PART II

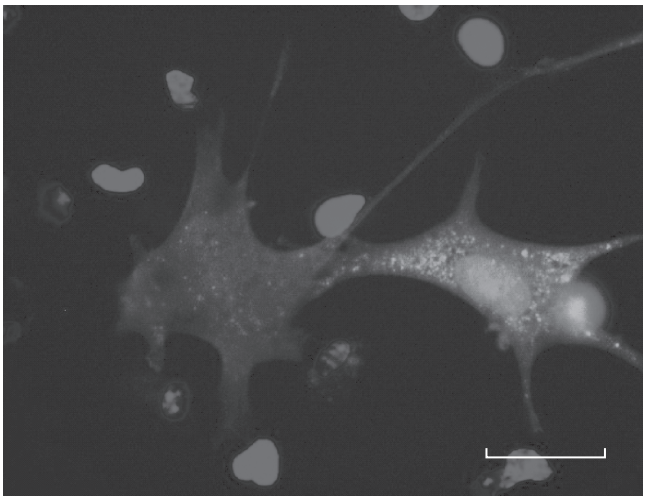
PRINCIPLES AND APPLICATIONS OF BIOREGENERATIVE ENGINEERING TO ORGAN SYSTEMS

SECTION 4

PRINCIPLES OF BIOREGENERATIVE ENGINEERING

10

MOLECULAR ASPECTS OF BIOREGENERATIVE ENGINEERING



Expression of green fluorescent protein (GFP) in cultured bone marrow stromal cells transfected with a GFP gene. Green: GFP. Blue: cell nuclei. Scale: 10 μm. See color insert.

The molecular aspects of bioregenerative engineering address the principles, technologies, and applications of bioregenerative engineering at the molecular level. Cellular activities, including cell proliferation, differentiation, and regeneration, are regulated by coordinated gene expression and protein activation. Thus, the regeneration of cells, tissues, and organs can be controlled by the modulation of gene expression and/or protein activities. The process of molecular modulation is often referred to as *molecular engineering.* Since protein modulation depends on, to a large extent, DNA modulation, this book will focus on DNA engineering.

Bioregenerative Engineering: Principles and Applications, by Shu Q. Liu

DNA engineering stems from the principles of natural genetic processes, such as gene replication, transcription, recombination, and mutation. Nature has designed and created one of the most elegant biological systems, the genome, for the storage and processing of hereditary information. The human genome is composed of about 3,000,000 nucleotide pairs and more than 50,000 genes, which are the fundamental units of genetic processing. A gene encodes the sequence of a polypeptide chain or protein. Each gene can be replicated to pass genetic information to the next generation and transcribed into messenger RNA (mRNA) to produce a protein. For the past half century, genetic research has demonstrated that a gene can be identified, removed, purified, modulated, replaced, reproduced, and transferred between cells or organisms. Methods for these preparations have provided a foundation for the establishment of the most influential technology in biomedical history: the recombinant DNA technology. Such a technology has revolutionized modern biomedical research, enabling the detection of gene mutants, the identification of genetic causes for pathological disorders, the reproduction and modeling of gene mutation-induced diseases and, more importantly, the potential improvement of cell regeneration and treatment of human diseases by using genetic approaches. Biomedical research with genetic engineering may hold the key to the understanding and treatment of deadly diseases, such as degenerative disorders, cancer, and atherosclerosis.

DNA engineering approaches are established on the basis of DNA recombination and transfection technologies. The ultimate goal of DNA engineering is to repair or replace defective or mutant genes, which cause pathological disorders and cannot be naturally repaired, and thus to enhance the regeneration of disordered cells, tissues, and organs. To achieve such a goal, it is necessary to understand the structure and function of genes and the role of gene mutation in the induction pathogenic disorders, and to establish engineering technologies for constructing therapeutic genes, which are used for replacing defect or mutant genes. It is also necessary to detect the function and effectiveness of the therapeutic genes delivered into target cells. In this chapter, the principles of molecular engineering are outlined.

DNA ENGINEERING

Gene Mutation [10.1]

In a broad sense, human diseases can be induced by gene mutation, environmental stimulation, or a combination of both. *Gene mutation* is defined as a heritable alteration in the structure of DNA. Gene mutation can occur spontaneously during evolution, which may play a role for evolutionary development of plant and animal species, and can also be induced by environmental factors, such as chemical (carcinogens), physical (radiation), and biological (viruses) factors. For instance, cytosine can be deaminated to form uracil spontaneously (Fig. 10.1) or under the action of certain chemicals, such as nitrous acid. There are various types of gene mutation. These types are summarized in Fig. 10.2. During the entire lifespan of an organism, each base pair may be mutated at a rate of 10^{-9}–10^{-10}. The rate of mutation for a gene is dependent on the number of the base pairs.

There are two forms of gene mutation: alterations in chromosomal structure and alterations in specific gene structure. Chromosomal alterations usually include apparent deletion or translocation of a segment of the involved chromosome(s). Each altered segment may

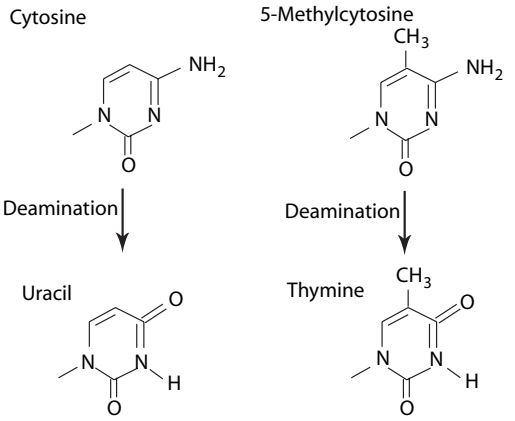


Figure 10.1. Schematic representation of cytosine mutation.

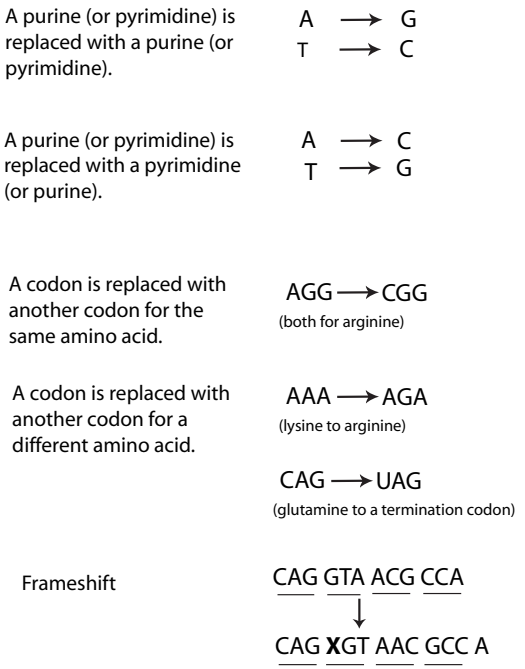


Figure 10.2. Common types of gene mutation found in the mammalian genomes. Based on bibliography 10.1.

contain multiple gene loci. In contrast, alterations in gene structure include mutation of single or multiple base pairs, which are often difficult to notice. In certain cases, a change in a single base pair, known as point mutation, may result in the translation of defective proteins, leading to severe clinical consequences. Sickle ell anemia is a typical example. This disorder is induced by the substitution of a single base C (cystine) with T (thymine) in the hemoglobin β chain gene. This substitution changes the codon CTC to TTC, resulting in the replacement of glutamic acid with lysine at the sixth residue of the hemoglobin

β chain. As a result, the erythrocytes exhibit an irregular shape and increased rigidity of the cytoskeleton, rendering the erythrocytes difficult to pass through capillaries. Capillary obstruction and oxygen deficiency are often found in peripheral tissues in sickle cell anemia. These pathological alterations are associated with clinical manifestations such as hemolysis and anemia, impaired system development and growth, increased susceptibility to infection, and micro- and macroinfarction.

