the 5′ UTR of the target genes (Box 18-2 Fig. 2). Later, other miRNAs were found in worms, some with homology to similar genes in animals and plants, suggesting that this mechanism of regulation was more widespread. Everything was tied together when it became clear that siRNAs and miRNAs work through the same pathways. Thus, the picture emerged of a world of tiny RNAs involved in gene regulation—some exogenously supplied, others inbuilt as part of gene regulatory programs employed during development. The field developed very rapidly, as the dates in this account reveal, moving from obscure phenomenology to a Nobel Prize and demanding of its own chapter in textbooks, in just 15 years. The accelerated progress was perhaps largely a consequence of the range of species (yeast, plant, and worm) studied and approaches (genetics, biochemistry, structural studies, and bioinformatics) employed.



BOX 18-2 FIGURE 2 microRNA *lin-4* binds within the 3′-untranslated region (UTR) of its target gene *lin-14*. (a) The *lin-4* pre-miRNA before processing by Dicer. The sequence of the miRNA is shown in blue. (b) The seven sequences within the *lin-4* 3′-UTR that can base-pair with the *lin-4* miRNA to various extents, as shown in part c. (Modified from Ha I. 1996. *Genes Dev.* 10: 3041–3050, Fig. 1. © Cold Spring Harbor Laboratory Press.)

largely unique sequence, flanked by a series of repeats common to all centromeres. The repeats are important to function and contribute to the formation of heterochromatin and the transcriptional silencing associated with the region, as we shall see. Histones within the heterochromatin carry repressing markers: low levels of acetylation, and methylation on lysine 9 of histone H3 tail (H3K9) (see Table 7-7).

S. pombe has only a single gene for each of the major components of the RNAi pathway—Dicer and Argonaute. Higher organisms have multiple Dicer and Argonaute genes with partially redundant functions, making genetic manipulation of the pathway more difficult. In addition, unlike the situation in flies and worms, loss of the RNAi pathway is not lethal to *S. pombe*, although it does make the cells grow poorly by, for example, disrupting chromosome segregation. It was a surprise, however, to discover that loss of any component of the RNAi pathway led to loss of histone H3K9 methylation and therefor also loss of gene silencing at the centromeres, particularly as this silencing was known to be transcriptional; until this discovery, RNAi had been thought to act only posttranscriptionally.

The key to understanding this transcriptional silencing seems to be the centromeric repeats themselves; these sequence elements are transcribed from both strands by RNA polymerase II, producing complementary transcripts that can hybridize to form dsRNAs. The RNAs are in turn acted on by the RNAi machinery to generate siRNAs that somehow—and quite how remains unclear—direct an Argonaute-containing RISC-like complex (called RNA-induced transcriptional silencing [RITS] complex) to the centromeres. The siRNAs could in theory do this by recognizing DNA at the centromeres, through sequence-specific base pairing directly with the DNA template. But more likely are models in which the siRNAs recruit RITS to transcripts tethered to the centromere by RNA polymerase II. Recruitment results in slicing of centromic transcripts, which is, in turn, required for spreading of the histone modification apparatus along the centromere (Fig. 18-13). Thus transcription itself may itself spread silencing, when transcripts are targeted by RNAi.

As we mentioned above, the mating-type loci of *S. pombe* are also transcriptionally silenced, and here the silencing is not lost in mutant strains defective for RNAi. It is believed that RNAi acts in this case as well, but only in initially establishing the silenced state—it is not required for maintaining silencing once it is established; other, protein-based mechanisms sustain the repressed state—just as they do in *S. cerevisiae* (Chapter 17). RNAi is also believed to play a part in heterochromatin silencing in other organisms, ranging from flies to plants. Silencing of unwanted transcription from transposons also appears to be RNA-mediated, as we describe below.

THE EVOLUTION AND EXPLOITATION OF RNAi

Did RNAi Evolve As an Immune System?

The RNAi machinery is widespread in eukaryotes, although not ubiquitous. It does not occur in *S. cerevisiae*, for example, as we just noted. It is believed, however, that at least the basic system existed in the most recent common ancestor to all eukaryotes, but was subsequently lost in some lineages.

