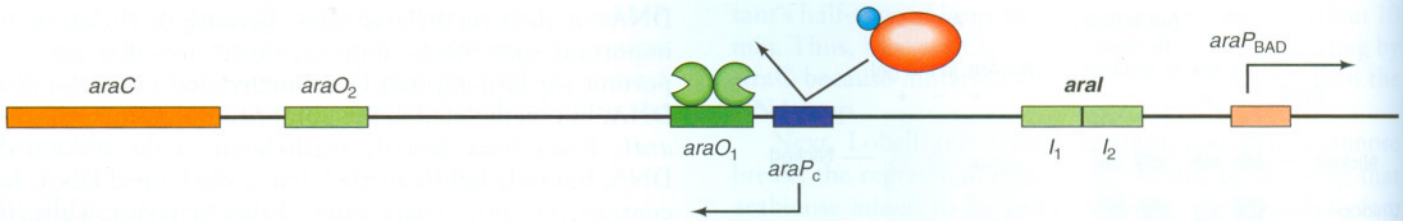

Figure 7.26 Autoregulation of *araC*. AraC (green) binds to $araO_1$ and prevents transcription leftward from P_C through the *araC* gene. This can presumably happen whether or not arabinose is bound to AraC, that is, with the control region either unlooped or looped.

conditions to the $araI_2$ mutant DNA. We infer that $araI_2$ is necessary for AraC binding in the unlooped state and is therefore contacted by AraC under these conditions.

These data suggest the model of AraC–DNA interaction that was depicted in Figure 7.21b and c. A dimer of AraC causes looping by simultaneously interacting with $araI_1$ and $araO_2$. Arabinose breaks the loop by changing the conformation of AraC so the protein loses its affinity for $araO_2$ and binds instead to $araI_2$.

Autoregulation of *araC*

So far, we have only mentioned a role for $araO_1$. It does not take part in repression of *araBAD* transcription; instead it allows AraC to regulate its own synthesis. Figure 7.26 shows the relative positions of *araC*, P_C, and $araO_1$. The *araC* gene is transcribed from P_C in the leftward direction, which puts $araO_1$ in a position to control this transcription. As the level of AraC rises, it binds to $araO_1$ and inhibits leftward transcription, thus preventing an accumulation of too much repressor. This kind of mechanism, where a protein controls its own synthesis, is called **autoregulation.**

SUMMARY The *ara* operon is controlled by the AraC protein. AraC represses the operon by looping out the DNA between two sites, $araO_2$ and $araI_1$, that are 210 bp apart. Arabinose can derepress the operon by causing AraC to loosen its attachment to $araO_2$ and to bind to $araI_2$ instead. This breaks the loop and allows transcription of the operon. CAP and cAMP further stimulate transcription by binding to a site upstream of *araI*. AraC controls its own synthesis by binding to $araO_1$ and preventing leftward transcription of the *araC* gene.

7.3 The *trp* Operon

The *E. coli* ***trp*** (pronounced "trip") **operon** contains the genes for the enzymes that the bacterium needs to make the amino acid tryptophan. Like the *lac* operon, it is subject to negative control by a repressor. However, there is a fundamental difference. The *lac* operon codes for **catabolic** enzymes—those that break down a substance. Such operons tend to be turned on by the presence of that substance, lactose in this case. The *trp* operon, on the other hand, codes for **anabolic** enzymes—those that build up a substance. Such operons are generally turned off by that substance. When the tryptophan concentration is high, the products of the *trp* operon are not needed any longer, and we would expect the *trp* operon to be repressed. That is what happens. The *trp* operon also exhibits an extra level of control, called attenuation, not seen in the *lac* operon.

Tryptophan's Role in Negative Control of the *trp* Operon

Figure 7.27 shows an outline of the structure of the *trp* operon. Five genes code for the polypeptides in the enzymes that convert a tryptophan precursor, chorismic acid, to tryptophan. In the *lac* operon, the promoter and operator precede the genes, and the same is true in the *trp* operon. However, the *trp* operator lies wholly within the *trp* promoter, whereas the two loci are merely adjacent in the *lac* operon.

In the negative control of the *lac* operon, the cell senses the presence of lactose by the appearance of tiny amounts of its rearranged product, allolactose. In effect, this inducer causes the repressor to fall off the *lac* operator and derepresses the operon. In the case of the *trp* operon, a plentiful supply of tryptophan means that the cell does not need to spend any more energy making this amino acid. In other words, a high tryptophan concentration is a signal to turn off the operon.

