DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY, 11(5), 1073-1088 (1985)

PRINCIPLES OF GENETIC ENGINEERING TECHNOLOGY

John M. Tabor, Ph.D.
Bristol-Myers, Industrial Division
Syracuse, New York

Genetic Engineering Technology has made possible the manipulation of genetic information and has been described as the new revolution in science. It is responsible for the development of new industrial applications and areas of basic research. In the pharmaceutical field, alone, it has applications to the production of natural and synthetic drugs (proteins/enzymes) at levels that can not be achieved by conventional methods. It can be applied diagnostically to detect disease or the potentiality of disease. Not only is it being used to improve industrial processes for the production of current market drugs, but will also generate novel drugs for the future. There is also the potential that this technology will correct defective genes in humans with genetic diseases. In addition, this technology has numerous applications for agricultural, chemical and other industries. Genetic engineering is not a single technique but represents a collection of interrelated techniques, including recombinant DNA technologies. The principles of genetic engineering will be highlighted in this review.

INTRODUCTION

Imagine yourself searching for a single sentence of approximately 20 words in a library of 2,500 books. Let us make it more difficult. Assume that you do not know in which book this sentence is located or even in what part of the library to begin looking. In fact, you do not know what the words are in the sentence; only the general meaning of the sentence is

 0363-9045/85/1105-1073$3.50/0

understood. Does this situation sound like a hopeless task? This is analogous to the problem of locating and isolating a single unique gene from a human cell. Before the advent of recombinant DNA techniques this problem was insoluble. However, today through the development of recombinant DNA technology it is possible to identify, enrich and amplify complete genes from the cell of any organism. Genetic engineering employs integrated techniques in microbiology, molecular biology and includes recombinant DNA technologies. The implication of genetic engineering is to deliberately alter the expression of genetic traits of a cell or an organism.

Through the application of recombinant DNA and genetic engineering technologies the approaches for detecting disease, for developing new drugs for treating disease and the methods for drug manufacturing have been advanced. The purpose of this review is to highlight the principles of these technologies and to describe some of the methods employed.

GENES AND GENE EXPRESSION

A fundamental knowledge of heredity and the expression of genetic traits is necessary in order to appreciate the principles of genetic engineering. DNA (deoxyribonucleic acid) is the molecule that encodes all of the genetic characteristics of a cell and of an organism. DNA is composed of a polymer of nucleotides attached to a ribose sugar-phosphate backbone (Figure I). There are four different nucleotides, deoxyadenylate, deoxyguanylate, deoxycytidylate and deoxythymidylate. DNA is normally in a double polynucleotide strand (double-stranded) configuration. The strands are held together by noncovalent interactions. Each nucleotide on one strand has a complementary nucleotide on the opposite strand. The complement of deoxyadenylate is deoxyguanylate the complement of deoxycytidylate is deoxythymidylate and vice-versa. It is the linear sequence of nucleotides that codes for all genetic traits. The sequence of nucleotides is organized into units of information called genes. Each gene encodes one trait. There are exceptions where one gene encodes several traits, but for the purposes of this review we will assume that this

