CHAPTER 4

Plasmids

What Is a Plasmid?

IN ADDITION TO A CHROMOSOME, BACTERIAL cells often contain plasmids. These DNA molecules are found in essentially all types of bacteria and, as discussed below, play a significant role in bacterial adaptation and evolution. They also serve as important tools in studies of molecular biology. We address such uses later in this chapter.

Plasmids, which vary widely in size from a few thousand to hundreds of thousands of base pairs (bp) (a size comparable to that of the bacterial chromosome), are most often circular molecules of double-stranded DNA. However, some bacteria have linear plasmids, and some plasmids, most often those from gram-positive bacteria, can accumulate single-stranded DNA owing to aberrant rolling-circle replication (discussed below). The number of copies also varies among plasmids, and bacterial cells can harbor more than one type. Thus, a cell can harbor two or more different types of plasmids, with hundreds of copies of some plasmid types and only one or a few copies of other types.

Like chromosomes, plasmids encode proteins and RNA molecules and replicate as the cell grows, and the replicated copies are usually distributed into each daughter cell when the cell divides. Plasmids even share some of the same types of functions for accurate partitioning (Par functions) and site-specific recombinases with the host chromosome (see below). By one definition, any independently replicating element in the cell that does not contain genes essential for bacterial growth (the so-called housekeeping genes) is called a plasmid. Plasmids probably persist because they very often provide gene products that can benefit the bacterium under certain circumstances. Consequently, isolates of bacteria taken from the environment often will lose some or all plasmids over time when cultured in the laboratory. There are a number of examples where a plasmid has taken on many

doi:10.1128/9781555817169.ch4

of the attributes of a chromosome, such as larger size and encoding multiple housekeeping genes. For example, the pSymB plasmid of some *Sinorhizobium* species is about half as big as the chromosome and carries essential genes, including a gene for an arginine tRNA and the *minCDE* genes involved in division site selection. Also, *Vibrio cholerae* has two large DNA molecules, both of which carry essential genes. *Agrobacterium tumefaciens* has two large DNA molecules, one circular and the other linear, both of which carry essential genes. In cases where a plasmid is almost as big as a chromosome and carries essential genes, which one is the chromosome and which is a plasmid? Probably a better criterion for whether a DNA molecule is a plasmid or the chromosome is the nature of its origin of replication. In all known cases, one of the large DNA molecules has a typical bacterial origin of replication, with an *oriC* site and closely linked *dnaA*, *dnaN*, and *gyrA* genes, among others, while the other DNA molecule has a typical plasmid origin, with *repABC*-like genes more characteristic of plasmids.

Naming Plasmids

Before methods for physical detection of plasmids became available, plasmids made their presence known by conferring phenotypes on the cells harboring them. Consequently, many plasmids were named after the genes they carry. For example, R-factor plasmids contain genes for resistance to several antibiotics (hence the name R, for resistance). These were the first plasmids discovered, when *Shigella* and *Escherichia coli* strains resistant to a number of antibiotics were isolated from the fecal flora of patients in Japan in the late 1950s. The ColE1 plasmid, from which many of the cloning vectors were derived, carries a gene for the protein colicin E1, a bacteriocin that kills bacteria that do not carry the plasmid. The Tol plasmid contains genes for the degradation of toluene, and the Ti plasmid of *A. tumefaciens* carries genes for tumor induction in plants (see Box 5.1). This system of nomenclature has led to some confusion, because plasmids carry various genes besides the ones for which they were originally named and because similar plasmids can contain very different sets of genes. Many plasmids have also been altered beyond recognition in the laboratory to make plasmid cloning vectors (see below) and for other purposes.

To avoid further confusion, the naming of plasmids is now standardized. Plasmids are given number-and-letter names much like bacterial strains. A small "p," for plasmid, precedes capital letters that describe the plasmid or sometimes give the initials of the person or persons who isolated or constructed it. These letters are often followed by numbers to identify the particular construct. When the plasmid is further altered, a different number is assigned to indicate the change. For example, plasmid pBR322 was constructed by Bolivar and Rodriguez from the ColE1 plasmid and is derivative number 322 of the plasmids they constructed. pBR325 is pBR322 with a chloramphenicol resistance gene inserted. The new number, 325, distinguishes the plasmid from pBR322.

Functions Encoded by Plasmids

Plasmids can encode a few or hundreds of different proteins, resulting in vast differences in their sizes. However, as mentioned above, plasmids rarely encode gene products that are always essential for growth, such as RNA polymerase, ribosomal subunits, or enzymes of the tricarboxylic acid cycle. Instead, plasmid genes usually give bacteria a selective advantage under only some conditions.

Table 4.1 lists a few naturally occurring plasmids and some traits they encode, as well as the host in which they were originally found. As mentioned above, gene products encoded by plasmids include enzymes for the utilization of unusual carbon sources, such as toluene; resistance to substances such as heavy metals and antibiotics; synthesis of antibiotics; and synthesis of toxins and proteins that allow the successful infection of other organisms. Plasmids, combined with their hosts, have also been an invaluable tool to investigate other organisms. Table 4.2 lists the major classes of plasmids used in *E. coli* and some of their relevant features for molecular genetics.

Table 4.1 Some naturally occurring plasmids and the traits they carry

Plasmid	Trait	Original source
ColE1	Bacteriocin which kills *E. coli*	*E. coli*
Tol	Degradation of toluene and benzoic acid	*Pseudomonas putida*
Ti	Tumor initiation in plants	*A. tumefaciens*
pJP4	2,4-D (2,4-dichlorophenoxyacetic acid) degradation	*Alcaligenes eutrophus*
pSym	Nodulation on roots of legume plants	*Sinorhizobium meliloti*
SCP1	Antibiotic methylenomycin biosynthesis	*Streptomyces coelicolor*
RK2	Resistance to ampicillin, tetracycline, and kanamycin	*Klebsiella aerogenes*

Table 4.2 General classes of plasmids commonly used in *E. coli*

Founding replicon	Common examples	Host range	Comments
pMB1/ColEI	pBR322, pUC vectors, pGEM vectors, pBluescript vectors	Narrow	pBR322 is a low-copy-number vector (~20 copies/cell) that has been adapted as very-high-copy-number vectors (>300 copies/cell).
p15A	pACYC177, pACYC184	Narrow	The pACYC vectors are low-copy-number vectors (~15/cell). p15A is similar to but compatible with the pMB1/ColEI replicon.
pSC101	pSC101	Narrow	Low-copy-number vector (~5 copies/cell) good for toxic genes; temperature-sensitive derivatives exist.
F plasmid	pBeloBAC11	Narrow	The original fertility (F) plasmid; the replication origin is utilized in BACs.
RK2 (RP4)	pSP329, pCM62, pCM66	Broad	IncP group
RSF1010	pJRD215, pSUP104, pSUP204	Broad	IncQ group
pSa	pUCD2	Broad	IncW group
R6K	R6K	Broad	IncX group
pBBR1MCS	pBBR1MCS-2, pBBR1MCS-3, etc.	Broad	Undefined Inc group

It is interesting to speculate about why certain types of genes are found on plasmids and others are found on the chromosome. It is easy to understand why genes that directly favor the plasmid would be encoded on the element, and this chapter goes into the details of some of these systems (see Thomas, Suggested Reading). However, there are many genes that favor the host and plasmid equally, and it is curious that certain broad classes of genes are normally encoded on the chromosome while others are on plasmids. One idea holds that plasmids tend to harbor genes that are locally advantageous (see Eberhard, Suggested Reading). For example, genes that encode such things as heavy metal resistance or antibiotic resistance may be advantageous only in certain transient situations in certain specific places. This selection might be sufficient to allow them to be maintained in some hosts in the environment even though selection is not pervasive enough to allow them to become associated with a lineage of bacteria. It would be impossible for every bacterium to maintain genes that could be advantageous for every environment. However, there is a strong selective advantage for plasmids to accumulate in hosts in these environments. Across nature, plasmids do allow bacteria to occupy a larger variety of ecological niches. This also explains why a large number of plasmids exist with the capacity to direct their own transfer between bacteria in a process called conjugation (see chapter 5) and why many plasmids are capable of replicating in different types of bacteria. As we will see below, some plasmids are also likely to be maintained, not because of the benefits they provide the host bacterium, but because the plasmid has a system to actively harm hosts that lose the plasmid (see "Toxin-Antitoxin Systems and Plasmid Maintenance" below).

Plasmid Structure

Most plasmids are circular, although linear plasmids also exist. In a circular plasmid, all of the nucleotides in each strand are joined to another nucleotide on each side by covalent bonds to form continuous strands that are wrapped around each other. Such DNAs are said to be **covalently closed circular** and, because there are no ends to rotate, the plasmid can maintain supercoiled stress. As discussed in chapter 1, nonsupercoiled DNA strands exhibit a helical periodicity of about 10.5 bp, as predicted from the Watson-Crick double-helical structure of DNA. In contrast, in a DNA that is supercoiled, the two strands are wrapped around each other either more or less often than once in about 10.5 bp. If they are wrapped around each other more often than once every 10.5 bp, the DNA is positively supercoiled; if they are wrapped around each other less often, the DNA is negatively supercoiled. Like the chromosome, covalently closed circular plasmid DNAs are usually negatively supercoiled (see chapter 1). Because DNA is stiff, the negative supercoiling introduces stress, and this stress is partially relieved by the plasmid wrapping up on itself, as illustrated in Figure 4.1A. This makes the plasmid more compact, so that it migrates more quickly in an agarose gel (Figure 4.1B). In the cell, the DNA wraps around proteins, which relieves some of the stress. The remaining stress facilitates some reactions involving the plasmid, such as separation of the two DNA strands for replication or transcription.

PURIFYING PLASMIDS

The structure of plasmids can be used to purify them away from chromosomal DNA in the cell. Cloning manuals give detailed protocols for these methods (see chapter 1, Suggested Reading), but we review them briefly

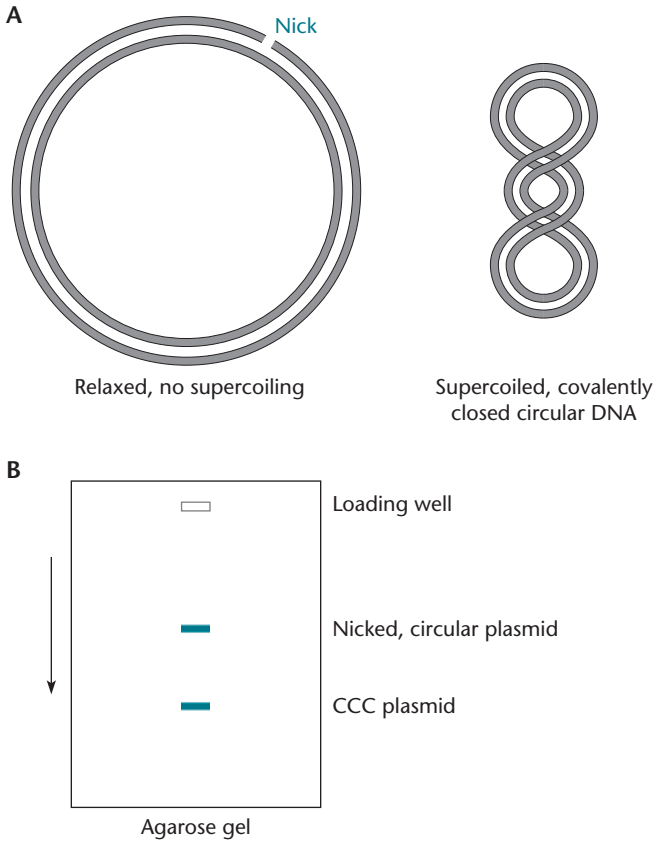


Figure 4.1 Supercoiling of a covalently closed circular (CCC) plasmid. **(A)** A break in one strand relaxes the DNA, eliminating the supercoiling and making the DNA less compact. **(B)** Schematic diagram of an agarose gel showing that the covalently closed supercoiled circles run faster on a gel than the nicked relaxed circles. Depending on the conditions, linear DNA and covalently closed circular DNA run in approximately the same position as nicked relaxed circles of the same length. The arrow shows the direction of migration.
doi:10.1128/9781555817169.ch4.f4.1

here. Many purification procedures are based on the relatively small size of most plasmids. Bacterial cells are usually lysed with a combination of a strong base (sodium hydroxide), the detergent sodium dodecyl sulfate (SDS) to solubilize the membranes, and a component to denature proteins. Following this treatment, a high-salt (potassium acetate) solution is added. Under these conditions, the chromosome precipitates with the cell debris and SDS upon centrifugation, while the small supercoiled plasmids remain in solution. Following these steps, a variety of procedures are used to further purify the plasmid DNA from the proteins and to concentrate the plasmid preparation. Historically, a solution of phenol and chloroform was added to the solution to separate the DNA from the proteins. After mixing of the aqueous lysate with phenol and chloroform, the components quickly separate into phases, similar in concept to oil and water. Plasmid DNA is separated from the proteins because the polar DNA molecules remain in the aqueous phase while the proteins remain at the interface of the solutions. Following extraction, plasmid DNA is concentrated using precipitation with a salt and alcohol solution and finally resuspended in water or a buffered solution. For convenience today, commercially available spin columns are usually used instead of separation using the phenol-chloroform solution. The plasmid solution is added to a resin matrix in a column, which binds DNA. An alcohol solution is used to wash away the residual proteins and salts from the column, and the plasmid DNA is finally eluted from the column using water or a buffered solution. The isolated plasmid DNA can then be visualized on a gel, digested with a restriction endonuclease, or sequenced, among other applications (see chapter 1)

The methods discussed above work well with plasmids that have many copies per cell and are not too large. However, large, low-copy-number plasmids are more difficult to isolate or even to detect. In some cases, plasmids are discovered by whole-genome sequencing of a bacterium when a series of DNA sequence reads assemble into a unique contiguous DNA structure separate from the host genome. Visual methods for detecting large plasmids involve separating them from the chromosome directly by electrophoresis on agarose gels (see chapter 1). The cells are often lysed directly on the agarose gel to avoid breaking the large plasmid DNA. The plasmid, because of its unique size, makes a sharp band on the gel, distinct from that due to chromosomal DNA, which is usually fragmented and so gives a more diffuse band. Also, methods such as pulsed-field gel electrophoresis have been devised to allow the separation of long pieces of DNA based on size. These methods depend on periodic changes in the direction of the electric field. The molecules attempt to reorient themselves each time the field shifts, and the longer molecules take longer to reorient than the shorter ones and thus move more slowly through the gel. Such methods have allowed the separation of DNA molecules hundreds of thousands of base pairs long and the detection of very large plasmids.

Properties of Plasmids

Replication

To exist free of the chromosome, plasmids must have the ability to replicate independently. DNA molecules that can replicate autonomously in the cell are called **replicons**. Plasmids, phage DNA, and the chromosomes are all replicons, at least in some types of cells.

To be a replicon in a particular type of cell, a DNA molecule must have at least one origin of replication, or *ori* site, where replication begins (see chapter 1). In

addition, the cell must contain the proteins that enable replication to initiate at this site. Plasmids encode only a few of the proteins required for their own replication. In fact, many encode only one of the proteins needed for initiation at the *ori* site. All of the other required proteins—DNA polymerases, ligases, primases, helicases, and so on—are borrowed from the host.

Each type of plasmid replicates by one of two general mechanisms, which is determined along with other properties by its *ori* region (see "Functions of the *ori* Region" below). The plasmid replication origin is often named *oriV* for *ori* vegetative, to distinguish it from *oriT*, which is the site at which DNA transfer initiates in plasmid conjugation (see chapter 5). Most of the evidence for the mechanisms described below came from electron microscope observations of replicating plasmid DNA.

THETA REPLICATION

Some plasmids begin replication by opening the two strands of DNA at the *ori* region, creating a structure that looks like the Greek letter θ—hence the name **theta replication** (Figure 4.2A and B). In this process, an RNA primer begins replication, which can proceed in one or both directions around the plasmid. In the first case, a single replication fork moves around the molecule until it returns to the origin, and then the two daughter DNAs separate. In the other case (bidirectional replication), two replication forks move out from the *ori* region, one in either direction, and replication is complete (and the two daughter DNAs separate) when the two forks meet somewhere on the other side of the molecule.

The theta mechanism is the most common form of DNA replication, especially in gram-negative bacteria like the proteobacteria. Commonly used plasmids, including ColE1, RK2, and F, as well as the bacteriophage P1, use this type of replication. While not called the theta mechanism in chapter 1, the form of replication initiated in the chromosome at *oriC* also occurs by this mechanism (see chapter 1).

ROLLING-CIRCLE REPLICATION

Other types of plasmids replicate by very different mechanisms. One type of replication is called **rolling-circle (RC) replication** because it was first discovered in a type of phage where the template circle seems to "roll" while a copy of the plasmid is made and processed to unit length. Plasmids that replicate by this mechanism are sometimes called **RC plasmids**. This type of plasmid is widespread and is found in the largest groups of bacteria, as well as archaea.

In an RC plasmid, replication occurs in two stages. In the first stage, the double-stranded circular plasmid DNA replicates to form another double-stranded circular DNA and a single-stranded circular DNA. This stage is analogous to the replication of the DNA of some single-stranded DNA phages (see chapter 7) and to DNA transfer during plasmid conjugation (see chapter 5). In the second stage, the complementary strand is synthesized on the single-stranded DNA to make another double-stranded DNA.

The details of the RC mechanism of plasmid replication are shown in Figure 4.2C. First, the Rep protein recognizes and binds to a palindromic sequence that contains the double-strand origin (DSO) on the DNA. Binding of the Rep protein to this sequence might allow the formation of a cruciform structure by base pairing between the inverted-repeated sequences in the cruciform. Once the cruciform forms, the Rep protein can make a nick in the sequence. It is important for the models that the Rep protein is also known to function as a dimer, at least in some plasmids. After the Rep protein has made a break in the DSO sequence, it remains covalently attached to the phosphate at the 5′ end of the DNA at the nick through a tyrosine in one copy of the Rep protein in the dimer, as shown. DNA polymerase III (the replicative polymerase [see chapter 1]) uses the free 3′ hydroxyl end at the break as a primer to replicate around the circle, displacing one of the strands. It may use a host helicase to help separate the strands, or the Rep protein itself may have the helicase activity, depending on the plasmid. Once the circle is complete, the 5′ phosphate is transferred from the tyrosine on the Rep protein to the 3′ hydroxyl on the other end of the displaced strand, producing a single-stranded circular DNA. This process is called a phosphotransferase reaction and requires little energy. The same reaction is used to re-form a circular plasmid after conjugational transfer (see chapter 5).

It is less certain what happens to the newly formed double-stranded DNA when the DNA polymerase III has made its way all around the circle and gets back to the site of the DSO. Why does it not just keep going, making a longer molecule with individual genomes linked head to tail in a structure called a concatemer? Such structures are created when some phage DNAs replicate by an RC mechanism (see chapters 7 and 8), and it may also occur in some plasmids. One idea is that the DNA polymerase III does keep going past the DSO for a short distance, creating another double-stranded DSO. The other copy of the Rep protein in the dimer may then nick the newly created DSO, transferring the 5′ end to itself as described above. This might inactivate the Rep protein, releasing it with a short oligonucleotide attached. Other reactions, probably involving host DNA ligase, then cause the nick to be resealed, resulting in a circular double-stranded DNA molecule.