Gene mutation may also induce premature termination of protein synthesis, resulting in the formation of incomplete or defective proteins. The translation of proteins is controlled by the sequence of mRNA, which contains the coding information for a specified protein. Three codons, including UAA, UAG, and UGA, serve as signals that induce the termination of protein translation. When a gene mutation occurs in a gene to change an amino acid codon to one of the termination codons, protein translation will be terminated prematurely, producing a defective protein. For instance, the codon for the amino acid tyrosine is UAU. The substitution of the last base U with A changes the tyrosine codon to a termination codon UAA, which signals the protein synthesis machinery to terminate protein translation. Since proteins are major components that participate in the constitution of cell structures and in the regulation of cellular activities, defective proteins often cause pathogenic alterations in cell structure and disorder in cell function.

Disorders Due to Gene Mutation [10.1]

Genetic disorders can be classified into three types: chromosomal defects, simple inherited genetic defects, and multifactorial defects. *Chromosomal defects* are disorders induced by deletion, duplication, or abnormal arrangement of partial or entire chromosome(s). Usually, a large amount of DNA or a large number of genes may be involved in chromosomal defects. Common chromosomal defects include trisomy (the presence of 47 chromosomes), monosomy (45 chromosomes), triplody (69 chromosomes), segmental duplication, and segmental deficiency. These chromosomal defects are often associated with abnormal anatomy of selected tissues and organs, mental retardation, behavioral disorders, and disorder of growth and development. Chromosomal defects can be detected by staining chromosomes with DNA dyes and optical microscopy. In normal cells, each of the 23 pairs of chromosomes exhibit distinct characteristics in terms of size, banding pattern, and the location of the centromere, which divides a chromosome into two arms. The sex chromosomes also show these characteristics. Women contain two XX sex chromosomes, while men contain one X and one Y sex chromosome. Any changes from the standard number and form of chromosome are considered chromosomal defects.

Simple inherited gene defects are disorders induced usually by a single mutant gene. On the basis of inheritance patterns, these defects have been classified into three categories: autosomal dominant, autosomal recessive, and X chromosome (or X)-linked gene defect or mutation. *Autosomal dominant gene defects* are those that cause clinical manifestations when one copy or allele of a gene in a genetic locus is mutated and the other copy or allele is normal, a condition referred to as *heterozygous gene mutation*. The gene defect is found in the autosomes, but not in the X and Y chromosomes. Because of the heterozygous nature of the gene defect (i.e., one chromosome of each pair contains the defective gene), 50% of the offspring will inherit the defective gene. The defective gene will be transmitted continuously through all generations by the individuals who carry the defective gene. Since the defect does not involve the sex

chromosomes, males and females are equally affected. Autosomal dominant gene disorders rarely cause infertility. Examples of autosomal dominant disorders include Huntington's chorea, Marfan's syndrome, familial hypercholesterolemia, myotonic dystrophy, familial hyperlipidemia, familial breast cancer, and hypertrophic obstructive cardiomyopathy.

Autosomal recessive gene defects are those that cause clinical manifestations only when both copies of a gene are defective at a gene locus. Individuals who carry only one copy of the defective gene do not exhibit clinical manifestations. While the defective gene is transmitted to offspring, the offspring do not show clinical symptoms. The condition that causes clinical manifestations is when the two parents are carriers of the same defective gene and the two defective copies of the affected gene are transmitted to the same individual. In that case, 25% offspring are not defective gene carriers, 50% are carrying one copy of the defective gene (heterozygotes), and the remaining 25% are homozygous carriers who express clinical manifestations. When a homozygote is married to a normal individual, the offspring are all heterozygous defective gene carriers, but do not expression clinical manifestations. The recessive type of gene defect occurs in the autosomes. Thus, males and females are both affected. Examples of common autosomal recessive disorders include sickle cell anemia, deafness, cystic fibrosis, hereditary emphysema, congenital adrenal hyperplasia, and familial Mediterranean fever.

X-linked gene defects are those that occur in the X chromosome. Although both males and females may have X-linked gene defects, the clinical expression of the disorder differs between the two genders. A female possesses two X chromosomes and may carry one or two copies of the defective gene. Thus, dominant and recessive forms of gene disorder can be found in females. In contrast, a male carries only one X chromosome. All male X-linked disorders are dominant. The X-linked defective gene will not be transmitted from the father to the son, because the son inherits only the Y chromosome from the father. A female can receive X chromosomes from both parents. When a male is a defective gene carrier, all his daughters will be defective gene carriers. When a female is a homozygous defective gene carrier and the father is normal, 50% of her offspring, including males and females, will become heterozygotes. Examples of common X-linked gene disorders include hemophilia A and B (factor VIII and IX deficiency, respectively), colorblindness, Duchenne muscular dystrophy, glucose 6-phosphate dehydrogenase deficiency, testicular feminization, and ocular albinism.

The third type of genetically related disorder is *multifactorial genetic disorders*. This is a group of disorders that are induced or enhanced by activation of more than one gene. The involved genes may act synergistically in the initiation and development of a disorder. Often, environmental factors are necessary components that stimulate the expression of the involved genes. When the genetic effect is accumulated to a critical level, one or more environmental factors may trigger the onset of the disorder. The environmental factors may also determine the severity of the disorder. These disorders do not express the inheritance features of the dominant or recessive gene defect-induced disorders. It is even difficult to clearly identify the genes that are involved in the pathogenesis of the disorder. Even for the same type of disorder, different genes may be involved in different individuals. Thus, it is difficult to predict the risk of inheritance for the offspring. A judgment is often made based on the basis of a family history and statistics. However, certain multifactorial genetic disorders may be dependent on other types of genetic disorder. For instance, familial hypercholesterolemia, a disorder induced by autosomal dominant gene defect, contributes significantly to the development of atherosclerosis, which is considered a

multifactorial genetic disorder. Examples of common multifactorial genetic disorders include atherosclerosis, diabetes mellitus, essential hypertension, cancer, epilepsy, and congenital heart disease, rheumatoid arthritis, and Parkinson's and Alzheimer's diseases.