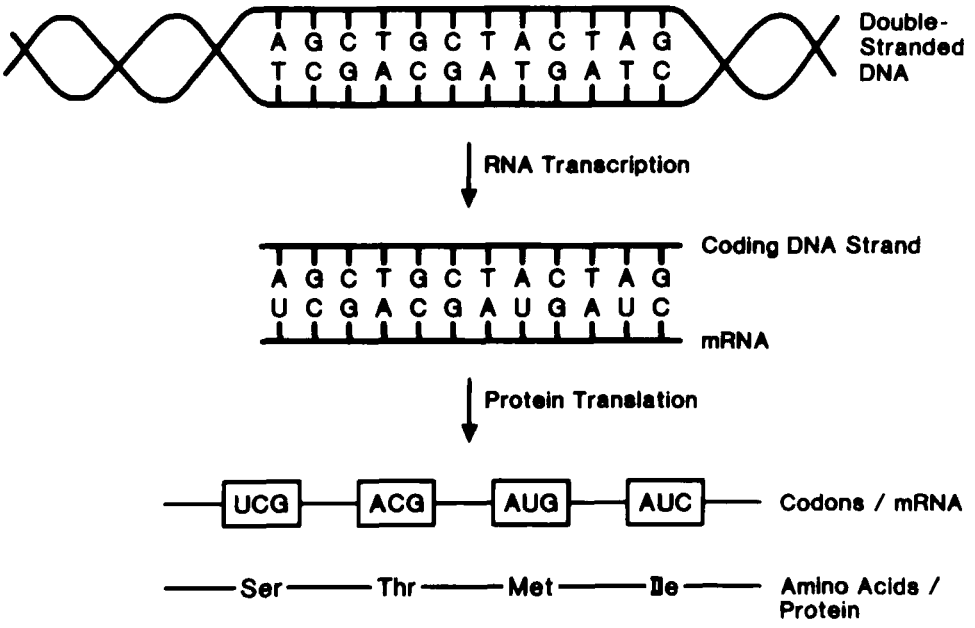

FIGURE I

Figure I depicts gene expression. The coding strand of double-stranded DNA is first transcribed into a mRNA molecule. The mRNA is subsequently translated into protein.

statement is correct. The total number of genes that describes an organism is referred to as the genome.

The genetic information is transferred faithfully from one cell generation to the next by a process of DNA replication. During DNA replication both strands of the DNA are copied by enzymes, DNA polymerases, into identical DNA strands. When a cell divides, each daughter cell receives identical DNA copies. In addition to retaining the genetic code, the DNA must also express this information at the appropriate time. The decoding process occurs in sequential steps, outlined in Figure I. The DNA becomes relaxed producing a localized region where it is single-stranded. The exposed gene is transcribed into a RNA (ribonucleic acid) molecule complementary to one of the DNA strands referred to as the

coding-strand. The enzymes that synthesize the RNA are RNA polymerases. RNA has a similar chemical composition to DNA. Because the RNA has a nucleotide sequence that is complementary to the DNA nucleotide sequence of the gene, it acts as a messenger of information and is called messenger RNA (mRNA). Ribosomes are subcellular structures that bind to the beginning of the mRNA and read the sequence in groups of 3 nucleotides (codon). Each codon is translated into a unique amino acid. Using all 4 nucleotides, in combinations of 3, there are 64 different codons that encode 20 different amino acids. The fact that each amino acid is encoded by more than one codon is indicative of a degenerate code. However, the genetic code is unambiguous in that each codon only encodes one amino acid. The amino acids are covalently linked by amide bonds that form the backbone of proteins. The function of each protein is determined by the sequence of amino acids which is in turn determined by the sequence of nucleotides in the gene. Some proteins are structural and contribute to cytoskeletal and tissue structure. Other proteins are enzymatic and control anabolic or catabolic pathways. Protein represents the end-product of gene expression or the decoding process.

In order to gain insight into the fundamentals of heredity and the regulation of gene expression it is often necessary to have large quantities of a pure gene for analysis. Because of the complexity of procaryotic and eucaryotic genomes, extremely powerful methods must be available in order to identify, enrich and amplify a gene of interest. For example, the genome of the bacteria Escherichia coli encodes approximately 4,000 different genes. In order to perform unambiguous analysis on one of these genes it is essential to remove the other 3,999 unwanted genes. In addition, it is also necessary to amplify the enriched gene to a level that permits detection. The amplification can be on the order of 10^{12} to 10^{14} copies depending upon the particular analyses being performed. The level of difficulty is orders of magnitude greater when dealing with a genome of a higher eucaryote such as man, which has the capacity to encode approximately 6 million genes. The techniques that identify, enrich and amplify specific DNA sequences, are collectively referred to as recombinant DNA

technology. These techniques have been developed, for the most part independently, during the last 30 years and continue to be refined.

GENE CLONING

This section will describe some of the techniques of recombinant DNA technology to illustrate the general scheme of gene cloning.

Having isolated genomic DNA from cell suspensions or tissues using conventional methods[1], the next step is to fragment the DNA to a smaller more manageable size. This is performed by restriction endonucleases (restriction enzymes) that recognize unique nucleotide sequences (restriction sites), usually 4 or 6 nucleotides in length and produce double-stranded breaks in the DNA[2]. Many restriction enzymes produce staggered breaks in the DNA resulting in single-stranded self-complementary ends, referred to as "sticky-ends." The "sticky-ends" can be annealed with "sticky-ends" on any DNA fragment generated by the same restriction enzyme. The DNA fragments generated by restriction endonuclease treatment can be separated by size fractionation using electrophoretic techniques. Figure II shows a restriction enzyme profile of a human α-interferon gene after treatment with different restriction enzymes. The fragmented DNA is electrophoresed in an acrylamide gel and is visualized by staining with ethidium bromide which fluoresces when subjected to ultra-violet light. The smaller DNA fragments localized at the bottom of the profile, migrate faster than larger fragments during electrophoresis, thus, affecting the size fractionation. As illustrated by this figure, different restriction enzymes recognize different restriction sites along the same DNA molecule and produce different numbers and sizes of fragments. The restriction profile is characteristic for each DNA molecule and provides a powerful tool for DNA analysis.

