

## Major concepts .

- Vectors for cloning large DNA inserts
  - Vectors derived from bacteriophage lambda
  - Cosmids
  - Bacterial artificial chromosomes
  - Yeast Artificial chromosomes
- M13 vectors (not in textbook)
  - replicative cycle of single stranded DNA viruses
  - site-directed mutagenesis
- cDNA
  - Isolation of polyadenylated mRNA
  - Reverse transcriptase
  - Removal of RNA template
  - Second strand synthesis
  - cDNA clones
- Recombinant DNA libraries.
  - "Shotgun" cloning
  - Hybridization probes
  - Selection of clones containing specific sequences
- Expression vectors
  - Promoter supplied by vector
  - Shine and Delgarno sequence supplied by vector
  - Control of promoter activity
  - Modifying eukaryotic cDNAs for expression in prokaryotic hosts.

**Vectors for large DNA inserts:** As shown in table 9.3 of the textbook, plasmid vectors do not work well for DNA inserts longer than about 10 kb in length. The abbreviation "kb" stands for sequence length in kilobases (actually kilobase pairs for double stranded DNA). Many different types of vectors have been developed for the cloning of longer DNA inserts. This lecture examines four such vectors:

1. Modified bacteriophage lambda, in which the genes that are only needed for lysogeny are replaced with a cloned insert of 12-20 kb;
2. Cosmids, which combine features from plasmids and bacteriophage lambda and can be used for inserts up to 46 kb;
3. Bacterial artificial chromosomes, which are highly engineered F factors that can carry up to 300 kb of inserted DNA;
4. Yeast artificial chromosomes, which have all essential elements of eukaryotic chromosomes, and can carry up to 500 kb

**Lambda phage vectors:** The life cycle of bacteriophage lambda was described on pages 202-205 and control over the choice between lysis and lysogeny was described on pages 233-237 (which we skipped over this year). In the replicative form, the bacteriophage lambda genome is a closed circle containing about 50 kb of DNA. However, before the genome is packed into phage heads, it is cut once by an endonuclease to generate a linear

DNA with cohesive (sticky) ends, which can be rejoined to form a new circle when the bacteriophage injects its DNA into a new host cell. The cohesive ends are called "*cos*-sites". The cut site essentially separates the early and late parts of the right operon, such that in the linearized genome, genes for head and tail protein are at the left, genes related to lysogeny are near the middle, and genes related to DNA synthesis and lytic cycle on the right. The lambda genome is converted to a vector by removing the genes related to lysogeny and replacing them with a "stuffer" sequence of about 15 kb that contains selectable markers, as well as restriction endonuclease cut sites at both ends (figure 9.10). The vector is no longer capable of lysogeny, but can readily initiate a lytic infection. Techniques have been developed for inserting the vector genome into phage heads, making it easy to infect bacteria with lambda phage vectors containing either the stuffer DNA or a cloned insert.

**Cloning in lambda vectors:** For cloning, the stuffer is removed and replaced with a DNA insert that has been cut with the same restriction endonuclease. The two arms and the insert are simply ligated together and packed into phage heads (figure 9.11). Packaging of the DNA into a phage head only works when the total length of the DNA is between 38 and 51 kb. The size of the DNA insert that can be successfully cloned in a lambda phage vector is dependent on the amount of phage DNA that has been replaced with the stuffer, with an upper limit of about 23 kb. Because successful packaging does not occur when there is not enough DNA, no infectious particles are obtained that lack an insert or contain an insert that is below a lower size limit for the particular vector. Inserts that cause the total amount of DNA to be too large are also not successfully cloned. For a typical lambda phage vector, the insert usually must be at least 12 kb and not over about 20 kb. This size range makes lambda phage vectors particularly useful for genomic libraries (collections of phage containing all of a genome in 12-20 kb fragments). Lambda phage vectors of this sort are sometimes called Charon vectors, a reference to Greek mythology, in which the boatman Charon ferries the souls of the dead across the River Styx.

**Selective markers for lambda phage vectors:** Wild-type bacteriophage lambda form turbid plaques because some of the infected bacteria enter the lysogenic state and are not lysed. However, lambda phage vectors always form clear plaques because the genes for lysogeny have been removed to make room for the temporary "stuffer" sequence. The stuffer contains an IPTG-inducible beta-galactosidase gene, which is eliminated from the vector when the stuffer is replaced with a cloned DNA insert. When induced with IPTG, plaques formed by vectors that still contain their stuffer turn blue with X-gal, whereas those with cloned DNA inserts remain colorless. Thus, cloned inserts will be found only in clear plaques that do not turn blue with X-gal.