Principles of DNA Engineering [10.2]

DNA engineering is applied to restore or modulate the structure of defective genes and thus to restore or improve the function of these genes. DNA engineering is established to study and treat disorders due to genetic defects. To achieve such goals, it is necessary to accomplish the following tasks: (1) identification of a cell type that is involved in the disorder of interest, (2) identification of a mutant gene that is responsible for or contributes to the disorder, (3) construction of a therapeutic gene, (4) transfection of target cells with the therapeutic gene for the treatment of the disorder; and (5) test of the effectiveness of gene transfection. These approaches are briefly discussed here.

Identification of Cell Types Involved in a Disorder. A pathological process inevitably involves cells, tissues, and organs, influencing the structure and functions of these biological systems. It may be fairly straightforward to identify the cell types involved in certain diseases, but may be difficult in others, depending on the nature of the disease and the accessibility of the involved organs. For instance, hepatoma may likely involve hepatocytes and other cell types in the liver, and atherosclerosis may involve blood and vascular cells, such as monocytes, macrophages, endothelial cells, and smooth muscle cells. For these diseases, cell types involved can be easily identified since the clinical signs and the anatomical locations of the pathological disorders can be easily recognized. However, it may not be that easy to identify a cell type involved in a neurological or psychological disorder since the functional anatomy of the brain is not well understood and the brain is not accessible. In any case, the first task in molecular engineering is to identify cell types that are involved in the disorder of interest. Although the genome is identical for all cell types, the pattern and level of gene expression can be vastly different. A disease is often associated with changes in gene expression. Thus, it is a critical step to identify cell types with altered gene expression. These cell types can be collected and prepared for the identification of genes involved in the disorder of interest.

Identification of Mutant Genes. One of the most important tasks in DNA engineering is the identification of a mutant gene that is responsible for the disorder of interest. Once a mutant gene is identified, a therapeutic gene can be established for replacement of the mutant gene and treatment of the disorder. It is important to note that a large number of genes have been identified, cloned, and analyzed in terms of their structure, function, and contribution to the pathogenesis of many diseases. It is strongly advised that, before starting an investigation, one should conduct a thorough literature search for information relevant to the genes potentially involved in the disorder of interest. It can save a tremendous amount of time and effort if the investigation starts with a known gene. However, if the cause of the disorder and the contribution of potential genes are not known, one may have to start with gene identification. A number of procedures are necessary for the identification of a mutant gene, including: (1) assessment of mRNA transcription or gene expression, (2) extraction of DNA from a selected cell sample and digestion DNA with restriction enzymes, (3) construction of recombinant DNA, (4) establishment of a DNA library, (5)

selection of the gene of interest, (6) amplification of the selected gene, (7) gene sequencing and analysis, and (8) test of the function of the selected gene.

Assessing mRNA Transcription. The identification of mutant genes may start with assessing the level of gene transcription. In general, a gene defect or mutation may result in an alteration in the level of mRNA transcription. The alteration may be a complete deficiency, increase, or decrease in mRNA transcription. Since a change in the regulatory activity of gene expression may also induce an increase or decrease in mRNA transcription, an alteration in gene expression may not be indicative of the presence of a mutant gene. However, such a change may serve as a clue for identifying potential mutant genes. Genes with altered expression can be selected for further analyses, such as DNA sequencing and functional test, which provide conclusive information for identifying the mutant gene.

To assess the level of mRNA transcription or gene expression, it is necessary to prepare mRNA and measure the level of mRNA transcription. Total mRNA can be extracted from target cells, and the transcription of mRNA can be assessed by analytical approaches such as Northern blotting and gene microarray analyses. Since the gene microarray analysis provides a profile of mRNA transcription covering hundreds and thousands of genes, this method is more efficient than Northern blotting analysis. Here, the principle of gene microarray analysis is briefly discussed.

Gene microarray analysis is an approach established to detect quantitatively the transcription levels of multiple mRNAs according to the rule of complementary hybridization with prearranged DNA or cDNA (complementary DNA) samples with known sequences. In this analysis, fluorochrome-conjugated mRNA or reverse-transcribed cDNA samples, known as *probes*, from a specified cell type are applied to arrays of microspots coated with selected DNA samples of known sequences, inducing hybridization reactions. The fluorescent levels of the spots with mRNA-hybridized DNA can be measured and used to represent the relative level of mRNA transcription in the selected cell type (Fig. 10.3). Because the DNA samples coated on the microarrays are selected with known structure, the hybridized mRNA or cDNA samples can be identified. With the information accumulated in the gene bank for the past decades, genes involved in most known diseases have been studied, identified, and cloned. Thus, it is not difficult to establish gene microarrays that can be used for the identification of genes involved in common diseases. A major advantage for the gene microarray analysis is that a single experiment can provide an expression profile for a large number of genes.

To conduct a gene microarray analysis, it is necessary to follow several procedures, including the preparation of gene microarrays, preparation of fluorochrome-conjugated mRNA or cDNA probes, probe-DNA hybridization, detection of the relative levels of mRNA or cDNA probes hybridized to DNA targets, and data analysis. Gene microarrays can be prepared by coating denatured genomic DNA, cDNA (DNA prepared by reverse-transcribing mRNA), or synthesized oligonucleotide samples to arrays of DNA-binding spots created by photolithography on a glass slide of a centimeter size. Each spot can be as small as 50–350 μm in dimension. Thus, a large number of different DNA samples can be arranged in a single slide (>5000 spots/cm^2), which is also known as a gene "chip." In practice, it is considerably easier to prepare microarrays with oligonucleotides compared to those with DNA and cDNA. Oligonucleotides can be synthesized in single-strand format. An appropriate size of oligonucleotides for gene microarray analysis should be 20–120 bases.

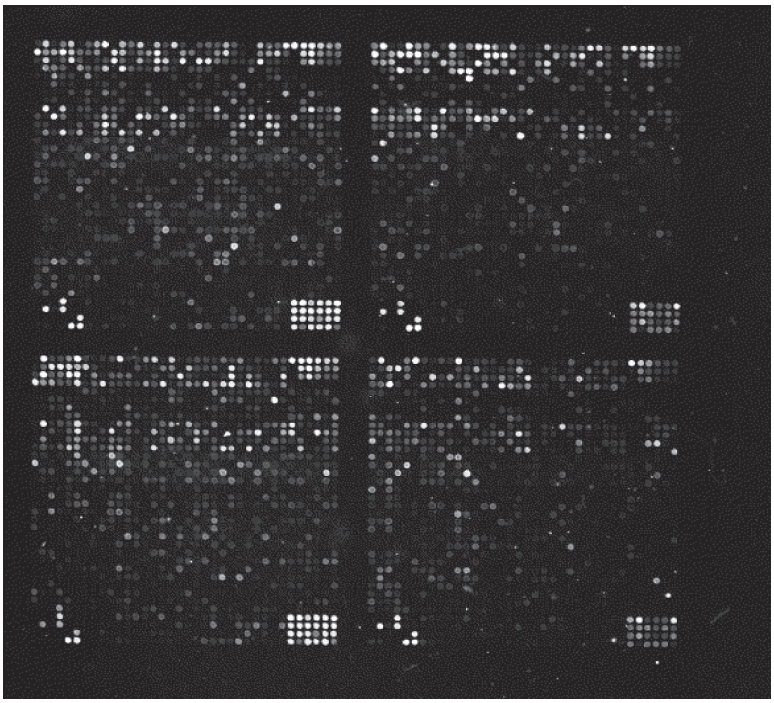


Figure 10.3. An image of a microarray containing >5000 genes. Each spot features a pool of identical single-stranded DNA molecules representing a single gene. The brightness of the spot is proportional to the amount of fluorescent mRNA hybridized to the DNA of the spot. The fluorescence spots can be identified by automated image analysis. The fluorescence intensity from each spot can be measured and compared to the background fluorescence. The images are further compared to images obtained from control measurements and transformed into a gene expression matrix, which can be analyzed by numerical methods. (Reprinted by permission of the Federation of the European Biochemical Society from Brazma A, Vilo J: Gene expression data analysis, *FEBS Lett* 480:17–24, copyright 2000.)