After the genomic DNA has been digested with restriction enzyme the DNA fragments are joined to a vector DNA molecule cut with the same restriction enzyme. The vector is capable of autonomous DNA replication in an appropriate cellular host. DNA fragments that are joined to the vector DNA are passively replicated, or copied, along with the

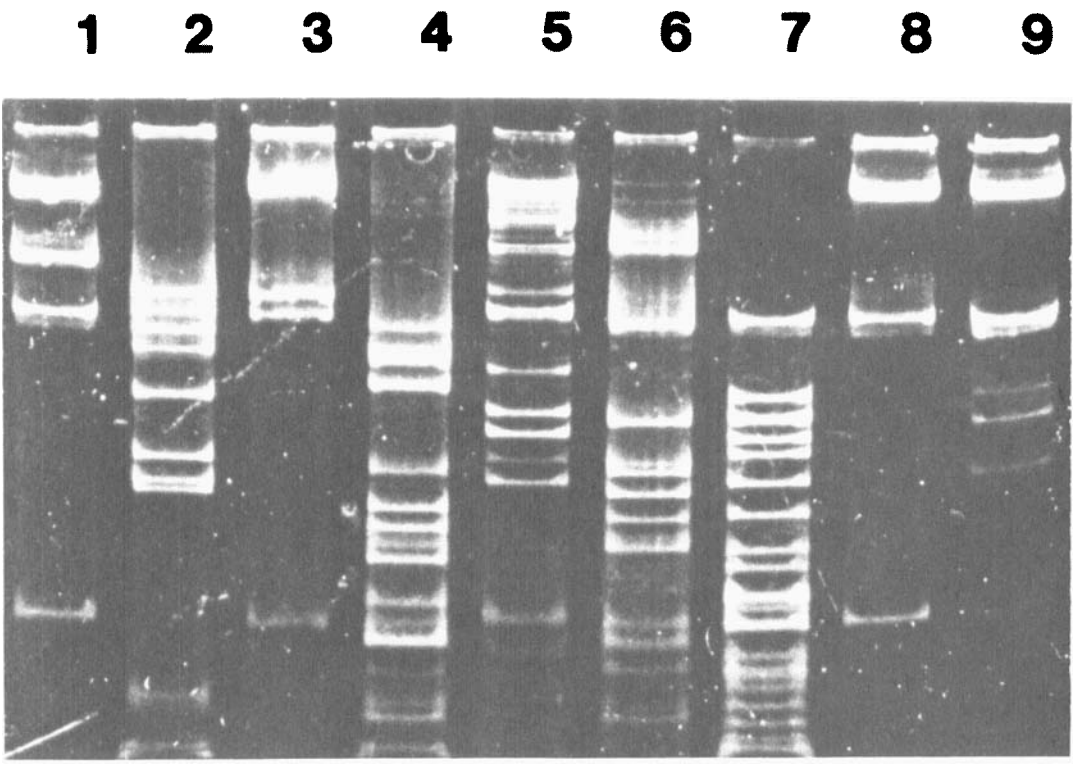

FIGURE II

A restriction endonuclease profile of human α-interferon DNA. Restriction endonucleases used: Bam HI, lane 1; Alu I, lane 2; Hind III, lane 3; Hae III, lane 4; Dde I, lane 5; Sau 96I, lane 6; Cfo I, lane 7; Pst I, lane 8; Pst I and Bgl II, lane 9.

vector. In addition to being self-replicating, vectors usually have several other important features. A unique restriction site in the vector allows for simple reconstruction after the vector is cleaved with the restriction enzyme. Reconstruction of the vector is more complex if the vector has multiple restriction sites recognized by the same restriction enzyme. Selectable genes encoded by the vector, such as drug resistance markers, are particularly valuable for detection and selection of the vector in the host. In bacterial, fungal and plant cell culture systems, the vector is either a plasmid[3,4,5,6] or phage[7] DNA molecule. In animal cell culture systems the vector is usually a modified animal virus DNA.[8] Vectors have been constructed that contain portions of DNA from bacterial, fungal and

animal vectors. These mixed vectors, referred to as shuttle vectors[9], have the ability to replicate in hosts from which the vector DNA is derived. They are useful for transferring genetic information between species and Kingdoms.