**Cosmids:** Linearized lambda-phage genomes contain *cos* sites at their ends (cohesive ends). It is possible to create an artificial plasmid consisting of two *cos* sites plus a plasmid origin of replication and up to 46 kb of foreign DNA. This will replicate as a plasmid and then can be packaged into lambda-phage heads for infection into bacteria. These highly engineered vectors are a sort of cross between a plasmid and a lambda

phage and are capable of carrying 30-46 kb of foreign genes with only a little genetic material of their own (figure 9.13).

**Bacterial artificial chromosomes (BAC):** The F factor plasmid has the ability to continue to function even when integrated into a complete bacterial chromosome. Highly modified F plasmids have been generated that are capable of cloning very large inserts of up to 300,000 base pairs. One interesting feature is the incorporation of cut sites for restriction endonucleases with eight base cut sites. Such endonucleases cut DNA less frequently and thus generate larger fragments for cloning. Bacterial artificial chromosomes are sometimes introduced into their host cells by electroporation, which consists of a brief treatment with high voltage electric current that momentarily disrupts the cell membranes and facilitates entry of large DNA molecules. Once in the cell, the BAC replicate like F plasmids.

**Yeast Artificial Chromosomes (YAC):** Our textbook delays a discussion of yeast artificial chromosomes until after it has presented a general discussion of eukaryotic chromosomes. In brief summary, a yeast artificial chromosome (YAC) contains a yeast origin of replication, a centromere, a telomere at each end, and a large inserted DNA sequence of up to about 500 kb (figure 10.25, page 311). Prior to insertion of the foreign DNA, the essential components of the YAC are maintained in bacterial cells as circular plasmids.

---

**The material that follows on single-stranded cloning vectors and site-directed mutagenesis is not described in our current textbook. You can find a brief discussion of M13 vectors on pages 434-436 of Klug and Cummings, Concepts of Genetics, 5th Edition (Norlin Reserve) and a discussion of site-directed mutagenesis on pages 414-415 of that book. However, it is not necessary for you to know more about these topics than is discussed in these notes.**

**Bacteriophage M13:** Bacteriophage M13 has a single-stranded DNA genome. After M13 infection into a bacterial cell, a complementary DNA strand is synthesized, generating a double-stranded replicative form (RF) of the bacteriophage genome. The complementary strand then serves as a template for synthesis of new single-stranded viral DNA by a rolling circle mechanism (like figure 2.30, except that the second strand is not synthesized until a new cycle of replication is initiated). The single stranded DNA is cut into genome length fragments and extruded from the cell. This single-stranded DNA is useful in sequencing studies on foreign DNA cloned into M13, as will be discussed in Lecture 17. In addition, M13 clones can easily be subjected to site-directed mutagenesis, as described below.

**Cloning into an M13 vector:** M13 vectors have been engineered to contain restriction endonuclease cut sites in the double-stranded replicative form. For cloning, the replicative form can be isolated from infected bacteria or generated artificially from the single-stranded form with a complementary synthetic oligonucleotide primer and DNA

polymerase plus ligase. The double stranded form is cut with a restriction endonuclease and the DNA is inserted much like any other cloning procedure. The double-stranded replicative form with the insert is infected into bacteria and generates single stranded clones by the normal process of single-stranded genomic replication.

**Site-directed mutagenesis:** A synthetic oligonucleotide is prepared that is complementary to the region that is to be mutated except that it contains the desired mutation (usually a one nucleotide change). The oligonucleotide containing the mutation is hybridized to the single-stranded cloned gene and used as a primer to make a complete complementary strand that is identical to complementary strand of the replicative form, except for the mutation introduced by the synthetic primer. The double stranded replicative form with the mutation in its complementary strand is then infected into *E. coli*, where the modified complementary strand serves as a template for production of new single-stranded vectors carrying the mutated cloned insert. If desired, the phage DNA can be made double-stranded and the mutated insert can be cut out of the M13 vector, and put into any other convenient double-stranded DNA vector. The net result of this procedure is to selectively change one base pair in the coding sequence and thus to change a single amino acid in the coded protein (see figure 14.19 of Klug and Cummings). This is an extremely powerful tool for detailed analysis of the roles of individual amino acids in the overall function of a protein.