But what does RNAi do, biologically? There are miRNAs of course—and the RNAi machinery is required to produce and use those regula-

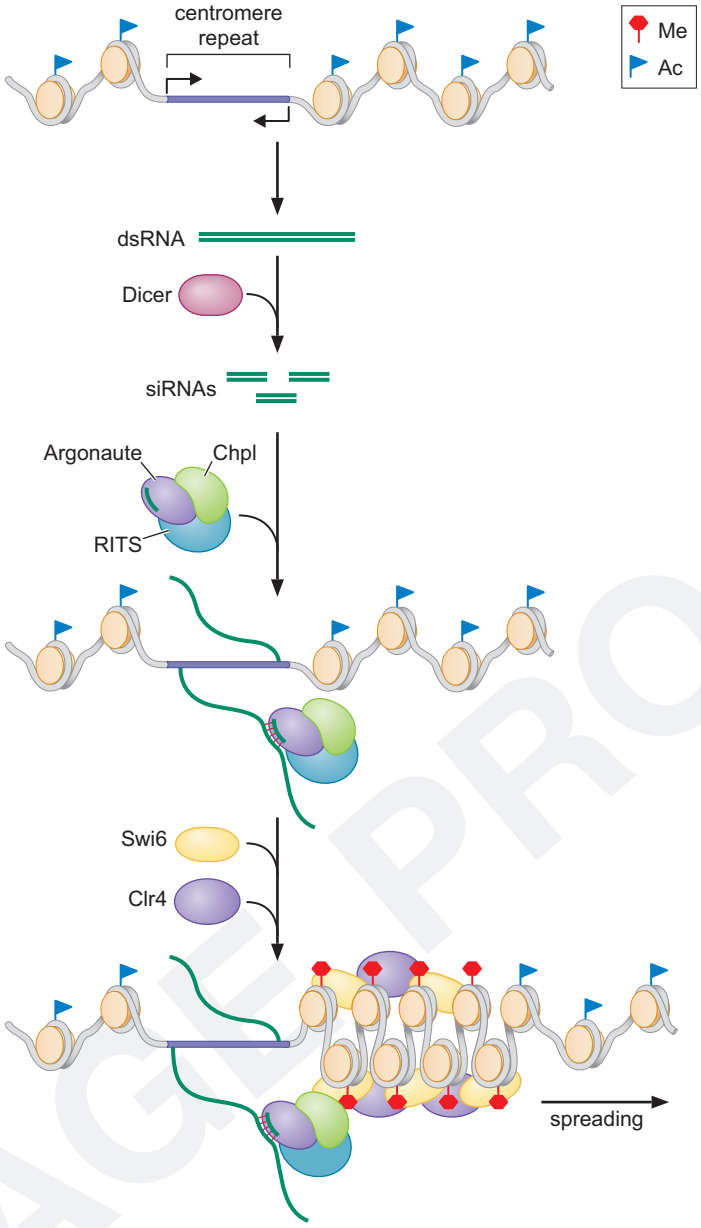


FIGURE 18-13 A model for RITS recruitment and the silencing of centromeres. Shown at the top are nucleosomes around the repeat sequences at a centromere in *S. pombe*. The repeat sequences are transcribed from both strands by RNA polymerase II, generating dsRNA that is a substrate for Dicer. The siRNAs thereby produced are loaded into the Argonaute-containing complex RITS. As shown in the middle, the siRNA-containing RITS is then recruited to the PolII-tethered transcripts being generated by continued transcription of the centromeric repeats, through complementarity between the siRNA and the transcript. This complex then recruits factors (Clr4 and Swi6) that locally modify nucleosomes by adding the H3K9 silencing markers. Another subunit of RITS, Chp1, contains a chromodomain (Fig. 7-40), which, by binding to the methylated nucleosomes, likely stabilizes the binding of RITS. Although not shown in the figure, "slicing" of the transcripts by Argonaute (within RITS) generates substrate RNAs for the RdRP, which synthesizes a complementary strand and thus generates further substrate for Dicer. This process is required for the nucleosome modification—and thus the region silenced—to spread.

tors—but some organisms have the RNAi machinery and no miRNAs (including *S. pombe*). It is, in fact, believed that miRNAs evolved to take advantage of the existence of the RNAi machinery rather than being the reason that machinery exists. One ancient function the RNAi machinery might have served (and still serves) is protecting organisms from transposons and viruses.

Transposons are found in all eukaryotes and, in some cases, make up a substantial amount of a genome. In humans, for example, about 45% of our genome is made of sequences that were once transposons. Transposons are often transcriptionally silent and packaged into heterochromatin. In some RNAi mutants, however, the histone modifications associated with transposon silencing are lost. In addition, in plants and worms, several siRNAs have been identified that correspond to transposons. And in some cases in both these organisms, the loss of RNAi reactivates transposons, causes them to jump, and leads to high levels of spontaneous

mutagenesis. Not as many transposons are reactivated as are known to generate siRNAs, however. This might reflect a situation similar to that described above for the mating-type silencing in *S. pombe*: RNAi might be essential for initiating silencing of some transposons, but the silencing then becomes self-sustaining without further need for the siRNAs.