The restriction enzyme cleaved genomic and vector DNA are mixed and the "sticky-ends" are annealed (Figure III). The hybrid, or recombinant molecules can then be covalently joined by treatment with DNA ligase.[10] The hybrid vector DNA is introduced into the cellular host by a process of DNA transformation. For bacterial systems the cells are treated with calcium and "heat-shocked." In the case of a plasmid vector it forms a stable replicating associate within the host. The phage vector, on the other hand, usually packages its replicated DNA into protein particles that are released from the bacterium as infectious phage. A similar DNA transformation procedure is employed for animal cells.[12] The vector also forms a stable replicating associate in the animal cell.

Because the efficiency of DNA transformation is low the probability of a host receiving more than one hybrid vector is extremely unlikely. Therefore, each clone from a cellular host transformed with DNA contains only one kind of hybrid vector and, thus, only one genomic fragment. By insertion of a single genomic DNA fragment into a vector and amplification within the host the specific genomic fragment is greatly enriched. From this concept the term gene cloning is derived. However, since all the DNA fragments of a genome are ligated to a vector, the product of DNA transformation is thousands of different recombinant cell clones. The actual number of clones depends on the size of the genome. If, for example, the E. coli genome is treated with a restriction enzyme that cuts the DNA into 1,000 fragments and these fragments are joined to a vector, then there will be 1,000 different clones after transformation of the host with the hybrid vector DNA. Only one of the recombinant clones will contain the genomic fragment of interest. The next obvious step is to devise a screening procedure to detect the clone of interest. The screen is often the most labor-intensive step in the cloning procedure.

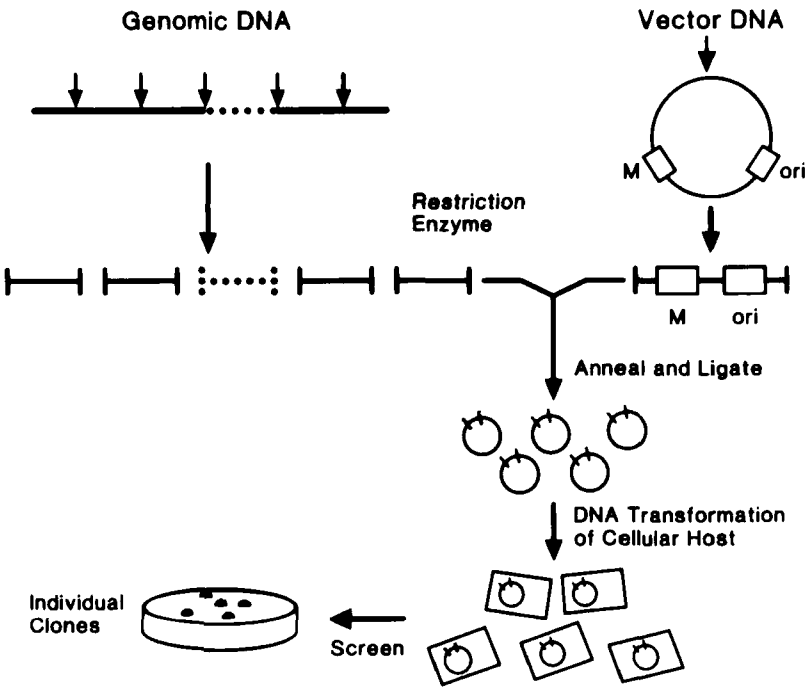

FIGURE III

Figure III is an outline of the manipulations used during cloning and selecting restriction endonuclease fragments from genomic DNA. On the vector, M represents a selectable gene marker and ori represents the origin of DNA replication for the vector. Arrows indicate restriction endonuclease sites.

To develop a successful screen requires a fundamental knowledge of the gene or gene product. Two general screening procedures are employed to identify a specific recombinant clone. One procedure relies upon detection of the protein product of the cloned gene[7] and the other relies upon detecting the nucleotide sequence of the gene of interest.[13] If a screen relies on the detection of the gene product it is reasonable to assume that the cloned gene must be expressed in the host-vector system. However, because of complications related to the expression of heterologous genes in host-vector systems, this assumption can not always be made. Nevertheless this strategy has been successful for identifying many recombinant

clones. For example, a simple screen that would identify the gene that encodes resistance to ampicillin could rely on detection of the gene product. Transformation of an ampicillin sensitive bacterial host with cloned genomic DNA from ampicillin resistant bacteria and culturing these transformants in the presence of the antibiotic will select for clones that are resistant to the drug. Clones that have acquired ampicillin resistance carry the hybrid vector that expresses the ampicillin resistance gene. Alternatively, a DNA or RNA probe can be isolated or synthesized that is complementary to the nucleotide sequence of the gene of interest. The probe can be radiolabeled and annealed to the DNA of individual clones. Clones that retain the radioactive probe contain DNA sequences complementary to the probe and, therefore, carry the gene of interest.