The displaced circular single-stranded DNA now replicates by a completely different mechanism using only host-encoded proteins. The RNA polymerase first makes

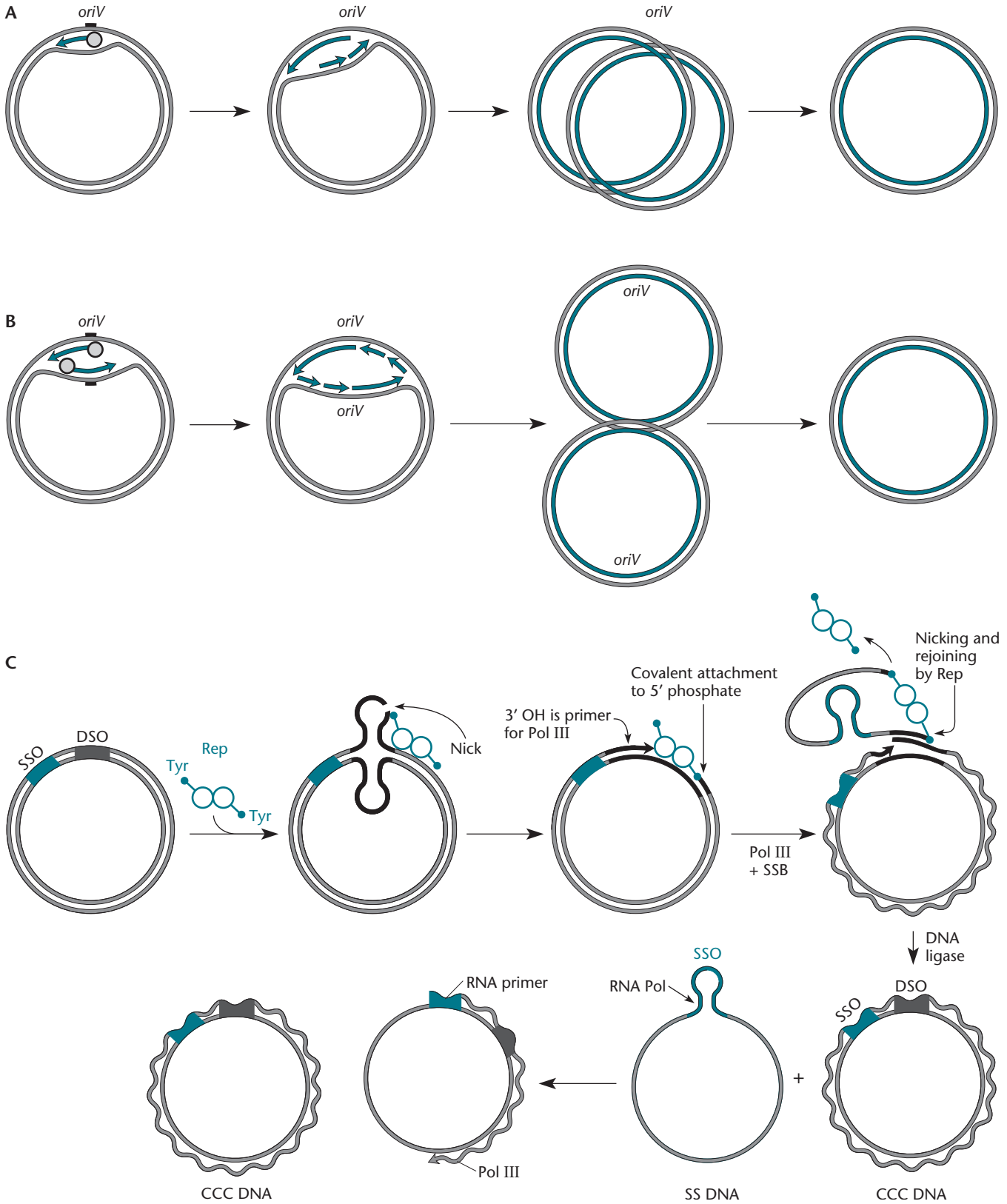


Figure 4.2 Some common schemes of plasmid replication. **(A)** Unidirectional replication. The origin region is designated *oriV.* Replication terminates when the replication fork gets back to the origin. **(B)** Bidirectional replication. Replication terminates when the replication forks meet somewhere on the DNA molecule opposite the origin. **(C)** RC replication. A nick is made at the DSO by the plasmid-encoded Rep protein, which remains bound to the 5′ phosphate end at the nick. The free 3′ OH end then serves as a primer for the DNA polymerase III (Pol III) that replicates around the circle, displacing one of the old strands as a single-stranded DNA. The Rep protein then makes another nick, releasing the single-stranded circle, and also joins the

a primer at a different origin, the single-strand origin (SSO), and this RNA then primes replication around the circle by DNA polymerase III. However, the RNA polymerase does not make this primer until the single-stranded DNA is completely displaced during the first stage of replication. This delay is accomplished by locating the SSO immediately counterclockwise from the DSO (Figure 4.2C), so that the SSO does not appear in the displaced DNA until the displacement of the single-stranded DNA is almost complete. After the entire complementary strand has been synthesized, the DNA polymerase I removes the RNA primer with its 5′ exonuclease activity while simultaneously replacing it with DNA, and host DNA ligase joins the ends to make another double-stranded plasmid. The net result is two new double-stranded plasmids synthesized from the original double-stranded plasmid.

In order for the complementary strand of the displaced single-stranded DNA to be synthesized, the RNA polymerase of the host cell must recognize the SSO on the DNA. In some hosts, the SSO is not well recognized, and single-stranded DNA accumulates. For this reason, some RC plasmids were originally called single-stranded DNA plasmids, although we now know that this is not their normal state. Broad-host-range RC plasmids presumably have an SSO that is recognized by the RNA polymerases of a wide variety of hosts, which allows them to make the complementary strand of the displaced single-stranded DNA in a variety of hosts.

The Rep protein is used only once for every round of plasmid DNA replication and is destroyed after the round is completed. This allows the replication of the plasmid to be controlled by the amount of Rep protein in the cell and keeps the total number of plasmid molecules in the cell within narrow limits. A little later in this chapter, we discuss how the copy numbers of other types of plasmids are controlled.

REPLICATION OF LINEAR PLASMIDS

As mentioned above, some plasmids and bacterial chromosomes are linear rather than circular (Box 4.1). In general, linear DNAs face two problems in all organisms. One issue with linear DNAs is that the cell must have a way to distinguish the "normal" DNA ends at the ends of the linear fragments from ends formed when DNA double-strand breaks occur, which would otherwise be lethal to the cell and must quickly be repaired. A second problem with linear plasmids and chromosomes has to do with replicating the lagging-strand template, the strand that ends with a 5′ phosphate, all the way to the end of the DNA. This has been called the "primer problem" because DNA polymerases cannot initiate the synthesis of a new strand of DNA. They can only add nucleotides to a preexisting primer, and in a linear DNA, there is no upstream primer on this strand from which to grow. Different linear DNAs solve the primer problem in different ways. Some linear plasmids have hairpin ends, with the 5′ and 3′ ends joined to each other. These plasmids replicate from an internal origin of replication to form dimeric circles composed of two plasmids joined head to tail to form a circle, as shown in the figure in Box 4.1. These dimeric circles are then resolved into individual linear plasmid DNAs with closed hairpins at the ends. The hairpin ends are presumably not recognized as DNA double-strand breaks, because they are not targets for exonucleases in the cell. A completely different mechanism is also used to maintain linear plasmids in some systems. With this mechanism, a special enzyme called a terminal protein attaches to the 5′ ends of the plasmid DNA (Box 4.1). It is interesting that bacteria with linear plasmids also often have linear chromosomes, and the two DNAs replicate by similar mechanisms.

Functions of the *ori* Region

In most plasmids, the genes for proteins required for replication are located very close to the *ori* sequences at which their products act. Thus, only a very small region surrounding the plasmid *ori* site is required for replication. As a consequence, the plasmid still replicates if most of its DNA is removed, provided that the *ori* region remains and the plasmid DNA is still circular. Smaller plasmids are easier to use as cloning vectors, as discussed below, so often the only part of the original plasmid that remains in a cloning vector is the *ori* region.

In addition, the genes in the *ori* region often determine many other properties of the plasmid. Therefore, any DNA molecule with the *ori* region of a particular plasmid will have most of the characteristics of that plasmid. The following sections describe the major plasmid properties determined by the *ori* region.

ends to form a circle by a phosphotransferase reaction (see the text). The DNA ligase then joins the ends of the new DNA to form a double-stranded circle. The host RNA polymerase makes a primer on the single-stranded DNA origin (SSO), and Pol III replicates the single-stranded (SS) DNA to make another double-stranded circle. DNA Pol I removes the primer, replacing it with DNA, and ligase joins the ends to make another double-stranded circular DNA. CCC, covalently closed circular; SSB, single-strand-DNA-binding protein.
doi:10.1128/9781555817169.ch4.f4.2

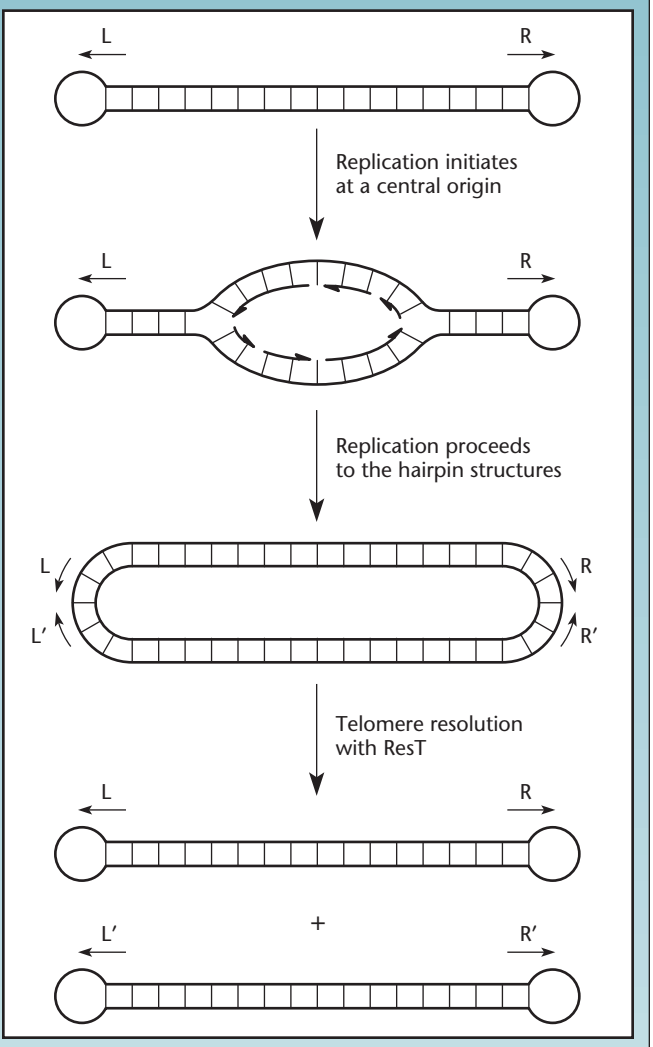
BOX 4.1

Linear Chromosomes and Plasmids in Bacteria

Not all bacteria have circular chromosomes and plasmids. Some, including *Borrelia burgdorferi* (the causative agent of Lyme disease), *Streptomyces* spp., *A. tumefaciens*, and *Rhodococcus fasciens*, have linear chromosomes and often multiple linear plasmids. As mentioned in the text, the replication of the ends of linear DNAs presents special problems because DNA polymerases cannot prime their own replication. This means that they cannot replicate all the way to a 3′ end in a linear DNA. If RNA at the end of a linear DNA primes the last Okazaki fragment and the RNA primer is then removed, there is no upstream primer DNA to prime its replacement with DNA as there is in a circular DNA. Eukaryotic chromosomes, which are linear, solve this problem by having special DNA regions called telomeres at their ends. Most telomeres do not need complementary sequences to be synthesized from the template as during normal DNA replication. Most use an enzyme called telomerase. The Nobel Prizes in physiology and medicine in 2010 went to people who determined how these enzymes work. Telomerase contains an RNA that is complementary to the repeated sequences at the ends of the DNA. This enzyme makes reiterated copies of the repeated telomeric sequences at the ends. When the linear chromosome replicates, some of these repeated sequences at the 3′ end are lost, but this is not a problem, because they will be resynthesized by the telomerase before the DNA replicates again.

Telomeres solve the problem of replicating linear DNAs without losing sequence information. However, bacteria with linear chromosomes are not known to have telomeres made by telomerases and must use different strategies. Two different strategies for dealing with linear chromosomes in bacteria are exemplified in *Streptomyces* and *Borrelia*. In both examples, the chromosome replicates from an *ori* sequence located toward the middle of the chromosome, from where replication occurs bidirectionally using the DnaA initiator protein in a system that is likely similar to those of other bacteria. However, the ways in which the linear ends are maintained in these organisms are very different.

The very large linear chromosome of *Streptomyces* has inverted-repeat sequences at its ends and a protein, terminal protein (TP), attached to the 5′ ends. Replication to the 3′ ends of the chromosome is thought to involve both these inverted repeats and TP in a process called "patching." After the linear DNA replicates, the 3′ end of each DNA remains single stranded, which allows the inverted-repeat sequences to form hairpins. Replication of these hairpins, combined with some sort of slippage, then allows complete replication of the ends by a process that is not completely understood.



doi:10.1128/9781555817169.ch4.Box4.1.f

The mechanism used by *Borrelia* to replicate its linear chromosome is very different from that found in *Streptomyces* spp. and is illustrated in the figure. In *Borrelia*, the 5′ phosphate and 3′ OH at each end of its linear chromosome are joined to each other to form hairpins, as shown. When replication initiated in the center of the chromosome gets to the ends, the linkage between the 5′ phosphate and 3′ OH hairpins forms a dimerized chromosome, with two copies of the chromosome forming a circle containing two copies of the chromosome linked end to end, as shown. An enzyme called telomere resolvase protein (ResT) (in analogy to the telomerase of eukaryotes, even though it does not work in the same way) then recreates the original hairpin ends from

BOX 4.1 (continued)

Linear Chromosomes and Plasmids in Bacteria

these double circles by making a staggered break in the two strands where the original ends were and then rejoining the 3′ end of one strand to the 5′ end of the other strand to form a hairpin. The ResT enzyme works somewhat like some topoisomerases and tyrosine recombinases (see chapter 9) in that the breaking and rejoining process goes through a 3′ phosphoryltyrosine intermediate, where the 3′ phosphate end is covalently joined to a tyrosine (Y [see inside front cover]) before it is joined to the 5′ hydroxyl end.

Bacteria that have linear chromosomes may contain linear or circular plasmids, and the same is true for bacteria with circular chromosomes. This suggests that minimal cellular adaptation may be needed to maintain a chromosome in either the linear or circular form. Experiments in *E. coli* have provided a clever and clear way to help address this question. A lysogenic *Siphoviridae* (lambda-like) bacteriophage called N15 that was found in an isolate of *E. coli* uses the same mechanism for maintaining its linear genome as that found in *Borrelia*, with a *cis*-acting telomere region called *tos* and a *trans*-acting resolvase protein, TelN. Amazingly, it was found that the normally circular *E. coli* chromosome could be converted into a linear chromosome simply by moving the *tos* region into the terminus region of *E. coli* and expressing the TelN protein; no other adaptations were needed. The *E. coli* strain replicated normally and showed essentially no changes in gene expression. As would be predicted, the *E. coli* strain with a linear genome no longer required the Xer dimer resolution system that is normally needed to resolve circular dimer chromosomes. Supporting the finding that supercoiling is constrained into multiple domains in the chromosome, the *E. coli* strain with the linear chromosome still required the topoisomerases Topo IV and gyrase for its replication.

Linear chromosome ends also need to be differentiated from lethal double-strand break damage that must be recognized and repaired in the chromosome. Having a hairpin structure and/or a terminal protein at the end of the chromosome likely allows the cellular machinery to distinguish normal linear chromosome ends from random double-stranded breaks in *Borrelia* and *Streptomyces.*

References

Cui, T., N. Moro-Oka, K. Ohsumi, K. Kodama, T. Ohshima, N. Ogasawara, H. Mori, B. Wanner, H. Niki, and T. Horiuchi. 2007. *Escherichia coli* with a linear genome. *EMBO Rep.* **8**:181–187.

Kobryn, K., and G. Chaconas. 2002. ResT, a telomere resolvase encoded by the Lyme disease spirochete. *Mol. Cell* **9**:195–201.

Yang, C. C., C. H. Huang, C. Y. Li, Y. G. Tsay, S. C. Lee, and C. W. Chen. 2002. The terminal proteins of linear *Streptomyces* chromosome and plasmids: a novel class of replication priming proteins. *Mol. Microbiol.* **43**:297–305.

HOST RANGE

The **host range** of a plasmid includes the types of bacteria in which the plasmid can replicate; it is usually determined by the *ori* region. Some plasmids, such as those with *ori* regions of the ColE1 plasmid type, including pBR322, pET, and pUC, have a **narrow host range** (Table 4.2). These plasmids replicate only in *E. coli* and some other closely related bacteria, such as *Salmonella* and *Klebsiella* species. Work with plasmids has historically been biased by work in *E. coli*, but presumably there are also types of plasmids that will replicate only in other closely related groups of bacteria that would also technically qualify as possessing a narrow host range. In contrast, plasmids with a **broad host range** include the RK2 and RSF1010 plasmids, as well as the RC plasmids, like pBBR1MCS (Table 4.2). The host ranges of these plasmids are truly remarkable. Plasmids with the *ori* region of RK2 can replicate in most types of gram-negative proteobacteria, and RSF1010-derived plasmids even replicate in some types of gram-positive bacteria, like the *Firmicutes*. There are also plasmids that are used for work in the low-G+C gram-positive *Firmicutes*. Some plasmids used for expressing genes in various examples from the *Firmicutes* are shown in Table 4.3.

It is perhaps surprising that the same plasmid can replicate in bacteria that are so distantly related to each other. Broad-host-range plasmids must encode all of their own proteins required for initiation of replication, and therefore, they do not have to depend on the host cell for any of these functions. They also must be able to express these genes in many types of bacteria. Apparently, the promoters and ribosome initiation sites for the replication genes of broad-host-range plasmids have evolved so that they can be recognized in a wide variety of bacteria.

Determining the Host Range

The actual host ranges of most plasmids are unknown because it is sometimes difficult to determine if a plasmid can replicate in other hosts. First, we must have a way of introducing the plasmid into other bacteria. Transformation systems (see chapter 6) have been developed for

Table 4.3 Plasmids used in *B. subtilis*

Plasmid	Use	*B. subtilis ori*	*E. coli ori*	*B. subtilis* drug resistance	*E. coli* drug resistance
pUB110	Cloning vector	pUB110		Neo^r	
pMK3	Cloning vector	pMK3		Cam^r	
pDG148	Shuttle vector Inducible expression	pUB110	pBR322	Neo^r	Amp^r
pMUTIN	Integration vector Gene disruption Inducible expression *lacZ* fusion		pBR322	Erm^r	Amp^r

some, but not all, types of bacteria, and if one is available, it can be used to introduce plasmids into the bacterium. Electroporation can often be used to introduce DNA into cells. Plasmids that are self-transmissible or mobilizable (see chapter 5) can sometimes be introduced into other types of bacteria by conjugation, a process in which DNA is transferred from one cell to another using plasmid-encoded transfer functions.

Even if we can introduce the plasmid into other types of bacteria, we still must be able to select cells that have received the plasmid. Most plasmids, as isolated from nature, are not known to carry a convenient selectable gene, such as one for resistance to an antibiotic, and even if they do, the selectable gene may not be expressed in the other bacterium, since most genes are not expressed in bacteria distantly related to those in which they were originally found. Sometimes we can introduce a selectable gene, chosen because it is expressed in many hosts, into the plasmid. For example, the kanamycin resistance gene, first found in the Tn*5* transposon, is expressed in most gram-negative bacteria, making them resistant to the antibiotic kanamycin. We can either clone a marker gene into the plasmid or introduce a transposon carrying a selectable marker into the plasmid by methods discussed in chapter 9.

If all goes well and we have a way to introduce the plasmid into other bacteria, and the plasmid carries a marker that is likely to be expressed in other bacteria, we can see if the plasmid can replicate in bacteria other than its original host. Care must also be taken to ensure that the plasmid has not recombined into the host chromosome. Since the mechanisms for introducing DNA into different types of bacteria differ and because there are many barriers to plasmid transfer between species, determining the host range of a new plasmid is a very laborious process. Therefore, the host ranges of plasmids are often extrapolated from only a few examples.

REGULATION OF COPY NUMBER

Another characteristic of plasmids that is determined mostly by their *ori* region is the **copy number**, or the average number of that particular plasmid per cell. More precisely, we define the copy number as the number of copies of the plasmid in a newborn cell immediately after cell division. Copy number control must have been an important early step in the evolution of plasmids. All plasmids must regulate their replication; otherwise, they would fill up the cell and become too great a burden for the host, or their replication would not keep up with the cell replication and they would be progressively lost during cell division. Some plasmids, such as the F plasmid of *E. coli*, replicate only about once during the cell cycle. Naturally, all plasmids have a somewhat low copy number, but plasmids have also been engineered to allow a much higher copy number per cell to facilitate biotechnology. Copy number information for some plasmids is shown in Table 4.2.

The regulation mechanisms used by plasmids with higher copy numbers often differ greatly from those used by plasmids with lower copy numbers. Plasmids that have high copy numbers, such as the modified derivatives of the ColE1 plasmid origin, need only have a mechanism that inhibits the initiation of plasmid replication when the number of plasmids in the cell reaches a certain level. Consequently, these molecules are called **relaxed plasmids**. In contrast, low-copy-number plasmids, such as F, must replicate only once or very few times during each cell cycle and so must have a tighter mechanism for regulating their replication. Hence, they are called **stringent plasmids**. Much more is understood about the regulation of replication of relaxed plasmids than about the regulation of replication of stringent plasmids.

The regulation of relaxed plasmids falls into three general categories. Some plasmids are regulated by an antisense RNA, sometimes called a countertranscribed RNA (ctRNA) because it is transcribed from the same region of the plasmid DNA as an RNA essential for plasmid replication but from the opposite strand. The ctRNA is therefore complementary to the essential RNA and is able to hybridize to the essential RNA and inhibit its function. In many cases, the ctRNA inhibits

the translation of a protein essential for replication. The ctRNA of these plasmids is often assisted in its inhibitory role by a protein. Other plasmids are regulated by a ctRNA alone. Still others are regulated by a protein alone, which binds to repeated sequences in the plasmid DNA called iterons, thereby inhibiting plasmid replication. Examples of these three types of regulation are discussed below.

INCOMPATIBILITY

Another function of plasmids that is controlled by the *ori* region is **incompatibility**. Incompatibility refers to the inability of two plasmids to coexist stably in the same cell. Many bacteria, as they are isolated from nature, contain more than one type of plasmid. These plasmid types coexist stably in the bacterial cell and remain there even after many cell generations. In fact, bacterial cells containing multiple types of plasmids are not cured of each plasmid any more frequently than if the other plasmids were not there.

