Constructing DNA Libraries with Phage and Other Cloning Vectors

Most DNA cloning is done with *E. coli* plasmid vectors because of the relative simplicity of the cloning procedure. However, the number of individual clones that can be obtained by plasmid cloning is limited by the relatively low efficiency of *E. coli* transformation and the small number (only a few hundred) of individual transformed colonies that can be grown on a typical culture plate. These limitations make plasmid cloning of all the genomic DNA of higher organisms impractical. For example, ≈1.5×10^5 clones carrying 20-kb DNA fragments are required to represent the total human haploid genome, which contains ≈ 3×10^9 base pairs. Fortunately, cloning vectors derived from various bacteriophages have proved to be a practical means for obtaining the required number of clones to represent large genomes. A collection of clones that includes *all* the DNA sequences of a given species is called a *genomic DNA library,* or simply *genomic library.* Once a genomic library is prepared, it can be screened for clones containing a sequence of interest.

Nearly Complete Genomic Libraries of Higher Organisms Can Be Prepared by λ Cloning

With the availability of λ phage cloning vectors, preparation of genomic libraries for higher organisms, including humans, is feasible. A genomic library is a set of λ (or plasmid) clones that collectively contain every DNA sequence in the genome of a particular organism. Figure 7-12 summarizes the general procedure for constructing a λ genomic library. The λ DNA first is treated with a restriction enzyme to produce fragments

called λ *vector arms,* which have sticky ends and together contain all the genes necessary for lytic growth. This step frees the nonessential region in the middle of the λ genome; this region is separated from the λ arms and discarded. Genomic DNA then is extracted from a cell type that contains all the genetic information of the organism under study. Sperm cells or cells of an early embryo often are used as sources of mammalian DNA. The extracted DNA then is cleaved by a restriction enzyme to produce ≈20-kb fragments with sticky ends complementary to the sticky ends on the λ vector arms being used.

The λ arms and the collection of genomic DNA fragments are mixed in about equal amounts. The complementary sticky ends on the fragments and λ arms hybridize and then are joined covalently by DNA ligase. Each of the resulting recombinant DNA molecules contains a foreign DNA fragment located between the two arms of the λ vector DNA. The ligated recombinant DNAs then are packaged into λ virions in vitro as described above. Only DNA molecules of the correct size can be packaged to produce fully infectious recombinant λ virions.

Finally, the recombinant λ virions are plated on a lawn of *E. coli* cells to generate a large number of recombinant λ plaques. Since each plaque arises from a single recombinant virion, all the progeny λ phages that develop are genetically identical and constitute a clone carrying a particular genomic DNA insert. The different plaques correspond to distinct phage clones, each carrying a different DNA insert, and collectively they constitute a λ genomic library. Multiple λ vectors have been constructed containing different restriction sites, so that restriction fragments generated by a variety of restriction enzymes can be cloned in λ vectors.

In constructing a library of genomic DNA, the DNA commonly is cleaved with a restriction enzyme that recognizes a 4-bp restriction site

(e.g., *Sau*3A, as shown in Figure 7-12). A specific 4-bp sequence will occur on average once every $4^4 = 256$ base pairs. Complete digestion of the human haploid genome, which contains $\approx 3 \times 10^9$ base pairs, would yield somewhat more than 10^7 nonoverlapping different fragments. However, to increase the probability that all regions of the genome are successfully cloned and represented in the λ genomic library, the genomic DNA usually is only partially digested to yield overlapping restriction fragments of ≈20 kb (Figure 7-13). In a large λ library constructed from such overlapping restriction fragments, a specific sequence of genomic DNA (e.g., the gene encoding -globin) often is contained in or extends over several "overlapping" clones. By a technique called chromosome walking, overlapping genomic clones can be ordered.

The size of a genomic library for a given organism depends on the amount of DNA in that organism's haploid genome. If the human genome of about 3×10^9 base pairs is cleaved into 20-kb fragments for insertion into a λ vector, then roughly 1.5×10^5 different recombinant λ phage virions would be required to constitute a complete library. Because the restriction fragments of human DNA are incorporated into phages randomly, about 10^6 recombinant phages are necessary to assure that each region of human DNA has a 90 –95 percent chance of being included.