The consequences of these cloning procedures are enrichment and amplification of a genomic DNA sequence. In typical E. coli plasmid vector or phage vector preparations this means that the genomic sequence is amplified from 10^{10} to 10^{11} copies per milliliter of unconcentrated bacterial lysate. The degree of enrichment, as would be anticipated, is dependent upon the genome size and is usually appreciable due to the elimination of unwanted background DNA. For example, a DNA fragment isolated from a human cell (3×10^{12} daltons of DNA) and inserted into the plasmid vector pBR322 (2.3×10^{6} daltons of DNA) would be enriched approximately 10^{6}-fold. Collectively, these techniques can isolate a single gene from any cell source and produce enough copies of that gene to make almost any analysis possible.

The general method of gene cloning described above is effective when the genomic DNA is from a procaryotic organism. Procaryotic gene sequences are relatively simple and colinear with the protein sequences they encode (Figure IV). However, in higher eucaryotic organisms, the coding sequences (exons) are sometimes interrupted by noncoding DNA (introns). The eucaryotic gene is transcribed first into a mRNA molecule that retains the introns. Splicing and joining enzymes subsequently remove the introns to generate a mature mRNA. These enzymes are not synthesized in procaryotes. The mature mRNA is then translated into

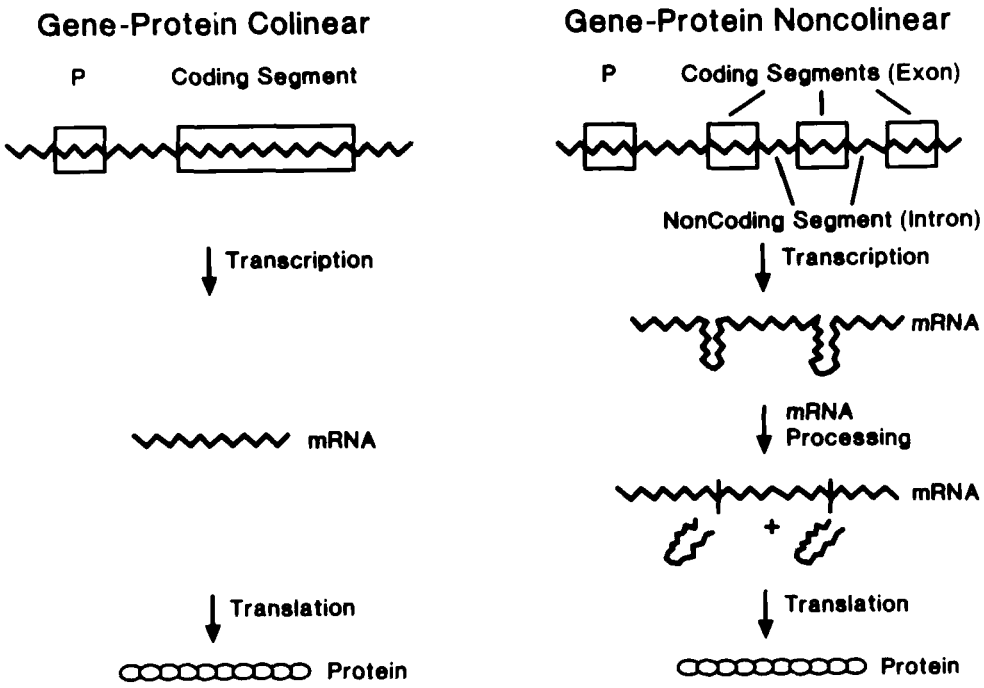

FIGURE IV

Figure IV illustrates uninterrupted coding sequences where the gene and protein are colinear and interrupted coding sequences where the gene and protein are not colinear. P represents the transcriptional promoter for the gene.

protein that is colinear with the mature mRNA but not with the gene sequence. The difficulty arises when the eucaryotic gene is cloned using the above approach into a procarotic host. Because the procaryotic host does not have splicing and joining enzymes to remove introns, the mRNA will encode additional amino acids that will produce a nonfunctional protein. In order to circumvent this problem a complementary DNA (cDNA) copy can be made from mature mRNA isolated from the eucaryotic cell. The cDNA copy is then inserted into a vector and cloned. The cDNA because it is a copy of the mature mRNA expresses a functional protein regardless of the host.