To detect the level of mRNA transcription from a cell sample, two types of probe can be synthesized from extracted mRNA: cDNA and RNA. cDNA probes can be synthesized by reverse transcription using extracted mRNAs as templates and the reverse transcriptase as a catalytic enzyme. Such a procedure produces single-stranded cDNA molecules. The cDNA probes can be conjugated with fluorochromes for probe identification. Alternatively, RNA probes can be synthesized from cDNA templates reverse-transcribed from mRNAs. In this preparation, T7 RNA polymerase is used to synthesize RNA molecules from double-stranded cDNA templates in the presence of fluorochrome-labeled nucleotides. The synthesized RNAs can be used as probes for gene microarray analysis.

The next step is to apply the probe, either single-stranded cDNAs or RNAs, to gene microarrays. The probing cDNA or RNA molecules hybridize with target DNA molecules via hydrogen bonds on the basis of the DNA complementary rule. Excessive and unhybridized probes can be removed by washing. The fluorescence of hybridized probes can be imaged and recorded from the microarrays by a laser scanner, and the intensity of

fluorescence can be measured and analyzed. Such intensity represents the relative level of mRNA transcription. It is important to note that positive and negative controls should be introduced to the analysis. A positive control can be a selected cDNA probe that is known to hybridize to a target DNA, whereas a negative control is a cDNA probe that does not hybridize to a given target DNA.

The profile of mRNA transcription derived from a gene microarray analysis can be analyzed and compared to the profile of mRNA transcription from a normal or specified control cell type. Thus, genes with altered transcription can be identified from the microarray analysis. Although changes in the level of mRNA transcription may not indicate a gene defect or mutation, an increased, decreased, or null gene transcription suggests a candidate gene for further investigation. Once a gene with an altered transcription level is identified, the gene can be cloned and analyzed for the identification of structural mutation. If no mutation is found, the alterations in gene expression reflect changes in the regulatory activity for gene expression.

Extracting and Digesting DNA. Once a gene with altered mRNA transcription is identified as described above, DNA can be isolated and collected for gene analyses, including gene sequencing and identification of gene mutation. To obtain DNA, a selected cell sample can be lysed (for cultured cells) or homogenized (for a tissue sample), and DNA can be extracted and purified with established methods. Since DNA is a very large molecule, it should be digested into short fragments for DNA manipulation. DNA digestion can be accomplished by using restriction enzymes, which are found in bacteria and are capable of cleaving DNA at specific sequences. Although restriction enzymes are originated from bacteria, they can cleave DNA molecules from all known species as long as the DNA contains digestion target sequences. Each DNA molecule contains a large number of restriction sites for various types of restriction enzyme. These restriction sites are randomly distributed. DNA fragments with desired lengths can be generated by selecting appropriate restriction enzymes based on the locations of restriction sites. The digestion of target DNA at specified sites is one of the most important genetic approaches used for DNA manipulation. It is impossible to clone genes without the assistance of the restriction enzymes. It should be noted that restriction enzymes are not present in mammals. Thus, DNA molecules are not digested in the body systems under physiological conditions.

Restriction enzymes can cleave DNA into fragments with two different forms of end—sticky end and blunt end—depending on the type of restriction enzyme, but not the sequence of DNA. Some restriction enzymes, such as EcoRI, EcoRII, and HindIII, can cut the two strands of a DNA molecule unevenly, generating DNA ends with uneven strand length (the 3′ end is longer than the 5′ end or vise versa). The uneven end is also known as "sticky" end, so named because two such ends can hybridize to each other based on the complementary rule. Other restriction enzymes, such as HindII, HaeIII, and SmaI, cut the two strands of a DNA molecule evenly, leaving DNA ends with even strand length. The even DNA ends are also called "blunt" ends. In most cases, the substrate sequence for a restriction enzyme is identical for the two DNA strands at a given restriction site, but oriented in antiparallel directions. Thus, a "sticky" cut can produce two identical ends at a restriction site. This is a very useful feature for the creation of recombinant DNA. A "blunt" DNA end can be remodeled into a "sticky" end by an enzymatic treatment. Once DNA fragments are prepared, these fragments can be integrated into DNA vectors and used to create recombinant DNAs.

Constructing Recombinant DNA. The extracted and digested DNA fragments need to be further processed to generate DNA structures that can be cloned, screened for the desired gene, analyzed, and tested. An effective approach for such purposes is the construction of *recombinant DNA*. Recombinant DNA is a complex of DNA constructed by integrating a DNA fragment selected from a donor organism into a DNA vector collected from a different organism. The process of constructing a recombinant DNA is defined as *DNA recombination*. The purpose of DNA recombination is to generate functional genes in vitro, so that the gene of interest can be cloned, amplified, tested, used for gene transfection, and expressed in the transfected cells and tissues.

Gene cloning is a process of gene replication, generating multiple copies of the same gene. For cloning a gene, it is necessary to prepare a *gene cloning vector*. A vector is a bacterial DNA structure that contains restriction sites, and can accept foreign gene inserts and replicate in living bacteria. A vector is often engineered to add multiple restriction sites for the insertion of foreign gene fragments and to add functional genes for the selection, identification, and collection of the vector. A number of vector types, such as plasmids, λ phages, and cosmids, have been established and used in gene cloning. A typical plasmid vector carrying a LacZ gene (β-galactosidase gene) is shown in Fig. 10.4. The LacZ gene is often used as a reporter gene for gene selection or gene transfection.

To make recombinant genes, a selected donor DNA or cDNA fragment and a selected vector are treated with an identical type of restriction enzyme to generate the same type of "sticky" DNA ends. The enzyme-treated DNA fragment and vector are then incubated under a desired condition to hybridize the DNA fragment with the vector. At this stage, although the donor DNA and vector are joined with hydrogen bonds, the sugar–phosphate backbone is not linked together with phosphodiester bonds. An enzyme known as DNA ligase is needed to link the backbone, completing the recombinant process.