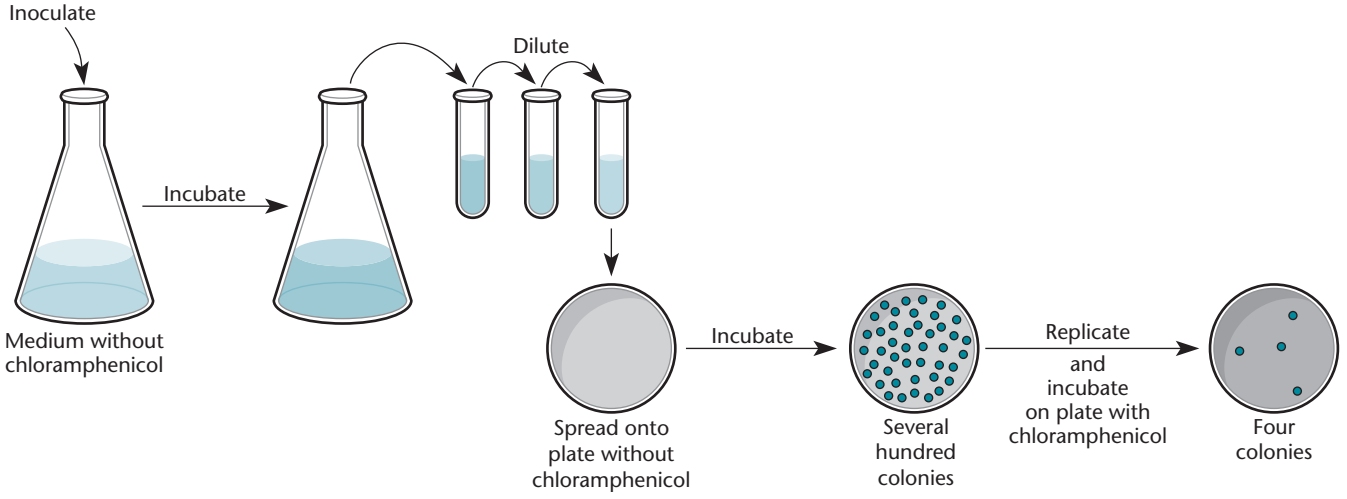
However, sometimes two plasmids of different types cannot coexist stably in the same cell. In this case, one or the other plasmid is lost as the cells multiply; this loss is more frequent than would occur if the plasmids were not occupying the cells together with the other plasmid. If two plasmids cannot coexist stably, they are said to be members of the same **incompatibility (Inc) group**. If two plasmids can coexist stably, they belong to different Inc groups. There are a number of ways in which plasmids can be incompatible. One way is if they can each regulate the other's replication. Another way is if they share the same partitioning (*par*) functions, which are often closely associated with replication control in the *ori* region. There may be hundreds of different Inc groups, and plasmids are usually classified by the Inc group to which they belong. For example, RP4 (also called RK2) is an IncP (incompatibility group P) plasmid. In contrast, RSF1010 is an IncQ plasmid; it can therefore be stably maintained with RP4 because it belongs to a different Inc group but cannot be stably maintained with another IncQ plasmid.

Determining the Inc Group

To classify a plasmid by its Inc group, we must determine if it can coexist with other plasmids of known Inc groups. In other words, we must measure how frequently cells are cured of the plasmid when it is introduced into cells carrying another plasmid of a known Inc group. However, we can know that cells have been cured of a plasmid only when it encodes an easily testable trait, such as resistance to an antibiotic. Then, the cells become sensitive to the antibiotic if the plasmid is lost.

The experiment shown in Figure 4.3 is designed to measure the curing rate of a plasmid that contains the Cam^r gene, which makes cells resistant to the antibiotic chloramphenicol. To measure the frequency of plasmid curing, we grow the plasmid-containing cells in medium with all the growth supplements and no chloramphenicol. At different times, we take a sample of the cells, dilute it, and plate the dilutions on agar containing the same growth supplements but, again, no chloramphenicol. After incubation of the plates, we replicate the plate onto another plate containing chloramphenicol (see chapter 3). If we do not observe any growth of a colony "copied" from the master plate, the bacteria in that colony must all have been sensitive to the antibiotic, and hence, the original bacterium that had multiplied to form the colony must have been cured of the plasmid.

Figure 4.3 Measurement of the curing of a plasmid carrying resistance to chloramphenicol. See the text for details. doi:10.1128/9781555817169.ch4.f4.3



The percentage of colonies that contain no resistant bacteria is the percentage of bacteria that were cured of the plasmid at the time of plating.

To apply this test to determine if two plasmids are members of the same Inc group, the two plasmids must contain different selectable genes, for example, genes encoding resistance to different antibiotics. Then, one plasmid is introduced into cells containing the other plasmid. Resistance to both antibiotics is selected for. Then, cells containing both plasmids are incubated without either antibiotic and finally tested on antibiotic-containing plates, as described above in the example with chloramphenicol. The only difference is that the colonies are transferred onto two plates, each containing one or the other antibiotic. If the percentage of cells cured of one or the other plasmid is no higher than the percentage cured of either plasmid when it was alone, the plasmids are members of different Inc groups. We continue to apply this test until we find a known plasmid, if any, that is a member of the same Inc group as our unknown plasmid.

Technically, even two plasmids in the same Inc group could be maintained if they were maintained with a high copy number and had distinct antibiotic resistances that could be selected at the same time. However, this is not advisable, because it can select for plasmid fusion events via recombination between the vectors.

Incompatibility Due to Shared Replication Control

One way in which two plasmids can be incompatible is if they share the same mechanism of replication control. The replication control system does not recognize the two as different, so either plasmid may be randomly selected for replication. At the time of cell division, the total copy number of the two plasmids will be the same, but one may be represented much less than the other. Figure 4.4 illustrates this by contrasting the distributions at cell division of plasmids of the same Inc group with plasmids of different Inc groups. Figure 4.4A shows a cell containing two types of plasmids that belong to different Inc groups and use different replication control systems. In the illustration, the two plasmids exist in equal numbers before cell division, but after division, the two daughter cells are not likely to get the same number of each plasmid. However, in the new cells, each plasmid replicates to reach its copy number, so that at the time of the next division, both cells again have the same numbers of the plasmids. This process is repeated each generation, so very few cells will be cured of either plasmid.

Now, consider the situation illustrated in Figure 4.4B, in which the cell has two plasmids that belong to the same Inc group and therefore share the same replication control system. As in the first example, both plasmids originally exist in equal numbers, but when the cell divides, it is unlikely that the two daughter cells will receive the same number of the two plasmids. Note that in the original cell, the copy number of each plasmid is only half its normal number; both plasmids contribute to the total copy number, since they both have the same *ori* region and inhibit each other's replication. After cell division, the two plasmids replicate until the total number of plasmids in each cell equals the copy number. The underrepresented plasmid (recall that the daughters may not receive the same number of plasmids if the plasmid is high copy number) does not necessarily replicate more than the other plasmid, so that the imbalance of plasmid numbers might remain or become even worse. At the next cell division, the underrepresented plasmid has less chance of being distributed to both daughter cells, since there are fewer copies of it. Consequently, in subsequent cell divisions, the daughter cells are much more likely to be cured of one or the other of the two plasmid types by chance alone.

Incompatibility due to copy number control is probably more detrimental to low-copy-number plasmids than to high-copy-number plasmids. If the copy number is only 1, then only one of the two plasmids can replicate; each time the cell divides, a daughter is cured of one of the two types of plasmids.

Incompatibility Due to Partitioning

Two plasmids can also be incompatible if they share the same Par (partitioning) system. Par systems help segregate plasmids or chromosomes into daughter cells upon cell division (see below). Normally this helps ensure that both daughter cells get at least one copy of the plasmid and neither daughter cell is cured of the plasmid. If coexisting plasmids share the same Par system, one or the other is always distributed into the daughter cells during division. However, sometimes one daughter cell receives one plasmid type and the other cell gets the other plasmid type, producing cells cured of one or the other plasmid. We discuss what is understood about the mechanisms of partitioning below.

Plasmid Replication Control Mechanisms

The mechanisms used by some plasmids to regulate their copy number have been studied in detail. Some of the better-understood mechanisms are reviewed in this section.

ColE1-DERIVED PLASMIDS: REGULATION OF PROCESSING OF PRIMER BY COMPLEMENTARY RNA

The mechanism of copy number regulation of the plasmid ColE1 was one of the first to be studied. Figure 4.5 shows a partial genetic map of the original ColE1 plasmid. This plasmid has been put to use in numerous molecular biology studies, and many vectors have been

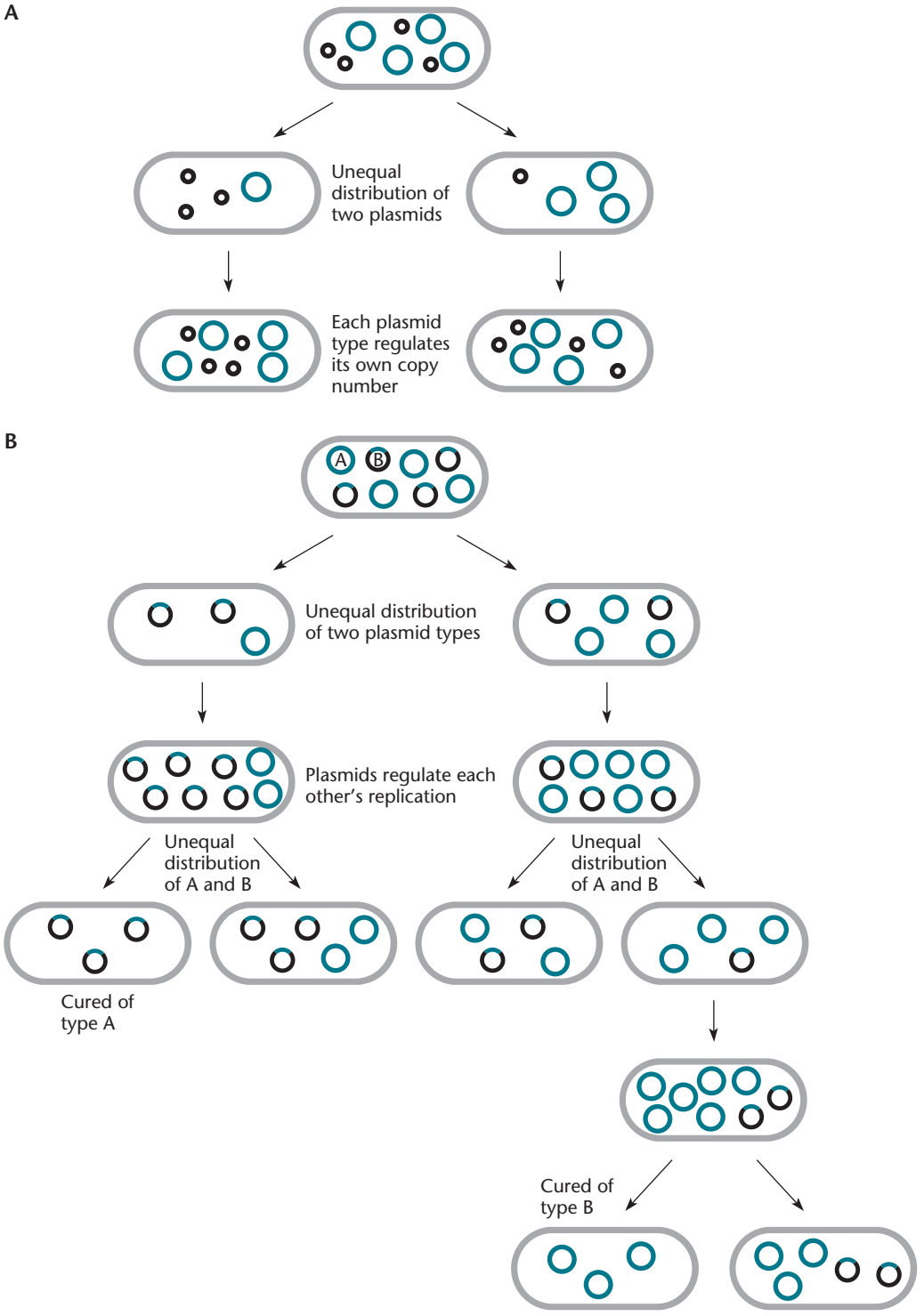


Figure 4.4 Coexistence of two plasmids from different Inc groups. **(A)** After division, both plasmids replicate to reach their copy numbers. **(B)** Curing of cells of one of two plasmids when they are members of the same Inc group. The sum of the two plasmids is equal to the copy number, but one may be underrepresented and lost in subsequent divisions. Eventually, most of the cells contain only one or the other plasmid. The light-blue region in the smaller plasmid indicates that it shares the *ori* region with the larger plasmid.
doi:10.1128/9781555817169.ch4.f4.4

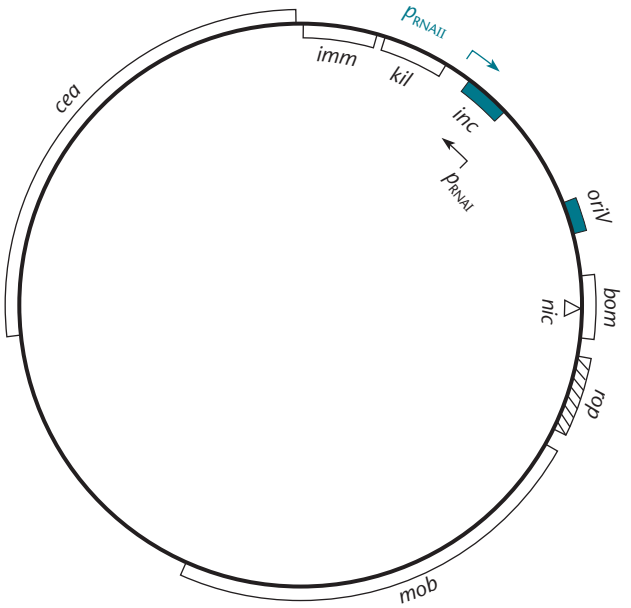


Figure 4.5 Genetic map of plasmid ColE1. The plasmid is 6,646 bp long. On the map, *oriV* is the origin of replication; p_{RNAII} is the promoter for the primer RNA II, *inc* encodes RNA I, *rop* encodes a protein that helps regulate the copy number, *bom* is a site that is nicked at *nic, cea* encodes colicin ColE1, and *mob* encodes functions required for mobilization (discussed in chapter 5).
doi:10.1128/9781555817169.ch4.f4.5

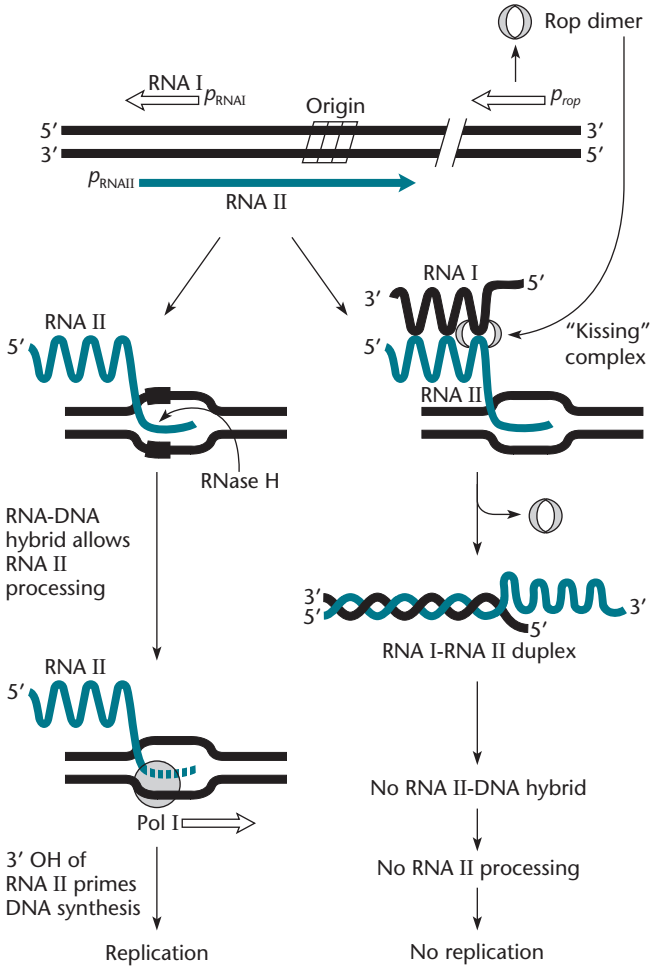


Figure 4.6 Regulation of the replication of ColE1-derived plasmids. RNA II must be processed by RNase H before it can prime replication. "Origin" indicates the transition point between the RNA primer and DNA. RNA I binds to RNA II and inhibits the processing, thereby regulating the copy number. p_{RNAI} and p_{RNAII} are the promoters for RNA I and RNA II transcription, respectively. RNA II is shown in blue. The Rop protein dimer enhances the initial pairing of RNA I and RNA II.
doi:10.1128/9781555817169.ch4.f4.6

derived from it or its close relative, pMB1 (Table 4.2). These vectors include the commonly used pBR322 plasmid and plasmids with modified forms of the pMB1/ColE1 origin, like the pUC plasmids, the pBAD plasmids, and the pET series of plasmids discussed below and in chapter 7. Expression control using the p_{BAD} promoter that is found in the pBAD plasmids is described in chapter 12. Although the genetic maps of these cloning vectors have been changed beyond recognition for the pMB1/ColE1 vectors, they all retain the basic properties of the original ColE1 *ori* region, and hence, they share many of its properties, including the mechanism of replication regulation. However, these derivatives often have modifications to vastly increase the copy number of the vector to allow greater amounts of plasmid DNA to be easily isolated (Table 4.2).

The mechanism of regulation of ColE1-derived plasmids is shown in Figure 4.6. Replication is regulated mostly through the effects of a small plasmid-encoded RNA called RNA I. This small RNA inhibits plasmid replication by interfering with the processing of another RNA called RNA II, which forms the primer for plasmid DNA replication. In the absence of RNA I, RNA II forms an RNA-DNA hybrid at the replication origin. RNA II is then cleaved by the RNA endonuclease RNase H, releasing a 3′ hydroxyl group that serves as the primer for replication first catalyzed by DNA polymerase I. Unless RNA II is processed properly, it does not function as a primer, and replication does not ensue.

RNA I inhibits DNA replication through interference with RNA II primer formation by forming a double-stranded RNA with it, as illustrated in Figure 4.6. It can do this because the two RNAs are transcribed from opposite strands in the same region of DNA. Figure 4.7 illustrates how any two RNAs transcribed from the same region of DNA but from opposite strands are complementary. The regulatory capacity of small complementary RNAs was first shown in this system, but small RNAs are now known to be very important as mechanisms for controlling gene expression in many groups of

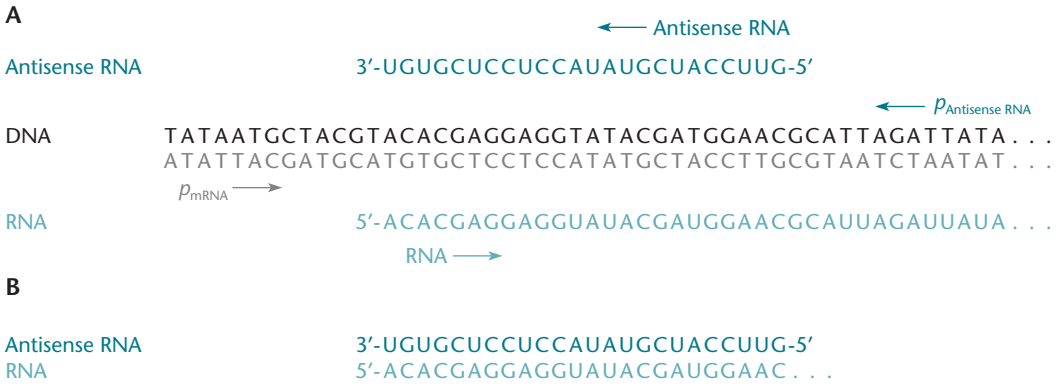


Figure 4.7 Pairing between an RNA and its antisense RNA. **(A)** An antisense RNA is made from the opposite strand of DNA in the same region. **(B)** The two RNAs are complementary and can base pair with each other to make a double-stranded RNA.
doi:10.1128/9781555817169.ch4.f4.7

bacteria (see chapter 12). Initially, the pairing between RNA I and RNA II occurs through short exposed regions on the two RNAs that are not occluded by being part of secondary structures. This initial pairing is very weak and therefore has been called a "kissing complex." The protein named Rop (Fig. 4.5) helps stabilize the kissing complex, although it is not essential. The kissing complex can then extend into a "hug," with the formation of the double-stranded RNA, as shown. Formation of the double-stranded RNA prevents RNA II from forming the secondary structure required for it to hybridize to the DNA before being processed by RNase H to form the mature primer.

Even though Rop (sometimes called Rom) is known to help RNA I to pair with RNA II and therefore help inhibit plasmid replication, it is not clear how Rop works, nor is the protein essential to maintain the copy number. Mutations that inactivate Rop cause only a moderate increase in the plasmid copy number.

