The DNA in our cell nuclei contains all the information necessary to direct our biological functions. The gene is 1 unit of such information that usually specifies the expression of 1 protein product. A detailed knowledge of each gene will enable us to understand disease at the molecular level. To study each gene we must purify and study it in isolation. This process is usually described as gene cloning. However, this is a difficult task since most of the human genome is junk DNA that does not encode genes. Messenger RNA (mRNA), on the other hand, is representative of only that DNA that has been transcribed. More to the point, within each tissue or cell type, only mRNA that is important for the function of the selected cells will be present. Indeed, different cell types, developmental stages, and many disease states arise because of differential gene expression. The problem is that mRNA itself cannot be easily manipulated to clone genes. On the other hand, DNA can be digested with restriction enzymes and propagated in a bacterial vector. Therefore, to isolate specific genes, DNA (which can be easily manipulated) is made from the mRNA by a process called reverse transcription. The resulting DNA is called complementary DNA (cDNA). When the mRNA from any specific cell type is used to generate cDNA, that collection of cDNA corresponds to the genes that are expressed in those cells at that time. This is especially valuable for the cloning of genes that are expressed specifically in a particular cell type or expressed differentially between normal and disease cells.

**CLONING A GENE USING A cDNA LIBRARY**

The first step in cloning a gene is the production of a cDNA library. A library in this case is simply a collection of different cDNA molecules that can be physically separated and manipulated. Since cDNA is made from mRNA, it is imperative that mRNA of the gene of interest be highly expressed in whichever tissue type is selected to serve as the starting material for the library. When the mRNA has been isolated and prepared, the first strand of the cDNA is produced. In generating a cDNA library, reverse transcriptase, an RNA-dependent DNA polymerase, is used to catalyze the reaction (Figure 1). First-strand cDNA synthesis is followed by second-strand synthesis. To do this, the mRNA template is removed and a second, complementary strand of DNA is synthesized using a DNA-dependent DNA polymerase.
Once double-stranded cDNA has been synthesized, the collection of molecules must be inserted into vectors so that they can be further manipulated. Vectors, most often bacteriophages, are merely self-replicating entities. They contain the information necessary to reproduce themselves, as well as any foreign stretch of DNA (e.g., cDNA) that has been inserted into their own genome. A single bacteriophage particle can be detected by the production of a plaque (the result of cell lysis) on a bacterial lawn (Figure 1). Therefore, if a phage suspension is greatly diluted and spread over a bacterial lawn, individual phage clones can be detected and isolated. Each phage particle contains only 1 cDNA molecule; thus, by isolating a single phage, a pure cDNA clone is isolated. The vector most commonly selected for use in cDNA libraries is bacteriophage λ, usually phage λ ZAP or phage λ gt11. The phage λ DNA is cut so that it can accept a cDNA insert. Once the phage λ DNA has incorporated a cDNA insert and has been religated, it can be assembled easily into a phage particle that is ready to infect a bacterium.

When the collection of phage has been made, it can be screened for the desired clone. First, the library is plated onto a bacterial lawn of Escherichia coli in a Petri dish. Each plaque is the result of the infection of bacteria with a single phage. Since each phage particle contains only 1 cDNA molecule, each plaque represents a single cDNA clone.

Figure 2. Three methods of screening a complementary DNA (cDNA) library. After the bacteria that have been infected with phages form plaques, a nitrocellulose filter is placed on top of the bacterial lawn. The cell lysate in the plaque binds the nitrocellulose with high affinity, creating a precise spatial replica of the plaques in the Petri dish. The cDNA library can now be screened by a number of methods. (1) Screening by DNA-DNA annealing. Synthetic-labeled DNA oligonucleotides are incubated with the nitrocellulose filter. Only those oligonucleotides that are specifically bound to the cDNA will remain and give a positive signal on an autoradiograph after the nitrocellulose is washed. (2) Screening for an antigen with serum. The filter is incubated with serum and the irrelevant antibodies are washed off. The bound antibody is detected using iodine 125 protein A. (3) Screening by interaction cloning. The filter is incubated with a radiolabeled protein. The unbound protein is washed off and the positive plaques identified using autoradiography.
mRNA. However, because the triplet genetic code is degenerate, great care must be taken in designing the DNA. Although any 3 bases correspond to only 1 amino acid, 18 of the 20 amino acids can be encoded by more than 1 set of 3 bases. Thus, the DNA can be designed either to be short but with a high degree of homology to the cDNA in question, or it can be somewhat longer with a lesser degree of homology. Both strategies have their advantages and drawbacks, and in the end it is generally smart to make a mixture of different oligonucleotides to try under varying conditions.