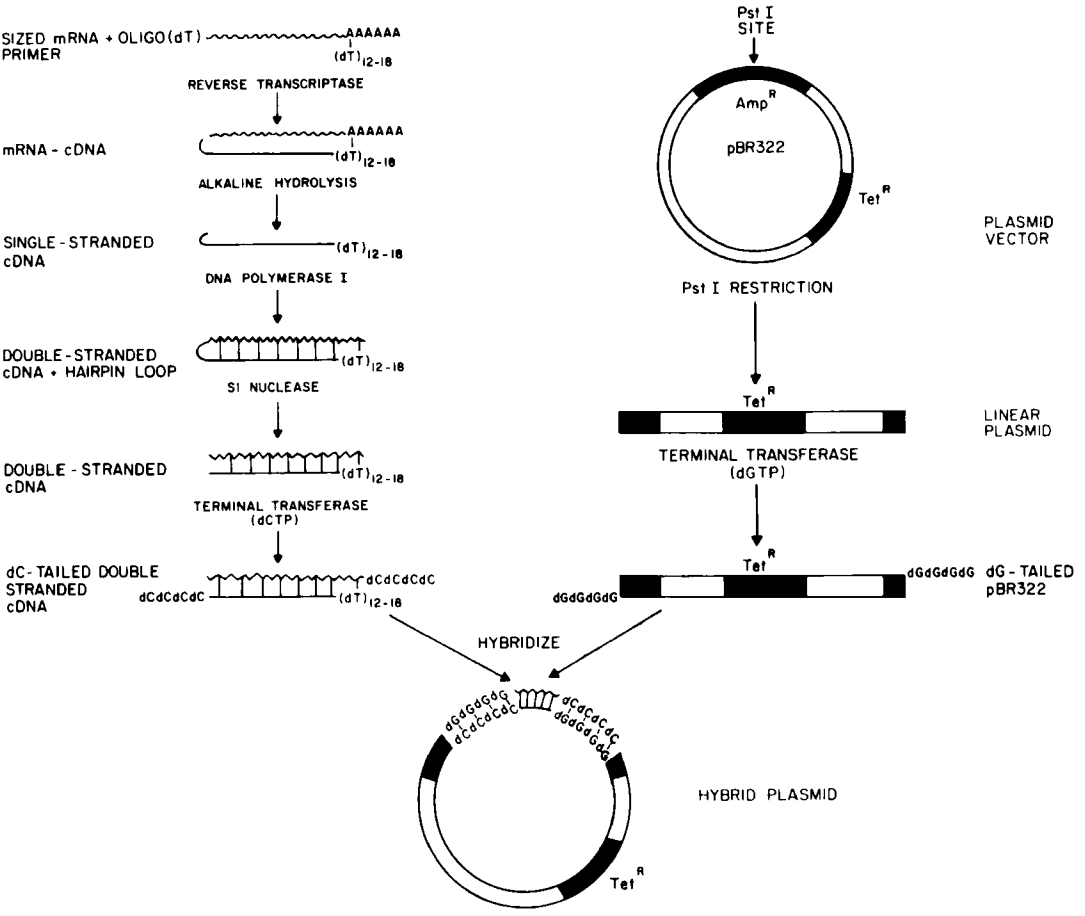

FIGURE V

Schematic Representation of steps used during cDNA cloning.

The methods of cDNA cloning are outlined in Figure V.[14] Mature mRNA is extracted from cells or tissues. The majority of these mRNA species have a homopolymer stretch of deoxyadenylate at the end of the molecules. To the mRNA Is added an oligonucleotide of deoxythymidylate (oligo-dT_{12-18}) that anneals to the homopolymer end of the mRNA. The oligonucleotide acts as a primer to synthesize the first strand of cDNA. Reverse transcriptase is an RNA dependent DNA polymerase that synthesizes a DNA molecule complementary to the mRNA. The mRNA is then

removed by an alkaline treatment. The second cDNA strand is synthesized complementary to the first using DNA polymerase I. The hairpin-loop formed during cDNA synthesis must be removed in order to insert the cDNA molecule into a vector. S1 nuclease is an enzyme that hydrolyzes single-stranded nucleic acids and will remove the hairpin structure. "Sticky-ends" are then synthesized by the action of terminal deoxynucleotidyl transferase. Usually the "sticky-end" is a homopolymer of deoxycytidylate. In parallel, a vector DNA is linearized by the action of a restriction enzyme (Pst I) and a homopolymer of deoxyguanylate is synthesized complementary to the "sticky-ends" of the cDNA. The cDNA and vector are annealed and the hybrid molecules are introduced into the appropriate host by DNA transformation. Selection of the correct recombinant clone is as described above.