Each plaque produced by a recombinant bacteriophage λ contains large numbers of recombinant virions and, consequently, large numbers of identical cloned DNA fragments. Hybridization methods used for identifying recombinant λ clones of interest. Because these methods allow specific detection of very small plaques, as many as 5×10^4 plaques can be screened on a single culture plate. Thus only 20 –30 petri dishes, each containing about 5×10^4 λ plaques, are sufficient to represent the entire human genome. In contrast, to screen 10^6 recombinant plasmids carrying the entire human genome would require about 5000 petri dishes,

because only 200 or so transformed *E. coli* colonies can be detected on a typical petri dish..

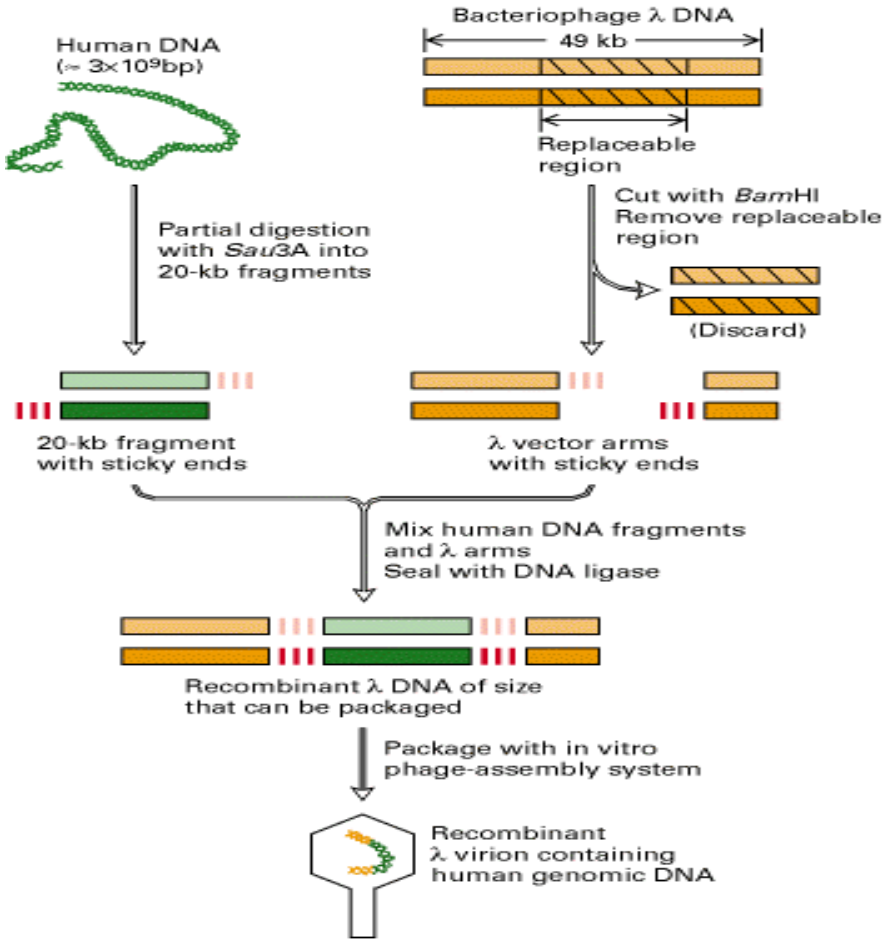

Figure 7-12. Construction of a genomic library of human DNA in a bacteriophage l vector. The nonessential regions in the right half of the l genome (dotted areas in Figure 7-10b) usually are deleted to maximize the size of the exogenous DNA fragment that can be inserted. Then the l DNA is treated to remove the central replaceable region. In this example, the replaceable region is cut out with *Bam*HI, and the total DNA from human cells is partially digested with *Sau*3A. These two restriction enzymes produce fragments with complementary sticky ends (red lines). The l vector arms and ≈20-kb genomic fragments are mixed, ligated, and packaged in vitro to produce recombinant l phage virions, which are plated on a lawn of *E. coli* cells. In the diagrams of DNA regions, light and dark shades of the same color indicate complementary strands.