There are potential problems for DNA recombination. One problem is that the open ends of a vector may rejoin together without the insertion of donor DNA. Such vectors should not be included in the analysis. A gene selection method is usually used to remove the vectors that do not contain the inserted gene of interest. For this method, a selection gene, such as the β-galactosidase gene, is inserted into a selected vector in a cloning region with multiple restriction sites, also known as a *polylinker* or a *multiple cloning site*. The selection mechanism is that the β-galactosidase gene can be expressed when transfected into bacterial cells, only if the β-galactosidase gene is not interrupted by another gene. The insertion of a donor DNA fragment into the polylinker interrupts the structure of the β-galactosidase gene and renders the β-galactosidase gene unfunctional. When vectors without the donor DNA fragment are transfected into bacteria, the β-galactosidase gene

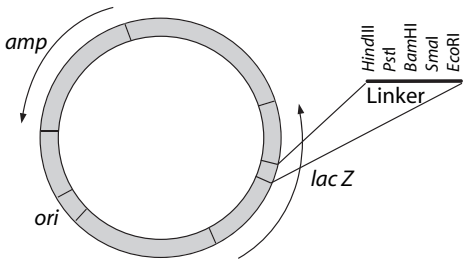


Figure 10.4. Schematic representation of the pUC18 plasmid vector containing a Lac Z gene.

can be expressed. Expressed β-galactosidase can react with a reagent known as X-gal to produce a blue-colored substance, which can be visualized under an optical microscope. However, the β-galactosidase gene in the vector with the donor DNA insert cannot be expressed and no blue-colored substance is generated when reacted with X-gal. Thus, clones with inserted donor genes can be selected on the basis of the color expression.

Another problem is that dominant "blunt" ends may be generated for the donor DNA, reducing the efficiency of donor DNA integration into the cloning vectors. To resolve such a problem, the donor DNA can be treated with a specific exonuclease that degrades the DNA at only the 5′ end, thus creating uneven "sticky" ends. However, the "sticky" ends of the donor DNA may not be complementary to the ends of a vector. To resolve such a problem, the open ends of the donor DNA fragment and the vector can be modified by synthesizing a short sequence at the ends. Such a process can be induced by treating both the donor DNA and the vectors separately with terminal transferase in the presence of different yet complementary nucleotides. For instance, the donor DNA can be treated with terminal transferase in the presence of dTTP, whereas the vector can be treated with terminal transferase in the presence of dATP. The added poly-T ends of the donor DNA are thus complementary to the poly-A ends of the vector.

Establishing a DNA Library. To identify and isolate the gene of interest from the gene pool containing all genes from a genome, it is necessary to establish a *DNA library*, which is defined as a collection of all gene clones from a cell sample of interest. A gene library can be established by using the procedures as described above, namely, extracting all DNA molecules, digesting the DNA with selected restriction enzymes into fragments of appropriate lengths, and inserting the DNA fragments into gene-carrying vectors. The vectors are then transfected into *Escherrichia coli* to establish *E. coli* colonies. Each *E. coli* colony is supposed to contain a specific DNA fragment. A DNA library is composed a large number of DNA fragments (Fig. 10.5). It is expected that one of these fragments contains the gene of interest. Such a gene can be identified and isolated by using a specific DNA or RNA probe as discussed on page 433.

For establishing a DNA library, two issues should be taken into account: vector type used in the cloning process and the DNA source. Vectors can accommodate DNA fragments with different sizes. Thus, vectors should be selected on the basis of the DNA fragment size. Plasmids and λ-phages are suitable for small DNA fragments, whereas cosmids and yeast artificial chromosomes (YACs) are used for large DNA fragments. Various vectors give different forms of DNA library. A plasmid or cosmid library is a collection of *E. coli* bacteria containing the cloned genes, a λ-phage library is a collection of λ-phages, whereas a YAC library is a collection of yeasts. The choice of vectors does not affect the structure of the cloned genes.

Another factor to consider in the construction of a DNA library is the source of DNA. DNA from two sources can be used for the construction: genomic DNA and complementary DNA (cDNA). The later is DNA synthesized from mRNA by reverse transcription. There are distinct features for the two types of DNA library. Genomic DNA contains introns and exons for all genes from a genome. A genomic DNA library is suitable for investigating the regulation of gene transcription, which requires the identification and understanding of the regulatory *cis* elements located in the introns. In contrast, a cDNA library contains only exons that encode proteins. In addition, cDNA does not contain genes that are not transcribed into mRNA at a given time and state. Thus, a cDNA library is suitable for the identification and isolation of expressed genes. Since most pathological

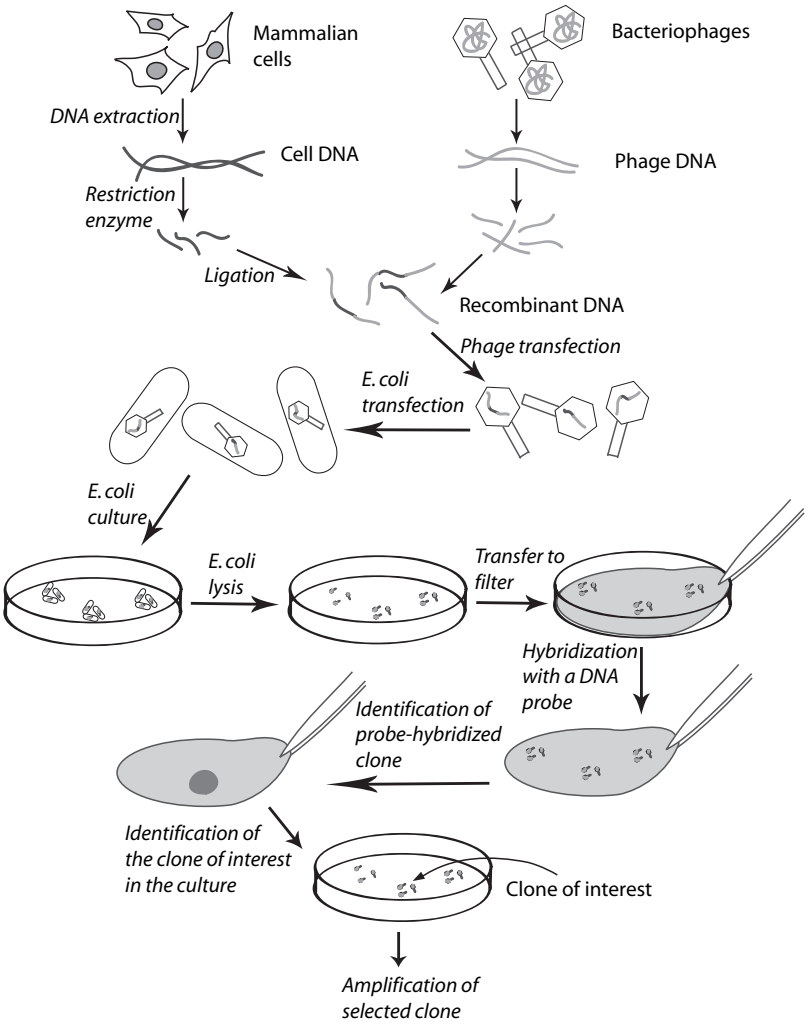


Figure 10.5. Preparation of a gene library. Based on bibliography 10.2.

disorders involve gene expression, creating a cDNA library is a practical approach for identifying genes that contribute to pathological disorders.

