Identifying, and Analyzing Cloned DNA

Suppose you have isolated a particular protein and want to isolate the gene that encodes it. A complete genomic λ library from mammals contains at least a million different clones; a cDNA library must contain as many clones to include the sequences of scarce mRNAs. How are specific clones of interest identified in such large collections? The most common method involves screening a library by **hybridization** with radioactively labeled DNA or RNA **probes.** In an alternative method, a specific clone in a library of cloned DNA is identified based on some property of its encoded protein.

The most complete characterization of a cloned DNA requires determination of its nucleotide sequence; from this sequence, the amino acid sequence of an encoded protein can be deduced. The sequence of genomic DNA includes introns as well as exons; it also includes regions that control gene expression by determining the type of cell in which the encoded protein is expressed, the stage in development it is expressed, and the amount of protein produced.

Libraries Can Be Screened with Membrane-Hybridization Assay

Under the conditions of temperature and ion concentration found in cells, DNA is maintained as a duplex (double-stranded) structure by the hydrogen bonds between A · T and G · C base pairs. DNA duplexes can be denatured (melted) into single strands by heating them in a dilute salt solution (e.g., 0.01 M NaCl), or by raising the pH above 11. If the temperature is lowered and the ion concentration in the solution is raised, or if the pH is lowered to neutrality, the A · T and G · C base pairs re-form between complementary single strands. This process goes by many names: renaturation, reassociation, hybridization, annealing. In a mixture of nucleic acids, only complementary single strands (or strands containing complementary regions) will reassociate; the extent of their reassociation is virtually unaffected by the presence of noncomplementary strands. Such *molecular hybridization* can take place between two complementary strands of either DNA or RNA, or between an RNA strand and a DNA strand.

To detect specific DNA clones by molecular hybridization, cloned recombinant DNA molecules are denatured and the single strands attached to a solid support, commonly a nitrocellulose filter or treated nylon membrane. When a solution containing single-stranded nucleic acids is dried on such a membrane, the single strands become irreversibly bound to the solid support in a manner that leaves most of the bases available for hybridizing to a complementary strand.

The procedure for screening a λ library with this membrane-hybridization technique is outlined in Figure 7-18. The recombinant λ virions present in plaques on a lawn of *E. coli* are transferred to a nylon membrane by placing the membrane on the surface of the petri dish. Many of the viral particles in each plaque adsorb to the surface of the membrane, but many virions remain in the plaques on the surface of the nutrient agar in the petri dish. In this way a ***replica*** of the petri dish containing a large number of individual λ clones is reproduced on the surface of the membrane. The original petri dish is refrigerated to store the collection of λ clones. The membrane is then incubated in an alkaline solution, which disrupts the virions, releasing and denaturing the encapsulated DNA. The membrane is then dried, fixing the recombinant l DNA to the membrane's surface. Next, the membrane is incubated with a radiolabeled probe under hybridization conditions. Unhybridized probe is washed away, and the filter is subjected to autoradiography.

The appearance of a spot on the autoradiogram indicates the presence of a recombinant λ clone containing DNA complementary to the probe. The position of the spot on the autoradiogram corresponds to the position on the original petri dish where that particular clone formed a plaque. Since the original petri dish still contains many infectious virions in each plaque, viral particles from the identified clone can be recovered for replating by aligning the autoradiogram and the petri dish and removing viral particles from the clone corresponding to the spot. A similar technique can be applied for screening a plasmid library in *E. coli* cells.

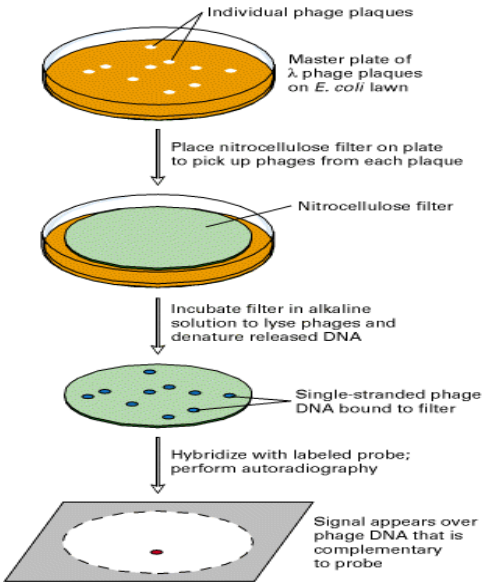

Figure 7-18. Identification of a specific clone from a λ phage library by membrane hybridization to a radiolabeled probe.