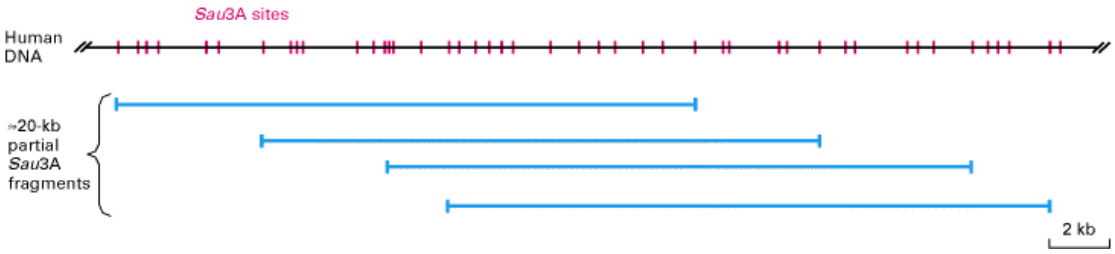

Figure 7-13. Production of overlapping restriction fragments by partial digestion of human genomic DNA with *Sau*3A. This restriction endonuclease recognizes the 4-bp sequence GATC and produces fragments with single-stranded sticky ends with this sequence on the 5′ end of each strand. A hypothetical region of human genomic DNA showing the *Sau*3A recognition sites (red) is shown at the top. Partial digestion of this region of DNA would yield a variety of overlapping fragments (blue) ≈20 kb long. Use of such overlapping fragments increases the probability that all sequences in the genomic DNA will be represented in a λ library.

cDNA Libraries Are Prepared from Isolated mRNAs

In higher eukaryotes, many genes are transcribed into mRNA only in specialized cell types. For example, mRNAs encoding globin proteins are found only in erythrocyte precursor cells, called *reticulocytes.* Likewise, the mRNA encoding albumin, the major protein in serum, is produced only in liver cells where albumin is synthesized. The specific DNA sequences expressed as mRNAs in a particular cell type can be cloned by synthesizing DNA copies of the mRNAs isolated from that type of cell, and then cloning the DNA copies in plasmid or bacteriophage λ vectors.

DNA copies of mRNAs are called **complementary DNAs(cDNAs);** clones of such DNA copies of mRNAs are called *cDNA clones.* In addition to representing only the sequences expressed as mRNAs in a particular cell type, cDNA clones lack the noncoding introns present in genomic DNA clones. Thus the amino acid sequence of a

protein can be determined directly from the nucleotide sequence of its corresponding cDNA. Many genes in higher eukaryotes are too large to be included in a single λ clone because of their large introns. In contrast, all full-length cDNAs, containing the entire protein-coding sequence, can be included in a single λ clone. However, because of methodological difficulties, not all cDNA clones are full length when initially produced; to obtain a full-length cDNA, it often is necessary to isolate several overlapping cDNA clones and then ligate them at rare restriction sites. Just as a large collection of clones containing fragments of genomic DNA representing the entire genome of a species is called a *genomic library,* a large collection of cDNA copies of all the mRNAs in a cell type is called a *cDNA library.*

Isolation of mRNAs and Synthesis of cDNAs

The first step in preparing a cDNA library is to isolate the total mRNA from the cell type or tissue of interest. Nature has greatly simplified the isolation of eukaryotic mRNAs: the 3′ end of nearly all eukaryotic mRNAs consists of a string of 50 –250 adenylate residues, the *poly(A) tail.* Because of their poly(A) tail, mRNAs can be easily separated from the much more prevalent rRNAs and tRNAs present in a cell extract by use of a column to which short strings of thymidylate (oligo-dTs) are linked to the matrix (Figure 7-14). When a cell extract is passed through an oligo-dT column, the mRNA poly(A) tails base-pair with the oligo-dTs, binding the mRNAs to the column. Since rRNAs, tRNAs, and other molecules do not bind to the column, they can be washed away. The bound mRNAs are recovered by elution with a low-salt buffer.