This mechanism provides an explanation for how the copy number of ColE1 plasmids is maintained. Since RNA I is synthesized from the plasmid, more RNA I is made when the concentration of the plasmid is high. A high concentration of RNA I interferes with the processing of most of the RNA II, and replication is inhibited. The inhibition of replication is almost complete when the concentration of the plasmid reaches about 16 copies per cell, the copy number of the original ColE1 plasmid.

We can predict from the model what the effect of mutations in RNA I should be. Formation of the kissing complex involves pairing between very small regions of RNA I and RNA II. However, these regions must be completely complementary for this pairing to occur and for plasmid replication to be inhibited. Changing even a single base pair in this short sequence makes the mutated RNA I no longer complementary to the RNA II of the original nonmutant ColE1 plasmid, so it is no longer able to "kiss" it and regulate its replication. However, a mutation in the region of the plasmid DNA encoding RNA I also changes the sequence of RNA II made by the same plasmid in a complementary way, since they are encoded in the same region of the DNA, but from the opposite strands. Therefore, the mutated RNA I should still form a complex with the mutated RNA II made from the same mutated plasmid and prevent its processing; it just cannot interfere with the processing of RNA II from the original nonmutant plasmid. Therefore, a single-base-pair mutation in the RNA I coding region of the plasmid should effectively change the Inc group of the plasmid to form a new Inc group, of which the mutated plasmid is conceivably the sole member! In fact, the naturally occurring plasmids ColE1 and its close relative p15A, from which other cloning vectors, such as pACYC177 and pACYC184 have been derived, are members of different Inc groups, even though they differ by only 1 base in the kissing regions of their RNA I and RNA II.

R1 AND ColIb-P9 PLASMIDS: REGULATION OF TRANSLATION OF Rep PROTEIN BY COMPLEMENTARY RNA

The ColE1-derived plasmids are unusual in that they do not require a plasmid-encoded protein to initiate DNA replication at the *oriV* region, only an RNA primer synthesized from the plasmid. Most plasmids require a plasmid-encoded protein, often called Rep, to initiate replication. The Rep protein is required to separate the strands of DNA at the *oriV* region, often with the help of host proteins, including DnaA (see chapter 1). Opening the strands is a necessary first step that allows the replication apparatus to assemble at the origin. The Rep proteins are very specific in that they bind only to the *oriV* of the same type of plasmid because they bind to certain

specific DNA sequences within *oriV*. The amount of Rep protein is usually limiting for replication, meaning that there is never more than is needed to initiate replication. Therefore, the copy number of the plasmid can be controlled, at least partially, by controlling the synthesis of the Rep protein.

The R1 Plasmid

One type of plasmid that regulates its copy number by regulating the amount of a Rep protein is the R1 plasmid, a member of the IncFII family of plasmids. Like ColE1 plasmids, this plasmid uses a small complementary RNA to regulate its copy number, and this small RNA forms a kissing complex with its target RNA (see Kolb et al., Suggested Reading). Also like ColE1 plasmids, the more copies of the plasmid in the cell, the more of this antisense RNA is made and the more plasmid replication is inhibited. However, rather than inhibiting primer processing, the R1 plasmid uses its complementary RNA to inhibit the translation of the mRNA that encodes the Rep protein and thereby to inhibit the replication of the plasmid DNA.

Figure 4.8 illustrates the regulation of R1 plasmid replication in more detail. The plasmid-encoded protein RepA is the only plasmid-encoded protein that is required for the initiation of replication. The *repA* gene can be transcribed from two promoters. One of these promoters, called p_{copB}, transcribes both the *repA* and *copB* genes, making an mRNA that can be translated into the proteins RepA and CopB. The second promoter, p_{repA}, is in the *copB* gene and so makes an RNA that can encode only the RepA protein. Because the p_{repA} promoter is repressed by the CopB protein, it is turned on only immediately after the plasmid enters a cell and before any CopB protein is made. The short burst of synthesis of RepA from p_{repA} after the plasmid enters a cell causes the plasmid to replicate until it reaches its copy number. Then, the p_{repA} promoter is repressed by CopB protein, and the *repA* gene can be transcribed only from the p_{copB} promoter.

Once the plasmid has attained its copy number, the regulation of synthesis of RepA, and therefore the replication of the plasmid, is regulated by the small CopA RNA. The *copA* gene is transcribed from its own promoter, and the RNA product affects the stability of the mRNA made from the p_{copB} promoter. Because the CopA RNA is made from the same region encoding the translation initiation region (TIR) for the *repA* gene, but from the other strand of the DNA, the two RNAs are complementary and can pair to make double-stranded RNA. Then, an RNase called RNase III, a chromosomally encoded enzyme that cleaves some double-stranded RNAs (see chapter 2), cleaves the CopA-RepA duplex RNA.

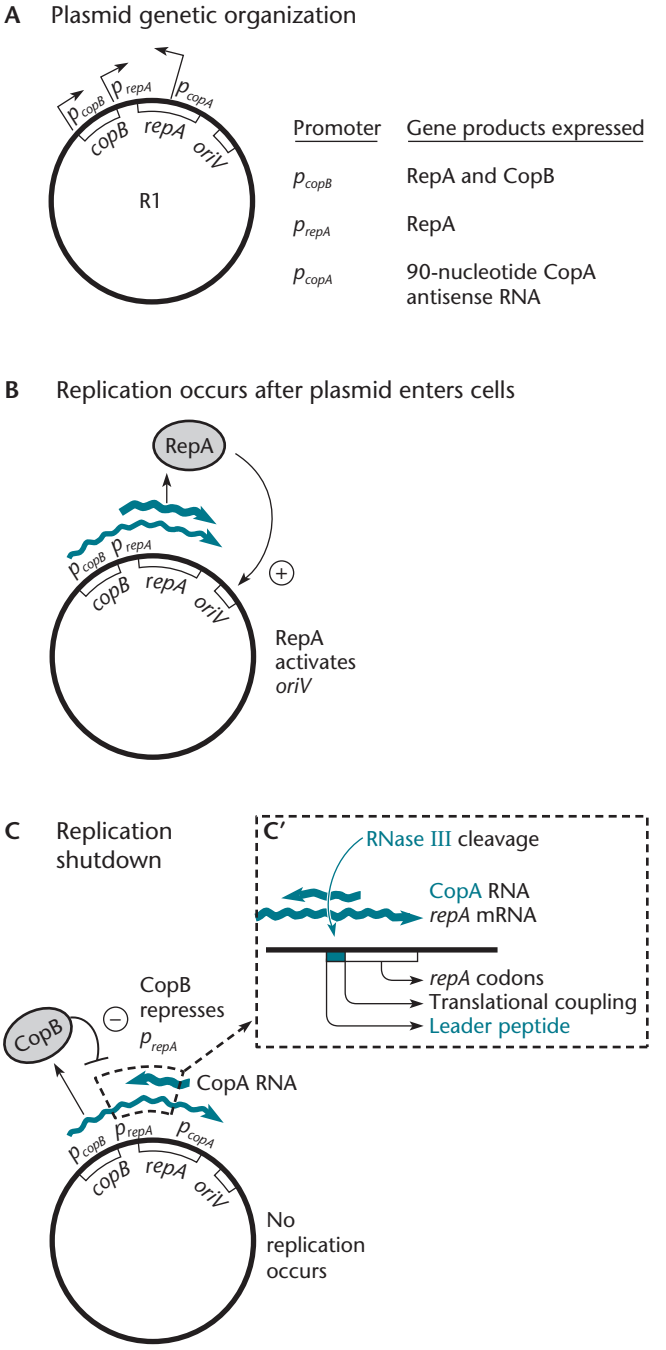


Figure 4.8 Regulation of replication of the IncFII plasmid R1. **(A)** Locations of promoters, genes, and gene products involved in the regulation. **(B)** Immediately after the plasmid enters the cell, most of the *repA* mRNA is made from promoter p_{repA} until the plasmid reaches its copy number. **(C)** Once the plasmid reaches its copy number, the CopB protein represses transcription from p_{repA}. Now, *repA* is transcribed only from p_{copB}. **(C′)** The antisense RNA CopA hybridizes to the leader peptide coding sequence in the *repA* mRNA, and the double-stranded RNA is cleaved by RNase III. This prevents translation of RepA, which is translationally coupled to the translation of the leader peptide.
doi:10.1128/9781555817169.ch4.f4.8

The reasons why cleavage of this RNA prevents the synthesis of RepA are a little complicated. The *5′* leader region of the mRNA, upstream of where the RepA protein is encoded, encodes a short leader polypeptide that has no function of its own but simply exists to be translated. The translation of RepA is coupled to the translation of this leader polypeptide (see chapter 2 for an explanation of translational coupling). Cleavage of the mRNA by RNase III in the leader region interferes with the translation of this leader polypeptide and, by blocking its translation, also blocks translation of the downstream RepA. Therefore, by having the CopA RNA activate cleavage of the mRNA for the RepA protein upstream of the RepA coding sequence, the plasmid copy number is controlled by the amount of CopA RNA in the cell, which in turn depends on the concentration of the plasmid. The higher the concentration of the plasmid, the more CopA RNA is made and the less RepA protein is synthesized, maintaining the concentration of the plasmid around the plasmid copy number.

The ColIb-P9 Plasmid

Yet another level of complexity of the regulation of the copy number by a complementary RNA is provided by the ColIb-P9 plasmid (Figure 4.9) (see Azano and Mizobuchi, Suggested Reading). As in the R1 plasmid, the Rep protein-encoding gene (called *repZ* in this case) is translated downstream of a leader peptide open reading frame, called *repY*, and the two are also translationally coupled. The translation of *repY* opens an RNA secondary structure that normally occludes the Shine-Dalgarno (S-D) sequence of the TIR of *repZ*. A sequence in the secondary structure then can pair with the loop of a hairpin upstream of *repY*, forming a pseudoknot (see Figure 2.2 for an example of a pseudoknot), thus permanently disrupting the secondary structure and leaving the S-D sequence for *repZ* exposed. A ribosome can then bind to the TIR for *repZ* and translate the initiator protein. The small complementary Inc RNA pairs with the loop of the upstream hairpin and prevents hairpin formation, leaving the S-D sequence of the *repZ* coding sequence blocked and preventing translation of *repY*.

THE pT181 PLASMID: REGULATION OF TRANSCRIPTION OF THE *rep* GENE BY A SMALL COMPLEMENTARY RNA

Not all plasmids that have an antisense RNA to regulate their copy numbers use it to inhibit translation or primer processing. Some plasmids of gram-positive bacteria, including the *Staphylococcus* plasmid pT181, use antisense RNAs to regulate transcription of the *rep* gene, in this case called *repC*, through a process called attenuation (Figure 4.10) (see Novick et al., Suggested Reading). The pT181 plasmid replicates by an RC mechanism, and the RepC protein is required to initiate replication of the leading strand at *oriV*. Also, the RepC protein is inactivated each time the DNA replicates (see above). This makes the RepC protein rate limiting for replication, i.e., the more RepC protein there is, the more plasmids are made. The antisense RNA binds to the mRNA for the RepC protein as the mRNA is being made and prevents formation of a secondary structure. This secondary structure would normally prevent the formation of a hairpin that is part of a factor-independent transcriptional terminator (see chapter 2). Therefore, if the secondary structure does not form, the hairpin forms and transcription terminates (i.e., is attenuated). Transcriptional regulation by attenuation is discussed in more detail in chapter 12.

This regulation works well only because the antisense RNA is so unstable that its concentration drops quickly if the copy number of the plasmid decreases, allowing fine-tuning of the replication of the plasmid with the copy number. In other plasmids of gram-positive bacteria, the antisense RNA is much more stable. These plasmids also use a transcriptional repressor to regulate transcription of the *rep* gene.

THE ITERON PLASMIDS: REGULATION BY COUPLING

Many commonly studied plasmids use a very different mechanism to regulate their replication. These plasmids are called iteron plasmids because their *oriV* regions contain several repeats of a certain set of DNA bases called an **iteron sequence**. The iteron plasmids include pSC101, F, R6K, P1, and the RK2-related plasmids (Table 4.2). The iteron sequences of these plasmids are typically 17 to 22 bp long and exist in about three to seven copies in the *ori* region. In addition, there are usually additional copies of these repeated sequences a short distance away that contribute to lowering the plasmid copy number.

One of the simplest of the iteron plasmids is pSC101. For our purposes, the essential features of the *ori* region of this plasmid (Figure 4.11) are the gene *repA*, which encodes the RepA protein required for initiation of replication, and three repeated iteron sequences, R1, R2, and R3, where RepA binds to regulate the copy number. The RepA protein is the only plasmid-encoded protein required for the replication of the pSC101 plasmid and many other iteron plasmids. It serves as a positive activator of replication, much like the RepA protein of the R1 plasmid. The host chromosome encodes the other proteins that either bind to this region or otherwise act to allow initiation of replication; they include DnaA, DnaB, DnaC, and DnaG (see chapter 1).

Iteron plasmid replication is regulated by two superimposed mechanisms. The first is control of RepA synthesis.

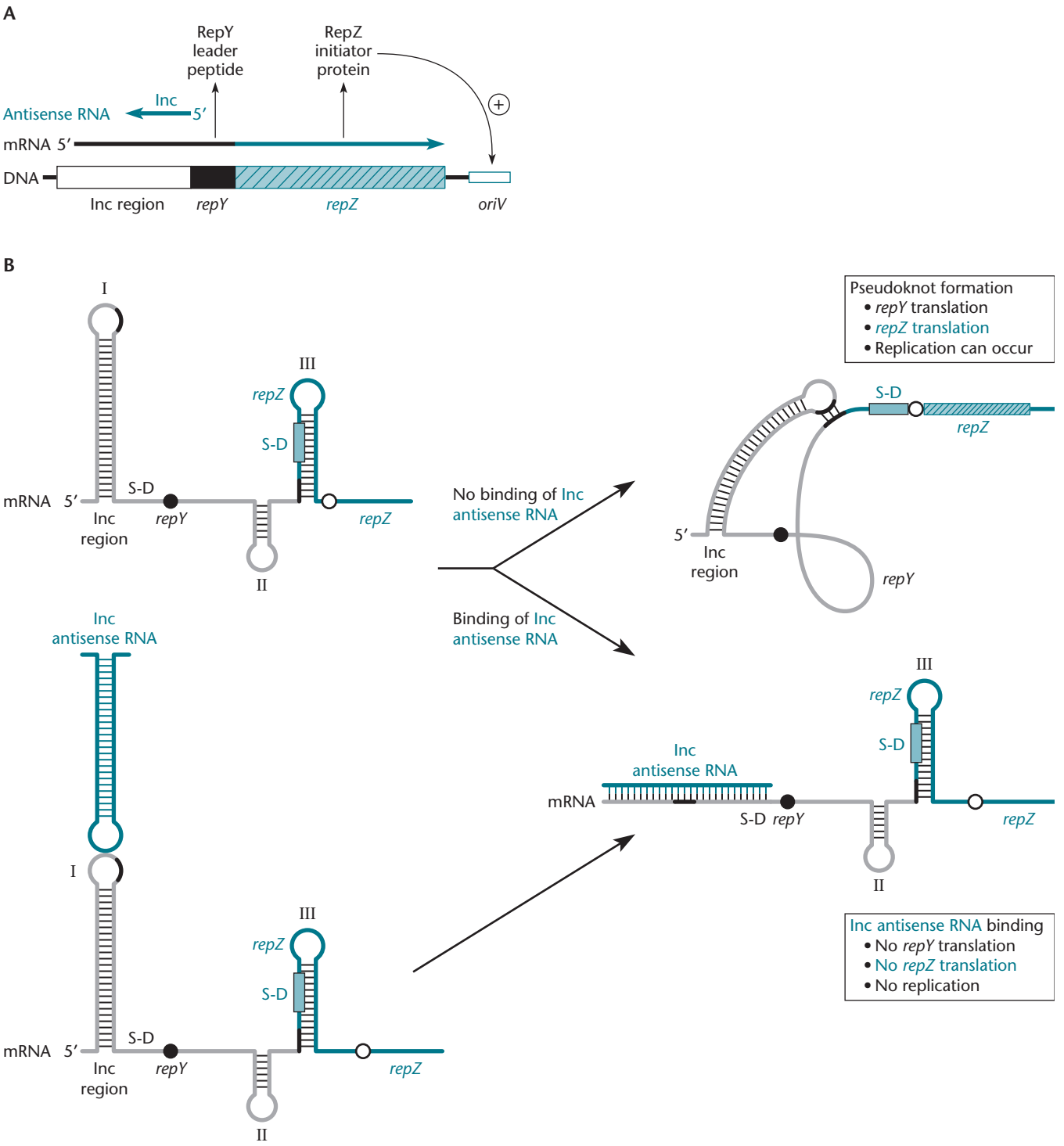
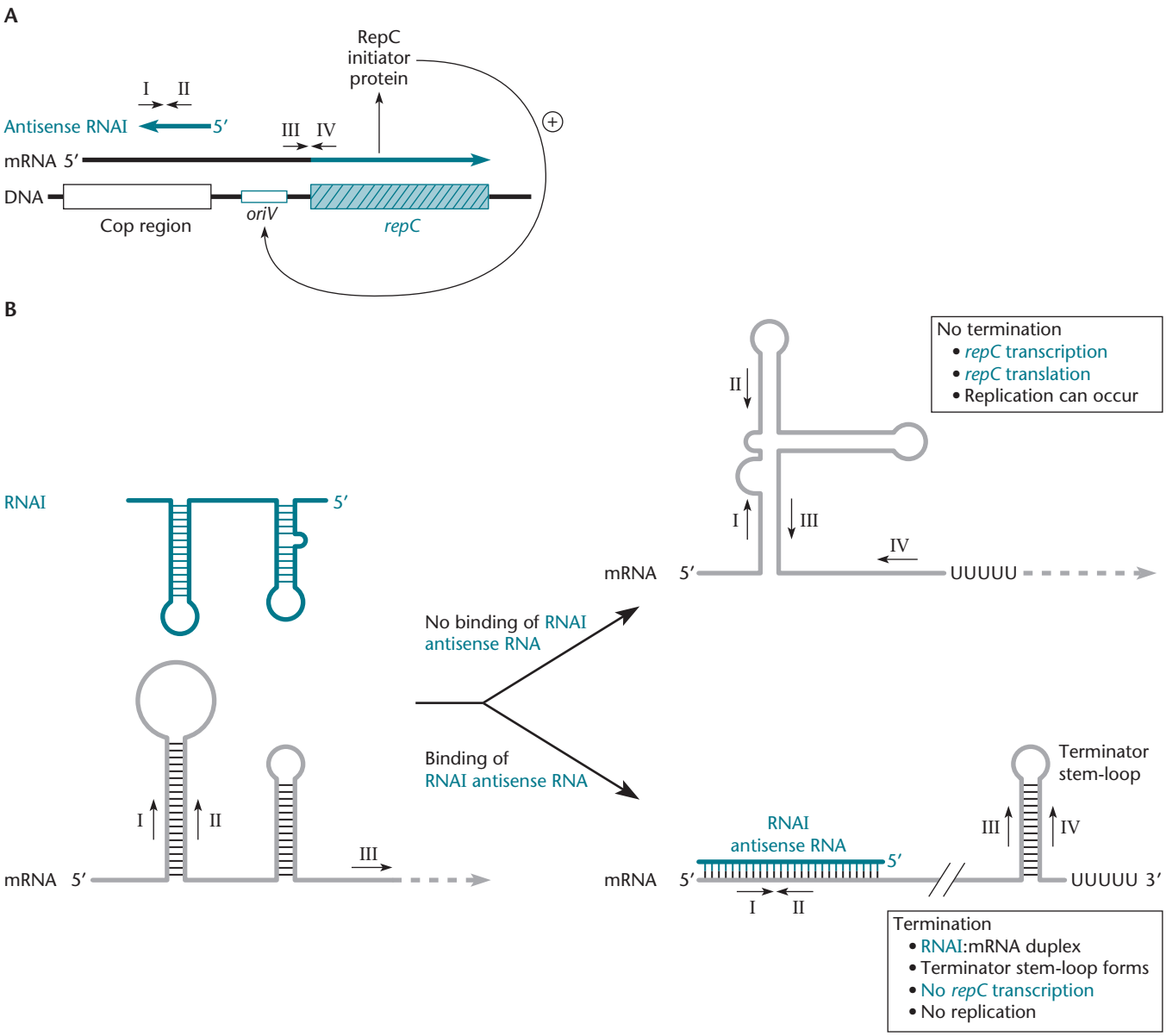


Figure 4.9 Regulation of plasmid ColIb-P9 copy number by antisense RNA inhibition of pseudoknot formation. **(A)** The minimal replicon with the *repY* (leader peptide) (black box) and *repZ* genes is shown. The Inc region encodes both the 5′ end of the *repYZ* mRNA and the antisense RNA. **(B)** The *repY* and *repZ* genes are translationally coupled. On the mRNA, the *repY* S-D sequence is exposed, whereas structure III sequesters the *repZ* S-D sequence (blue rectangle) and thereby prevents *repZ* translation. Also shown by thick black bars are regions in structures I and III that are complementary and so can pair, resulting in pseudoknot formation. The solid circle indicates the *repY* start codon; the open circle indicates the *repY* stop codon. Unfolding of structure II by the ribosome stalling at the *repY* stop codon results in the formation of a pseudoknot by base pairing between the complementary sequences and allows the ribosome to access the *repZ* S-D sequence. Binding of Inc antisense RNA to the loop of structure I directly inhibits formation of the pseudoknot, and the subsequent Inc RNA-mRNA duplex inhibits RepY translation, and consequently RepZ translation, since the two are translationally coupled. doi:10.1128/9781555817169.ch4.f4.9

Most commonly, the RepA protein represses its own synthesis by binding to its own promoter region and blocking transcription of its own gene. Therefore, the higher the concentration of plasmid, the more RepA protein is made and the more it represses its own synthesis. Thus, the concentration of RepA protein is maintained within narrow limits and the initiation of replication is strictly regulated. This type of regulation, known as **transcriptional autoregulation**, is discussed in chapter 12.

