MOLECULAR MEDICINE

STALKING THE GENE — DNA LIBRARIES

NADIA ROSENTHAL, PH.D.

FOR molecular biologists in search of clues to cell function, the ultimate object of inquiry is the gene. A scientist looking for a particular gene faces a formidable challenge, because each cell contains more than 100,000 different genes scattered over billions of nucleotides. To identify a gene, it is necessary to make some order out of the original material and then to devise a way of distinguishing the desired gene from all the others. Recombinant DNA libraries fulfill these requirements.

In theory, the use of restriction enzymes to cut DNA into predictable lengths, which was reviewed in the preceding article in this series, may seem a quick and practical way to isolate short fragments of DNA from chromosomes. This approach has severe limitations, however. We can see the problem if we cut DNA with a restriction enzyme and then try to separate the fragments by electrophoresis. So many pieces are of similar or identical length — even after digestion with enzymes that have very rare cleavage-recognition sites — that it is impossible to isolate one fragment from another. If we insert the DNA restriction fragments into cloning vectors and then introduce the vectors into microbial cells, however, we have a way of separately propagating and storing fragments of identical size. A collection of such cloned molecules in their microbial hosts is a recombinant DNA library (Fig. 1).

To be useful, a DNA library must be as complete as possible. The clones in the library must be sufficiently numerous to include all the sequences in an individual genome. A library of 300,000 clones in which each clone contained an average of 20,000 base pairs of DNA would contain all the sequences in the human

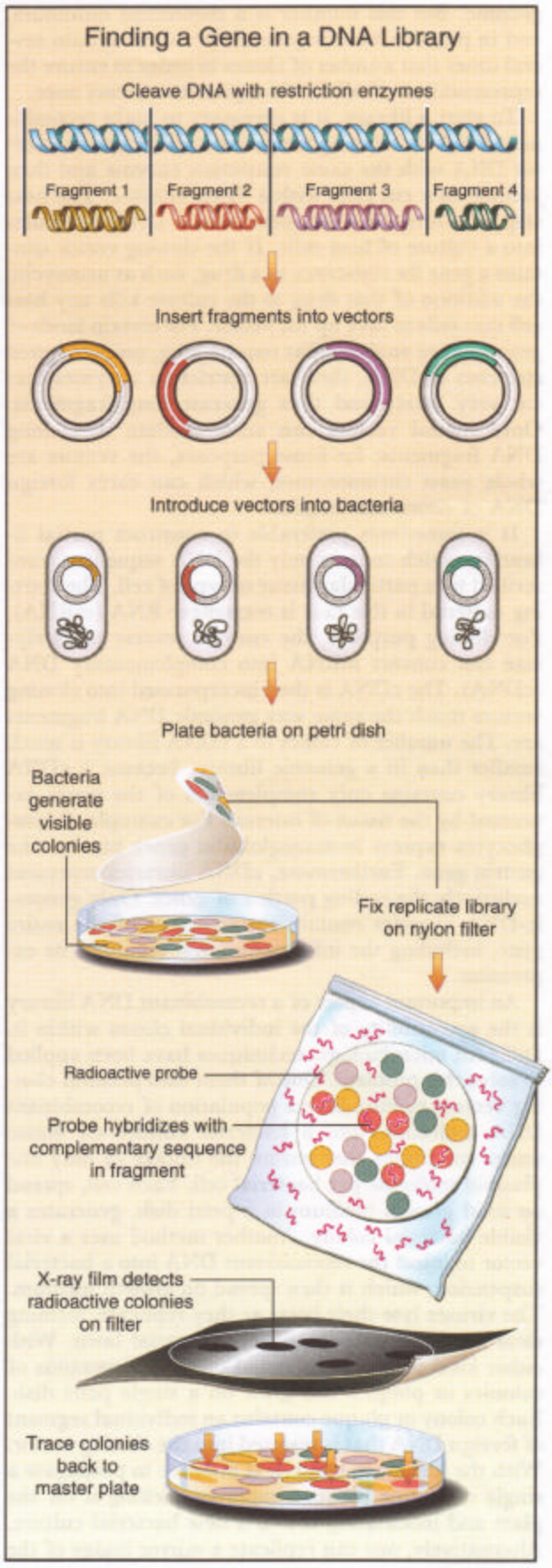

Figure 1. Screening a DNA Library.

The first step in making a library of DNA sequences is to cut DNA into fragments with restriction enzymes. These DNA fragments, when inserted into vectors, form recombinant molecules with the DNA of the vector (this diagram shows a plasmid vector, but viral vectors are also used). Bacteria carrying the vectors can replicate on an agar-coated petri dish, where they grow to form colonies. Each colony originates from a single bacterial cell and thus contains a single type of recombinant DNA fragment. A nylon filter placed on the surface of the petri dish picks up a portion of each colony. Chemical treatment of the filter lyses the bacterial cells, denaturing the DNA and fixing it in place. A radioactive probe for a known sequence of nucleotides can reveal the desired fragment on the filter. The filter, with the replicate library of the colonies on its surface, is incubated with a solution containing the radioactive DNA probe in a plastic bag (or glass dish), and after the unbound probe is washed away, an x-ray film can locate the radioactive colonies (black ovals). The position of the signals on the film serves as a map with which to locate the corresponding colonies on the original master plate. Once identified, these colonies can then be amplified in culture to produce large quantities of the desired recombinant DNA molecule.

genome. But this number is a theoretical minimum, and in practice human genomic libraries contain several times that number of clones in order to ensure the representation of all DNA sequences at least once.