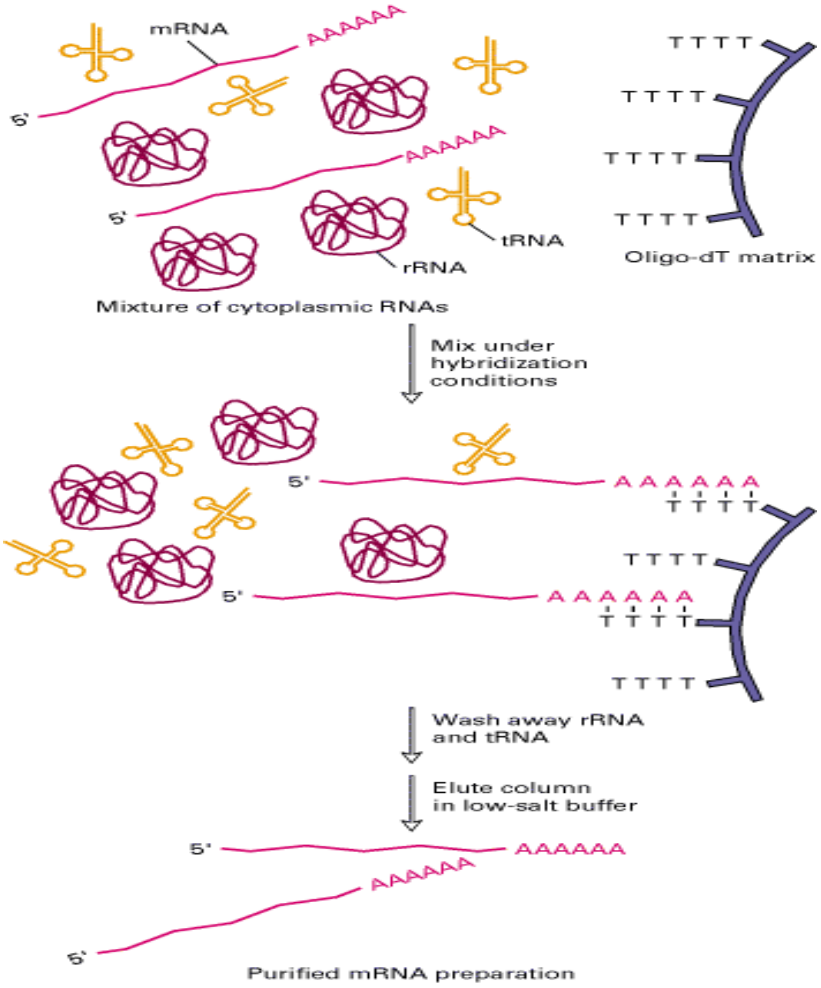

Figure 7-14. Isolation of eukaryotic mRNA by oligo-dT column affinity chromatography. Isolated cytoplasmic RNA consists mostly of ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs). The much less abundant mRNAs (red) have 3′ poly(A) tails, which hybridize to oligo-dT covalently coupled to the column matrix. After hybridization, the rRNAs and tRNAs are washed out of the column; then the mRNAs are eluted with a low-salt buffer. The resulting purified mRNA preparation contains many different mRNA molecules encoding different proteins.

The enzyme reverse transcriptase, which is found in retroviruses, is then used to synthesize a strand of DNA complementary to each mRNA molecule. This enzyme can polymerize deoxynucleoside triphosphates into a complementary DNA strand using an RNA molecule as template. Like other DNA polymerases, reverse transcriptase can add nucleo-tides only to the 3′ end of a preexisting primer base-paired to the template. Added free oligo-dT serves this function by hybridizing to the 3′ poly(A) tail of each mRNA template.

Larger DNA Fragments Can Be Cloned in Cosmids and Other Vectors

Both λ phage vectors and the more commonly used *E. coli* plasmid vectors are useful for cloning DNA fragments up to ≈20 –25 kb. However, cloning of much larger fragments is desirable for sequencing of extremely long DNAs such as the DNA in a eukaryotic chromosome. Also, because of the common occurrence of large introns in genes from higher eukaryotes, it is often necessary to clone DNA fragments greater than 25 kb in order to include an entire gene in one clone. Consequently, additional types of cloning vectors have been developed for cloning larger fragments of DNA (Table 7-2).