The protective effect of RNAi on viral infection of plants has been widely observed. Indeed, the effects were recognized long before RNAi was known to be an underlying mechanism. When one leaf on a plant is infected by a virus, a factor able to silence replication of the virus is spread systematically throughout the whole plant. This factor does not protect that originally infected leaf, but it does stop the infection from spreading. In plants mutated in the Argonaute or Dicer genes, infection spreads unhindered and viral replication is much higher. The protecting signal comprises siRNAs generated from the viral genome itself. Viruses have retaliated: they often carry genes whose function is to protect the infecting virus from host RNAi. One example is HcPro from potato virus Y, which acts to reduce production or stability of siRNAs. Other viral products effect other steps in the defense mechanism, including the systemic spreading of siRNAs.

Links between RNAi and human disease are described in Box 18-3, RNAi and Human Disease.

RNAi Has Become a Powerful Tool for Manipulating Gene Expression

The discovery of RNAi arose from observations made by investigators attempting to manipulate gene expression (see Box 18-2). In the cases of both cosuppression in plants and antisense RNA in worms, it was attempts to understand unexpected blips in those manipulations that led to the discovery of RNAi. It was therefore perhaps not surprising that once understood, RNAi was quickly exploited as a tool for manipulating gene expression. In worms, this is now routinely done. Libraries exist that encode dsRNAs that can target any gene in the worm genome. These libraries can be used to screen worms for the consequences of inhibiting expression of any given gene. The general way in which this is done is shown in Figure 18-14. Worms feed on bacteria. In the lab, they are fed *E. coli*, and it turns out that the quickest route to a worm's genome is through its stomach: any desired dsRNA can be expressed in the *E. coli* on which the worms feed, and this delivers enough substrate for the RNAi response to be triggered in cells of the worm, switching off genes homologous to the original dsRNA.

It would, of course, be of great benefit to screen for genes in this way in mammalian cells, where traditional genetic screens are not feasible. It was established that artificially synthesized siRNAs, made in vitro and introduced into mammalian cells in culture, trigger an RNAi response and down-regulate appropriate target genes, but the efficiency of transfection (getting the RNAs into the cells) was low. Longer dsRNA molecules are also problematic because they trigger a response that shuts down all translation in the cell, a response evolved to block viral replication; many viruses have RNA genomes. Despite these drawbacks this approach can still be fruitfully employed.

Being able to express the RNAs in cells (rather than having to introduce them in high concentrations) would get around the problems mentioned above and have a further benefit: the RNAs would be pro-