To start a library, it is necessary to make recombinant DNA by cleaving the genomic DNA and the vector DNA with the same restriction enzyme and then joining their cut ends with a ligase enzyme. The next step is to introduce the recombinant DNA molecules into a culture of host cells. If the cloning vector contains a gene for resistance to a drug, such as neomycin, the addition of that drug to the culture kills any host cell that fails to take up the vector. For certain kinds of gene-linkage analysis that require long, uninterrupted stretches of DNA, there are restriction enzymes that cut very rarely and thus generate long fragments. Only special vectors can accommodate these long DNA fragments; for some purposes, the vectors are whole yeast chromosomes, which can carry foreign DNA of enormous lengths.

It is sometimes preferable to construct partial libraries, which contain only the DNA sequences transcribed by a particular tissue or type of cell. The starting material in this case is messenger RNA (mRNA). For cloning purposes, the enzyme reverse transcriptase can convert mRNA into complementary DNA (cDNA). The cDNA is then incorporated into cloning vectors much the same way genomic DNA fragments are. The number of clones in a cDNA library is much smaller than in a genomic library, because a cDNA library contains only complements of the genes expressed by the tissue of interest. For example, B lymphocytes express immunoglobulin genes but not the gastrin gene. Furthermore, cDNA libraries represent exclusively the coding portion of genes. Only genomic-DNA libraries contain the structure of the entire gene, including the information that regulates its expression.

An important aspect of a recombinant DNA library is the accessibility of the individual clones within it. Different microbiologic techniques have been applied to solve this problem. One of them uses plasmid cloning vectors to introduce a population of recombinant DNA molecules into a bacterial culture en masse under conditions that ensure the uptake of only one plasmid molecule per bacterial cell. Each cell, spread on solid growth medium in a petri dish, generates a visible bacterial colony. Another method uses a viral vector to inject the recombinant DNA into a bacterial suspension, which is then spread on growth medium. The viruses lyse their hosts as they replicate, forming clear spots, or plaques, on the bacterial lawn. With either kind of library, plasmid or viral, thousands of colonies or plaques can grow on a single petri dish. Each colony or plaque contains an individual segment of foreign DNA that is inserted into the cloning vector. With the library in hand, it is possible to propagate a single colony or plaque simply by picking it off the plate and inoculating it into a new bacterial culture. Alternatively, one can replicate a mirror image of the whole collection by placing a nylon filter on the surface of the growth medium and allowing the bacteria or viruses to adhere. Denaturation immobilizes the transferred recombinant DNA molecules on the filter (Fig. 1).

Once the replicate library has been fixed on a nylon filter, the desired clone must be identified. Many screening procedures rely on the complementarity of DNA base pairing. Even if only a small portion of the sought-after DNA sequence is available, it is possible to synthesize a single-stranded DNA or RNA molecule with a complementary sequence. This short molecule, known as a probe, carries a tag in the form of a radioactive element or a group that can be recognized by an antibody. The nylon filter is immersed in a solution containing the probe. The probe diffuses to its target clone on the filter and hybridizes to it, marking its location (Fig. 1). The probe's unique sequence of nucleotides ensures that it hybridizes only to a nucleic acid molecule with the complementary sequence. The probing process destroys the bacterial or viral hosts on the nylon filter, but the position of the marked clone can be visualized by autoradiography or revealed by an antibody and traced back to the master plate, which harbors live cells.

How can one probe for the desired clone if nothing is known of its nucleotide sequence? There are numerous strategies to solve this problem, each tailored to the specific situation. Most protein-encoding DNA sequences occur only once in the genome of an organism. Therefore, it is possible to deduce a unique nucleotide sequence if a small stretch of the protein sequence is known. Alternatively, an antibody to the protein, or some functional test of the protein, can identify clones from cDNA "expression" libraries. In these libraries, special vectors allow the bacterial host cells to express the foreign genes as proteins, through transcription and translation. An advantage of expression vectors is that once the gene is isolated, it is easy to generate large quantities of the protein it encodes by growing the host cells in mass culture. In some cases, the only way to find a gene is to "walk" toward it, probing with a DNA sequence that lies nearby on the chromosome and painstakingly characterizing all the intervening DNA sequences until one hits the gene of interest.

Constructing DNA libraries has become a routine part of molecular biology, and distributing them to other researchers has greatly facilitated progress. Often, the process of identifying the library member we seek is far from trivial. New ways to scan libraries are constantly being devised, since it is only by analyzing the structure of a particular DNA segment that we can understand its normal function or detect the molecular basis for its malfunction.

Recommended Reading

Rosenthal N. Tools of the trade — recombinant DNA. N Engl J Med 1994;331:315-7.

Watson JD, Gilman M, Witkowski J, Zoller M. Recombinant DNA. 2nd ed. New York: Scientific American Books, 1992.