---

**cDNA cloning:** An alternative cloning procedure called complementary DNA (cDNA) cloning uses mRNA (usually of eukaryotic origin) as a starting point for cloning a coding sequence. The first step is to isolate the mRNA. This is often done by using a column containing immobilized poly (dT), which anneals to the poly (A) tails of most mRNAs. The mRNA can then be eluted by modifying the conditions such that the base pairing no longer holds the mRNA immobilized. The next step is to make a DNA copy of the RNA. The first strand of DNA is templated from the RNA by a viral enzyme known as *reverse transcriptase*. The RNA template is then digested away with *ribonuclease H*, which is specific for digesting the RNA part of an RNA:DNA hybrid. It is also possible to use selective alkaline hydrolysis with NaOH to remove the RNA without damaging the DNA, which is more resistant. The 3'-end of the single-stranded DNA tends to fold back on itself and find enough complementary base pairing to form a hairpin loop, which serves as a primer for second strand synthesis. The *Klenow fragment* of *DNA polymerase I* (which lacks endonuclease activity) is then used to synthesize the second DNA strand and the loop is cut with *S1-nuclease*. The so-called cDNA (complementary DNA) is then ligated into a vector. This can be done either by blunt end ligation or by adding sticky ends artificially. cDNA cloning is often done in an **expression vector** that contains a promoter that allows synthesis of mRNA and expression of the coded protein in the host cell, as will be discussed later in this lecture.

---

**Libraries:** As an alternative to selective cloning of specific DNA sequences, it is possible to construct a library of cloned sequences that is sufficiently complex so that it is statistically expected to include all of the sequences in the DNA that was used as a starting point. This can be the entire genome of an organism or all of the genes carried on one particular chromosome, or all of the mRNA sequences in a particular tissue that have first been converted to cDNA as described above. Libraries are typically constructed in vectors that accept only a limited range of sizes of inserts. To avoid cutting some genes into fragments that are too small to be cloned in such vectors, the digestion process is often stopped before it has been carried to completion. This generates cloned sequences that still contain some cut sites for the enzyme used for the cloning procedure, and usually results in overlapping clones, which are very useful for locating adjacent sequences. Alternatively, it is sometimes possible to use enzymes with less frequent cut sites if the sizes of their digestion fragments match the range that can be cloned in the vector that is being used.

**Hybridization probes** Complementary strands of DNA, RNA, or DNA plus RNA hybridize readily to form double stranded helical structures when placed under suitable annealing conditions. This property is used extensively in molecular genetics to identify specific nucleic acid sequences. A probe consisting of radioactively labeled DNA (or RNA) is hybridized to denatured DNA (or naturally single-stranded RNA) immobilized on a support, such as a nitrocellulose membrane. Hybridization is normally done at a temperature about 25°C below the melting (denaturation) temperature for the DNA. Probe sequences that do not hybridize because there is no immobilized complementary strand for them are washed off. The sites that contain sequences capable of hybridization are now radioactive and can be detected by autoradiography or by direct counting with scanning counters. When combined with electrophoretic separation of DNA (or RNA) fragments by size, hybridization probes play major roles in many different molecular biology procedures, including Northern and Southern blotting (lecture 17), DNA fingerprinting, etc.

**Cloned DNA probes:** Any DNA (or RNA) that can be prepared as a single uniform sequence can be used as a probe. Cloned DNA is particularly useful, because it consists of multiple copies of a single sequence, and is generally carried in a vector that is sufficiently "foreign" so that it will not react with any of the DNA that is being analyzed.

**Screening a library:** It is possible to transfer replicas of bacterial colonies containing cloned vectors to a nitrocellulose membrane, followed by lysis of the cells and fixing the DNA onto the membrane. Hybridization with a radioactive probe followed by washing off of unhybridized probe then reveals which colonies contain the desired cloned DNA. From the position of the radioactivity, which is detected by laying an X-ray film over the membrane and then developing it after an appropriate exposure period, it is possible to go back to the original plate containing the colonies of bacteria and pick the colony (or colonies) that contain plasmids with the desired cloned DNA inserts (figure 9.15). It is also possible to transfer bacteriophage from plaques onto nitrocellulose membranes and use a similar hybridization process to identify those plaques that contain specific cloned DNA sequences.