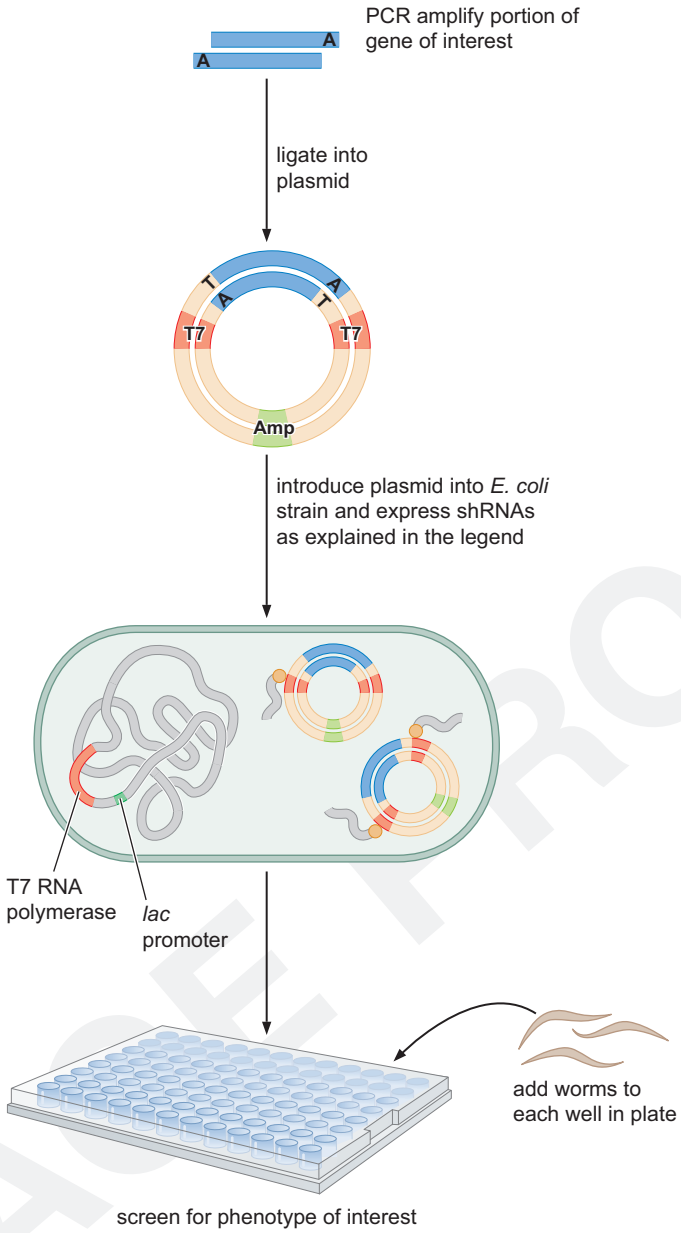


FIGURE 18-14 RNA interference can be induced in worms by feeding bacteria expressing dsRNAs. See Chapter 21 for details of the molecular manipulations required in the first steps of this scheme. The shRNA expression from the plasmid is under the control of a promoter recognized not by *E. coli* RNA polymerase, but rather by a single-subunit RNA polymerase from a phage called T7. The gene for that polymerase is expressed artificially in the cells used in this scheme, under control of the *lac* promoter (Chapter 16). Thus, production of the shRNAs can be controlled using an inducer of the *lac* promoter.

duced in a sustained manner rather than being present transiently as they are with transfection. But this approach presents its own formidable hurdles: how can one ensure that both complementary strands are expressed in the same cell and find each other efficiently to form the active dsRNA substrate for RNAi?

Instead of trying to deliver dsRNA into cells, investigators found it more fruitful to mimic miRNAs. To this end, libraries have been generated in which short genes are synthesized as oligonucleotides and cloned in plasmids. Each short gene is designed to give a transcript that will fold into a stem loop. These are processed by dicer in the cell to generate an siRNA that will direct silencing of its target genes. These short synthetic genes are called **short hairpin RNA** genes (**shRNAs**). By using an appropriately designed shRNA, any individual gene in the genome can be targeted. Or, with a suitable library, a ge-

BOX 18-3 RNAi and Human Disease

Cancer

A general decrease in levels of many miRNAs is often seen in cancers. This decrease has been taken to indicate that those miRNAs usually have a tumor-suppressing effect. Despite this general trend, other specific miRNAs are up-regulated in some cancers. Analogous to protein-coding genes implicated in cancer, the miRNAs in question are described as being tumor suppressors (if their absence increases cancer) or oncogenic (if their increased expression leads to cancer). Their targets tend to be genes involved in cell cycle progression (proliferation) or apoptosis.

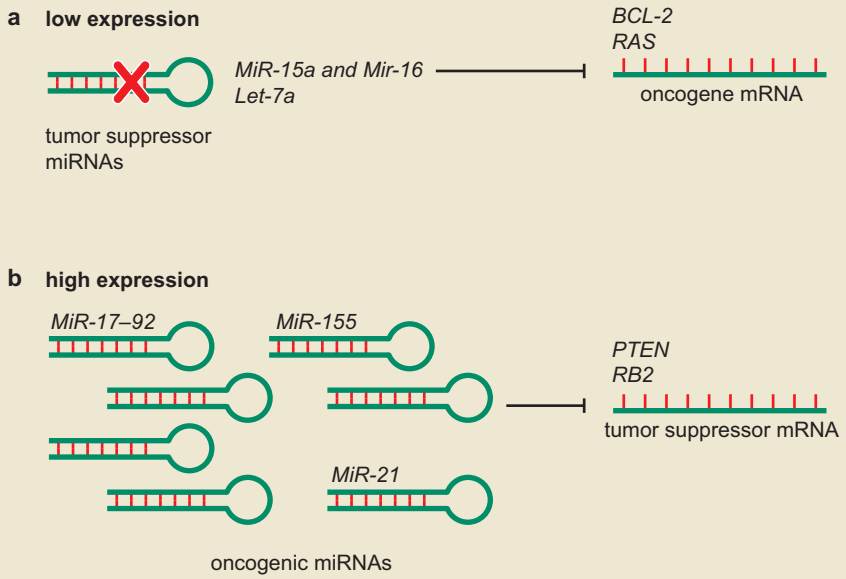
Of the 300 miRNAs identified in humans, more than one-half are located in regions of the genome regularly disrupted in cancers. Thus, in many cancers, the genes for these miRNAs are deleted or amplified, depending on the nature of the chromosomal rearrangement. So, for example, the miRNAs *miR-15* and *miR-16* induce apoptosis of cells by down-regulating the *BCL2* gene (see Box 18-3 Fig. 1). The most common form of adult leukemia in the western hemisphere is chronic lymphocytic leukemia (CLL), a disease associated with deletions in a region of chromosome 13 (13q14). This region of the genome contains the miRNA *miR-15a* and *miR-16a* genes; indeed, these are the only two genes included in the smallest deletions associated with CLL. Thus, when these genes are deleted, apoptosis is down-regulated and tumors can more readily arise and develop.