Table 7-2. Approximate Maximum Length of DNA That Can Be Cloned in Vectors

Vector Type	Cloned DNA (kb)
Plasmid	20
λ phage	25
Cosmid	45
P1 phage	100
BAC (bacterial artificial chromosome)	300
YAC (yeast artificial chromosome)	1000

One common method for cloning larger fragments makes use of elements of both plasmid and λ phage cloning. In this method, called *cosmid cloning,* recombinant plasmids containing inserted fragments up to 45 kb long can be efficiently introduced into *E. coli* cells. A cosmid vector is produced by inserting the COS sequence from λphage DNA into

a small *E. coli* plasmid vector about 5 kb long. Like other plasmid vectors discussed earlier, cosmid vectors contain a replication origin (ORI), an antibiotic-resistance gene (e.g., amp^r), and a polylinker sequence containing numerous restriction-enzyme recognition sites (Figure 7-16). Next, the cosmid vector is cut with a restriction enzyme and then ligated to 35- to 45-kb restriction fragments of foreign DNA with complementary sticky ends. If the concentration of foreign DNA is high enough, the ligation reaction generates long DNA molecules containing multiple restriction fragments of the foreign DNA separated by the 5-kb cosmid DNA. These ligated DNA molecules, which resemble the concatomers that form during replication of λ phage in a host cell, can be packaged in vitro as described earlier.

In the packaging reaction, the λ Nu1 and A proteins bind to COS sites in the ligated DNA and direct insertion of the DNA between two adjacent COS sites into empty phage heads. Packaging will occur so long as the distance between adjacent COS sites does not exceed about 50 kb (the approximate size of the λ genome). Phage tails then are attached to the filled heads, producing viral particles that contain a recombinant cosmid DNA molecule rather than the λ genome. When these virions are plated on a lawn of *E. coli* cells, they bind to phage receptors on the cell surface and inject the packaged DNA into the cells.

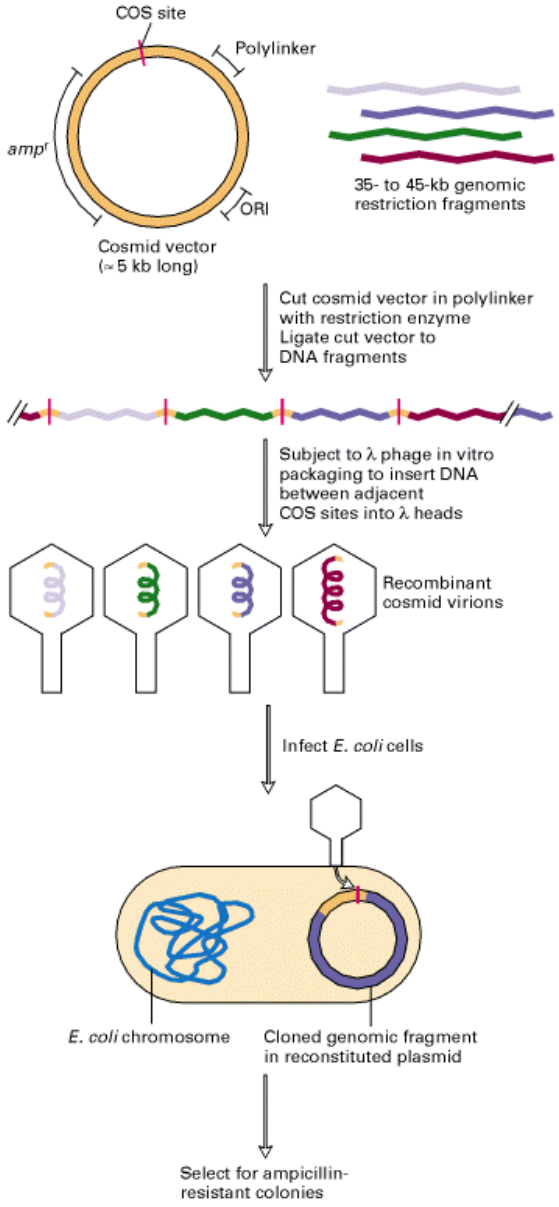

Figure 7-16. General procedure for cloning DNA fragments in cosmid vectors. This procedure has the high efficiency associated with λ phage cloning and permits cloning of restriction fragments up to ≈45 kb long. In this example, four different types of recombinant cosmid virions could be generated, each carrying one of the genomic fragments indicated by different colors. Plating of the recombinant virions on *E. coli* cells would yield four different types of colonies, but only one is depicted. Note that the lengths of vector DNA and genomic fragments are not to scale.
REFERENCE:

Molecular Cell Biology
Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E.
New York: W. H. Freeman & Co. ; c1999