Oligonucleotide Probes Are Designed Based on Partial Protein Sequences

Identification of specific clones by the membrane-hybridization technique depends on the ability to prepare specific, radiolabeled probes. Specific oligonucleotide probes for the gene encoding a protein of interest can be synthesized chemically if a portion of the amino acid sequence of the protein is determined. For an oligonucleotide to be useful as a probe, it must be long enough for its sequence to occur uniquely in the clone of interest and not in any other clones. For most purposes, this condition is satisfied by oligonucleotides containing about 20 nucleotides. This is because a specific 20-nucleotide sequence occurs once in every 420 (≈1012) nucleotides. Since all genomes are much smaller (≈3 × 109 nucleotides for humans), a specific 20-nucleotide sequence in a genome usually occurs only one time. In principle, probes this length can be prepared based on only a 7-aa sequence out of a protein's total sequence (20 nucleotides ÷ 3 nucleotides per codon ≈ 7 amino acids). However, a somewhat longer amino acid sequence usually is determined to allow the preparation of several probes for use in cloning a gene of interest.
Here we describe two different approaches for preparing oligonucleotide probes. Generally, a radiolabeled oligonucleotide probe is used to screen a λ cDNA library using the membrane-hybridization technique. Once a cDNA clone encoding a particular protein is obtained, the full-length radiolabeled cDNA can be used to probe a genomic library for clones that contain fragments of the gene encoding the protein.

Degenerate Probes

Based on the genetic code, the oligonucleotide sequences encoding the determined peptide sequences can be predicted. Recall, however, that the genetic code is degenerate; that is, many amino acids are encoded by multiple codons. Since the specific codons used to encode the protein of interest are unknown, oligonucleotides containing all possible combinations of codons must be synthesized to assure that one of them will match the gene perfectly.

To be certain of obtaining a probe based on this amino acid sequence that hybridizes perfectly to the unique sequence present in the gene, all 48 of the 20-mer probes must be synthesized. A mixture of 20-mer probes based on any other portion of this peptide sequence would have to contain considerably more than 48 oligonucleotides because of the presence of leucine or serine residues, each encoded by six different codons.

A mixture of all the oligonucleotides that can encode a selected portion of a peptide sequence is called a *degenerate probe.* Screening of a cDNA library with a degenerate probe using the membrane-hybridization technique will identify clones that hybridize to the perfectly complementary oligonucleotide present in the probe mixture.

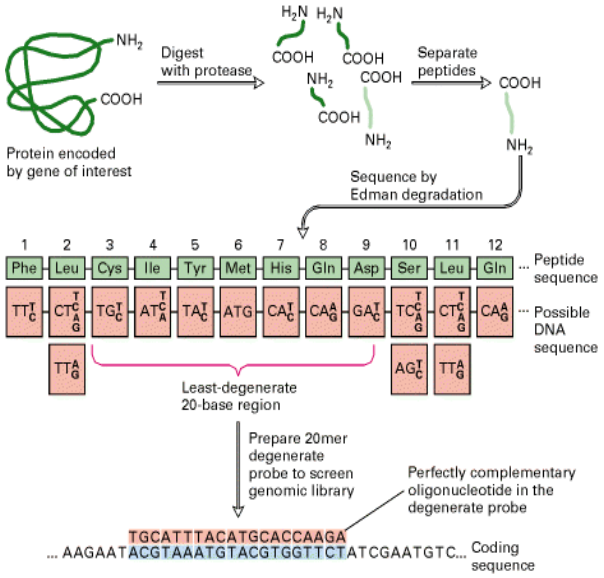

Figure 7-19. Designing oligonucleotide probes based on protein sequence. The determined sequences then are analyzed to identify the 6- or 7-aa region that can be encoded by the smallest number of possible DNA sequences. Because of the degeneracy of the genetic code, the 12-aa sequence (light green) shown here theoretically could be encoded by any of the DNA triplets below it, with the possible alternative bases at the same position indicated. For example, Phe-1 is encoded by TTT or TTC; Leu-2 is encoded by one of six possible triplets (CTT, CTC, CTA, CTG, TTA, or TTG). The region with the least degeneracy for a sequence of 20 bases (20-mer) is indicated by the red bracket. There are 48 possible DNA sequences in this 20-base region that could encode the peptide sequence 3 –9.

Antibody screening

An alternative approach to isolating a gene on the basis of its protein product is the use of antibodies as probes to screen expression libraries. In this case a cDNA library is generated in an expression vector that drives protein synthesis in *E. coli*. Phage plaques or bacterial colonies are then transferred to a filter, but the filter is then reacted with an antibody to identify clones that are producing the protein of interest.