In another region of chromosome 13 (13q31) is found *miR-17-92*, an oncogenic miRNA. Compared to normal tissue, expression of this gene is significantly increased in many cancers, including lung cancer, and especially in its most aggressive forms (e.g., small-cell lung cancer). In addition, overexpression of this miRNA in transgenic mice drives tumorigenesis. Among the many predicted targets of *miR-17-92* are two tumor suppressor genes, *PTEN* and *RB2*. One definite target is the cell cycle progression regulator E2F1. Both these and other examples of miRNAs in cancer are shown in Box 18-3 Figure 1.

Fragile X Mental Retardation

Through biochemical analysis of the RISC complex, a number of associated proteins have been identified. One of these is the Fragile X mental retardation protein (FMRP). The gene encoding this protein (*FMR1*) is X-linked and its mutation is the cause of the most common inherited form of mental retardation. FMRP is an RNA-binding protein involved in gene regulation. Patients lacking FMRP have a range of developmental defects, as well as the mental retardation, due to disrupted gene expression.

Drosophila has an FMRP homolog. In flies deficient for this gene, unusual synaptic connections between neurons and muscles were observed. One of the *Drosophila* Argonaute proteins was found to be associated with FMRP, whereas separate studies of Argonaute found that FMRP was bound to that component of the RNAi machinery. Similar findings followed in human cells as well, indicating an intriguing connection between the Fragile X condition and the RNAi gene regulatory pathway.



BOX 18-3 FIGURE 1 miRNAs as tumor suppressors of oncogenes. (a) In this model, an miRNA that normally down-regulates an oncogene can function as a tumor suppressor gene. The loss of function of the miRNA by mutation or deletion, for example, might result in an abnormal expression of the target oncogene, which would then contribute to tumor formation. (b) Here, the amplification or overexpression of an miRNA that down-regulates a tumor suppressor or other important genes involved in differentiation may contribute to tumor formation by stimulating proliferation, angiogenesis, and invasion. (Redrawn, with permission, from Garzon R. et al. 2006. *Trends Mol. Med.* 12: 580–587, Fig. 2. © Elsevier.)

netic screen can be carried out. In such a library, for example, each plasmid would encode a shRNA directed against a different gene. The whole library is transfected into cells such that each cell receives a different shRNA. Cells with a particular phenotype are chosen, and the gene whose repression led to that phenotype can be identified.

REGULATORY RNAs AND X-INACTIVATION

X-inactivation Creates Mosaic Individuals

Female mammals have two X chromosomes, whereas males have only one X and a Y chromosome. Although this is the basis of sex determination—what enables males to differ from females—it also poses a problem: any gene encoded by the X chromosome would, if left unchecked, be expressed at twice the level in females as in males. This imbalance would potentially cause disruption to metabolic and other cellular processes. Avoiding such problems requires what is called **dosage compensation**. In mammals this is achieved by females **inactivating** one of their two X chromosomes. This action results in none of the genes on that copy of the chromosome being expressed. In placental mammals, inactivation occurs at the 32- to 64-cell stage, and the choice of which X chromosome to inactivate—the maternal or paternal copy—is apparently made at random in each cell at that time. Once selected in each cell, the same copy remains inactivated in all the descendants of that cell.

A consequence of inactivation being random in each cell is that females are mosaics—some of their cells express the paternal and others the maternal X chromosome. This is usually of little consequence, although it can influence the severity of symptoms of X-linked diseases, depending on the proportion of cells in which the mutated gene is expressed or silenced. A more familiar example is the calico (or tortoiseshell) cat (Fig 18-15). In cats, one gene on the X chromosome influences whether fur is orange or black—one allele of that gene gives rise to orange fur, another allele gives black. In cats heterozygous for this gene, the different patches of black and orange fur reveal regions made up of cells in which one or the other X chromosome was inactivated. This observation also explains why all calico cats are female. The white comes from effects of an autosomal gene.

FIGURE 18-15 Visualizing X-inactivation: the calico cat. The patches of orange and black fur provide an indirect visualization of X-chromosome inactivation, as described in the text. (Courtesy VG.)