It is also possible to chemically synthesize a DNA sequence or gene. If all or part of the amino acid sequence of a protein is available, the DNA sequence can be determined from the genetic code. There are several chemical methods for synthesizing a DNA sequence.[15] If the protein is relatively small, the entire gene can be synthesized. This has been performed for the hormone somatostatin,[16] for human insulin[17] and for the lymphokine interleukin-2.[18] If the protein is very large, it may not be practical to synthesize the entire gene. However, a portion of the gene can be synthesized and used as a probe to screen for the recombinant clone or used as a specific primer in cDNA synthesis. The chemistry has been automated ("gene-machines") which makes this science available to a larger group of researchers.

Ultimately, determination of the nucleotide sequence is a major goal, regardless of how the gene or DNA sequence is cloned and selected. Chemical methods[19] and enzymatic methods[20] are available that distinguish the four nucleotides and determines their sequential order. Nucleotide sequence information is valuable for demonstrating that the presumptive cloned gene encodes the gene product of interest. Comparison of the sequence calculated from amino acid sequence data of

the protein will confirm that the correct DNA sequence has been amplified and selected. In addition, there are specialized regulatory sequences that are responsible for the transcription of a gene and subsequent translation into protein. DNA sequence information will determine if the regulatory sequences have been retained or lost during the cloning procedures. Strategies can then be developed to eliminate, replace or modify the regulatory elements.

APPLICATION

The intended application of the cloned gene will determine what steps are subsequently performed. In the pharmaceutical industry the cloned gene can be used directly in a diagnostic test. For example, it may be possible to clone a portion of the genome of a virus and use the cloned DNA as a probe to screen tissue explants for the presence of virus. Recent work has resulted in the identification of a cloned DNA sequence that can be used in prenatal diagnosis and carrier detection of phenylketonuria.[21] Alternatively, the intended application may require the expression of the cloned gene product. The utilization of recombinant DNA techniques for the isolation and expression of genes that encode human interferon proteins has made production of large amounts of interferon more economical than conventional production methods.[22] A disease state that is caused by a dysfunction in a gene or the expression of that gene may be alleviated by gene therapy. The strategy is to introduce and express a normal gene analog in the appropriate tissue of the diseased organism using genetic engineering technology.[23]

In order to express a cloned gene in a foreign host, such as E. coli, it is necessary to understand the regulatory sequences that are required for expression of genes in that host. In simple organisms there are three essential elements. First, a transcriptional promoter encoded by the DNA immediately preceeds the gene. RNA polymerase initiates and regulates RNA transcription at the promoter. Second, the ribosome binding site lies between the transcriptional promoter and the gene and is the site where ribosomes bind to mRNA and initiate protein translation. The gene

sequence is followed by a termination singal, third, that prevents read-through of ribosomes into adjacent gene sequences. In addition to these regulatory elements, there are additional sites and cofactors that can affect the expression of genes in procaryotes and eucaryotes. To express the heterologous cloned gene it must be placed in such a way that it is under the transcriptional control of a promoter on an expression vector. The cloned gene must also have an appropriate ribosome binding site and a termination signal. Frequently, the intact cloned gene is removed from the vector in which it was cloned, is modified and introduced into an expression vector to achieve high level expression of the gene. Numerous procaryotic and eucaryotic genes have been cloned and expressed in E. coli using this scheme.[22] The final goal is to scale-up from the laboratory to a production scale. It is, therefore, critical that the expression vector and host system developed is appropriate in a production mode. Some of the obvious but nontrivial requirements for such a host-vector system include: an expression vector that is stably carried by the host; high level expression of the cloned gene product; cost effective and efficient selective pressures to maintain the vector and express the gene; and, suitable host-vector system for extraction and purification of a functional gene product.