Selecting a Gene of Interest from a DNA Library. A DNA library may contain thousands of gene fragments. To find and collect the gene of interest, it is necessary to screen the DNA library. There are two approaches that can be used for DNA screening: directly probing the gene of interest or probing the protein encoded by the gene integrated into the expression vector, which is designed to express proteins in cultured *E. coli* cells. For the direct DNA approach, it is necessary to establish a DNA probe, which is a single-stranded DNA fragment or oligodeoxynucleotide, capable of hybridizing to its complimentary strand. With the tag of an identification marker to the probe, the target gene strand can be identified and isolated. For the protein-probing approach, an antibody

specific to the protein produced by the gene of interest can be used to identify the protein, which leads to the identification of the encoding gene.

As discussed on page 426, gene selection can start from the detection of mRNA transcription in a cell sample selected from a disordered organ by gene microarray analysis. An altered transcription level (increase, decrease, or null) may help to identify a potential gene involved in the pathogenesis of a disorder. It is important to point out that a gene microarray analysis does not provide information about alterations in the gene structure. It merely demonstrates the level of mRNA transcription, suggesting potential genes for further analyses. It is the DNA library screening and probing that provide gene clones, which can be used for sequencing and functional analyses.

For direct gene probing, a key issue is how to design a DNA probe. A probe can be constructed according to a selected sequence within the gene of interest (identified by gene microarray analysis), if the gene has been identified. Usually, information for most human genes is available in the GenBank database, from which a gene sequence can be selected. A DNA probe or oligodeoxynucleotide can be synthesized based on the selected sequence. A DNA probe can also based on the structure of a protein, which may potentially contribute to a pathological disorder of interest and is identified via protein analysis. Once the identity of the protein is known (via sequencing or mass spectrometry), an oligodeoxynucleotide can be synthesized based on the amino acid sequence of the protein. It is important to note that most amino acids are coded by more than one gene codon and there may exist a number of DNA sequences for each given amino acid sequence. For instance, the amino acid tyrosine (Tyr or Y) is encoded by gene codons TAT and TAC, and serine (Ser or S) is encoded by TCT, TCC, TCA, and TCG. Thus, it may be difficult to define the exact oligonucleotide sequence on the basis of a known protein sequence. To resolve such a problem, a fragment of protein with the minimal gene codon redundancy should be used. Furthermore, all possible DNA sequences for a selected protein fragment should be synthesized and mixed together. For a peptide sequence of 5 amino acids, including Tyr, Ser, Asp (asparagines), Cys (cysteine), and His (histidine), which are encoded by 2, 4, 2, 2, and 2 gene codons, respectively, 64 oligonucleotide strands need to be synthesized. The mixture of all these oligonucleotide strands should be applied to a DNA library for hybridization. The correct oligonucleotide fragment will hybridize to the gene of interest.

In addition to DNA probes, an antibody can be used as a probe to identify the protein product of a gene, thus identifying the gene indirectly. A key procedure for this approach is to induce protein expression. Such a task can be accomplished by using an expression vector. A cDNA fragment can be inserted into a selected expression vector in frame with the gene of a bacterial protein to generate a fusion protein. When *E. coli* cells are transformed with the vector, the cells will produce the fusion protein that contains the protein encoded by the inserted cDNA and the bacterial protein. In the DNA library, the fusion protein is expressed within or near the gene clone. Since the structure of the bacterial protein is known, an antibody can be developed and used to screen the DNA library. Any clones marked with the antibody should contain the gene of interest. Such a clone can be excised and collected for further analyses.

Amplification of the Selected Gene. The procedures described above can only produce a small amount of recombinant DNA. To carry out further DNA analyses such as DNA sequencing and transfection, the recombinant DNA must be amplified to generate a sufficient amount of DNA. An effective approach for DNA amplification is to transfect *E. coli*

cells with the identified and isolated recombinant gene from the DNA library. Each recombinant gene can replicate into multiple copies once transfected into *E. coli* cells. These recombinant gene copies can be rapidly amplified when *E. coli* cells grow. After cell culture for a day or two, billions of copies of each recombinant gene can be produced. The recombinant genes can be purified and isolated. The copies of the gene are referred to as gene clones.

DNA can also be amplified by using polymerase chain reaction (PCR), a method for in vitro DNA synthesis in the presence of a DNA template, DNA polymerase, DNA primers, and deoxynucleotides (dNTPs). This method can be used to select and amplify a DNA fragment of interest from genomic DNA by using an appropriate set of primers. In practice, a template DNA/cDNA fragment or a plasmid that contains a DNA fragment of interest is incubated in a PCR buffer supplemented with a temperature-resistant DNA polymerase (e.g., Teq DNA polymerase), DNA primers (specific to each selected DNA fragment), and dNTPs. The reaction mix is subject to about 30–35 thermal cycles with three alternating temperature levels for each cycle, including 95ºC, 60ºC, and 70ºC. The temperature 95ºC is for denaturing DNA, 60ºC is for primer annealing, and 70ºC is for DNA synthesis. After the PCR reaction, a million-fold of amplification can be achieved. It is important to note that PCR is suitable for the amplification of small DNA fragments (hundreds of base pairs) before the gene is inserted into a plasmid and for the selection of a gene fragment from a plasmid. PCR is not capable of amplifying large DNA molecules, such as plasmids or cosmids.

DNA Sequencing and Analysis. Once a gene of interest is identified, cloned, and isolated, the next steps are to confirm the isolated gene, detect the sequence of the gene, and assess possible structural changes or gene mutation. A common approach used for the confirmation of the isolated gene is Southern blotting analysis, named after E. M. Southern, who initially developed the technique. To prepare for Southern blotting analysis, the isolated gene-containing vector should be amplified as described on page 432 if the samples are not sufficient, and the gene of interest should be removed from the cloning vector by digestion with restriction enzymes that cut and release precisely a selected DNA fragment. The removed DNA fragment can be fractionated by gel electrophoresis, transferred to a filter membrane, and detected with the probe used for DNA library screening or a different probe designed on the basis of the same gene. The probe-reacted band of DNA on the filter membrane is the gene of interest. The identified band can be excised from the gel and used for gene sequencing and analyses.