Xist Is an RNA Regulator That Inactivates a Single X Chromosome in Female Mammals

How is an X chromosome inactivated, and how is inactivation inherited through the remainder of development? The initiating regulator is an RNA molecule called *Xist*. This RNA is encoded within the locus known to be vital for X-inactivation, the *Xic* (X-inactivation center) on the X chromosome. *Xist* RNA coats the X chromosome from which it is expressed. This is shown in the in situ in Figure 18-16a. It is not clear what causes this coating nor how it is restricted to one X chromosome (i.e., why it acts only in *cis*). It is, however, known that the action of *Xist* is central to inactivation and does not require other X-chromosomal sequences beyond *Xic*: when expressed ectopically from an autosomal location (i.e., from a non-sex chromosome), *Xist* can, to varying extents, silence genes along that chromosome. That is, it "inactivates" the autosome from which it is expressed.

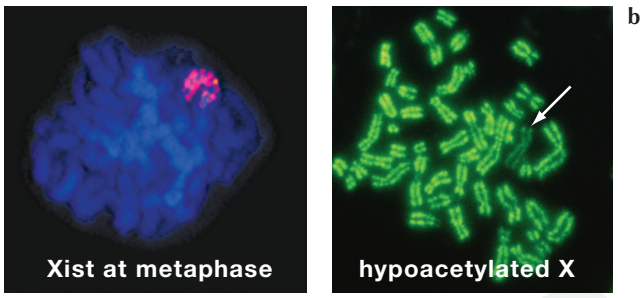


FIGURE 18-16 Visualizing X-inactivation: molecular markers. (a) Localization of *Xist* RNA along the inactive X-chromosome is shown by in situ hybridization in metaphase cells. (b) Chromosomes are stained for acetylation on histone H4. The arrow points to the inactivated X chromosome, which has much lower levels of acetylation than the other chromosomes. (Reprinted, with permission, from Brockdorff N. and Turner B.M. 2007. *Epigenetics* [ed. Allis et al.], p. 327. © Cold Spring Harbor Laboratory Press.)

Xist RNA itself does not cause silencing, but it recruits other factors that modify and condense chromatin and perhaps methylate DNA as well (just as we saw in other examples of silencing in Chapter 17). It is these modifications that cause silencing and ensure it is inherited; once firmly established, *Xist* itself is no longer required. One difference in histone modification of the inactivated X chromosome is shown in Figure 18-16b. There, the single inactivated X chromosome is much less acetylated than is the rest of the genome. As we saw in earlier chapters (7 and 17), deacetylated histones are associated with regions of the genome that are not transcribed.

How does a cell choose which X chromosome to inactivate? The answer is not yet known. But another RNA regulator might be key. This other RNA is also encoded by the *Xic* locus but on the opposite strand and overlapping the *Xist* gene. It is called *Tsix* (*Xist* spelled backward) and acts as a negative regulator of *Xist*. Indeed, if *Tsix* is mutated on a given X chromosome, it is that chromosome that will be chosen for inactivation. Thus, a balance between the production and stability of the *Xist* and *Tsix* RNAs may tilt the outcome one way or the other in each cell.

Dosage compensation is necessary in all animals—for example, worms and flies—just as it is in mammals. But in each case the mechanisms for achieving compensation are different. For example, in *Drosophila*, it is achieved by *increasing* expression of *X*-linked genes in the male (rather than decreasing them in the female). But there, too, the mechanism involves noncoding regulatory RNAs. In this case, the RNAs (called *roX1* and *roX2*) are involved in recruiting chromatin-modifying complexes to genes on the *X* chromosome in males, where they help activate transcription.

Noncoding RNAs are also found within some clusters of imprinted genes in the mouse. We discussed imprinting in gene expression in Chapter 17. In addition to the mechanisms described there, it is now believed that in some cases regulatory RNAs play a part, working in a manner analogous to *Xist* and *Tsix* in X-inactivation. The noncoding RNAs in question are regulated by the imprinting control region (ICR; see Fig. 17-26) found at each cluster, and they silence nearby genes on the same chromosome. Both DNA and histone methylation have been implicated in the silencing mechanism.

SUMMARY

Despite it being proposed as long ago as 1961 that RNA molecules were likely agents of gene regulation, it is only in the last few years that their widespread occurrence and significance in that role have come to light. Before that, attenuation of the *trp* operon in *E. coli* was a rare case where RNA sequences in the 5′ region of an mRNA were known to control expression of the downstream genes. In that case, alternative patterns of intramolecular base pairing within that region of the RNA give rise to alternative secondary structures that communicate different outcomes to the genes. In one conformation, transcription is terminated before it enters the coding region of the downstream genes, whereas in another conformation, it allows that transcription to continue, and the genes are expressed. Those genes encode enzymes involved in the synthesis of tryptophan, and the decision of whether or not to express them is based on the availability of uncharged $tRNA^{trp}$: if the uncharged species is rare in the cell, more Trp is needed and the genes are expressed; if there is ample charged $tRNA^{trp}$ in the cell, transcription of the *trp* genes is terminated.

