Recombinant DNA

Any DNA entering the cell may contain valuable genes; more likely, however, the DNA fragment will be either valueless or actively deleterious. Prokaryotes use restriction endonucleases (discussed below) to limit their uptake of foreign DNA. Eukaryotes generally lack restriction enzymes; instead, they sequester their DNA in the nucleus as one major mechanism for limiting the exposure of the eukaryotic genetic material to DNA from outside sources.

It is likely that cells could degrade foreign DNA much more effectively than they do. Some scientists have proposed that cells need at least limited access to foreign DNA as an additional (somewhat risky) method for evolving to meet changing environments. Because cells occasionally do take up foreign DNA, it is possible to perform experiments in which DNA uptake is crucial for the experiment to work.

General procedure

DNA from one organism is obtained and modified in some way to introduce any necessary changes. The modified DNA is then transferred to the host organism. Host cells that have taken up the DNA properly must be found and grown separately from those that did not. Finally, the modified organism must be tested to determine if the procedure worked as intended.

Performing this process requires a variety of techniques and reagents. The requirements include:

1. A method for purifying DNA
2. A method for copying a DNA strand
3. A method for determining the sequence of a DNA fragment
4. A method for cutting DNA specifically
5. A method for connecting two DNA fragments
6. A method for inserting DNA into an organism
7. A method for determining which cells contain the DNA

If these requirements are fulfilled, a wide variety of new properties can be introduced into any organism. Molecular biology would be impossible without the benefit of enzymes and other molecules that organisms originally evolved for their own purposes.

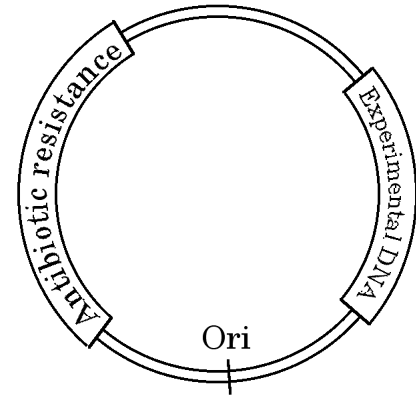
Molecular biology in its modern meaning (*i.e.* referring to genetic manipulation and analysis techniques) is a fairly new science; nearly all of the techniques used were invented after 1970. This means that techniques are still being invented, and many of the procedures have changed (often dramatically) during the last few years.

Plasmids

One method for generating large amounts of DNA is to allow bacteria to replicate the DNA and then purify the replicated DNA from the bacteria. In the majority of cases, the DNA to be replicated for this method is plasmid DNA. A **plasmid** is a double-stranded DNA molecule that will replicate in an organism. A typical plasmid used for molecular biology contains at least four features.

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.

1) The plasmid must be **circular**, because bacteria generally will not replicate linear DNA. 2) The plasmid must contain a sequence that functions as an **origin of replication** (ori). 3) The plasmid must contain a **selection mechanism** that will force the bacteria to retain the DNA; the most common type of selection mechanism used in bacteria is a gene for resistance to an antibiotic such as ampicillin. 4) The plasmid must contain a region for the insertion of the **experimental DNA**. A generic plasmid exhibiting these features is shown at right (the insertion sites are presumed to be present in the regions lacking other functions.

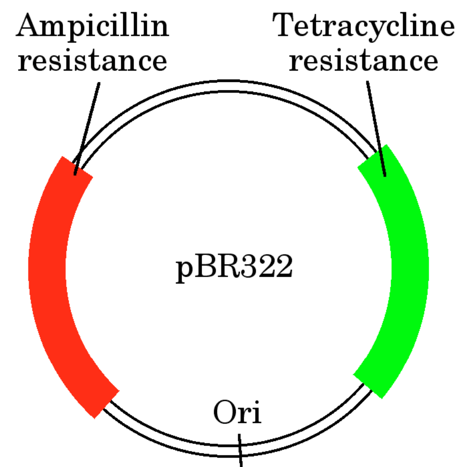


An **expression plasmid** is a specific type of plasmid used to allow expression of heterologous DNA. An expression plasmid must therefore have, in addition to the features listed above, a **strong promoter** element that will drive transcription of the foreign DNA[1] in the host organism, and an effective ribosome binding site that will allow efficient translation of the transcribed RNA.