To determine whether altered expression of a selected gene is due to gene defect or mutation, or due to merely a change in the regulatory activity of gene transcription, it is necessary to detect the sequence of the gene and compare it with that of a control gene. The sequence of a gene can be determined by using two methods: *base destruction sequencing* and *dideoxy sequencing*. For the first method, DNA strands are labeled at the 5′ ends with ^{32}P, the double-stranded DNA is denatured (separated), and one strand is discarded. Copies from the remaining strand are separated into four groups and used for sequencing. Each group is treated with a chemical reagent that selectively breaks one or two of the four bases. The four groups are treated with four different chemicals. For instance, group 1 is treated with a chemical that degrades the bases A and G, group 2 with a chemical for the base G, group 3 with a chemical for the base C, and group 4 with a chemical for the bases C and T. For each group, the degradation of the selected base results in the disruption of the DNA strand at all locations with the selectively degraded

base. When the degrading chemical is prepared in an appropriate concentration and reaction is carried out for an appropriate period, only a fraction of the target bases is degraded for each copy of DNA. Since the chemical reaction is completely random and a DNA strand contains a large number of the same base, DNA strands can be disrupted at different locations of the target base. Since a large number of DNA copies are present for each reaction with a selected degrading chemical, multiple DNA copies with an identical disruption location can be generated. For a group of 6 identical DNA copies, each contains 3 Gs, a treatment with a degrading chemical specific to the Gs under a well-controlled chemical concentration may induce only one disruption for each copy of DNA with an equal degradation probability for the three Gs, possibly resulting in an equal number of DNA copies disrupted at each G location. On the basis of such a principle, the four groups of single-stranded DNA samples can be disrupted randomly at distinct locations. Each group is composed of DNA fragments disrupted at a given base type with an equal disruption probability for each location of the same base. Resulting DNA fragments from each group can be run in a designated lane in gel electrophoresis. All DNA bands in each lane lost the same base at the 3′ ends, which is the target of a selected degrading chemical. The 5′ ends of the DNA fragments are tagged with ^{32}P for the purpose of identification. By comparing electrophoretic results from all four groups, all ends of the DNA fragments can be read and analyzed, giving a sequence of the entire DNA strand.

For the dideoxy sequencing method, a single stranded DNA is prepared and used as a template for synthesizing DNA in the presence of dideoxy nucleotides (ddNTPs). A ddNTP contains a deoxyribose sugar that lacks the 3′-hydroxyl group. A ddNTP can be incorporated into a DNA chain, but DNA synthesis is terminated whenever a ddNTP is incorporated because of the lack of the 3′-hydroxyl group, which is necessary for DNA extension. In practice, single-stranded template DNA is divided into four groups, and DNA synthesis is carried out in the presence of DNA polymerase, primers labeled with a radioisotope or fluorescent marker for identification, and the four types of standard deoxynucleotides (dATP, dTTP, dCTP, and dGTP). In addition, each reaction group is mixed with a distinct type of ddNTPs. Group 1 is mixed with ddATP, group 2 with ddTTP, group 3 with ddCTP, and group 4 with ddGTP. For each reaction group, the ddNTP is mixed in such a percentage with respect to the dNTP that the ddNTP molecules are incorporated into only a fraction of growing DNA chains, while the remaining DNA chains incorporate the standard dNTPs at any given time. This gives an equal probability of ddNTP incorporation into each site of a corresponding dNTP. DNA synthesis continues when a dNTP is incorporated to the respective site, but stops when a ddNTP is incorporated. As a result, various sizes of DNA chains can be synthesized in each reaction group with all chains ending at the same ddNTP site. Synthesized DNA fragments of different sizes from each reaction group can be fractionated by gel electrophoresis in a distinct lane. Each DNA band indicates the relative location of a designated ddNTP. By comparing all four reaction groups (each reacted with a different ddNTP), the locations of all four ddNTPs can be determined. By reading along the electrophoretic lanes, the sequence of the DNA fragment can be resolved.

Testing the Function of the Selected Gene. Based on the gene sequence, gene mutation, if any, can be identified by comparing with the sequence of a respective control gene from a gene database or from a specified control cell sample. The next step is to design a physiological test and assess whether the identified gene mutation influences the physiological function of the gene. A test may be prepared with the following procedures:

(1) constructing a recombinant vector containing the gene of interest; (2) establishing a physiological testing system, which is usually cultured cells or an in vivo animal model; (3) transferring the vector into the physiological testing system; and (4) assessing the influence of the transferred gene on the function of corresponding proteins and related cellular activities.

A recombinant vector with a mutant gene can be constructed as shown in Fig. 10.6. Since gene expression is a critical process for the test, it is necessary to integrate into the vector a gene promoter, which drives the gene expression. Commonly used promoters include cytomegalovirus (CMV) promoter and simian virus (SV)40 promoter, but any promoter that drives the expression of the selected gene can be used. In addition, one may integrate a reporter gene, such as a β-galactosidase gene and a green or red fluorescent protein gene, into the recombinant gene. The protein products of these reporter genes can serve as markers for the identification of the cells with positive gene transfection and expression. The constructed vector can be amplified by transforming and growing *E. coli* cells as described on page 432.

A functional test can be carried out in an in vitro cell culture system or an in vivo organ system. The in vitro system is easier to use and usually provides reliable information. It should be kept in mind that the endogenous wildtype gene in the testing cells may likely interfere with the transferred exogenous gene, rendering it difficult to identify the influence of the transferred exogenous gene. To resolve such a problem, a cell line with null gene mutation (complete loss of the function of a wildtype gene, also referred to as *gene knockout*) should be used. Such a cell line can be obtained from a transgenic animal

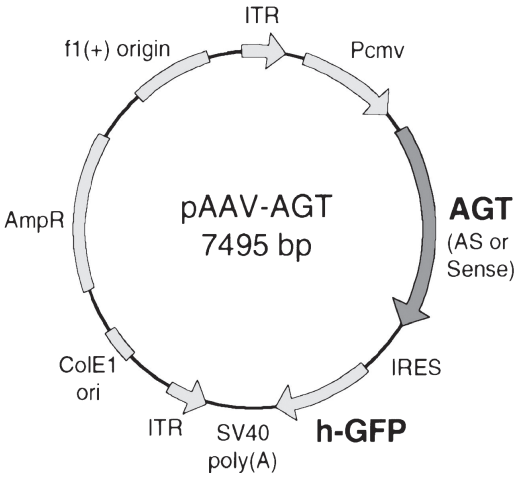


Figure 10.6. Schematic representation of the rat angiotensinogen (AGT) gene construct based on the adeno-associated virus (AAV)-derived plasmid. Plasmid pAAV-AGT-AS and pAAV-AGT-S were constructed by inserting a full-length cDNA (1.65 kb) of AGT into the unique *Hind* III site of the AAV-derived plasmid pTR-UF3 in antisense (AS) or sense (S) direction, respectively. AGT cDNA, indicated by the shaded arrow, is driven by a cytomegalovirus (CMV) early promoter (Pcmv). ITR, AAV inverted terminal repeat; IRES, polio virus type 1 internal ribosomal entry site; h-GFP, "humanized" victoria green fluorescent protein gene. Other open bars and arrows represent other basic elements of the vector. (Reprinted from Tang X et al: *Am J Physiol* 277:H2392–9, copyright 1999 by permission of the American Physiological Society.)

model without the gene of interest. Since the function of the selected wildtype gene is completely deficient, the function of the transferred recombinant gene can be assessed by comparing to a control with the function of the corresponding wildtype gene.