How does the cell sense the presence of tryptophan? In essence, tryptophan helps the *trp* repressor bind to its operator. Here is how that occurs: In the absence of tryptophan, no *trp* repressor exists—only an inactive protein called the **aporepressor.** When the aporepressor binds tryptophan, it changes to a conformation with a much higher affinity for the *trp* operator (Figure 7.27b). This is another allosteric transition like the one we encountered in our discussion of the *lac* repressor. The combination of aporepressor plus tryptophan is the ***trp* repressor;** therefore,

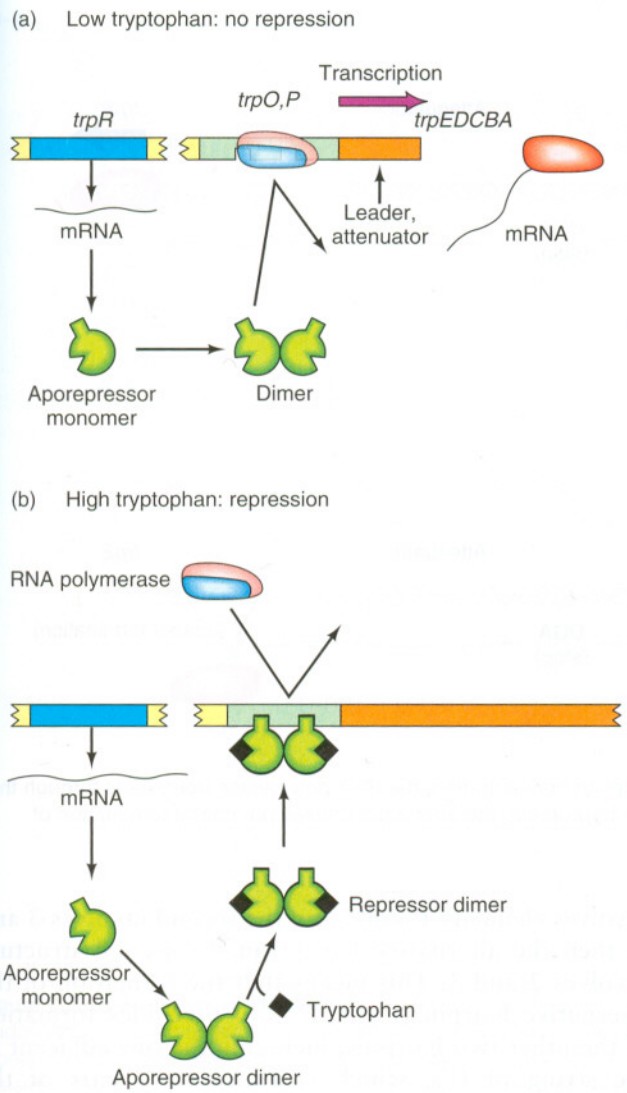

Figure 7.27 Negative control of the *trp* operon. (a) Derepression. RNA polymerase (red and blue) binds to the *trp* promoter and begins transcribing the structural genes (*trpE, D, C, B,* and *A*). Without tryptophan, the aporepressor (green) cannot bind to the operator. **(b)** Repression. Tryptophan, the corepressor (black), binds to the inactive aporepressor, changing it to repressor, with the proper shape for binding successfully to the *trp* operator. This prevents RNA polymerase from binding to the promoter, so no transcription occurs.

tryptophan is called a **corepressor.** When the cellular concentration of tryptophan is high, plenty of corepressor is available to bind and form the active *trp* repressor. Thus, the operon is repressed. When the tryptophan level in the cell falls, the amino acid dissociates from the aporepressor, causing it to shift back to the inactive conformation; the repressor–operator complex is thus broken, and the operon is derepressed. In Chapter 9, we will examine the nature of the conformational shift in the aporepressor that occurs on binding tryptophan and see why this is so important in operator binding.

SUMMARY The negative control of the *trp* operon is, in a sense, the mirror image of the negative control of the *lac* operon. The *lac* operon responds to an inducer that causes the repressor to dissociate from the operator, derepressing the operon. The *trp* operon responds to a repressor that includes a corepressor, tryptophan, which signals the cell that it has made enough of this amino acid. The corepressor binds to the aporepressor, changing its conformation so it can bind better to the *trp* operator, thereby repressing the operon.

Control of the *trp* Operon by Attenuation

In addition to the standard, negative control scheme we have just described, the *trp* operon employs another mechanism of control called **attenuation.** Why is this extra control needed? The answer probably lies in the fact that repression of the *trp* operon is weak—much weaker, for example, than that of the *lac* operon. Thus, considerable transcription of the *trp* operon can occur even in the presence of repressor. In fact, in attenuator mutants where only repression can operate, the fully repressed level of transcription is only 70-fold lower than the fully derepressed level. The attenuation system permits another 10-fold control over the operon's activity. Thus, the combination of repression and attenuation controls the operon over a 700-fold range, from fully inactive to fully active: (70-fold [repression] × 10-fold [attenuation] = 700-fold). This is valuable because synthesis of tryptophan requires considerable energy.

Here is how attenuation works. Figure 7.27 lists two loci, the ***trp* leader** and the ***trp* attenuator,** in between the operator and the first gene, *trpE*. Figure 7.28 gives a closer view of the leader–attenuator, whose purpose is to attenuate, or weaken, transcription of the operon when tryptophan is relatively abundant. The attenuator operates by causing premature termination of transcription. In other words, transcription that gets started, even though the tryptophan concentration is high, stands a 90% chance of terminating in the attenuator region.