Because plasmids are much smaller than chromosomal DNA (for *E. coli*, a typical plasmid contains 5 to 10 kilobase pairs (kb), while *E. coli* chromosomal DNA contains about 4,800 kb), separating the two types of DNA molecule is relatively straightforward. In addition, most plasmids used in molecular biology are "high copy number plasmids"; in other words, each bacterium contains many copies (usually >100) of the plasmid. Therefore, although each plasmid molecule is much smaller than the chromosome, plasmid DNA often comprises ~10% or more of the total DNA in the bacterium.

Plasmids are derived from naturally occurring extrachromosomal DNA that bacteria used for exchanging genetic material. In many cases, the naturally occurring plasmids contained genes that coded for antibiotic resistance.

Once of the first cloning vectors was **pBR322**. The name is derived from the names of its creators, F. Bolivar and R. Rodriguez. The plasmid pBR322 contains two antibiotic resistance genes. One of these genes codes for resistance to ampicillin and the other for resistance to tetracycline. This dual antibiotic resistance was designed to allow positive selection, in which only the cells containing the plasmid would be able to grow. In addition, pBR322 was designed so that foreign DNA could be inserted such that it would disrupt the other antibiotic resistance gene and inactivate it. This would allow a negative screen that would kill any cells that had



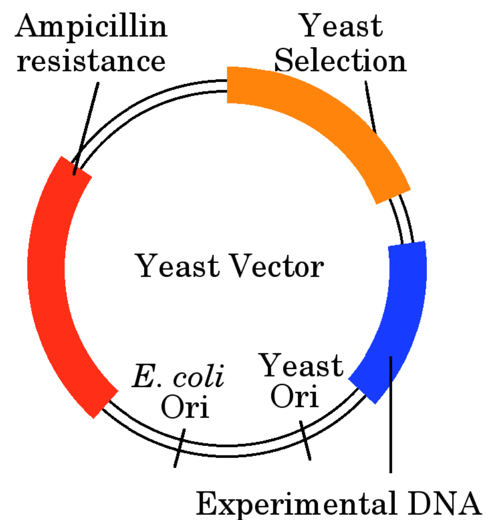
[1]Note that "foreign DNA" can be from any organism, including the organism used as a host. Thus, *E. coli* can be used to express *E. coli* proteins from expression plasmids. In most cases, however, the DNA being expressed is from a different organism, and is being expressed in *E. coli* for convenience.

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.

taken up the plasmid with the foreign DNA. (This negative screening technique is obsolete because much simpler methods of screening for DNA sequences are currently available.)

More recent plasmids generally contain only a single antibiotic resistance gene. In addition, more recent plasmids generally contain multiple cloning sites (MCS), which are relatively short regions containing large numbers of unique restriction sites. A well-designed example of a cloning vector is pUC18, which is distantly related to pBR322, but is somewhat smaller. Vectors such as pUC18 are high copy number plasmids, because they contain highly efficient origins of replication that result in more than 100 copies of the plasmid per cell.

In working with DNA to be put into yeast or eukaryotic organism, it is frequently much easier to work with the DNA using *E. coli* until last step. The plasmid must therefore contain a selectable marker for both *E. coli* and the other organism. In addition, the vector needs an origin of replication for *E. coli*, and a second origin of replication for the other organism. These plasmids are sometimes called **shuttle vectors**, because they allow DNA to be shuttled from one organism to another. (Note, however, that the DNA must be purified from the *E. coli*, and then transfected into the other organism; the DNA does not "shuttle" on its own.)