At this point, the nitrocellulose membranes with the cDNA bound to them are incubated with radiolabeled DNA, which should only anneal to cDNA. The membranes are then washed to remove unbound DNA and autoradiographed. The autoradiograph is aligned with the original plate and any positive phase selected. These plaques are replated until all the phages are positive. Finally, the DNA is purified from the positive phage and digested to release the cDNA insert.

This type of approach was taken by the group that cloned the gene for the human prion protein (PrP). Prions are transmissible pathogens now known to be involved in neurodegenerative diseases of animals and humans. In animals, the predominant disease is scrapie, while in man the group includes Creutzfeldt-Jakob disease, Gerstmann-Straussler syndrome, and kuru. These diseases are all transmissible and have common histopathological and clinical features. Primarily, spongiform degeneration of central nervous system neurons is observed, along with intense restrictive astrocytic gliosis and amyloid plaque formation. Prions contain a protein, called PrPSc, that is required for infectivity. Initially, the origin of PrPSc was unknown; whether it was the result of a viral or bacterial infection, or something altogether different, had not been determined.

To investigate the origin of PrPSc, a cDNA library was constructed from scrapie-infected hamster brains. Hamsters were used as experimental hosts because they have the shortest incubation period. First, prions were purified from scrapie-infected hamster brains. These prions were observed to contain 1 major protein constituent, called PrP 27-30, in which an amino acid sequence was determined via gas-phase microsequencing. To ensure that an abundant supply of the mRNA coding for this disease-related protein would be present in the starting material, scrapie-infected hamster brain tissue from the exponential phase of prion formation (35 days after inoculation) was used as a basis for the library. A set of 32 icosameric oligonucleotides was synthesized to screen the library. These oligonucleotides were derived from the reverse translation of a 7 amino acid sequence of PrP 27-30. A library containing 150,000 cDNA clones was screened, yielding 1 positive clone.

The next question to be answered was whether the isolated cDNA was derived from cellular DNA or from an infectious agent. DNA from healthy and scrapie-infected hamsters was digested with restriction enzymes and annealed with a radiolabeled copy of the cDNA insert. No difference was observed between these 2 preparations, suggesting that a single PrP gene exists in the hamster genome. In addition, no difference was observed in hamster PrP mRNA levels harvested at different times after inoculation. The information gathered in these experiments was used to detect and isolate related sequences in human and mouse DNA, and eventually led to the cloning of the human PrP gene.

The most important conclusion was that PrP is encoded by a cellular gene. Thus, PrP is present in healthy cells. Knowing the sequence of PrP has allowed researchers to study its biochemical properties in depth and determine what other factors are involved in these neurodegenerative diseases. The current hypothesis is that these neurodegenerative diseases are caused by a conformational change in the structure of PrP.

Another use of cDNA libraries is using human serum to screen for specific antigens. A slightly different kind of library than the one discussed earlier has to be constructed. In what is termed an expression library, vectors are used that can express the cDNA as fusion proteins. That is, the cDNA sequence is inserted within the coding region of another known protein so that the final product is the known protein attached, or fused, directly adjacent to the peptide encoded by the cDNA. This screening technique is similar to that described earlier. In this case, after removal from the Petri dish the nitrocellulose filter is blocked to prevent nonspecific adsorption of proteins. The blocked filter is then incubated with the human serum and washed to remove nonspecific interactions. Any bound antibodies are visualized by incubation with labeled protein A and autoradiography. Positive clones are selected and replated until all phages are positive.

This technique was used in the study of the neurological disorder paraneoplastic encephalomyelitis sensory neuronopathy. A clinical observation was made that certain patients with small cell lung cancer developed this disorder, which is characterized by dementia, sensory loss, and other neurological disorders. The sera of most of these patients reacted with a group of antigens that are expressed specifically in small cell lung tumors and neurons. The antibody responsible for these reactions is called anti-Hu. Cloning of the gene that encodes these proteins has enabled a more rigorous analysis of the antigens’ role in this disorder.