However, this mechanism of regulation by itself is not sufficient to regulate the copy number of the plasmid, especially that of low-copy-number stringent plasmids,

Figure 4.10 Regulation of plasmid pT181 copy number by antisense RNA regulation of transcription of the *repC* gene. **(A)** The genetic structure of the minimal replicon of pT181. Shown are the mRNA that encodes the RepC protein that initiates leading-strand replication at *oriV*; the Cop region that encodes the antisense RNA (RNA I) that regulates the copy number; and regions in the mRNA and antisense RNA, indicated by arrows labeled I, II, III, and IV, that can pair to form alternative secondary structures. **(B)** Formation of an antisense RNA-mRNA duplex regulates RepC expression by a transcriptional attenuation mechanism. The antisense RNA I can form a duplex with the 5′ end of the mRNA that encodes RepC and disrupt a secondary structure, allowing instead the formation of a terminator loop that causes transcription termination. doi:10.1128/9781555817169.ch4.f4.10



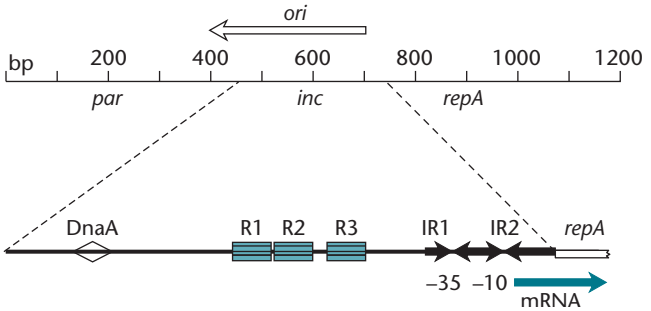


Figure 4.11 The *ori* region of pSC101. R1, R2, and R3 are the three iteron sequences (CAAAGGTCTAGCAGCAGAATTTACAGA for R3) to which RepA binds to handcuff two plasmids. RepA autoregulates its own synthesis by binding to the inverted repeats IR1 and IR2. The location of the partitioning site, *par* (see "Partitioning"), and the binding sites for the host protein DnaA are also shown.
doi:10.1128/9781555817169.ch4.f4.11

such as F and P1. Iteron plasmids must have another mechanism to regulate their copy numbers within narrow limits. This other form of regulation has been hypothesized to be the coupling of plasmids through the Rep protein and their iteron sequences (see McEachern et al., Suggested Reading). The **coupling hypothesis** for regulation of plasmid replication is illustrated in Figure 4.12. When the concentration of plasmids is high enough, they couple with each other via bound RepA proteins. This inhibits the replication of both coupled plasmids. The coupling mechanism allows plasmid replication to be controlled, not only by how much RepA protein is present in the cell, but also by the concentration of the plasmid itself or, more precisely, the concentration of the iteron sequences on the plasmid. Direct support for the coupling model in the replication control of iteron plasmids has come from electron micrographs of purified iteron plasmids mixed with the purified RepA protein for that plasmid. In these pictures, two plasmid molecules can often be seen coupled by RepA protein. In vitro and in vivo work also supports coupling as an important mechanism to prevent plasmid overreplication (see Das et al., Suggested Reading).

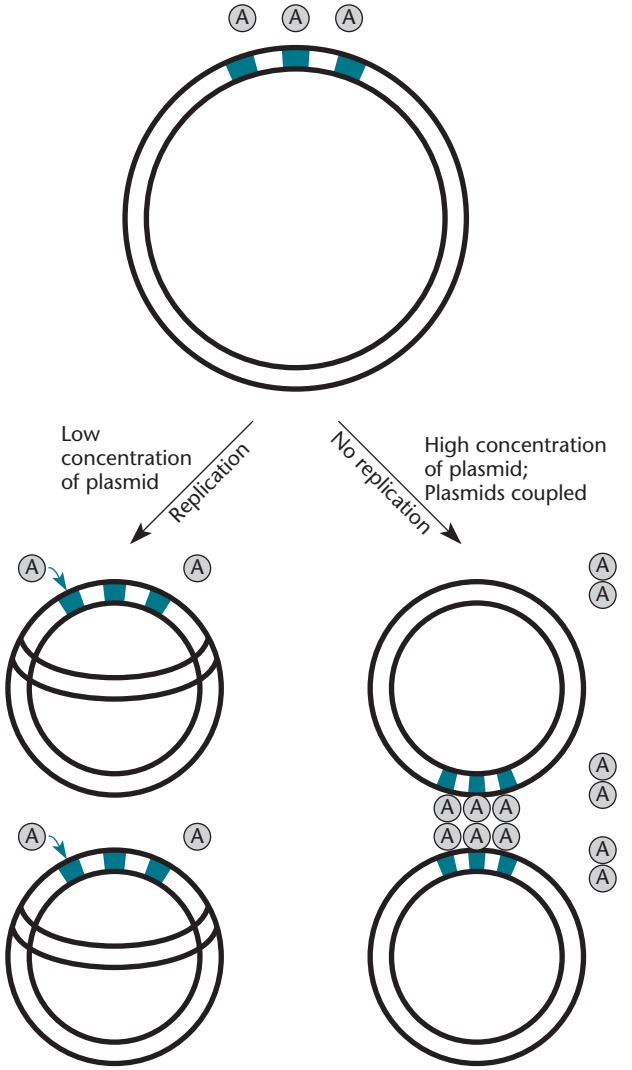


Figure 4.12 The "handcuffing" or "coupling" model for regulation of iteron plasmids. At low concentrations of plasmids, the RepA protein binds to only one plasmid at a time, initiating replication. At high plasmid and RepA concentrations, the RepA protein may dimerize and bind to two plasmids simultaneously, handcuffing them and inhibiting replication.
doi:10.1128/9781555817169.ch4.f4.12

HOST FUNCTIONS INVOLVED IN REGULATING PLASMID REPLICATION

As mentioned above, in addition to Rep, many plasmids require host proteins to initiate replication. For example, some plasmids require the DnaA protein, which is normally involved in initiating replication of the chromosome, and have *dnaA* boxes in their *oriV* regions to which DnaA binds (Figure 4.11). The DnaA protein may also directly interact with the Rep proteins of some plasmids. This may explain why some broad-host-range plasmids, such as RP4, make two Rep proteins. The different forms of the Rep protein might better interact with the DnaA proteins of different species of bacteria (see Caspi et al., Suggested Reading). The DnaA protein is involved in coordinating replication of the chromosome with cell division (see chapter 1); making their replication dependent on DnaA may allow plasmids to better coordinate their own replication with cell division. Like the chromosome origin (*oriC*), some *E. coli* plasmids also have Dam methylation sites close to their *oriV*. These methylation sites presumably help to further coordinate their replication with cell division. As with the chromosomal origin of replication, both strands of DNA at these sites must be fully methylated for initiation to occur. Immediately after initiation, only one

strand of these sites is methylated (hemimethylation), delaying new initiations at the sites (see the discussion of sequestration of chromosome origins in chapter 1). Despite substantial progress, however, the method by which the replication of very stringent plasmids, such as P1 and F (with a copy number of only 1), is controlled to within such narrow limits is still something of a mystery and the object of current research.

Mechanisms To Prevent Curing of Plasmids

Cells that have lost a plasmid during cell division are said to be **cured** of the plasmid. Several mechanisms prevent curing, including toxin-antitoxin systems (Box 4.2), site-specific recombinases that resolve multimers, and partitioning systems. The last two are reviewed below.

RESOLUTION OF MULTIMERIC PLASMIDS

The possibility that a cell will lose a plasmid during cell division is increased if the plasmids form dimers or higher multimers during replication. A dimer consists of two individual copies of the plasmid molecules linked head to tail to form a larger circle, and a multimer links more than two such monomers. Such dimers and multimers probably occur as a result of recombination between monomers. Recombination between two monomers forms a dimer, and subsequent recombination can form higher and higher multimers. Also, RC replication of RC plasmids can form multimers if termination after each round of replication is not efficient. Multimers may replicate more efficiently than monomers, perhaps because they have more than one origin of replication, so they tend to accumulate if the plasmid has no effective way to remove them. The formation of multimers creates a particular problem when the plasmid attempts to segregate into the daughter cells on cell division. One reason is that multimers lower the effective copy number. Each multimer segregates into the daughter cells as a single plasmid, and if all of the plasmid is taken up in one large multimer, it can segregate into only one daughter cell. Also, the presence of more than one *par* site on the multimer may cause it to be pulled to both ends of the cell at once, much like a dicentric chromosome can lead to nondisjunction in higher organisms. Therefore, multimers greatly increase the chance of a plasmid being lost during cell division.

To avoid this problem, many plasmids have site-specific recombination systems that resolve multimers as soon as they form. These systems can be either chromosomally encoded or encoded by the plasmid itself (see chapter 9). A site-specific recombination system promotes recombination between specific sites on the plasmid if the same site occurs more than once in the molecule, as it would in a dimer or multimer. This recombination has the effect of resolving multimers into separate monomeric plasmid molecules.

A well-studied example of a plasmid-encoded site-specific recombination system is the Cre-*loxP* system encoded by P1 phage. This phage is capable of lysogeny, and its prophage form is a plasmid, subject to all the problems faced by other plasmids, including multimerization due to recombination. The Cre protein, a tyrosine (Y) recombinase, promotes recombination between two *loxP* sites on a dimerized plasmid, resolving the dimer into two monomers. This system is very efficient and relatively simple and has been useful in a number of studies, including demonstrations of the interspecies transfer of proteins (see chapter 5). It has also been used as a model for Y recombinases, since the recombinase has been crystallized with its *loxP* DNA substrate. Y recombinases and their mechanism of action are discussed in chapter 9.

The best-known examples of host-encoded site-specific recombination systems used to resolve plasmid dimers are the *cer*-XerCD and the *psi*-XerCD site-specific recombination systems used by the ColE1 plasmid and the pSC101 plasmid, respectively. The XerCD system is mentioned in chapter 1 in connection with segregation of the chromosome of *E. coli*. The XerC and XerD proteins are part of a site-specific recombinase that acts on a site, *dif*, close to the terminus of replication of the chromosome to resolve chromosome dimers created during chromosome replication. The plasmids have commandeered this site-specific system of the host to resolve their own dimers by having sites at which the recombinase can act. The site on the ColE1 plasmid is called *cer*, and the site on pSC101 is called *psi*. As described in chapter 1, resolution of dimer chromosomes at *dif* sites requires an interaction between XerCD and the FtsK protein. In contrast, sites on plasmids do not share this requirement because of differences in the sequence where recombination occurs; however, these plasmid base sites have a different requirement. The *cer* site on the ColE1 plasmid is not recognized as such but is recognized only if two other host proteins, called PepA and ArgR, are bound close by in the DNA, as shown in Figure 4.13. A similar situation occurs in the pSC101 plasmid, but there the auxiliary host proteins are PepA and ArcA~P (phosphorylated ArcA), which binds close to where ArgR binds in *cer*. Apparently, these other proteins bind to XerCD recombinase at the plasmid *cer* or *psi* sites and help orient it for the recombination process. However, it is not clear how these particular host proteins came to play this role. The only thing these accessory proteins have in common is that they all normally bind to DNA because they are involved in regulating transcription in the host. This is yet another example of a case where plasmids commandeer host functions for their own purposes, in this case, a site-specific recombination system normally used for resolving chromosome dimers, as well as transcriptional

BOX 4.2

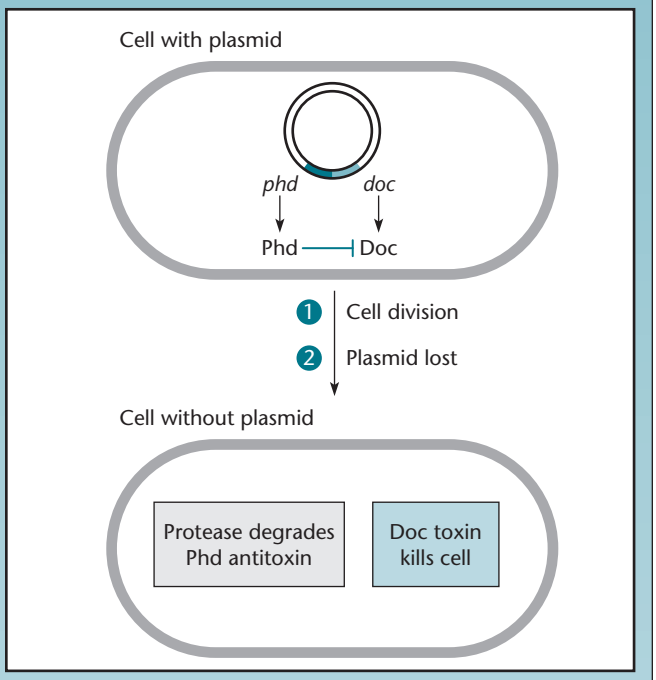
Toxin-Antitoxin Systems and Plasmid Maintenance

As extrachromosomal elements, plasmids can easily be lost during division. As described in the text, numerous adaptations beyond encoding beneficial functions have evolved in plasmids to guard against loss, including copy number control, dimer resolution systems, and partitioning systems. However, plasmids have an additional trick for being maintained on plasmids, called toxin-antitoxin systems. In what seems like revenge, some plasmids encode proteins that kill a cell if it is cured of the plasmid. Such functions are relatively common on mobile plasmids, including the F plasmid, the R1 plasmid, and the P1 prophage, which replicates as a plasmid (see chapter 8). These systems have been called plasmid addiction systems because they cause the cell to undergo severe withdrawal symptoms and die if it is cured of the plasmid to which it is addicted.

Plasmid addiction systems all use basically the same strategy. They consist of two components, which can be either proteins or RNA. One component functions as a toxin, and the other functions as an antitoxin or antidote. While the cell contains the plasmid, both the toxin and the antitoxin are made, and the antitoxin somehow inactivates the effect of the toxin, either by binding to the toxin and inactivating it directly or by somehow indirectly alleviating its effect. The PhD-Doc system is an example of the former type (see the figure). The PhD protein is the antitoxin that binds to the Doc toxin and inactivates it. Restriction-modification systems are an example of the latter type. The restriction endonuclease component of the restriction system is the toxin that cuts the chromosome and kills the cell, but not if the modification component, the antitoxin, has methylated a base in its recognition sequence (see chapter 1). Once the cell is cured of the plasmid, neither the toxin nor the antitoxin continues to be made. The toxin, however, is more stable than the antitoxin, so eventually, the antitoxin is degraded and the cured cells contain only the toxin. Without the antitoxin to counteract it, the toxin interrupts one of the essential processes in the cell.

The toxic proteins can affect any one of many essential processes in bacteria, depending on the system. For example, the toxic protein of the F plasmid, Ccd, alters DNA gyrase so that it causes double-strand breaks in the DNA. The Hok protein of plasmid R1 destroys the cellular membrane potential, causing loss of cellular energy. The Doc toxin of P1 inhibits translation elongation through and interaction with the 30S ribosomal subunit. Revenge aside, it is hard to understand the evolutionary advantage of killing cells after plasmid loss. There may be two interrelated answers to why mobile plasmids may contain toxin-antitoxin systems. One is that plasmid loss would likely lead to better growth of the plasmidless cell. Therefore, the newly cured cells could otherwise overgrow the plasmid-containing cells in the same population, and the toxin-antitoxin systems take away this advantage. A second benefit to having the toxin-antitoxin system on a mobile element is that the cells affected by the toxin could be saved from the toxic effect of plasmid loss if they again obtained the mobile plasmid from an adjacent cell. Therefore, part of the advantage may not be killing by the toxin, but instead, a type of paralysis until the mobile plasmid can return to the cell that lost the plasmid.

There is a good rationale for plasmid addiction systems in that they prevent cells cured of the plasmid from accumulating and thus help ensure survival of the plasmid. Therefore, it was a surprise to discover that similar toxin-antitoxin systems also occur in the chromosome. Some of these are on exchangeable DNA elements, such as genetic islands and superintegrons (see chapter 9), and may play a role similar to that of the addiction systems of plasmids, preventing loss of the DNA element. They could be considered selfish genes



The P1 phage-encoded plasmid addiction system Phd-Doc. Cells containing the plasmid contain both Phd and Doc; Phd is the antidote to Doc, binding to it and inactivating it. If the cell is cured of the plasmid, neither Phd nor Doc is made, but Doc is more stable than Phd and outlives it. Once Phd is degraded by a protease, Doc kills the cell (actually, Doc only inhibits translation, and MazE kills the cell in response).
doi:10.1128/9781555817169.ch4.Box4.2.f

BOX 4.2 (continued)

Toxin-Antitoxin Systems and Plasmid Maintenance

that prevent themselves, and therefore the DNA element in which they reside, from being lost from the cell. However, other toxin-antitoxin modules seem to be encoded by normal genes in the chromosome. Two examples of these are the MazEF and RelBE systems, both found in *E. coli* K-12. These two systems work in remarkably similar ways. MazF and RelE are the toxins and are RNases that cleave mRNA in the ribosome and block translation, killing the cell. MazF and RelE are the antitoxins, which bind to the toxins MazF and RelE, respectively, and inactivate them. The toxins and antitoxins are not made if translation is inhibited, but the toxin is more stable and longer lived than the antitoxin. These systems could therefore be considered suicide modules in that they cause the cell to kill itself if translation is inhibited, for example, by antibiotics or a plasmid addiction module, such as PhD-Doc, if the cell is cured of the P1 plasmid. A number of hypotheses have been proposed to explain the existence of these suicide modules. One is that they help prevent the spread of phages that inhibit host translation. Another is that they help shut down cellular metabolism in response to starvation and help ensure the long-term survival of some of the cells. It might be relevant that bacteria with a free-living lifestyle in the environment tend to have many more such suicide modules than do obligate parasitic bacteria, which can live only in the more stable, nurturing environment of a eukaryotic host.

Another apparent suicide system in *B. subtilis* lends credence to the idea that the purpose of suicide systems might be to kill some bacteria so that others may live (see Ellermeier et al., References). This system is much more complex than the others we have mentioned and consists of many genes in two operons, *skf* and *sdp.* To summarize, when a population of *B. subtilis* cells is starved for nutrients, some of them begin to sporulate. These cells produce a toxin that kills other cells that were slow to start sporulating. The killed cells are then devoured by the cells producing the toxin, which can then reverse their sporulation process. This buys the sporulating cells more time, in case the situation changes and they do not really need to sporulate, allowing them to avoid the need to employ a drastic measure to ensure survival.

References

Ellermeier, C. D., E. C. Hobbs, J. E. Gonzales-Pastor, and R. Losick. 2006. A three-protein signaling pathway governing immunity to a bacterial cannibalism toxin. *Cell* **124**:549–559.

Engelberg-Kulka, H., R. Hazan, and S. Amitai. 2005. *mazEF*: a chromosomal toxin-antitoxin module that triggers programmed cell death in bacteria. *J. Cell Sci.* **118**:4327–4332.

Greenfield, T. J., E. Ehli, T. Kirshenmann, T. Franch, K. Gerdes, and K. E. Weaver. 2000. The antisense RNA of the *par* locus of pAD1 regulates the expression of a 33-amino-acid toxic peptide by an unusual mechanism. *Mol. Microbiol.* **37**:652–660.

Hazan, R., B. Sat, M. Reches, and H. Engelberg-Kulka. 2001. Postsegregational killing by the P1 phage addiction module *phd-doc* requires the *Escherichia coli* programmed cell death system *mazEF. J. Bacteriol.* **183**:2046–2050.

regulatory proteins that are used for host cell gene regulation. XerCD is also a Y recombinase, and its mechanism of action is discussed in more detail in the section on tyrosine recombinases in chapter 9.

PARTITIONING

A very effective mechanism that plasmids have to avoid being lost from dividing cells is **partitioning** systems. These systems ensure that at least one copy of the plasmid is present in each daughter cell after cell division. The functions involved in these systems are called **Par functions**, and in many ways they are analogous to the Par functions involved in chromosome segregation. In fact, the discovery of Par systems in plasmids preceded their discovery in chromosome segregation.