**Reverse genetics with degenerate oligonucleotide probes:** In certain cases, it is possible to generate a hybridization probe for a coding sequence based on the amino acid sequence of the coded protein. This is dependent on finding a stretch of at least six amino acids whose codon redundancies are relatively low. The procedure is to artificially synthesize a mixture of all of the possible nucleotide sequences that could code for the amino acid sequence in question, and to use that mixture as a hybridization probe (see boxed example 9.4). One of the members of the mixture is expected to contain the exact coding sequence and thus to hybridize stringently with the message sequence. In addition, others with a single base mismatch may hybridize sufficiently so that they can be seen to be associated with the message sequence if the stringency of the hybridization reaction is reduced somewhat by adjusting temperature and/or salt concentrations so that slightly mismatched probes are not washed off of their immobilized target sequences.

---

**Expression vectors:** It is often desirable to be able to obtain expression of the protein coded by a cloned gene or cDNA. Expression vectors are useful for production of the coded protein in various types of host organisms, including commercial production of proteins that are difficult to obtain in adequate quantities from natural sources. They also permit the proteins that are produced to be used to identify the cloned genes that code for those proteins. They can also be used to produce fusion proteins that are useful in the isolation of previously unknown protein products.

Expression vectors of many different types have been designed for use in various types of host cells. Typically they contain either constitutively strong promoters or promoter constructs that are capable of regulated expression. They also usually contain a ribosome-binding site, such as the bacterial Shine-Dalgarno sequence, to insure vigorous translation of the transcripts that are produced from them. In many cases, they also contain an ATG start codon, followed by a few amino acids from a host protein. In such cases, the cloned gene must be in-frame and inserted in the right direction. In cases where one wishes to produce a eukaryotic protein using a bacterial vector, it is necessary to start with a cDNA clone, which already has intron sequences spliced out, since splicing does not occur in prokaryotic cells.

**Inducible expression:** Production of large amounts of a foreign protein can be toxic to a host cell. Many expression vectors have been designed with inducible expression to allow the vector to be grown up to a high level in the cell without expression, followed by a burst of intense expression to generate as much as possible of the product before the host cell ceases to function. The host cell and the vector are often engineered to work together. One example is use of a bacterial strain that produces large amounts of the lac repressor protein and a vector with the cloned gene under the control of the lac promoter/operator system. The repressor inhibits expression while the vector population is being expanded. Expression is activated at the appropriate time by adding a synthetic analog of allolactose called IPTG that is not metabolized by the cells and thus provides a strong stable induction signal. Another control mechanism that can be utilized involves the lambda left promoter, which when combined with a temperature sensitive lambda repressor (cI) gene,

allows large amounts of bacteria containing vectors to be grown up at a permissive temperature, followed by massive expression of the cloned gene when the temperature is raised enough to inactivate the heat-sensitive lambda repressor protein, as described in the textbook for the pPLa2311 expression vector. Promoters that respond to hormones or to heavy metals are often used for inducible expression in eukaryotic cells. We will revisit expression vectors in the lecture on biotechnology near the end of the semester.

**Use of antibodies to identify cloned genes:** Cloned genes (or cDNAs) that code for proteins that can be identified with antibodies are frequently detected through the use of an expression vector derived from bacteriophage lambda. It is called lambda-gt11 (usually written with the Greek letter lambda). This vector produces a fusion protein under control of the inducible promoter for the *E. coli* lactose (*lac*) operon. The vector lyses the bacterial cells, forming plaques on a lawn of *E. coli*. When expression is induced with IPTG at a critical time during plaque formation, the fusion protein is released into the plaques and can be detected by blotting onto nitrocellulose, followed by binding of antibody and then binding of radioactive Protein A, which attaches to the antibody. This allows plaques that express proteins capable of binding to the antibody to be located by autoradiography. The detection procedure is similar to that used in Western blots (Lecture 16).

**Heteroduplex analysis:** Hybridization of two nucleic acids that contain a mixture of complementary and non-complementary sequences will cause the formation of a series of loops that are similar in principle to the loops that are formed when processed mRNA is hybridized to genomic DNA that contains introns (figure 3.22) Such structures can be visualized quite readily with an electron microscope. Thus, it is possible to see a cloned DNA sequence in a vector that has been denatured and hybridized with a vector that lacks the insert.