Riboswitches control genes in a similar way: alternative secondary structures in the 5′-untranslated region of genes determine whether transcription of those genes continues (or, in other cases, whether translation is initiated). With riboswitches, the choice of alternative secondary structure depends on the direct binding to the RNA of the ligands that control the given gene. These are small metabolites (vitamin B12, SAM, amino acids) and the genes regulated by each one encode enzymes that use the metabolite in question.

E. coli also encodes small RNAs (called sRNAs) that act in *trans* to regulate genes. Thus, small genes encode short RNAs that base-pair with mRNAs bearing complementary sequences. This situation either inhibits translation of those target mRNAs, triggers their destruction, or even, in some cases, *stimulates* their translation. The actions of bacterial small RNAs are similar in many regards to those of small RNAs that regulate genes in eukaryotic cells, although the machinery used to produce these eukaryotic RNA regulators and the machinery used in carrying out their effects on target genes are quite different.

Eukaryotes use dsRNA as an agent of gene silencing, in a process called RNAi. Special enzymes (Drosha and, most generally, Dicer) recognize dsRNA and generate from that short (21–22-nucleotide) RNAs that are the active species for gene silencing. The dsRNA substrates from which these are generated can arise from two complementary strands base-pairing or a single molecule folding into a secondary structure with a characteristic double-strand region in an appropriate context. The former are acted on by Dicer, and the short products are called siRNAs; the latter require the action of Drosha followed by Dicer, and the products are called miRNAs. Both Dicer and Drosha have RNase domains and cut the substrate RNA on the basis of size and structure, rather than specific sequence.

Once produced, siRNAs and miRNAs act in essentially the same way. They are incorporated into a machine called RISC where one of the RNA strands is selected as the so-called guide RNA and directs the mature RISC complex to target RNAs with complementarity to that guide RNA. Once there, RISC either "slices" the RNA (through its catalytic subunit Argonaut, which includes an RNaseH-related domain) or inhibits the translation of the mRNA. Which route to silencing is chosen depends largely on how good the base-pairing match is between the guide RNA and the target—the higher the match, the more likely it is to trigger slicing. The guide RNA can also direct RISC with associated histone-modifying complexes to promoter regions where it silences genes transcriptionally by modifying their promoters. It is likely that even in these cases, recruitment to the promoter is through base pairing between the guide RNA and an mRNA, but in this instance, one still being made and thus still associated with RNA polymerase II at the gene.

miRNAs are encoded by genes within organisms where they typically act as regulators of genes involved in development—those from worms and plants are well-studied examples. miRNAs have also been associated with cancer, with some miRNAs being classified as tumor suppressors and others as oncogenes. The dsRNAs that give rise to siRNAs can arise from various sources ranging from infecting viruses, to transcribed repeat regions (centromeres or transposons), to dsRNA introduced into a cell deliberately by an experimenter who wants to down-regulate expression of a specific gene. This latter use of RNAi has become a regular tool and is particularly useful in systems where traditional genetics is not feasible.

Finally, we saw another case where noncoding RNAs regulate an important biological process. Female animals have two X chromosomes, whereas males have just one (and a Y chromosome). To ensure both sexes express comparable amounts of X-chromosome gene products, a mechanism of dosage compensation must correct for this unequal chromosome number. Mammals do this by inactivating one of the X chromosomes in females. An RNA molecule (*Xist*) encoded on the X chromosome regulates this process.

BIBLIOGRAPHY

Books

Hannon G.J. 2003. *RNAi: A guide to gene silencing*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Regulatory RNAs. 2006. *Cold Spring Harbor Symp. Quant. Biol.*, vol. 71. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Bacterial Small RNAs

Coppins R.L., Hall K.B., and Groisman E.A. 2007. The intricate world of riboswitches. *Curr. Opin. Microbiol.* **10:** 176–181.

Gottesman S. 2005. Micros for microbes: Non-coding regulatory RNAs in bacteria. *Trends Genet.* 21: 399–404. Review.

Storz G. and Haas D. 2007. A guide to small RNAs in microorganisms. *Curr. Opin. Microbiol.* **10:** 93–95.