Restriction enzymes

Restriction enzymes are crucial reagents in molecular biology, because they only cleave specific DNA sequences. Molecular biology would be extremely difficult without restriction enzymes. Fortunately, different bacteria have evolved a large variety of restriction enzymes, and enzymes with specificity for large numbers of sequences are now commercially available.

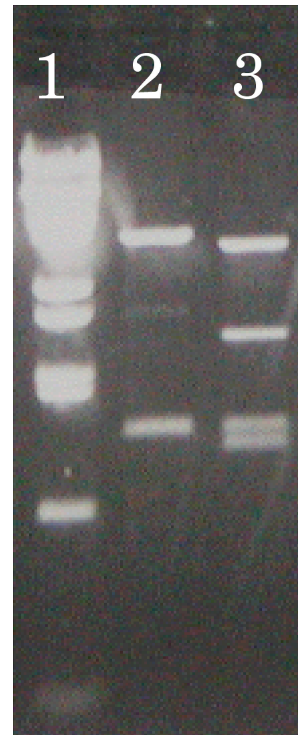
Cells protect their own DNA by methylating it. This works because the restriction enzyme from a particular cell type will not cleave DNA with the correct methylation pattern at its recognition site. Different species have different restriction enzymes and methylases, and therefore limit their ability to use foreign DNA.

In order to decide which restriction enzymes to use for a particular process, you will need to know the sequence of the DNA molecules of interest. A number of computer programs automate the process of analyzing DNA sequences for the presence of restriction sites.

Restriction digestion is typically used for two purposes: restriction mapping and specific DNA cleavage for the production of new constructs.

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.

Plasmid preparation procedures are non-specific: they can be used to purify any plasmid present within the bacteria. This is a major advantage, because it means that the protocol does not need to be changed for different plasmids. However, it also means that it is possible to purify the wrong plasmid if the cells used did not contain the intended plasmid DNA. (For example, the identity of the plasmid harbored by a cell is frequently in doubt during a construction procedure.) A simple method for checking the identity of plasmids is a technique called **restriction mapping**, in which the plasmid is subjected to digestion with different restriction endonucleases. Because restriction enzymes cleave DNA at specific sequences, the size of the resulting fragments can be predicted. If the observed restriction fragment sizes match the predicted sizes, it is likely that the plasmid is indeed correct.



In the gel shown at right, lane 1 shows molecular weight standards, while lanes 2 and 3 show the results of cleaving two different plasmids with a set of restriction enzymes. Note the difference in band patterns resulting from the absence of some of the restriction sites in Plasmid 2 compared to Plasmid 3.

In performing plasmid construction, cleavage of DNA molecules is usually necessary to create the fragments to be ligated together. In some cases, restriction sites will be present in appropriate places in the parent DNA molecules. On the other hand, in many cases, no restriction sites will be available; in these cases, one important feature of PCR is the ability to readily engineer restriction sites at the ends of the PCR products.

Restriction enzymes must cleave both strands of the double stranded DNA. They can do this in a number of ways. Some restriction enzymes cleave both strands at the same location, resulting in a "blunt" end. Other restriction enzymes cleave at different locations on the different strands, leaving short stretches of single stranded DNA. For commercially available restriction enzymes, these short stretches of single stranded DNA are typically either two bases or four bases long (see the comments on compatible enzymes below).

Most, although not all, restriction enzymes cleave palindromic sequences. In molecular biology, a palindrome is a sequence that reads the same on one strand as it does on the opposite strand. For example, the sequence 5´-GAATTC will have the identical sequence on the complementary strand:

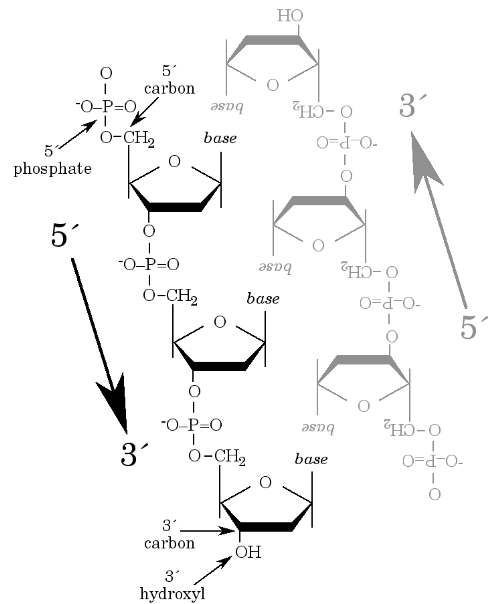
The palindromic nature of most restriction enzymes means that for enzymes that do not leave blunt ends, the overhangs for both strands are identical. These ends are considered compatible (see below).

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.

Ligase

The process of connecting two pieces of DNA together is called **ligation**, and is catalyzed by an enzyme called **ligase**. The ligase used most often in molecular biology is derived from the T4 bacteriophage, and uses ATP to supply the energy necessary for the reaction. In addition to ATP, *T4 ligase requires DNA with a 5´-phosphate group and free 3´-hydroxyl groups*.

The drawing at right shows a (very short) region of double stranded DNA. Both strands of the DNA molecule contain 5´-phosphates and free 3´-hydroxyl groups; this DNA molecule is therefore capable of being ligated. This DNA fragment is blunt ended (*i.e.* all of the bases are paired with bases from the opposite strand); note that some restriction enzymes leave blunt ends, while others leave "overhangs", which are short stretches of single stranded DNA at either the 5´ or 3´ end.



Synthetic oligonucleotides contain free 5´-hydroxyl groups, and therefore must be subjected to phosphorylation prior to ligation. In contrast, most (although not all) restriction enzymes leave 5´-phosphate groups; most restriction fragments can be ligated immediately after digestion.

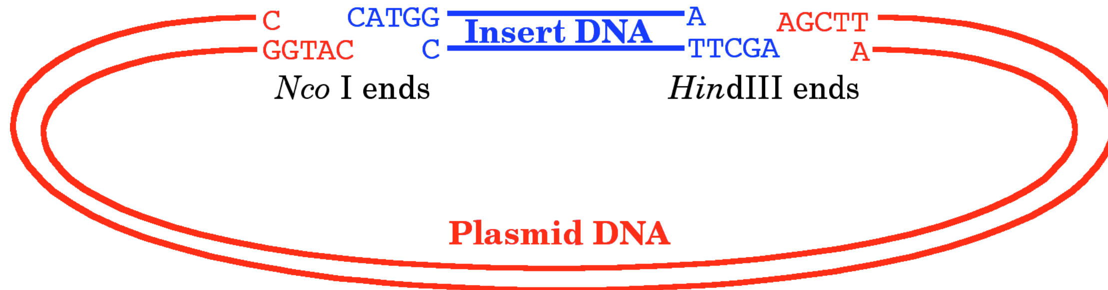
Ligation also requires compatible ends to the DNA. The drawings below show examples of different types of compatible and incompatible ends (N's imply that any arbitrary sequence would could be present).

Compatible Blunt ends		Compatible 4-base, 5´-sticky ends		Compatible 2-base, 3´-sticky ends	
5´–NNNNN	NNNNN	5´–N	CATGN	5´–NNNTG	NNN
3´–NNNNN	NNNNN	3´–NGTAC	N	3´–NNN	ACNNN

Incompatible (blunt/non-blunt ends)		Incompatible 4-base, 5´-overhangs		Incompatible 4-base, 3´-overhangs	
5´–NNNNN	N	5´–N	AGCTN	5´–NCTAG	N
3´–NNNNN	GTACN	3´–NGTAC	N	3´–N	AATTN

The drawing below shows DNA fragments generated from restriction digests that used *Nco* I and *Hin*dIII; these enzymes both leave 4-base 5´-overhangs. The end generated by digestion with one of these enzymes is compatible with other ends generated by the **same** enzyme, but not by the ends generated by the other enzyme. In attempting to construct a plasmid, scientists typically try to take advantage of the specificity of the restriction enzymes and of the ligase to force the creation a plasmid with the insert in the correct orientation.