The reason for this premature termination is that the attenuator contains a transcription stop signal (terminator): an inverted repeat followed by a string of eight A–T pairs in a row. Because of the inverted repeat, the transcript of this region would tend to engage in intramolecular base pairing, forming the "hairpin" shown in Figure 7.29. As we learned in Chapter 6, a hairpin followed by a string of U's in a transcript destabilizes the binding between the transcript and the DNA and thus causes termination.

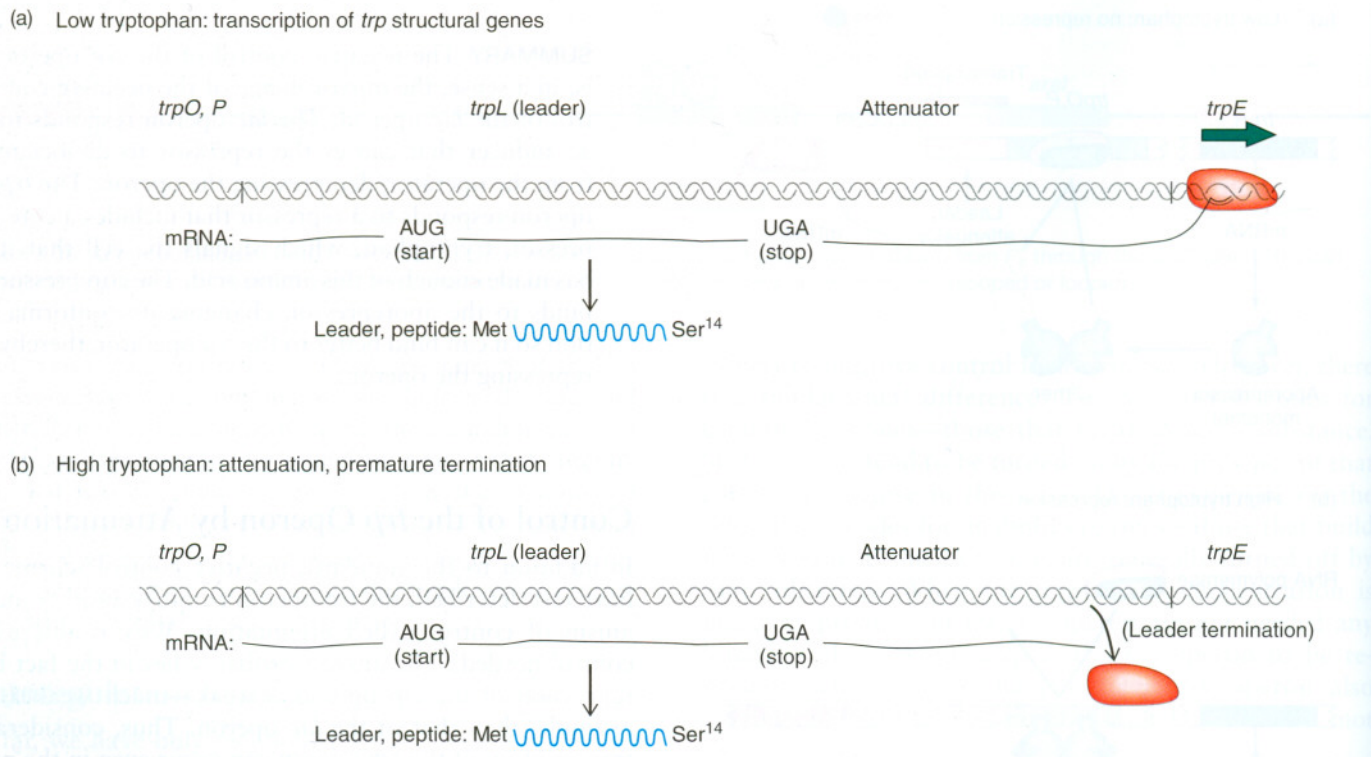

Figure 7.28 Attenuation in the *trp* operon. **(a)** In the presence of low tryptophan concentration, the RNA polymerase (red) reads through the attenuator, so the structural genes are transcribed. **(b)** In the presence of high tryptophan, the attenuator causes premature termination of transcription, so the structural genes are not transcribed.

SUMMARY Attenuation imposes an extra level of control on an operon, over and above the repressor–operator system. It operates by causing premature termination of transcription of the operon when the operon's products are abundant.