The final purified product of gene cloning is analyzed biochemically and tested in animal systems that model a specific disease. The data from these analyses is submitted as an Investigation Exemption for a New Drug Application (IND) and clinical trials are performed. If the drug is efficacious in the clinical trials, then a New Drug Application (NDA) is submitted and, if approved by the FDA, a new health care product can be marketed.

CONCLUSION

One final point, I feel compelled to make. There is considerable skepticism regarding the value of recombinant DNA technology and genetic engineering in industry today. This attitude, for the most part, is derived from the lack of apparent success of several products of this

technology. In particular, human interferons have not been demonstrated to be an effective treatment for multiple forms of cancer. I attribute part of this failure to our naivete regarding the mechanism of action of these proteins in the body and partly to "bio-hype" which has built unrealistic expectations for these drugs. Nevertheless, the recombinant DNA technology that produced these drugs has been a tremendous success. The potential application of these tools in the pharmaceutical industry is limited only by our imagination. However, the success of these endeavors will depend upon how carefully these projects are selected.

REFERENCES

1. J. Marmur, J. Mol. Biol., 3, 208 (1961).
2. R.J. Roberts, in "Methods in Enzymology," S.P. Colowick and N.O. Kaplan, eds., Vol. 68, Academic Press, New York, 1979, p. 27.
3. F. Bolivar, R.L. Rodriquez, P.J. Greene, M. Betlach, H.L. Heyneker, H.W. Boyer, J.H. Crosa and S. Falkow, Gene, 2, 95 (1977).
4. E.Katz, C.J. Thompson and D.A. Hopwood, J. Gen. Micro., 129, 2703 (1983).
5. V.M. Williamson, J. Bennetzen, E.T. Young, K. Nasmyth and B.D. Hall, Nature, 283, 214 (1980).
6. M.F. Thomashow, R. Nutter, K. Postle, M.D. Chilton, F.R. Blattner, A. Powell, M.P. Gordon and E.W. Nester, Proc. Natl. Sci. USA, 77, 6648 (1970).
7. J.M. Tabor and V.C. Bode, Gene, 5, 255 (1979).
8. A.F. Purchio and G.C. Fareed, in "Methods in Enzymology," S.P. Colowick and N.O. Kaplan, eds., Vol. 68, Academic Press, New York, 1979, p. 357.
9. K. Beckingham, Plasmid, 4, 354 (1980).
10. M. Gellert, Proc. Natl. Acad. Sci. USA, 57, 148 (1967).

11. S.N. Cohen, A.C.Y. Chang and C. Hsu, Proc. Natl. Acad. Sci. USA, 69, 2110 (1972).

12. M. Wigler, A. Pellicer, S. Silverstein and R. Axel, Cell, 14, 725 (1978).

13. M. Grunstein and D. Hogness, Proc. Natl. Acad. Sci. USA, 72, 3961 (1975).

14. A. Efstratiades, T. Maniatis, F.C. Kafatos, A. Jeffrey and J. Vournakis, Cell, 4, 367 (1975).

15. M.H. Caruthers, in "Methods of DNA and RNA Sequencing," S.M. Weissman, ed., Praeger, New York, 1983, p. 1.

16. K. Itakuru, T. Hirose, R. Crea, A.D. Riggs, H.S. Heyneker, F. Bolivar and H.W. Boyer, Science, 198, 1056 (1977).

17. R. Crea, A. Kraszewski, T. Hirose and K. Itakuru, Proc. Natl. Acad. Sci. USA, 75, 5765 (1978).

18. Y. Stabinsky and K. Alton, personal communication.

19. A.M. Maxam and W. Gilbert, Proc. Natl. Sci. USA, 74, 560 (1977).

20. F. Sanger, S. Nicklen and A.R. Coulson, Proc. Natl. Acad. Sci. USA, 5463 (1977).

21. S.L. Woo, A.S. Lidsky, F. Gutter, T. Chandra and K.J.H. Robson, Nature, 306, 151 (1983).

22. D.V. Goeddel, E. Yelverton, A. Ullrich, H.L. Heyneker, G. Miozzari, W. Holmes, P.H. Seeburg, T. Dull, L. May, N. Stebbing, R. Crea, S. Maeda, R. McCandliss, A. Sloma, J.M. Tabor, M. Gross, P. Familletti and S. Pestka, Nature, 287, 411 (1980).

23. M.J. Cline, in "Basic Biology of New Developments in Biotechnology," A. Hollander, A.I. Laskin and P. Rogers, eds., Plenum Press, New York, 1983, p. 77.