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.



In setting up a ligation reaction, it is usually desirable to use a molar excess of insert DNA relative to plasmid DNA. The insert DNA will have no effect on the cells because it lacks the features required for replication; having a molar excess of insert DNA makes it more likely that the ligase will find both plasmid DNA and insert DNA molecule to connect together, and reduces the chance that the incompatible ends of the plasmid DNA will ligate to form circular plasmid DNA.

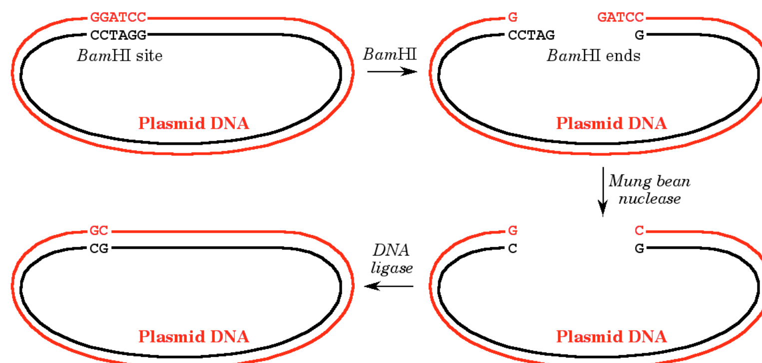
DNA modifying enzymes

Ligase requires a 5´ phosphate. It is therefore sometimes necessary to add a 5´ phosphate to a DNA molecule in order to allow the DNA fragment to be ligated. The T4 bacteriophage enzyme **polynucleotide kinase** is frequently used to phosphorylate DNA.

Alternatively, **alkaline phosphatase** is used to remove phosphate from fragments for which ligation is undesirable.

Finally, exonucleases cleave nucleotides from the ends of DNA fragments. This trimming may be useful in preparing a DNA fragment for a construction procedure.

An example of the use of exonucleases is a procedure intended to remove a restriction site from a plasmid. Mung bean nuclease cleaves single stranded linear DNA molecules beginning at the 5´ end. Cleavage of a DNA molecule with a restriction enzyme (*Bam*HI in the example below) leaves DNA with four base 5´ overhangs. Treatment with mung bean nuclease removes the overhangs, creating blunt ends that can be ligated together. Ligation of the blunt ends results in re-formation of the plasmid without the *Bam*HI site.



Copyright © 1999 – 2003 by Mark Brandt, Ph.D.

Polymerases

The DNA polymerases used for molecular biology are generally based on the enzymes responsible for DNA repair. The replication polymerases are typically large protein multimers exhibiting initiation conditions that are far too complex to be useful for molecular biological purposes; in contrast most of the DNA repair enzymes are monomeric enzymes that will initiate DNA synthesis whenever they are exposed to a primer, a denatured template, and substrate nucleotide triphosphates.

The first widely used DNA polymerase was the Klenow fragment of *E. coli* DNA polymerase I. This enzyme is still used, but more specialized enzymes are more frequently used. The first of these is the DNA polymerase derived from the T4 bacteriophage. The T4 DNA polymerase and some mutant forms with modified properties are widely used for DNA sequencing.

Other widely used polymerases include the heat stable enzymes used for PCR. The earliest, and still one of the most commonly used is the polymerase from *Thermus aquaticus*. The *Taq* polymerase is readily purified from *E. coli* expressing the enzyme, and efficiently catalyzes primer extension under a wide variety of conditions. The *Taq* polymerase has a significant error rate, which has led to the use of some other thermostable polymerases. The other polymerases generally have higher fidelity, but are also generally more sensitive to inhibition by minor contaminants in the reaction.

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.