The Par Systems of Plasmids

The Par systems have been studied extensively in plasmids. The Par systems of low-copy-number plasmids fall into at least two groups whose members are related to each other by sequence and function. At least one of these groups of plasmid Par functions is also related to the putative chromosomal Par systems of some bacteria, as described in chapter 1. One group is represented by the Par system of the R1 plasmid of *E. coli*, and the other, much larger group is represented by the Par systems of the F, P1, and broad-host-range RK2 plasmids, among others. The latter is also the group to which the chromosomal partitioning systems from the bacteria *Bacillus subtilis* and *Caulobacter crescentus* belong. The two groups of partitioning systems differ in the details of how they achieve the feat of plasmid partitioning, but evidence is accumulating that they do this by forming dynamic filaments in the cell. The R1 plasmid partitioning system is addressed first, since it seems to be the better understood of the two systems. It is also the best example to date of a dynamic filament-forming structure in bacteria that can move cellular constituents around, analogously to the actin filaments of eukaryotes (see "Bacterial Cell Biology and the Cell Cycle" in chapter 14).

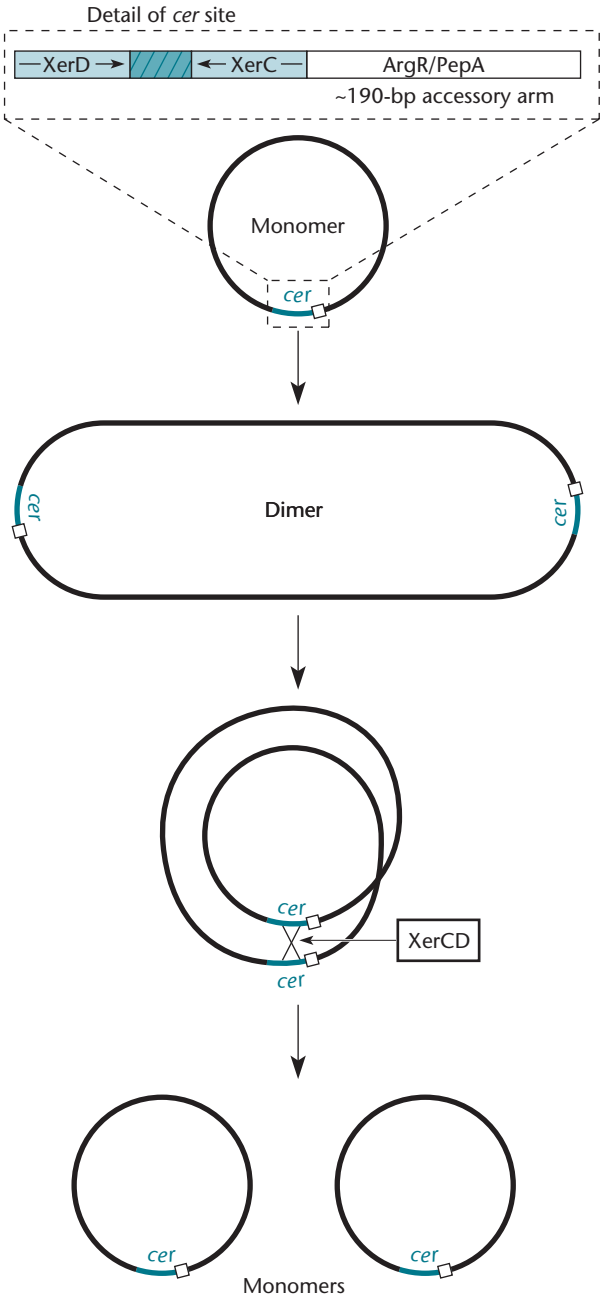


Figure 4.13 The Xer functions of *E. coli* catalyze site-specific recombination at the ColE1 plasmid *cer* site to resolve plasmid dimers. The sites of binding of the host proteins ArgR and PepA are shown.
doi:10.1128/9781555817169.ch4.f4.13

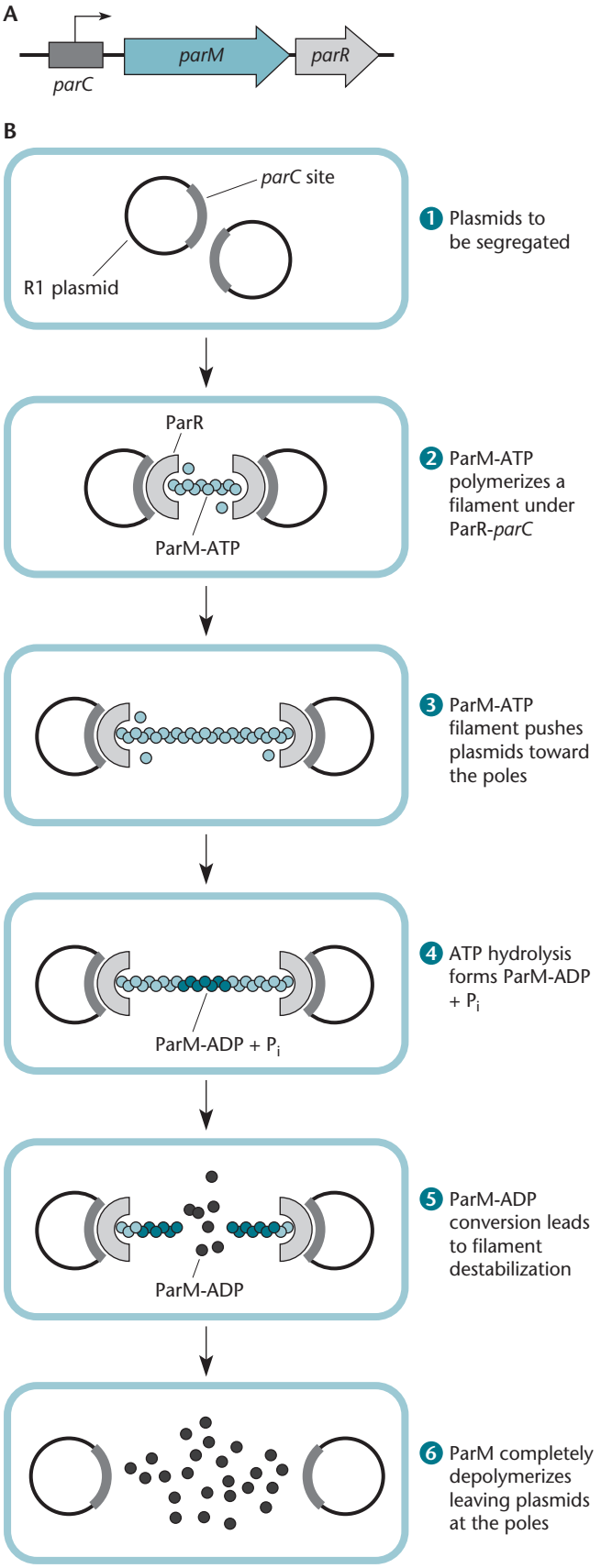
THE R1 PLASMID Par SYSTEM

The mechanism of partitioning by the R1 plasmid is illustrated in Figure 4.14. The partitioning system of the R1 plasmid consists of two protein-coding genes, *parM* and *parR*, as well as a centromere-like *cis*-acting site, *par*C. The actin-like ParM protein forms a filament that pushes the plasmids to the cell poles (see Campbell and Mullins, Suggested Reading). The polymerization process occurs quickly, rapidly pushing the plasmids to the cell poles (Figure 4.14). This is followed by depolymerization, leaving the plasmids at the poles. The polymerization and depolymerization process is regulated by the ATP-bound state of ParM; in the test tube, ParM proteins form a stable filament in the presence of nonhydrolyzable ATP, indicating that the shift from the ATP- to the ADP-bound state is important in the depolymerization process and the dynamic nature of the filaments. The prevailing model that supports the available data holds that polymerization of ParM-ATP occurs under the cap-like ParR-*par*C complex, propelling the complex toward the cell pole. Depolymerization would initiate from the base of the filament as ParM-bound ATP is hydrolyzed to ADP. Unlike other ATP-binding proteins that are also found in ADP-bound form, it appears that ParM can also be found in a third form, bound to ADP plus inorganic phosphate (P_i). The transition state of ParM-ADP plus P_i is thought to help produce a fragile intermediate state that completely collapses once ParM-ADP forms at the base of the filament.

THE P1 AND F PLASMID Par SYSTEMS

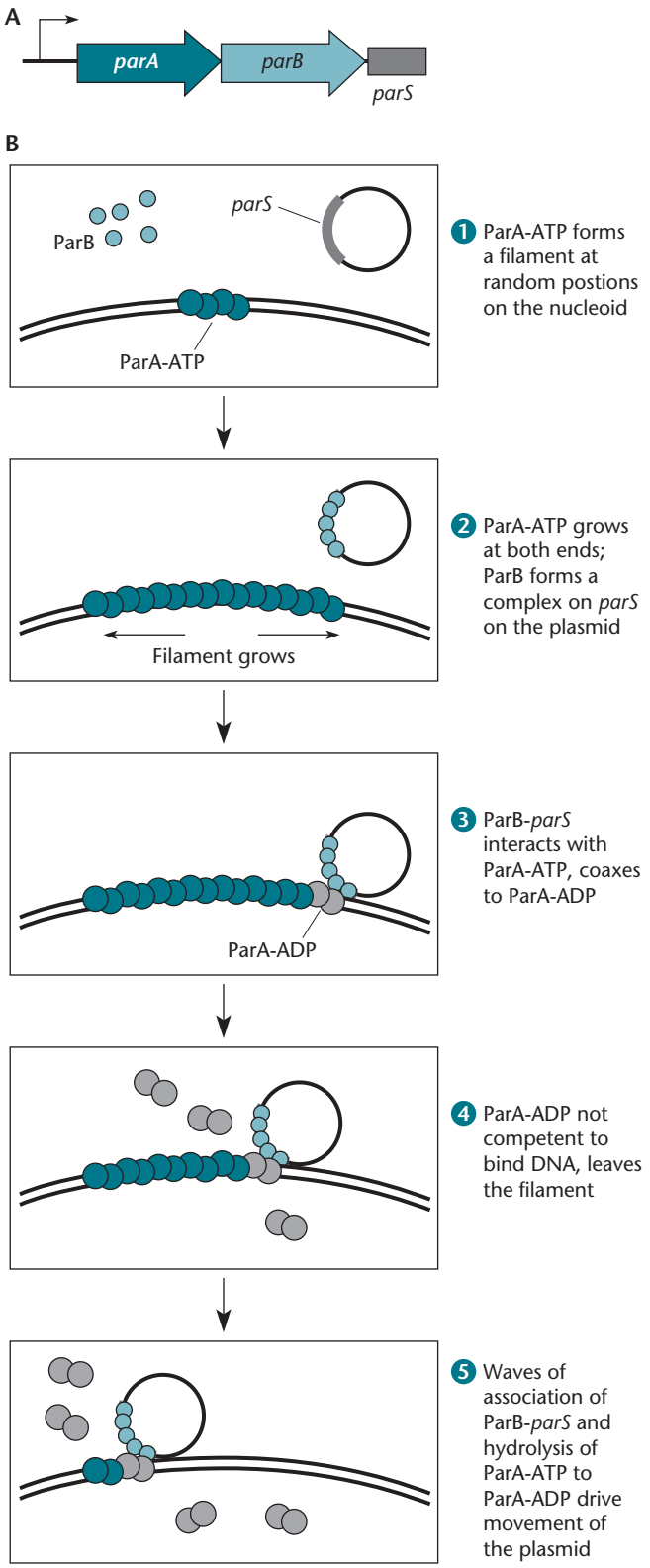
The larger group of plasmid-partitioning systems is related to the systems of the P1 plasmid prophage and the F plasmid. Also, some bacteria, including *B. subtilis* and *C. crescentus*, seem to use a similar Par system to aid in partitioning and organizing their chromosomes (see chapter 1). The composition of these Par systems is grossly similar to that of the Par system of the R1 plasmid in that they usually consist of two proteins and an adjacent *cis*-acting site on the DNA to which one of the proteins, often called ParB, binds. Also, the other protein, often called ParA, has ATP-binding motifs. However, the ATP-binding protein has no homology to actin or any other cytoskeletal protein and belongs to a special family of proteins called the P-loop ATPases. It is distantly related to MinD, a protein that helps select the division site in bacteria and, in *E. coli* at least, oscillates from one end of the cell to the other during the cell cycle (see chapter 1). In the F plasmid, the corresponding proteins are called SopA and SopB, and the site is called *sopC* (*sop*, for stability of plasmid). For simplicity, we refer to them as ParA, ParB, and *parS*. In some plasmids, the Par systems are encoded in the same region of the plasmid as the replication proteins; in rare instances, the system may lack an autonomous ParB protein, and perhaps a larger ParA-related protein may play both roles.

The mechanism of action of these Par systems appears to be fundamentally different from that of the system described above. Recent microscopic analyses of plasmid partitioning with these systems indicated that it involves directed movement across the nucleoid



for plasmid segregation. A model that accounts for the available observations holds that instead of being pushed, as in the R1 system, the ParB-*parS* complex is actually pulled by an attraction to ParA that polymerizes at random places on the nucleoid. In one of the models, this would involve the ParB-*parS* complex being pulled as it depolymerizes ParA filaments (see Ringgaard et al., Suggested Reading) (Figure 4.15). Repeated cycles of ParA filament assembly and disassembly would tend to maximize the distribution of plasmid copies to opposite cell halves, so that upon cell division, the plasmids are appropriately segregated. In this model, the partition system can be separated into two interacting subsystems. In the first subsystem, ATP-bound ParA dimers bind chromosomal DNA at random positions nonspecifically but cooperatively leading to ParA filament formation. In the second part of the system, dimers of ParB bind specifically to the *parS* site on the plasmid and nucleate the binding of new ParB dimers around this site on the plasmid. The ParB-*parS* complex on the plasmid then becomes competent for associating with a ParA filament on the chromosome via one of the ends of the filament. The interaction with ParB signals ParA dimers to hydrolyze ATP to ADP and release of ParA from DNA. As the first dimer of ParA is released, the ParB-*parS* complex then binds to the next ParA dimer on the filament. This behavior would allow the ParB-*parS* complex to be pulled along the nucleoid as it goes through cycles of ParA binding on the receding ParA filament. When the end of the filament is reached, the plasmid dissociates and is free to move to a new filament at another location. The ParA dimers released from disassociating filaments can also move to another location on the chromosome and make new filaments, but the exchange of ADP with ATP is a slow process that favors ParA diffusion to a place different from where it was released. Such a system will naturally distribute *parS*-containing plasmids away from

Figure 4.14 Model for partitioning of the R1 plasmid. **(A)** Structure of the *par* locus of R1, showing the positions of the *parM* and *parR* genes, as well as the *cis*-acting *parC* site. The transcription start site is in the *parC* site. **(B)** ParR binds to *parC*, making a site of ParM-ATP nucleation. Filaments grow by adding successive ParM-ATP subunits under the ParR-*parC* "cap," where growth of the filament pushes the plasmid containing *parC* to the cell poles. The filament is destabilized as ParM hydrolysis converts it to ParM-ADP, possibly through a ParM-ADP plus P_i intermediate, where rapid loss of the filament occurs as a wave of ParM-ADP-mediated instability passes through the filament, destabilizing it. ParM can then be recharged with ATP before the plasmids are partitioned again prior to the next cell division. The *parC* site on the plasmid is shown in grey.
doi:10.1128/9781555817169.ch4.f4.14



one another in the cell. ParA filament formation is difficult to demonstrate *in vitro*, and a variation on this model has been suggested in which the ParA protein accumulates under some other form of polymerization on the nucleoid (see Vecchiarelli et al., Suggested Reading). The ParB-*parS* complex would be attracted to the region where ParA accumulated and, in turn, act to mobilize ParA off the nucleoid by encouraging ATP hydrolysis. In this way, the Par proteins would go through a type of chasing behavior reminiscent of the activity of the MinE and MinD proteins but on a local level across the nucleoid instead of an oscillation across the entire length of the cell.

INCOMPATIBILITY DUE TO PLASMID PARTITIONING

If two plasmids share the same partitioning system, they will be incompatible, even if their replication control systems are different. Incompatibility due to shared partitioning systems makes sense, considering the models presented above. If two plasmids that are otherwise different share the same Par system, one plasmid of each type can be directed to opposite ends of the cell before the cell divides. In this way, one daughter cell can get a plasmid of one type while the other daughter cell gets the plasmid of the other type, and cells are cured of one or the other plasmid. However, even though shared partitioning systems can cause incompatibility, this is usually not the sole cause of their incompatibility. Usually, cells with the same partitioning system also share the same replication control system, since the two are often closely associated on the plasmid; therefore, the incompatibility is due to both systems. In fact, in some cases, the replication control genes and the partitioning genes are intermingled around the origin of plasmid replication.

Figure 4.15 Model for partitioning by *par* systems on P1, F, and RK2. **(A)** Structure of the *par* locus on P1, F, and RK2 showing the positions of the *parA* and *parB* genes and the *cis*-acting *parS* site. Transcription for the promoter (P*parAB*) is controlled by the ParA protein. **(B)** Plasmids containing the *parS* site bound by ParB are pulled across the chromosome as they mediate the depolymerization of ParA filaments across the nucleoid. ParA-ATP binds randomly as a dimer to the nucleoid and polymerizes in both directions, while ParB specifically associates with *parS* sites on the plasmid. Interaction between the ParB-*parS* complex and the ParA filament stimulates ParA to hydrolyze ATP, converting it to ParA-ADP, a form that dissociates from the end of the filament. The ParB-*parS* complex is then free to associate with the next dimer of ParA-ATP in the filament, causing net movement of the ParB-*parS* complex-containing plasmid. ParA-ADP can then convert to its ParA-ATP form and associate at the other end of the filament or elsewhere across the nucleoid.
doi:10.1128/9781555817169.ch4.f4.15

Plasmid Cloning Vectors

As discussed in chapter 1, a cloning vector is an autonomously replicating DNA (replicon) into which other DNAs can be inserted. Any DNA inserted into the cloning vector replicates passively with the vector, so that many copies (clones) of the original piece of DNA can be obtained. Plasmids offer many advantages as cloning vectors, and many types of plasmids have been engineered to serve as plasmid cloning vectors.

As described above, the physical properties of plasmids allow them to be purified easily for manipulation and reintroduction into bacteria. Because very few functions need to reside on the plasmid itself, they can be small, which, in addition to allowing them to be manipulated more easily, also reduces the burden on the cell. Given the universality of the genetic code, they offer a way of expressing proteins from other bacteria or those from other types of organisms. In fact, in one of the first cloning experiments, a frog gene was cloned into plasmid pSC101 (see Cohen et al., Suggested Reading). Cloning vectors have a wide variety of uses that make them important for the study of bacteria or as tools for a better understanding of a broad range of other organisms. For example, sometimes they are used as expression constructs to express a protein for study directly in the host, while at other times they may be used to overexpress a protein to facilitate isolating large amounts of the protein for work outside the cell, such as for structural studies, biochemical studies, or commercial applications (see chapters 2 and 7). Plasmid constructs are very commonly constructed using *E. coli* as a host; these constructs may later be moved into less tractable organisms, where they can also replicate, in what are called shuttle vectors (see below).

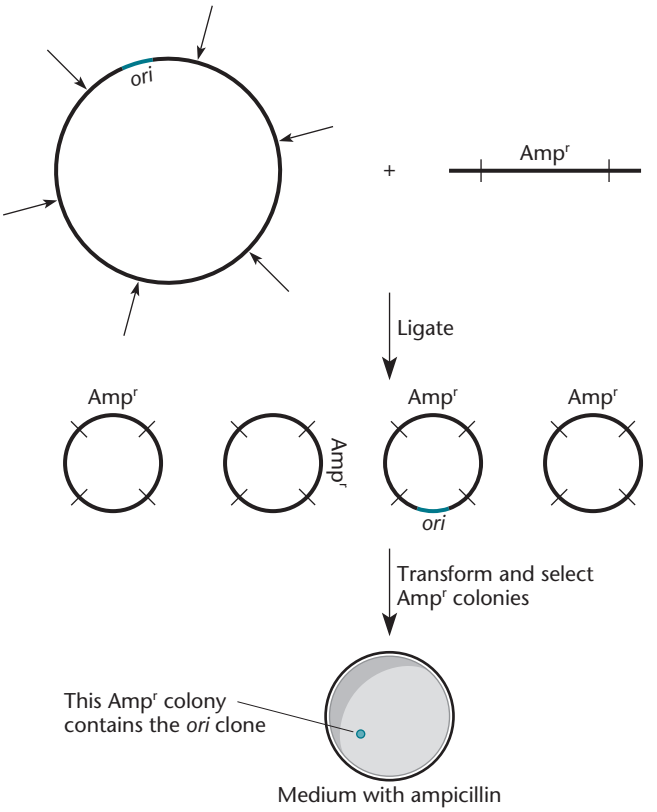
ANATOMY OF A PLASMID CLONING VECTOR

Most plasmids, as they are isolated from nature, are too large to be convenient as cloning vectors and/or do not contain easily selectable genes that can be used to move them from one host to another. Commonly used vectors have a number of attributes that make them more convenient to work with. Some of these features are explained below.