Defeating Attenuation

When tryptophan is scarce, the *trp* operon must be activated, and that means that the cell must somehow override attenuation. Charles Yanofsky proposed this hypothesis: Something preventing the hairpin from forming would destroy the termination signal, so attenuation would break down and transcription would proceed. A look at Figure 7.30a reveals not just one potential hairpin near the end of the leader transcript, but two. However, the terminator includes only the second hairpin, which is adjacent to the string of U's in the transcript. Furthermore, the two-hairpin arrangement is not the only one available; another, containing only one hairpin, is shown in Figure 7.30b. Note that this alternative hairpin contains elements from each of the two hairpins in the first structure. Figure 7.30 illustrates this concept by labeling the sides of the original two hairpins 1, 2, 3, and 4. If the first of the original hairpins involves elements 1 and 2 and the second involves 3 and 4, then the alternative hairpin in the second structure involves 2 and 3. This means that the formation of the alternative hairpin (Figure 7.30b) precludes formation of the other two hairpins, including the one adjacent to the string of U's, which is a necessary part of the terminator (Figure 7.30a).

The two-hairpin structure involves more base pairs than the alternative, one-hairpin structure; therefore, it is more stable. So why should the less stable structure ever form? A clue comes from the base sequence of the leader region shown in Figure 7.31. One very striking feature of this sequence is that two codons for tryptophan (UGG) occur in a row in element 1 of the first potential hairpin. This may not seem unusual, but tryptophan (Trp) is a rare amino acid in most proteins; it is found on average only once in every 100 amino acids. So the chance of finding two Trp codons in a row *anywhere* is quite small, and the fact that they are found in the *trp* operon is very suspicious.

In bacteria, transcription and translation occur simultaneously. Thus, as soon as the *trp* leader region is transcribed, ribosomes begin translating this emerging mRNA. Think about what would happen to a ribosome trying to translate the *trp* leader under conditions of tryptophan starvation (Figure 7.32a). Tryptophan is in short supply, and here are two demands in a row for that

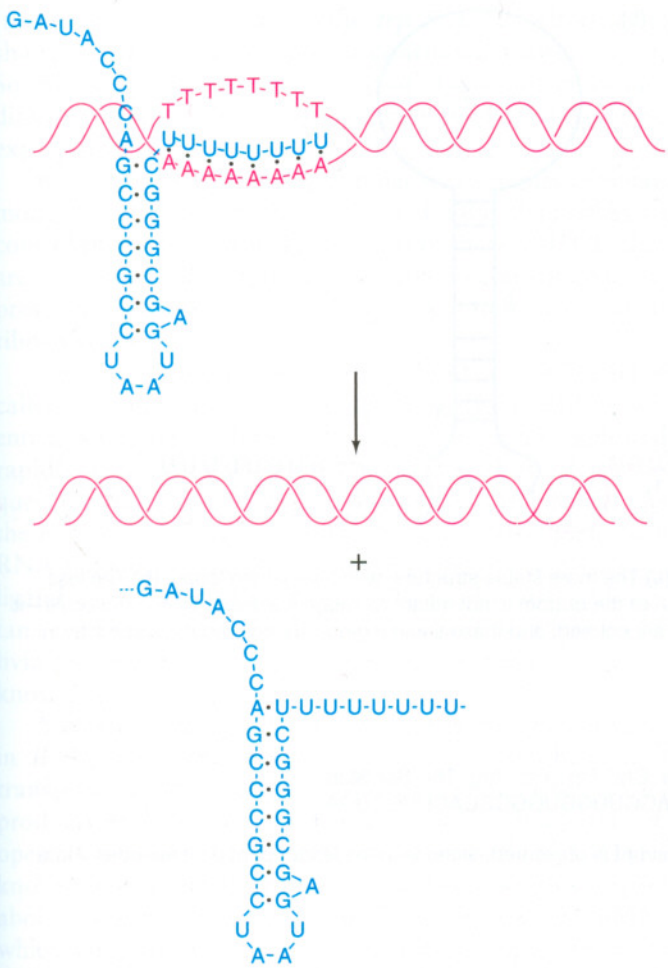

Figure 7.29 Mechanism of attenuation. The transcript of the leader–attenuator region (blue) contains a string of U's and an inverted repeat, so it can form a hairpin structure. When the RNA polymerase pauses at the string of U's, the hairpin forms, then the transcript is released, so termination occurs before transcription can reach the *trp* genes.

very amino acid. In all likelihood, the ribosome will not be able to satisfy those demands immediately, so it will pause at one of the Trp codons. And where does that put the stalled ribosome? Right on element 1, which should be participating in formation of the first hairpin. The bulky ribosome clinging to this RNA site effectively prevents its pairing with element 2, which frees 2 to pair with 3, forming the one-hairpin alternative structure. Because the second hairpin (elements 3 and 4) cannot form, transcription does not terminate and attenuation has been defeated. This is desirable, of course, because when tryptophan is scarce, the *trp* operon should be transcribed.

Notice that this mechanism involves a coupling of transcription and translation, where the latter affects the former. It would not work in eukaryotes, where transcription and translation take place in separate compartments. It also depends on transcription and translation occurring at about the same rate. If RNA polymerase outran the ribosome, it might pass through the attenuator region before the ribosome had a chance to stall at the Trp codons.