Riboswitches and Attenuation

Gilbert S.D. and Batey R.T. 2006. Riboswitches: Fold and function. *Chem. Biol.* **13:** 805–807.

Gollnick P., Babitzke P., Antson A., and Yanofsky C. 2005. Complexity in regulation of tryptophan biosynthesis in *Bacillus subtilis*. *Annu. Rev. Genet.* **39:** 47–68.

Winkler W.C. 2005. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr. Opin. Chem. Biol.* **9:** 594–602.

Winkler W.C. and Breaker R.R. 2005. Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.* **59:** 487–517.

Yanofsy C. 2000. Transcription attenuation: Once viewed as a novel regulatory strategy. *J. Bacteriol.* **182:** 1–8.

Mechanisms of RNAi

Baulcombe D. 2005. RNA silencing. *Trends Biochem. Sci.* **30:** 290–293.

Cook A. and Conti E. 2006. Dicer measures up. *Nat. Struct. Mol. Biol.* **13:** 190–192.

Hannon G.J. and Rossi J.J. 2004. Unlocking the potential of the human genome with RNA interference. *Nature* **431:** 371–378.

Pei Y. and Tuschl T. 2006. On the art of identifying effective and specific siRNAs. *Nat. Methods* **3:** 670–676.

Peters L. and Meister G. 2007. Argonaute proteins: Mediators of RNA silencing. *Mol. Cell* **26:** 611–623.

Rana T.M. 2007. Illuminating the silence: Understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell Biol.* **8:** 23–36.

Tang G. 2005. siRNA and miRNA: An insight into RISCs. *Trends Biochem. Sci.* **30:** 106–114.

Tolia N.H. and Joshua-Tor L. 2007. Slicer and the Argonautes. *Nat. Chem. Biol.* **3:** 36–43.

Valencia-Sanchez M.A., Liu J., Hannon G.J., and Parker R. 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* **20:** 515–524.

Zaratiegui M., Irvine D.V., and Martienssen R.A. 2007. Noncoding RNAs and gene silencing. *Cell* **128:** 763–776.

MicroRNAs

Ambros V. 2004. The functions of animal microRNAs. *Nature* **431:** 350–355.

Bushati N. and Cohen S.M. 2007. microRNA functions. *Annu. Rev. Cell Dev. Biol.* **23:** 175–205.

Hannon G.J., Rivas F.V., Murchison E.P., and Steitz J.A. 2006. The expanding universe of noncoding RNAs. *Cold Spring Harbor Symp. Quant. Biol.* **71:** 551–564.

History of miRNAs and RNAi

Baulcombe D. 1994. Replicase-mediated resistance: A novel type of virus resistance in transgenic plants? *Trends Microbiol.* **2:** 60–63.

Lee R., Feinbaum R., and Ambros V. 2004. A short history of a short RNA. *Cell* (suppl. 2) **116:** S89–S92, 1 p. following S96.

Ruvkun G., Wightman B., and Ha I. 2004. The 20 years it took to recognize the importance of tiny RNAs. *Cell* **116**(suppl. 2)**:** S93-S96, 2 p. following S96.

MicroRNAs and Disease

Calin G.A. and Croce C.M. 2006. MicroRNA signatures in human cancers. *Nat. Rev. Cancer* **6:** 857–866.

Dalmay T. and Edwards D.R. 2006. MicroRNAs and the hallmarks of cancer. *Oncogene* **25:** 6170–6175.

Garzon R., Fabbri M., Cimmino A., Calin G.A., and Croce C.M. 2006. MicroRNA expression and function in cancer. *Trends Mol. Med.* **12:** 580-587.

Zhang B., Pan X., Cobb G.P., and Anderson T.A. 2007. microRNAs as oncogenes and tumor suppressors. *Dev. Biol.* **302:** 1–12.

Xist RNA and X-inactivation

Alexander M.K. and Panning B. 2005. Counting chromosomes: Not as easy as 1, 2, 3. *Curr. Biol.* **15:** R834–R836.

Deng X and Meller V.K. 2006. Non-coding RNA in fly dosage compensation. *Trends Biochem. Sci.* **31:** 526–532.

Heard E. 2005. Delving into the diversity of facultative heterochromatin: The epigenetics of the inactive X chromosome. *Curr. Opin. Genet. Dev.* **15:** 482–489.

Ng K., Pullirsch D., Leeb M., and Wutz A. 2006. *Xist* and the order of silencing. *EMBO Rep.* **8:** 34–39.

Pauler F.M., Koerner M.V., and Barlow D.P. 2007. Silencing by imprinted noncoding RNAs: Is transcription the answer? *Trends Genet.* **23:** 284–292.