ORIGINS OF REPLICATION

The most basic feature of a plasmid cloning vector is an origin of replication that allows it to replicate independently of the chromosome. In early genetic experiments, the minimal features required for autonomous replication could be isolated with the help of a selectable gene product. For example, the origin of DNA replication region that allows autonomous replication can be cloned when it resides on the same DNA fragment as a selectable marker, such as a gene allowing antibiotic resistance. As shown in Figure 4.16, the plasmid is cut into several pieces with a restriction endonuclease (indicated by the arrows), and the pieces are ligated (joined) to another piece of DNA that has a selectable marker, such as resistance to ampicillin (Amp^r). For the experiment to work, the second piece of DNA cannot have a functional origin of replication. The ligated mixture is then introduced into bacterial cells by transformation, and the antibiotic-resistant transformants are selected by plating the mixture on agar plates containing growth medium and the antibiotic. The only DNA molecules able to replicate and also to confer antibiotic resistance on the cells are hybrids with both the *ori* region of the plasmid and the piece of DNA with the antibiotic resistance gene. Therefore, only cells harboring these hybrid molecules can grow on the antibiotic-containing medium. Using similar techniques, segregation systems could also be isolated as DNA regions that when cloned into these plasmid constructs would stabilize their maintenance after many generations of growth. Given

Figure 4.16 Finding the origin of replication (*ori*) in a plasmid. Random pieces of the plasmid are ligated to a piece of DNA containing a selectable gene but no origin of replication and introduced into cells. Cells that can form a colony on the selective plates contain the selectable gene ligated to the piece of DNA containing the origin.
doi:10.1128/9781555817169.ch4.f4.16



the broad availability of inexpensive DNA sequencing, DNA amplification using PCR has streamlined the manipulation of DNA, but these techniques are likely to continue to be important as more tools are needed for newly identified types of bacteria that lack a system that has been previously characterized.

SELECTABLE GENES

All plasmids exert some kind of load on the host and require a selectable gene to allow maintenance of the vector. In addition, a selectable marker is needed to select for cells that have received the plasmid, because of the low efficiency of the procedures used to introduce DNA. In some cases, it is useful to have plasmids with different selectable genetic markers, because more than one plasmid may need to be maintained at one time. Antibiotic resistance genes have historically been taken from transposons and other plasmids. Some antibiotic resistance genes that have been introduced into cloning vectors are the chloramphenicol resistance (Cam^r) gene of transposon Tn*9*; the tetracycline resistance (Tet^r) gene of plasmid pSC101; the Amp^r gene of transposon Tn*3*; and the kanamycin resistance (Kan^r) gene of transposon Tn*5*. The antibiotic resistance gene that is chosen depends on the uses to which the cloning vector will be put. Some antibiotic resistance genes, such as the Tet^r gene from pSC101, are expressed only in some types of bacteria closely related to *E. coli*, while others, such as the Kan^r gene from Tn*5*, are expressed in most gram-negative bacteria.

INTRODUCING UNIQUE RESTRICTION SITES

Since many applications of plasmid cloning vectors require that clones be introduced into restriction endonuclease cleavage sites, it is necessary that a cloning vector have some restriction sites that are unique. If a site is unique, the cognate restriction endonuclease cuts the vector at only that one site when it is used to cut the plasmid. One can then clone pieces of foreign DNA into the unique site, and the cloning vector will remain intact. Restriction sites for typical restriction endonucleases recognize a 6-bp sequence that on average will occur about once every 1,000 bp (see chapter 1), so that a cloning vector of 3,000 to 4,000 bp is apt to have more than one site for the restriction enzyme. Historically, many tricks have been used to eliminate the extra sites in the plasmid.

In many plasmid cloning vectors, the unique sites are located in a selectable gene, so that insertion of a foreign piece of DNA in the site inactivates the selectable gene. This is called **insertional inactivation** and is discussed below. Normally, during a cloning operation, only a small percentage of the cloning vector molecules pick up a foreign DNA insert. If those that have picked up an insert no longer confer the selectable trait, for example, resistance to an antibiotic, they can be more easily identified. Many cloning vectors also have the unique restriction sites for many different restriction endonucleases all grouped into one small region on the plasmid called a polyclonal or multiple restriction site. This offers the convenience of choosing among a variety of restriction endonucleases for cloning, and the cloned DNA is always inserted at the same general place in the vector, independent of the restriction endonuclease used. **Polyclonal sites** can also be used for **directional cloning**. If the cloning vector is cut by two different restriction endonucleases with unique sites within the polyclonal site, the resulting overhangs cannot pair to recyclize the plasmid. The plasmid can recyclize only if it picks up a piece of foreign DNA. If the piece of DNA to be cloned has overhangs for the two different sites at its ends, it is usually cloned in only one orientation into the polyclonal site.

The unique restriction sites can also be placed so that genes cloned into them will be expressed from promoters and TIRs on the plasmid. Plasmids with these features are called expression vectors and can be used to express foreign genes in *E. coli* and other convenient hosts. Such vectors can also be used to attach affinity tags to proteins to aid in their purification. Expression vectors and affinity tags are discussed in connection with translational and transcriptional fusions in chapter 2. Expression vectors have also been adapted from bacteriophage regulation systems, as discussed in chapter 7.

Examples of Plasmid Cloning Vectors

A number of plasmid cloning vectors have been engineered for special purposes. Almost all of these plasmids have at least some of the features mentioned above for a desirable cloning vector.

1. They are small, so that the plasmid can be easily isolated and introduced into various bacteria.
2. They have relatively high copy numbers, so that the plasmid DNA can be easily purified in sufficient quantities.
3. They carry easily selectable traits, such as a gene conferring resistance to an antibiotic, which can be used to select cells that contain the plasmid.
4. They have one or a few sites for specific restriction endonucleases, which cut DNA and allow the insertion of foreign DNA segments. Also, these sites usually occur in genes that can be easily screened for to facilitate the detection of plasmids with foreign DNA inserts by insertional inactivation.

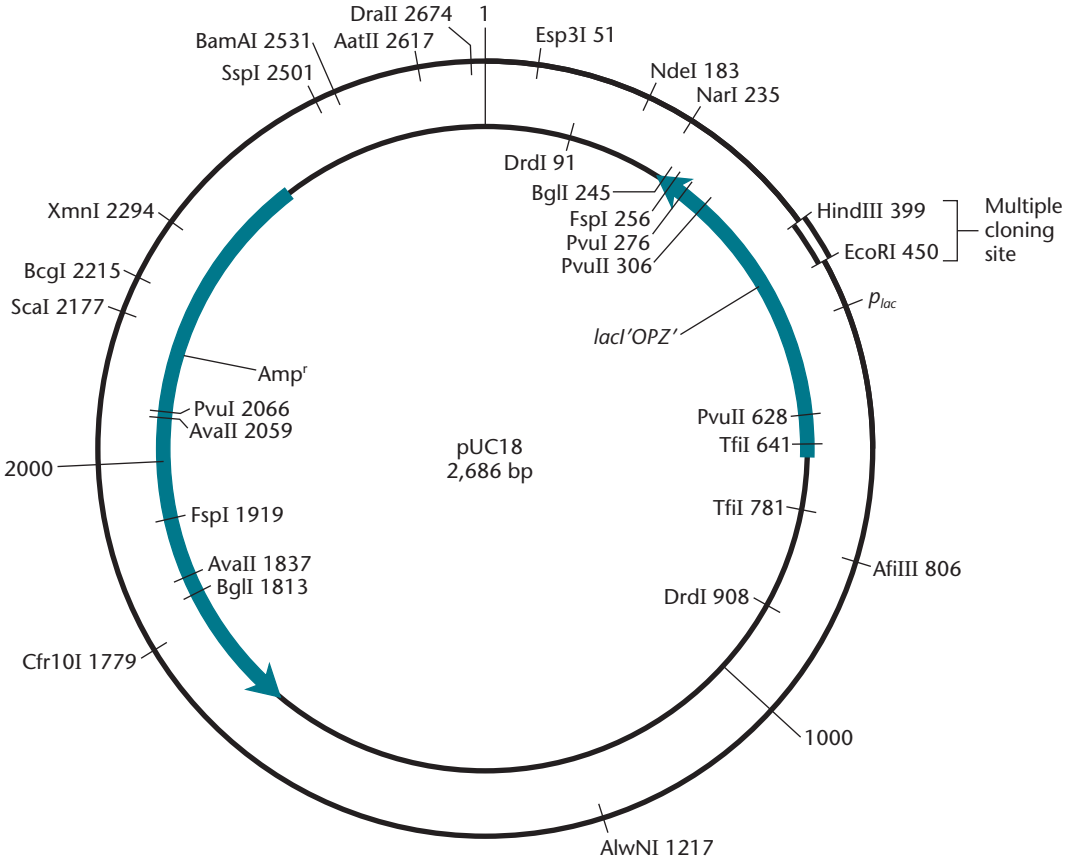
Many plasmid cloning vectors have other special properties that aid in particular experiments. For example, some contain the sequences recognized by

phage-packaging systems (*pac* or *cos* sites), so that they can be packaged into phage heads (see chapters 7 and 8). Expression vectors can be used to produce foreign proteins in bacteria. Mobilizable plasmids have mobilization (*mob*) sites, so they can be transferred by conjugation to other cells (see chapter 5). Some broad-host-range vectors have *ori* regions that allow them to replicate in many types of bacteria or even in organisms from different domains. Shuttle vectors contain more than one type of replication origin, so they can replicate in multiple unrelated organisms. These and some other types of specialty plasmid cloning vectors are discussed in more detail below and in later chapters.

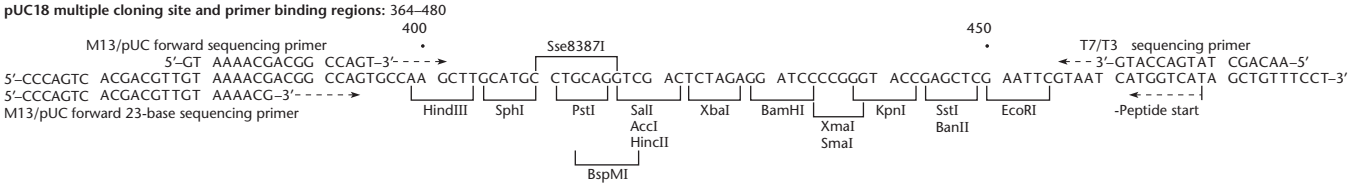
pUC PLASMIDS

Some of the most commonly used plasmid cloning vectors are the pUC vectors and vectors derived from them. One pUC vector, pUC18, is shown in Figure 4.17. The pUC plasmids are very small (only ~2,700 bp of DNA) and, as explained above, have been modified to have copy numbers in the hundreds, making them relatively easy to purify. They also have the easily selectable Amp^r gene.

Figure 4.17 pUC expression vector. A gene cloned into one of the restriction sites in the multiple-cloning site almost invariably disrupts the coding sequence for the *lacZ* α-peptide. If it is inserted in the correct orientation, the gene is transcribed from the *lac* promoter (p_{lac}). If the open reading frame for the gene is in the same reading frame as that for the *lacZ* coding sequence, the gene is also translated from the *lacZ* TIR, and the N-terminal amino acids of *lacZ* become fused to the polypeptide product of the gene. doi:10.1128/9781555817169.ch4.f4.17



pUC18 multiple cloning site and primer binding regions: 364–480



One of the most useful features of these plasmids is the ease with which they can be used for insertional inactivation. They encode the N-terminal region of the *lacZ* gene product, called the α-peptide, which is not active in the cell by itself but complements the C-terminal portion of the protein, called the *lacZ* ω-polypeptide, to make active β-galactosidase by a process termed α-complementation. This enzyme cleaves the indicator dye 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) to produce a blue color. Some host strains, such as *E. coli* JM109, have been engineered to produce the ω-polypeptide of *lacZ* but not the α-peptide. As a consequence, *E. coli* JM109 lacking a plasmid forms white colonies on plates containing X-Gal, but cells containing a pUC plasmid form blue colonies on X-Gal plates because the two segments, the α-peptide encoded in the plasmid and the ω-polypeptide encoded in the chromosome, are capable of interacting to form a functional product capable of cleaving the X-Gal substrate. The pUC plasmids have a multicloning site containing the recognition sequences for many different restriction endonucleases within the coding region for the α-peptide (Figure 4.17). If a foreign DNA fragment is cloned into any one of these sites, the bacterium does not make the α-peptide and the colonies are white on X-Gal plates; bacteria containing plasmids with inserts are therefore easy to identify. These plasmids are also transcription vectors, because the promoter, p_{lac} (Figure 4.17), that drives expression of the α-peptide can also be used to express what is encoded in a piece of DNA directionally cloned into the multicloning site on the plasmid. The *lac* promoter is also inducible and only expressed if an inducer, such as isopropyl-β-D-thiogalactopyranoside (IPTG) or lactose, is added (see chapter 12). Thus, the cells can be propagated before the synthesis of the gene product is induced, a feature that is particularly desirable if the gene product is toxic to the cell. Genes cloned into one of the multicloning sites in the *lacZ* gene can also be translated from the *lacZ* TIR on the plasmid, provided that there are no intervening nonsense codons and the gene is cloned in the same reading frame as the upstream *lacZ* sequences.

CONDITIONAL VECTORS

In some cases, it is useful to have a vector that can replicate only under certain conditions. Low-copy-number vectors with temperature-sensitive replication are often useful because they facilitate plasmid curing. The vector with temperature-sensitive replication is introduced at the permissive temperature of 30°C in *E. coli*. When the cell population is shifted to the nonpermissive temperature of 40°C, replication stops and the plasmid is lost by dilution. Temperature-sensitive vectors can be useful as tools for integrating DNA sequences into the chromosome for gene fusions, as described in chapter 3. In some cases, it is useful to have a plasmid system in which replication occurs only in a certain host background. For example, as described in chapter 9, an in vitro transposition system in which a mobile DNA transposon containing an antibiotic resistance gene moves from a donor DNA plasmid into a different target plasmid as a way to subject it to insertional mutagenesis has been developed. In such a procedure, it is convenient to transform all of the DNA products under conditions where only the donor plasmid can replicate. One popular system utilizes the plasmid R6K origin of replication, called γ (see Metcalf et al., Suggested Reading). The γ origin requires the *trans*-acting protein π for replication. Normally the π protein is encoded by a gene called *pir* on the plasmid. However, in this system, only the *cis*-acting origin remains on the plasmid. Because the π protein is not normally found in *E. coli*, these vectors replicate only in a conditional host where the *pir* product is expressed from the host chromosome. The *pir* gene is either inserted into a neutral site or is introduced through the use of phage λ. Under this system, the plasmid is maintained at a copy number of about 15 with the wild-type *pir* gene. However, a strain can also be used with a mutant *pir* gene, *pir-116*, which maintains the plasmid vector with a copy number of about 250, making it more amenable to plasmid purification.

BACTERIAL ARTIFICIAL CHROMOSOME VECTORS

One problem with using high-copy-number cloning vectors such as pUC vectors is that the clones are very unstable, particularly if they contain large DNA inserts. If the clone exists in many copies, ectopic recombination between repeated sequences in the copies can rearrange the sequences in the clone (see chapter 3). This is a particular problem in some applications, for example, the sequencing of large genomes, such as the human genome, where it is necessary to obtain plasmid libraries containing very large inserts. The DNA of higher eukaryotes, including humans, contains many repeated sequences. For this reason, **bacterial artificial chromosome (BAC) cloning vectors** have been designed (Figure 4.18). These plasmid vectors are based on the F plasmid origin of replication and have a copy number of only 1 in *E. coli*. They can also accommodate very large inserts, on the order of 300,000 bp of DNA. This was expected, since it was known that F′ factors can be very large and are quite stable, especially in a RecA$^-$ host. An F′ factor is a naturally occurring plasmid in which the F plasmid has incorporated a large region of the *E. coli* chromosome (see chapter 5).

The original pBAC vector, shown in Figure 4.18 (see Shizuya et al., Suggested Reading), contains the F plasmid origin of replication and partitioning functions and a selectable chloramphenicol resistance gene.

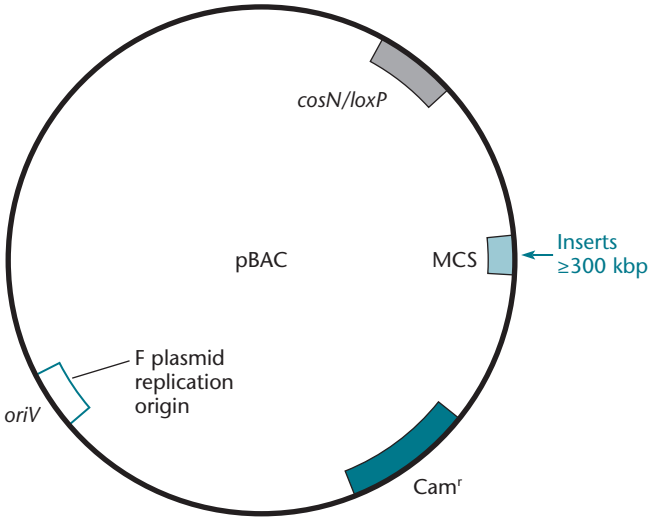


Figure 4.18 pBAC cloning vector for cloning large pieces of DNA for genome sequencing. The multiple cloning site (MCS) where clones are inserted is shown. Also shown are the *loxP* and *cosN* sites, where the plasmid can be cut by the Cre recombinase or λ terminase, respectively, for restriction mapping of the insert. These recognition sites are long enough that they almost never occur by chance in the insert.
doi:10.1128/9781555817169.ch4.f4.18

It also contains unique HindIII and BamHI restriction endonuclease cleavage sites into which large DNA fragments can be introduced, as well as a number of other features that are helpful in the cloning and sequencing of large fragments. Surrounding the cloning sites are the sites for other restriction endonucleases, chosen to be very GC rich so that they are not likely to exist in human DNA, which is relatively low in GC base pairs. This allows the DNA inserts to be excised from the cloning vector without (usually) cutting the DNA insert, as well. More contemporary vectors use sites recognized by an unusual class of enzymes called homing endonucleases, which have recognition sequences of around 30 bp. Even though these sites are not as stringently recognized as those recognized by restriction endonucleases, they still virtually eliminate the possibility that the site will be found in the DNA that is intended to be cloned. Two sites, *loxP* and *cosN*, allow the plasmid to be cut at unique sites for restriction site mapping of the clone.

Broad-Host-Range Cloning Vectors

As explained above, many of the common *E. coli* cloning vectors, such as pBR322, the pUC plasmids, and the pET plasmids, have modified *ori* regions derived from ColE1. These modified ColE1 derivatives maintain the very narrow host range found with the original plasmid (Table 4.2). They replicate only in *E. coli* and a few of its close relatives. However, some cloning applications require a plasmid cloning vector that replicates in other gram-negative bacteria, so cloning vectors have been derived from the broad-host-range plasmids RSF1010 and RK2, which replicate in most gram-negative bacteria. In addition to the broad-host-range *ori* region, these cloning vectors sometimes contain a *mob* site, which can allow them to be mobilized into other bacteria (see chapter 5). This trait is very useful, because ways of introducing DNA other than conjugation have not been developed for many types of bacteria, although electroporation works for many (see chapter 6).

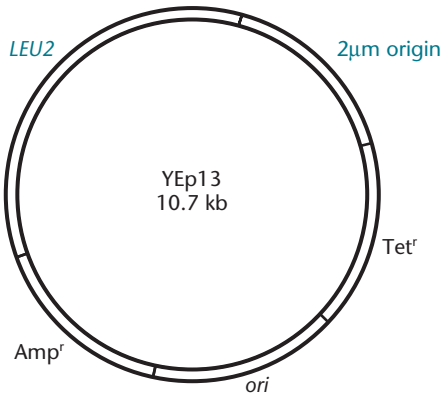
SHUTTLE VECTORS

Sometimes, an experiment requires that a plasmid cloning vector be transferred from one organism into another. If the two organisms are not related, the same plasmid *ori* region is not likely to function in both organisms. Such applications require the use of **shuttle vectors**, so named because they can be used to "shuttle" genes between the two organisms. A shuttle vector has two origins of replication, one that functions in each organism. Shuttle vectors also must contain selectable genes that can be expressed in each organism.

In most cases, one of the organisms in which the shuttle vector can replicate is *E. coli*. The genetic tests can be performed with the other organism, but the plasmid can be purified and otherwise manipulated by the refined methods developed for *E. coli*.