You may be wondering how the polycistronic mRNA made from the *trp* operon can be translated if ribosomes are stalled in the leader at the very beginning. The answer is that each of the genes represented on the mRNA has its own translation start signal (AUG). Ribosomes recognize each of these independently, so translation of the *trp* leader does not affect translation of the *trp* genes.

On the other hand, consider a ribosome translating the leader transcript under conditions of abundant tryptophan (Figure 7.32b). Now the dual Trp codons present no barrier to translation, so the ribosome continues through element 1 until it reaches the stop signal (UGA) between elements 1 and 2 and falls off. With no ribosome to interfere, the two hairpins can form, completing the transcription termination signal that halts transcription before it reaches the *trp* genes. Thus, the attenuation system responds to the presence of adequate tryptophan and prevents wasteful synthesis of enzymes to make still more tryptophan.

Other *E. coli* operons besides *trp* use the attenuation mechanism. The most dramatic known use of consecutive codons to stall a ribosome occurs in the *E. coli* histidine *(his)* operon, in which the leader region contains seven histidine codons in a row!

SUMMARY Attenuation operates in the *E. coli trp* operon as long as tryptophan is plentiful. When the supply of this amino acid is restricted, ribosomes stall at the tandem tryptophan codons in the *trp* leader. Because the *trp* leader is being synthesized just as stalling occurs, the stalled ribosome will influence the way this RNA folds. In particular, it prevents the formation of a hairpin, which is part of the transcription termination signal that causes attenuation. Therefore, when tryptophan is scarce, attenuation is defeated and the operon remains active. This means that the control exerted by attenuation responds to tryptophan levels, just as repression does.

7.4 Riboswitches

We have just seen an example of controlling gene expression by manipulating the structure of the 5′-untranslated region (UTR) of an mRNA (the *trp* mRNA of *E. coli*). In this case, a macromolecular assembly (the ribosome) senses the concentration of a small molecule

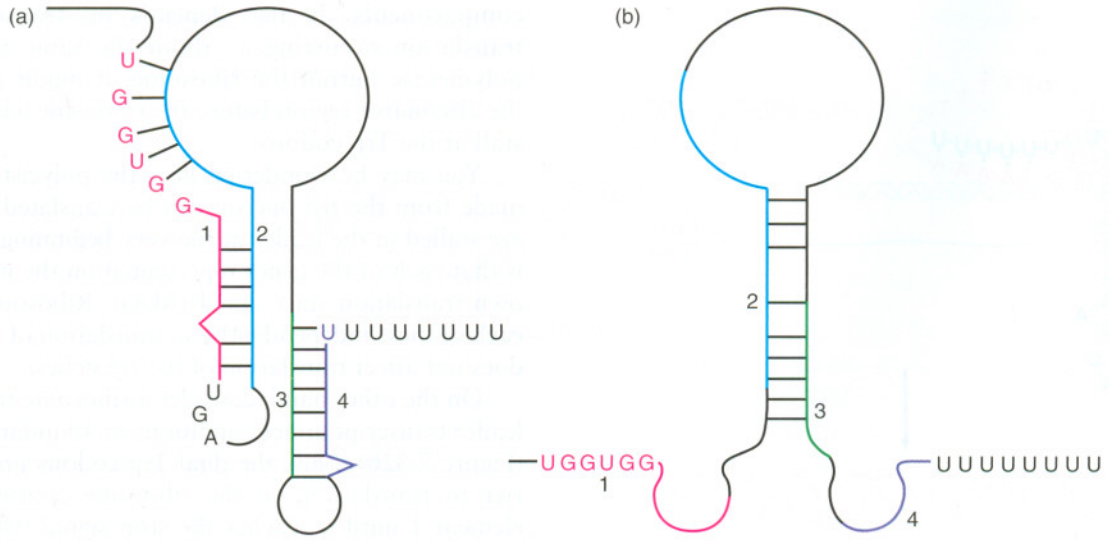

Figure 7.30 Two structures available to the leader–attenuator transcript. (a) The more stable structure, with two hairpin loops. **(b)** The less stable structure, containing only one hairpin loop. The curved shape of the RNA at the bottom is not meant to suggest a shape for the molecule—it is drawn this way simply to save space. The base-paired segments (1–4) in (a) are colored, and these same regions are colored the same way in (b) so they can be recognized.

Met Lys Ala Ile Phe Val Leu Lys Gly Trp Trp Arg Thr Ser Stop
pppA---AUGAAAGCAAUUUUCGUACUGAAAGGUUGGUGGCGCACUUCCUGA

Figure 7.31 Sequence of the leader. The sequence of part of the leader transcript is presented, along with the leader peptide it encodes. Note the two Trp codons in tandem (blue).