In this case, the method for screening a library included using the serum of a patient with the disorder, which contains the anti-Hu antibody. For the purposes of this experiment, an expression library was made from cerebellar tissue, in which mRNA for the antigen in question would be in relative abundance. When the library of 1 million clones was screened with the
anti-Hu serum, 8 positive clones were revealed, none of which were positive when screened with normal human serum. Three of the clones encoded 1 gene, which was called Hud. The other 5 encoded another highly similar gene, Huc.

The primary advance that has arisen directly from this work is a diagnostic test for paraneoplastic encephalomyelitis sensory neuropathy. In this test, the Hud protein is mass produced in bacteria and then purified, and a patient’s serum can be checked using Western blot analysis to see if it contains antibodies against Hud, signifying a positive diagnosis. Of particular importance is the finding that most patients with paraneoplastic encephalomyelitis sensory neuropathy also have a tumor, so a neurologist whose patient has a positive anti-Hu result should search the patient for an occult tumor. It is important to note that the recombinant protein used in this assay gives a definitive result that could not be obtained by using cellular extracts. Such extracts could contain a different 36-kd protein that may be recognized by other antibodies, producing a false-positive result.

INTERACTION CLONING: THE HUNTINGTIN-ASSOCIATED PROTEIN

In many hereditary diseases the pathogenic gene has been identified but the function of its protein product is unknown. One way to clarify the function of a disease gene product is to identify other proteins that interact with it. This is called interaction cloning. The genes encoding interacting proteins can be cloned by screening a cDNA library with a radioactively labeled protein.

This method has been used to identify genes that are involved in Huntington disease. This autosomal dominant neurodegenerative disorder usually has its onset in the fourth or fifth decade of life and is marked by motor disturbance, cognitive loss, and psychiatric manifestations. Huntington disease is caused by an expanded polyglutamine repeat in the gene that codes for the protein huntingtin. To study the role of huntingtin in the progression of Huntington disease, a rat brain cDNA expression library was screened for interacting proteins using the polyglutamine repeat domain. A new gene was discovered, called huntingtin-associated protein 1 (HAP1), which has been found to bind to mutated huntingtin with greater affinity than to the wild-type protein. After rat HAP1 cDNA had been isolated, the human gene was cloned from human caudate tissue using polymerase chain reaction. Human HAP1 is most consistently expressed in those parts of the brain that are primarily affected in Huntington disease, namely caudate and cortex. Since huntingtin is expressed ubiquitously throughout the body, the selective expression of HAP1 may explain the neural specificity of Huntington disease. Further screening of cDNA libraries using this protein-protein interaction approach may yield other proteins involved in Huntington disease, leading to a greater understanding of the molecular mechanisms that underlie this disorder.

SUMMARY AND FUTURE DEVELOPMENTS

In summary, screening cDNA libraries is a highly valuable approach for discovering important information about neurological functions and diseases. As illustrated herein, the advantages of knowing a gene’s sequence are profound and numerous. Cloning a gene also provides therapeutic opportunities. A gene that helps prevent or cure a disease can be overexpressed in cells. Alternatively, antisense DNA vectors can be constructed that may block expression of a gene that is responsible for causing a disease.

One recent development in the field is the undertaking of 2 groups to sequence as many cDNA clones as possible in the hope that they will become medically important at some point. Human Genome Sciences and The Institute for Genomic Research, both located in Rockville, Md, have set up facilities where the sole function is to produce cDNA libraries from different starting tissues, human or otherwise, and then sequence these cDNA strands. Robotic machines do much of the work, such as picking clones and sequencing them. Computers are responsible for lining up and organizing overlapping sequences and arranging the genes into likely functional classes. These databases are of great benefit to researchers who can quickly ascertain the complete sequence and likely function of the gene of their interest. In addition, it permits scientists to compare the sequence of the genes they have cloned with similar genes from other species. For more than 20 years, cDNA libraries have played an integral role in biomedical research. In light of the exciting developments in the field, they promise to be of even greater importance in the future.

Accepted for publication February 13, 1998.

The databases of the Human Genome Sciences and The Institute for Genomic Research are available at www.hgsi.com and www.tigr.org.

Reprints: Henry M. Furneaux, PhD, Box 20, Laboratory of Molecular Neuro-Oncology, Memorial Sloan Kettering Cancer Center, 1275 York Ave, New York, NY 10021 (e-mail: h_furneaux@ski.mskcc.org).

REFERENCES