Most plasmid replication functions and antibiotic resistance genes derived from *E. coli* are nonfunctional in gram-positive bacteria, such as *B. subtilis*, which has led to the development of a series of shuttle vectors (Table 4.3). Other shuttle vectors have been developed for other gram-positive bacteria and *E. coli*, whereas still others can be used in a wide variety of eukaryotes. For example, plasmid YEp13 (Figure 4.19) has the replication origin

Figure 4.19 Shuttle plasmid YEp13. The plasmid contains origins of replication that function in the yeast *S. cerevisiae* and the bacterium *E. coli*. It also contains genes that can be selected in *S. cerevisiae* and *E. coli*.
doi:10.1128/9781555817169.ch4.f4.19



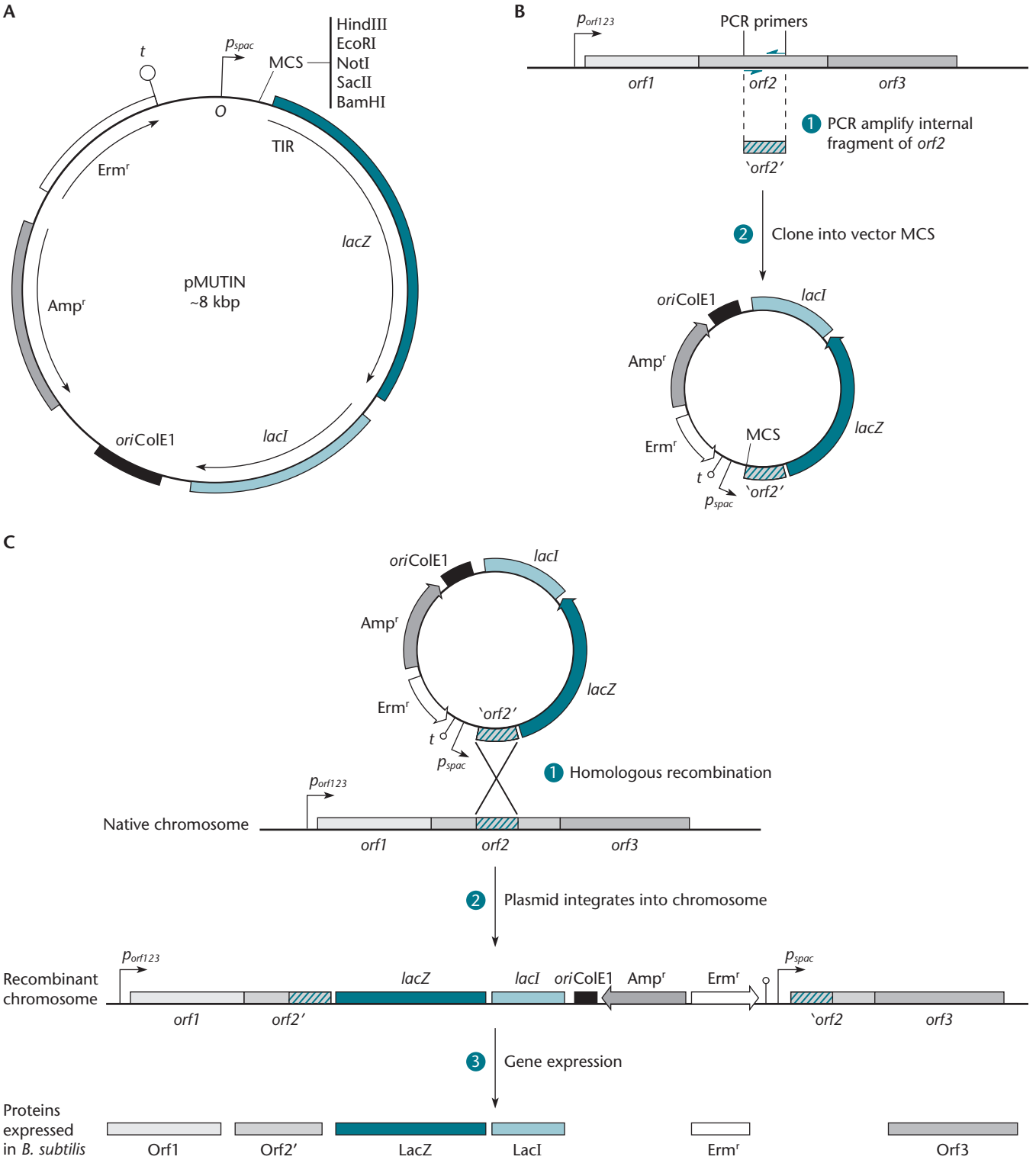


Figure 4.20 A plasmid-based method for genome-wide gene disruption in *B. subtilis.* **(A)** Map of the pMUTIN vector showing the Amp^r gene for selection in *E. coli* and the Erm^r gene for selection in *B. subtilis.* Also shown is the ColE1 origin of replication, which allows replication in *E. coli,* but not *B. subtilis.* The *lacZ* reporter gene includes a translation TIR of a *B. subtilis* gene and the multiple cloning site (MCS) into which PCR fragments can be cloned. The LacI repressor is made from the *lacI* gene of *E. coli,* modified so that it can be expressed in *B. subtilis.* p_{spac} is an inducible hybrid promoter that contains sequences of a promoter from the

of the 2μm circle, a plasmid found in the yeast *Saccharomyces cerevisiae*, so it can replicate in *S. cerevisiae*. It also has the pBR322 *ori* region and thus can replicate in *E. coli*. In addition, the plasmid contains the yeast gene *LEU2*, which can be selected in yeast, as well as an Amp^r gene, which confers ampicillin resistance on *E. coli*. Similar shuttle vectors that can replicate in mammalian or insect cells and *E. coli* have been constructed. Some of these plasmids have the replication origin of the animal virus simian virus 40 and the ColE1 origin of replication.

A VECTOR FOR GENOME-WIDE GENE DISRUPTION IN *B. SUBTILIS*

The advent of whole-genome sequencing has spurred the development of methods to determine systematically the function of each of the thousands of open reading frames of an organism. This requires methods to inactivate each of the open reading frames to begin to determine the function of the product of each of them. A plasmid vector, pMUTIN, that was used for such a systematic analysis has been developed for *B. subtilis* (see Vagner et al., Suggested Reading). Besides allowing gene-by-gene disruption, this vector allows measurement of the expression of the gene, as well as allowing the expression of downstream genes, thereby preventing polarity effects on the downstream genes (see "Polycistronic mRNA" in chapter 2). This analysis was facilitated by the highly efficient natural transformation system of *B. subtilis* (see chapter 6).

A diagram of pMUTIN is shown in Figure 4.20A. The plasmid can be grown in *E. coli*, with selection for the Amp^r gene on the plasmid. However, if the plasmid is transformed into *B. subtilis*, the plasmid is unable to replicate because the ColE1 origin of replication is not functional in the organism. The only way it can be maintained in *B. subtilis* is if it recombines with the chromosome. Recombination into the chromosome requires that the plasmid contain a region of DNA that is homologous to a region on the chromosome which in the experiment comes from the gene of interest. Selecting for the presense of the Erm^r gene by growth on medium containing erythromycin, which is expressed in *B. subtilis*, selects for the rare event where the plasmid has been introduced into the bacterium and recombined into the chromosome. Once integrated, a *lacZ* reporter gene on the plasmid, which has a TIR from *B. subtilis*, is transcriptionally fused to the promoter for the gene into which it has integrated and therefore makes β-galactosidase only under conditions where the target gene is normally transcribed (see chapter 2). An inducible p_{spac} promoter on the integrated plasmid also allows transcription of downstream genes in the operon. This promoter, which contains the *lac* operator, is active only in the presence of the inducer IPTG, because the *E. coli lacI* gene for the Lac repressor (modified for expression in *B. subtilis*) was also introduced into the plasmid (see chapter 12).

Figure 4.20B illustrates the cloning steps needed to use the pMUTIN vector to disrupt the middle gene, *orf2*, of a three-gene operon. A fragment internal to *orf2* is PCR amplified and cloned into the multiple cloning site just downstream of the p_{spac} promoter. Figure 4.20C illustrates what happens when the plasmid containing this clone is introduced into *B. subtilis* and Erm^r transformants are selected. Recombination between the cloned sequences on the plasmid and *orf2* promotes integration of the plasmid by a single crossover, making the cells Erm^r. Because the fragment was internal to *orf2*, both the upstream and downstream copies of *orf2* are incomplete and presumably inactive. If β-galactosidase is expressed, the promoter for the operon, called p_{orf123} (Figure 4.20C), must be active under the growth conditions used. If IPTG is added, *orf3* is also expressed, preventing polarity. Therefore, the only gene that is disrupted in the presence of IPTG is *orf2*, so that any phenotypes observed in the presence of IPTG are due to the disruption of *orf2*, allowing the function of the *orf2* gene to be deduced. An important caveat to this approach is that *orf2* is disrupted but not totally deleted, and the N terminus could still retain some activity. A second caveat is that the downstream *orf3* is expressed from a different promoter, so that its product could be made in larger or smaller amounts than normal; this also has the potential to cause phenotypes.

This vector has been used to disrupt more than 4,100 annotated open reading frames in *B. subtilis*. The effort involved a consortium of laboratories worldwide,

B. subtilis phage SP01 and three *lac* operators (*o*) to which the LacI repressor binds to make it inducible by IPTG. *t* is a strong hybrid transcriptional terminator from λ phage and an rRNA operon. **(B)** Cloning into pMUTIN. **(1)** A fragment internal to the target gene is PCR amplified with primers that add restriction sites compatible with those in the MCS for directional cloning (see chapter 1). **(2)** The fragment is cloned into the MCS on the plasmid. **(C)** Integration of a recombinant vector into the *B. subtilis* chromosome. **(1)** Homologous recombination into the native chromosomal locus. **(2)** Structure of the recombinant chromosome after plasmid integration. **(3)** Products of gene expression. The prime before or after an *orf* or protein indicates that only part of the *orf* or protein remains.
doi:10.1128/9781555817169.ch4.f4.20

especially in Europe and Japan. One outcome was to define a set of approximately 270 essential genes (see Kobayashi et al., Suggested Reading). Approximately 70% of the essential genes were found to have homologs in eukaryotes and archaea. This analysis would have missed essential genes that are redundant or expressed in media or under growth conditions different from those that were used, but more recent analyses have specifically studied duplicated genes (see Thomaides et al., Suggested Reading).

SUMMARY

1. Plasmids are DNA molecules that exist free of the chromosome in the cell. Most plasmids are circular, but some are linear. The sizes of plasmids range from a few thousand base pairs to almost the length of the chromosome itself. Probably the best distinguishing characteristic of a plasmid is that it has a typical plasmid origin of replication with an adjacent gene for a Rep protein rather than a chromosome origin with an *oriC* gene, along with a *dnaA* gene and other genes typical of the chromosomal origin of replication.

2. Plasmids usually carry genes for proteins that are necessary or beneficial to the host under some situations but are not essential under all conditions. Evolution probably selected for plasmids carrying nonessential or locally beneficial genes because it allows the chromosome to remain smaller but still allows bacterial populations to respond quickly to changes in the environment.

3. Plasmids replicate from a unique origin of replication, or *oriV* region. Many of the characteristics of a given plasmid derive from this *ori* region. They include the mechanism of replication, copy number control, partitioning, and incompatibility. If other genes are added to or deleted from the plasmid, it will retain most of its original characteristics, provided that the *ori* region remains.

4. Many plasmids replicate by a theta mechanism, with replication forks moving from a unique origin with leading and lagging template strands, much like circular bacterial chromosomes. Others use an RC mechanism, similar to that used to replicate some phage DNAs and during bacterial conjugation. In RC replication, the plasmid is cut at a unique site, and the Rep protein remains attached to the 5′ end at the cut through one of its tyrosines. The free 3′ end is used as a primer to replicate around the circle, displacing one of the strands. When the circle is complete, the 5′ phosphate is transferred from the Rep protein to the 3′ hydroxyl to form a single-stranded circle. A strand complementary to the single-stranded circle is made, using a different origin, and then the host ligase joins the ends to form two double-stranded circular DNAs. Linear plasmids replicate by more than one mechanism. Some have hairpin ends and replicate from an internal origin around the ends to form dimeric circles that are then processed to form two linear plasmids. Others have a terminal protein at both 5′ ends and extensive inverted-repeat sequences at their ends. They may replicate the ends by some sort of slippage mechanism.

5. The copy number of a plasmid is the number of copies of the plasmid per cell immediately after cell division.

6. Different types of plasmids use different mechanisms to regulate their initiation of replication and therefore their copy numbers. Some plasmids use small complementary RNAs (ctRNAs) transcribed from the other strand in the same region (countertranscribed) to regulate their copy numbers. In ColE1-derived plasmids, the ctRNA, called RNA I, interferes with the processing of the primer for leading-strand replication, called RNA II. In other cases, including the R1, ColIB-P9, and pT181 plasmids, the ctRNA interferes with the expression of the Rep protein required to initiate plasmid DNA replication.

7. Iteron plasmids regulate their copy numbers by two interacting mechanisms. They control the amount of the Rep protein required to initiate plasmid replication, and the Rep protein also couples plasmids through their iteron sequences.

8. Some plasmids have a special partitioning mechanism to ensure that each daughter cell gets one copy of the plasmid as the cells divide. These partitioning systems usually consist of two genes for proteins and a *cis*-acting centromere-like site.

9. If two plasmids cannot stably coexist in the cells of a culture, they are said to be incompatible or to be members of the same Inc group. They can be incompatible if they have the same copy number control system or the same partitioning functions.

10. The host range for replication of a plasmid is defined as all the different organisms in which the plasmid can replicate. Some plasmids have very broad host ranges and can replicate in a wide variety of bacteria. Others have very narrow host ranges and can replicate only in very closely related bacteria.

11. Many plasmids have been engineered for use as cloning vectors. They make particularly desirable cloning vectors for some applications because they do not kill the host, can be small, and are easy to isolate. Some plasmids can carry large amounts of DNA and are used to make BACs for eukaryotic-genome sequencing. Plasmids have been adapted to be used to express cloned genes in bacteria, to do gene replacements in the chromosome, and to perform systematic gene inactivation for functional genomics.

QUESTIONS FOR THOUGHT

1. Why are genes whose products are required for normal growth not carried on plasmids? List some genes that you would not expect to find on a plasmid and some genes you might expect to find on a plasmid.

2. Why do you suppose some plasmids have broad host ranges for replication? Why is it that not all plasmids have broad host ranges?

3. How do you imagine a partitioning system for a single-copy plasmid, such as F, could work? How might a copy number control mechanism work?

4. How would you find the genes required for replication of the plasmid if they are not all closely linked to the *ori* site?

5. How would you determine which of the replication genes of the host *E. coli* (e.g., *dnaA* and *dnaC*) are required for replication of a plasmid you have discovered?

6. The R1 plasmid has a leader polypeptide translated upstream of the gene for RepA, and cleavage of the mRNA by RNase III occurs in the coding sequence for this leader polypeptide. This blocks the translation of the leader polypeptide and also the translation of the downstream *repA* gene to which it is translationally coupled. Do you think it would have been easier just to have the cleavage occur in the coding sequence for the RepA protein itself? Why or why not?

7. Try to design a mechanism that uses inverted-repeat sequences at the ends to replicate to the ends of a linear plasmid without the DNA getting shorter each time it replicates.

PROBLEMS

1. The IncQ plasmid RSF1010 carries resistance to the antibiotics streptomycin and sulfonamide. Suppose you have isolated a plasmid that carries resistance to kanamycin. Outline how you would determine whether your new plasmid is an IncQ plasmid.

2. A plasmid has a copy number of 6. What fraction of the cells are cured of the plasmid each time the cells divide if the plasmid has no partitioning mechanism?

3. The ampicillin resistance gene of plasmid RK2 is unregulated. The more copies of the gene a bacterium has, the more gene product is made. In this case, the resistance of the cell to ampicillin is higher the more of these genes it has. Use this fact to devise a method to isolate mutants of RK2 that have a higher than normal copy number (copy-up mutations). Determine whether your mutants have mutations in the Rep-encoding gene.

4. Outline how you would determine whether a plasmid has a partitioning system.

5. What would be the effect of mutating one of the two complementary sequences in structures I and III of the ColIb-P9 plasmid origin region on the copy number of the plasmid? What would it be in the presence and absence of the Inc antisense RNA?

SUGGESTED READING

Azano, K., and K. Mizobuchi. 2000. Structural analysis of late intermediate complex formed between plasmid ColIb-P9 Inc RNA and its target RNA. How does a single antisense RNA repress translation of two genes at different rates? *J. Biol. Chem.* **275:**1269–1274.

Bagdasarian, M., R. Lurz, B. Ruckert, F. C. H. Franklin, M. M. Bagdasarian, J. Frey, and K. N. Timmis. 1981. Specific purpose plasmid cloning vectors. II. Broad host, high copy number, RSF1010-derived vectors, and a host vector system for cloning in *Pseudomonas*. *Gene* **16:**237–247.

Bao, K., and S. N. Cohen. 2001. Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. *Genes Dev.* **15:**1518–1527.

Campbell, C. S., and R. D. Mullins. 2007. In vitro visualization of type II plasmid segregation: bacterial actin filaments pushing plasmids. *J. Cell. Biol.* **179:**1059–1066.

Caspi, R., M. Pacek, G. Consiglieri, D. R. Helinski, A. Toukdarian, and I. Konieczny. 2001. The broad host range replicon with different requirements for replication initiation in three bacterial species. *EMBO J.* **20:**3262–3271.

Cohen, S. N., A. C. Y. Chang, H. W. Boyer, and R. B. Helling. 1973. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl. Acad. Sci. USA* **70:**3240–3244.

Das, N., M. Valjavec-Gratian, A. N. Basuray, R. A. Fekete, P. P. Papp, J. Paulsson, and D. K. Chattoraj. 2005. Multiple homeostatic mechanisms in the control of P1 plasmid replication. *Proc. Natl. Acad. Sci. USA* 22:2856–2861.

Eberhard, W. G. 1989. Why do bacterial plasmids carry some genes and not others? Plasmid **21:**167–174.

Funnell, B. E., and G. J. Phillips (ed.). 2004. *Plasmid Biology.* ASM Press, Washington, DC.

Hamilton, C. M., M. Aldea, B. K. Washburn, P. Babitzke, and S. R. Kushner. 1989. A new method for generating deletions and gene replacements in *Escherichia coli*. *J. Bacteriol.* **171:** 4617–4622.

Khan, S. A. 2000. Plasmid rolling circle replication: recent developments. *Mol. Microbiol.* **37:**477–484.

Kim, G. E., A. I. Derman, and J. Pogliano. 2005. Bacterial DNA segregation by dynamic SopA polymers. *Proc. Natl. Acad. Sci. USA* **102:**17658–17663.

Kobayashi, K., S. D. Ehrlich, A. Albertini, G. Amati, K. K. Andersen, M. Arnaud, et al. 2003. Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. USA* **100:**4678–4683.

Kolb, F. A., C. Malmgren, E. Westof, C. Ehresmann, B. Ehresmann, E. G. H. Wagner, and P. Romby. 2000. An unusual structure

formed by anti-sense target binding involves an extended kissing complex and a four-way junction and side-by-side helical alignment. *RNA* **6:**311–324.

McEachern, M. J., M. A. Bott, P. A. Tooker, and D. R. Helinski. 1989. Negative control of plasmid R6K replication: possible role of intermolecular coupling of replication origins. *Proc. Natl. Acad. Sci. USA* **86:**7942–7946.

Meacock, P. A., and S. N. Cohen. 1980. Partitioning of bacterial plasmids during cell division: a *cis*-acting locus that accomplishes stable plasmid maintenance. *Cell* **20:**529–542.

Metcalf, W. W., W. Jiang, L. L. Daniels, S. K. Kim, A. Haldimann, and B. L. Wanner. 1996. Conditionally replicated and conjugative plasmids carrying *lacZ* alpha for cloning, mutagenesis, and allele replacement in bacteria. *Plasmid* **35**:1–13

Novick, R. P., and F. C. Hoppensteadt. 1978. On plasmid incompatibility. *Plasmid* **1:**421–434.

Novick, R. P., S. Iordanescu, S. J. Projan, J. Kornblum, and I. Edelman. 1989. pT181 plasmid regulation is regulated by a countertranscript-driven transcriptional attenuator. *Cell* **59:**395–404.

Peters, J. E. 2007. Gene transfer in gram-negative bacteria, p. 735–755. *In* C. A. Reddy, T. J. Beveridge, J. A. Breznak, G. A. Marzluf, T. M. Schmidt, and L. R. Snyder (ed.), *Methods for General and Molecular Microbiology*, 3rd ed. ASM Press, Washington, DC.

Radloff, R., W. Bauer, and J. Vinograd. 1967. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells. *Proc. Natl. Acad. Sci. USA* **57:**1514–1521.

Ringgaard, S., J. von Zon, M. Howard, and K. Gerdes. 2009. Movement and equipositioning of plasmids by ParA filament disassembly. *Proc. Natl. Acad. Sci. USA* **106:**19369–19374.

Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using a F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89:**8794–8797.

Thomaides, H. B., E. J. Davison, L. Burston, H. Johnson, D. R. Brown, A. C. Hunt, J. Errington, and L. Czaplewski. 2007. Essential bacterial functions encoded by gene pairs. *J. Bacteriol. 189:*591–602.

Thomas, C. M. 2000. Paradigms of plasmid organization. *Mol. Microbiol.* **37:**485–491.

Vagner, V., E. Dervyn, and S. D. Ehrlich. 1998. A vector for systematic gene inactivation in *Bacillus subtilis*. *Microbiology* **144:**3097–3104.

Vecchiarelli, A. G., Y. W. Han, X. Tan, M. Mizuuchi, R. Ghirlando, C. Biertümpfel, B. E. Funnell, and K. Mizuuchi. 2010. ATP control of dynamic P1 ParA-DNA interactions: a key role for the nucleoid in plasmid partition. *Mol. Microbiol.* **78**:78–91.

Yao, S., D. R. Helinski, and A. Toukdarian. 2007. Localization of the naturally occurring plasmid ColE1 at the cell pole. *J. Bacteriol. 189:*1946–1953.