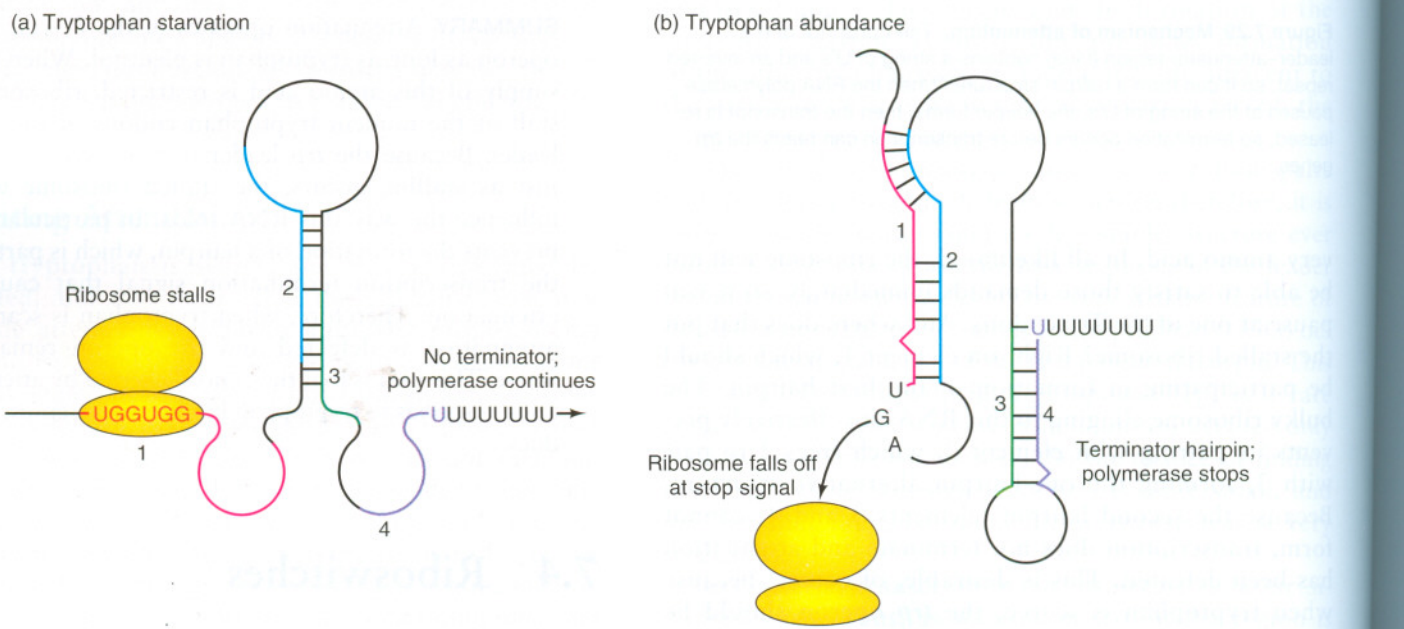

Figure 7.32 Overriding attenuation. (a) Under conditions of tryptophan starvation, the ribosome (yellow) stalls at the Trp codons and prevents element 1 (red) from pairing with element 2 (blue). This forces the one-hairpin structure, which lacks a terminator, to form, so no attenuation should take place. **(b)** Under conditions of tryptophan abundance, the ribosome reads through the two tryptophan codons and falls off at the translation stop signal (UGA), so it cannot interfere with base pairing in the leader transcript. The more stable, two-hairpin structure forms; this structure contains a terminator, so attenuation occurs.

(tryptophan) and binds to the *trp* 5′-UTR, altering its shape, thereby controlling its continued transcription. So this is an example of a group of macromolecules mediating the effect of a small molecule (or *ligand*) on gene expression.

We also have a growing number of examples of small molecules acting *directly* on the 5′-UTRs of mRNAs to control their expression. The regions of these 5′-UTRs that are capable of altering their structures to control gene expression in response to ligand binding are called **riboswitches.**

The region of a riboswitch that binds to the ligand is called an **aptamer.** Aptamers were first discovered by scientists studying evolution in a test tube, who exploited rapidly replicating RNAs to select for short RNA sequences that bind tightly and specifically to ligands. As the RNAs replicate, they make mistakes, producing new RNA sequences, and those that bind best to a particular ligand are selected. Experimenters found many such aptamers in these in vitro experiments and wondered why living things did not take advantage of them. Now we know that they do.

A classic example of a riboswitch is the *ribD* operon in *B. subtilis*. This operon controls the synthesis and transport of the vitamin riboflavin and one of its products, flavin mononucleotide (FMN). Bacterial *rib* operons contain a conserved element in their 5′-UTRs known as the ***RFN* element.** Mutations in this region abolish normal control of the *ribD* operon by FMN, which led to the hypothesis that this *RFN* element interacts with a protein that responds to FMN or, perhaps, with FMN itself.

To test the hypothesis that the *RFN* element is an aptamer that binds directly to FMN, Ronald Breaker and colleagues used a technique called **in-line probing.** This method relies on the fact that efficient hydrolysis (breakage) of a phosphodiester bond in RNA needs a 180° ("in-line") arrangement among the attacking nucleophile (water), the phosphorus atom in the phosphodiester bond, and the leaving hydroxyl group at the end of one of the RNA fragments created by the hydrolysis. Unstructured RNA can easily assume this in-line conformation, but RNA that is constrained by secondary structure (intramolecular base pairing) cannot. Thus, spontaneous cleavage of linear, unstructured RNA will occur much more readily than will cleavage of a structured RNA with lots of base pairing.

Thus, Breaker and colleagues incubated a labeled RNA fragment containing the *RFN* element in the presence and absence of FMN. Figure 7.33a shows that the patterns of spontaneous hydrolysis of the RNA were different in the presence and absence of FMN, suggesting that FMN binds directly to the RNA and causes it to shift its conformation. This is what we would expect of an aptamer bound to its ligand.

In particular, Breaker and colleagues found that FMN binding rendered certain phosphodiester bonds less susceptible to cleavage, whereas others retained their normal susceptibility (Figure 7.33b). Furthermore, the changes in susceptibility were half-maximal at an FMN concentration of only 5 nM. This indicates high affinity between the RNA and its ligand.

The patterns of decreased susceptibility to cleavage in the presence of FMN suggested the two alternative conformations of the *RFN* element depicted in Figure 7.33c. In the absence of FMN, the element should form an antiterminator, with the hairpin remote from the string of six U's. But FMN would cause the conformation of the element to shift such that it forms a terminator, blocking expression of the operon. This makes sense because, with abundant FMN, there is no need to express the *ribD* operon, so the proposed attenuation by FMN would save the cell energy.

To test this hypothesis, Breaker and colleagues performed an in vitro transcription assay with a cloned DNA template containing both the *RFN* element and the proposed terminator. They found that transcription terminated about 10% of the time at the terminator even in the absence of FMN, but FMN raised the frequency of termination to 30%. They mapped the termination site with a run-off transcription assay (Chapter 5) and showed that transcription terminated right at the end of the string of U's. Next, they used a mutant version of the DNA template that encoded fewer than six U's in the putative terminator. In this case, FMN caused no change in the frequency of termination, presumably because the shorter string of U's considerably lowered the efficiency of the terminator, even with FMN. Thus, with the wild-type gene, FMN really does appear to force more of the growing transcripts to form terminators that halt transcription.

We will see another example of a riboswitch in Chapter 17, when we study the control of translation. We will learn that a ligand can bind to a riboswitch in an mRNA's 5′-UTR, and can control translation of that mRNA by changing the conformation of the 5′-UTR to hide the ribosome-binding site.

These two examples of riboswitches both operate by depressing gene expression: one at the transcriptional level, and one at the translational level. Indeed, all riboswitches studies to date work that way, although there is no reason why a riboswitch could not work by stimulating gene expression. These two examples, among others, also lead to a general model for riboswitches (Figure 7.34). They are regions in the 5′-UTRs of mRNAs that contain two modules: an aptamer and another module, which Breaker and colleagues call an *expression platform*. The expression platform can be a terminator, a ribosome-binding site, or another RNA element that affects gene expression. By binding to its aptamer and

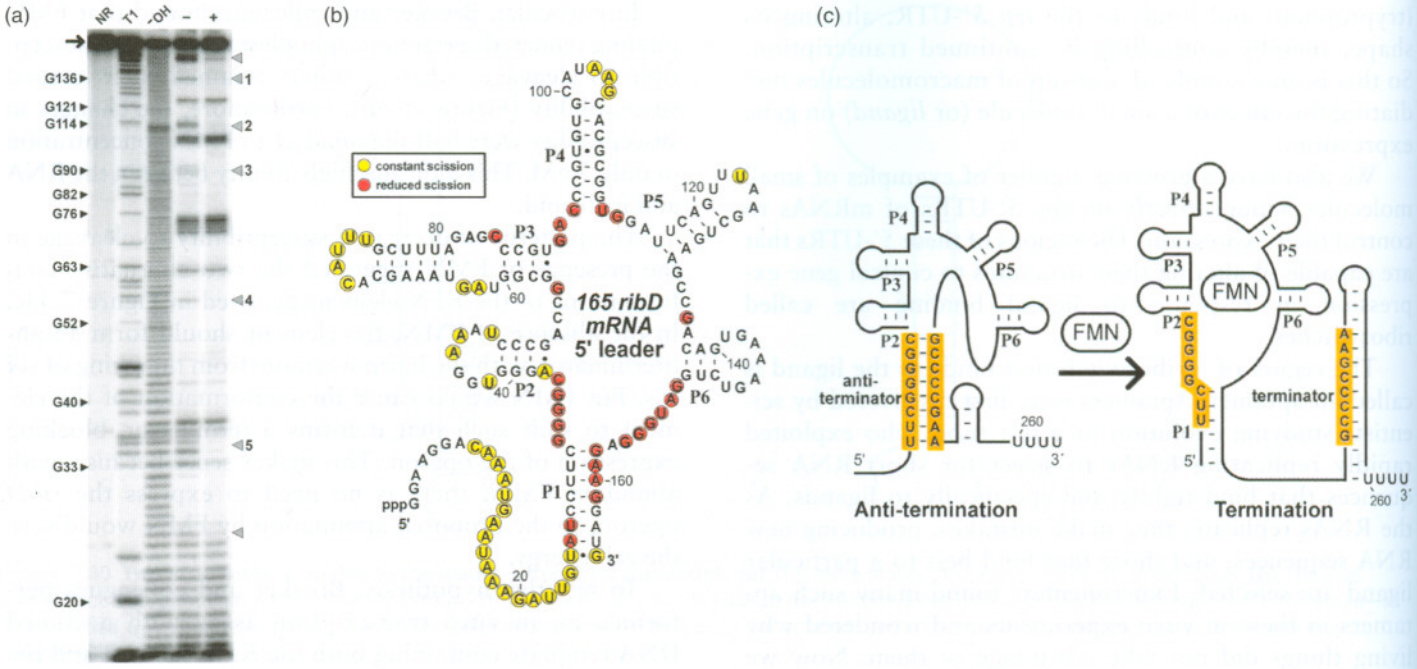

Figure 7.33 Results of in-line probing of *RFN* element and model for the action of the *ribD* riboswitch. (a) Gel electrophoresis results of in-line probing. Lane 1, no RNA; lane 2, RNA cut with RNase T1; lane 3, RNA cut with base; lanes 4 and 5, RNAs subjected to spontaneous cleavage in the absence (–) and presence (+) of FMN for 40 h at 25°C. Arrows at right denote regions of the RNA that became less susceptible to cleavage in the presence of FMN. **(b)** Sequence of part of the 5′-UTR of the *B. subtilis ribD* mRNA, showing the internucleotide linkages that became less susceptible to spontaneous cleavage upon FMN binding (red), and those that showed constant susceptibility (yellow). The secondary structure of the element is based on comparisons of sequences of many *RFN* elements. **(c)** Proposed change in structure of the riboswitch upon FMN binding. In the absence of FMN, base pairing between the two yellow regions forces the riboswitch to assume an antiterminator conformation, with the hairpin remote from the string of U's. Conversely, binding of FMN to the growing mRNA allows the GCCCCGAA sequence to base-pair with another part of the riboswitch, creating a terminator that stops transcription. *(Source: (a-c)* © 2002 National Academy of Science. Proceedings of the National Academy of Sciences, vol. 99, no. 25, December 10, 2002, p. 15908–15913 "An mRNA structure that controls gene expression by binding FMN," Chalamish, and Ronald R. Breaker, fig.1, p. 15909 & fig. 3, p. 15911.)

changing the conformation of the riboswitch, a ligand can affect an expression platform, and thereby control gene expression.

Note that a riboswitch is another example of allosteric control, that is, one in which a ligand causes a conformational change in a large molecule that in turn affects the ability of the large molecule to interact with something else. We encountered an allosteric mechanism earlier in this chapter in the context of the *lac* operon, where a ligand (allolactose) bound to a protein (*lac* repressor) and interfered with its ability to bind to the *lac* operator. In fact, many examples of allosteric control are known, but up until recently they all involved allosteric proteins. Riboswitches work similarly, except that the large molecule is an RNA, rather than a protein.

Finally, riboswitches may provide a window on the "RNA world," a hypothetical era early in the evolution of life, in which proteins and DNA had not yet evolved. In this world, genes were made of RNA, not DNA, and enzymes were made of RNA, not protein. (We will see modern examples of catalytic RNAs in Chapters 14, 17, and 19.) Without proteins to control their genes, life forms in the RNA world would have had rely on small molecules interacting directly with their genes. If this hypothesis is true, riboswitches are relics of one of the most ancient forms of genetic control.

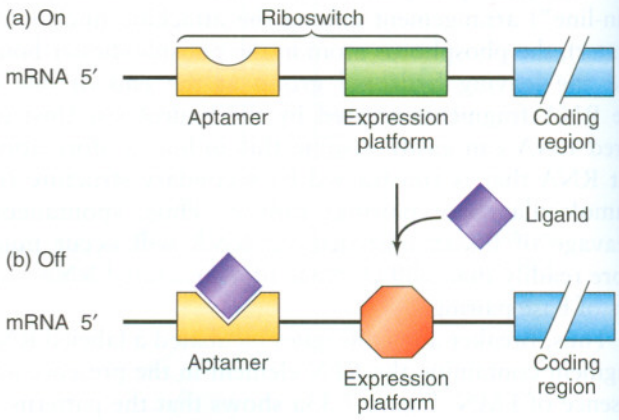

Figure 7.34 A model for riboswitch action. (a) Absence of the ligand. Gene expression is turned on. **(b)** Presence of the ligand. The ligand has bound to the aptamer in the riboswitch, causing a change in the conformation of the riboswitch, including the expression platform. This turns gene expression off.