The next step for testing the gene function is to transfer the gene construct into cultured cells. It is a natural property that cells can endocytose large molecules, including DNA, into the cytoplasm. A fraction of endocytosed DNA can go through the nucleus membrane and integrate into the genome. Several methods, including virus-, salt-, liposome-, receptor-, and electroporation-mediated gene transfer, have been established and used to facilitate gene transfection into mammalian cells. The exogenous gene can be expressed transiently, producing encoded protein. The function of the transferred gene can be tested in several aspects, including the expression and function of the encoded protein as well as the activity of the host cells. Tests can be conducted at selected times after 12 hs from gene transfer. Protein expression can be detected by Western blotting or immunoblotting, which is a sensitive method for detecting a small amount of protein. A comparison to a sample from control cells with the corresponding wildtype gene demonstrates alterations in the mutant protein in terms of molecular weight and expression level. A functional test can be designed according to the nature of the protein. If the protein of interest is an enzyme, the catalytic activity of the protein can be assessed by detecting changes in the substrate protein. In vitro tests can be carried out by using an isolated enzyme and substrates. The influence of the transferred gene on the host cells can be assessed by measuring changes in a cellular activity known to be related to the transferred gene.

Constructing a Recombinant Therapeutic Gene. Once the pathological role of a mutant gene is identified, the next step is to design and construct a therapeutic gene that can be used to restore the physiological function of the gene (Fig. 10.7). Methods described on page 429 can be used for this purpose. Recall that there exist several types of genetic disorders, including chromosomal defects, simple inherited genetic defects, and multifactorial defects. Different strategies should be used for these types of genetic disorder. For chromosomal defects, since a large amount of DNA is involved, it is difficult to restore the structure and function of the defect chromosomes. For a simple inherited genetic defect, sine a single gene is usually involved, a corresponding wildtype gene may be isolated, cloned, and used to construct a therapeutic gene, which can be used for correcting the defect gene. For multifactorial defects, it is necessary to identify the genes involved and determine the pathogenic mechanisms of the disorder, based on which therapeutic genes can be constructed to modulate the activity of the mutant genes. For instance, the upregulation of multiple growth factor genes may contribute to the initiation and development of atherosclerosis. Growth inhibitor genes can be constructed and transferred into vascular cells to inhibit the activity of growth factors and thus to suppress atherogenesis.

Transfection of Target Cells with a Therapeutic Gene. The primary goal of molecular engineering is to correct pathological disorders due to gene mutation or changes in gene activities. An effective approach is to transfer therapeutic genes, as constructed on page 429, into target cells to restore the structure and function of the mutant or altered genes. Such an approach is referred to as *gene therapy.* The transferred genes, once in the cytoplasm, can be transported to the cell nucleus, integrated into the genome, and expressed to produce corresponding proteins. It is expected that the newly expressed proteins can function in place of the corresponding mutant proteins, thus reducing or suppressing

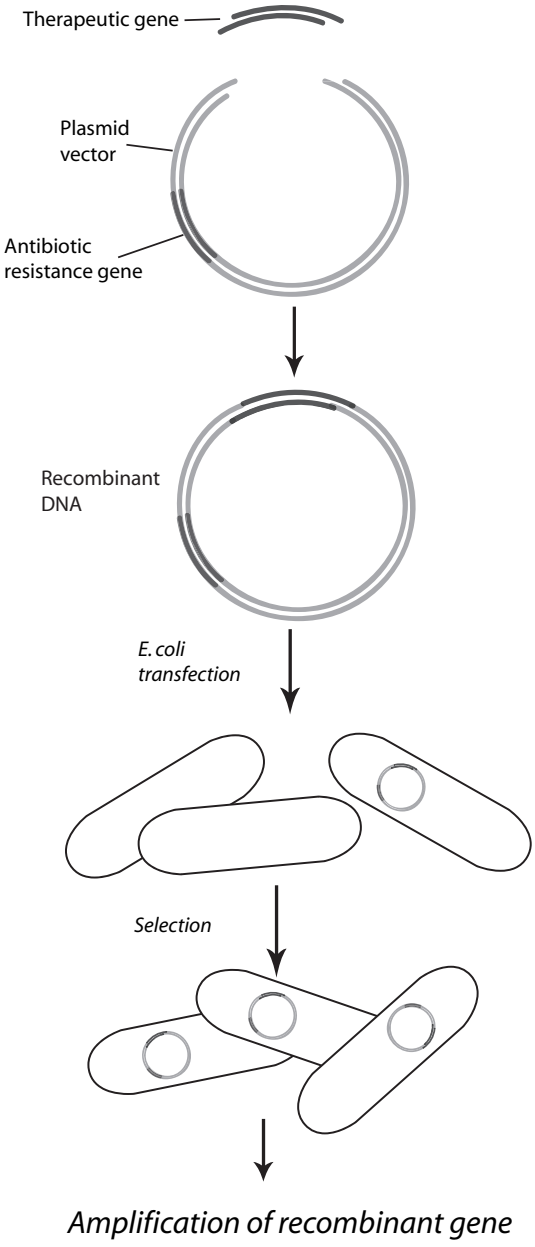


Figure 10.7. Construction of a therapeutic gene vector. Based on bibliography 10.2.

mutant protein-induced pathological disorders. Investigations for the past decade have provided tremendous evidence for the potential application of gene therapy to human disorders.

There are two general approaches that can be potentially applied to human gene therapy: embryonic gene transfer and somatic gene transfer. For *embryonic gene transfer*, a blastocyst is collected from a pregnant animal, and a therapeutic gene is injected into

the cells of the blastocyst to induce site-specific gene integration or homologous gene recombination. Embryonic cells from the blastocyst are stem cells. Once these stem cells carry the therapeutic gene, the gene can be passed to differentiated cells. Gametes may also carry the therapeutic gene and pass it to next generations. The transferred gene may function in place of the mutant gene. This is potentially an effective approach for the correction of pathological disorders due to hereditary gene defect or mutation. However, procedures for blastocyst collection induce cell injury and death. A safe, reliable, and effective technique has not been established for the treatment of hereditary human disorders. A technically similar technique has been used to modulate or remove selected genes in the embryonic stem cells of animals for the establishment of transgenic models. These models have been used extensively in scientific research.

Somatic gene transfer is an approach used for transferring therapeutic genes into the somatic cells of humans and animals after birth. Obviously, it is impossible to transfer a gene into all somatic cells in the body. Thus, this approach is used for gene transfer into selected target cells. Although mammalian cells are capable of taking up DNA, a process known as *endocytosis* or "naked gene transfer," the rate of DNA taking up is very low. A number of gene transfer-mediation methods, including virus-, liposome-, receptor-, electroporation-, salt-mediated gene transfer, have been established to facilitate gene transfer into mammalian cells. Each of these gene transfer strategies has strengths and weaknesses. The therapeutic efficacy and effectiveness differ between these mediation methods. The viral approach usually results in a higher rate of gene transfection and longer period of gene expression than do the nonviral approaches. However, viral carriers may cause infectious disorders and gene mutation in transfected cells. In contrast, the nonviral approaches are safe and easy to carry out, and exhibit low immunogenicity. However, their efficacy is relatively low and the expression duration of the transferred gene is short compared to the viral approach.

The choice of a gene transfer approach may also be dependent on the structure, function, and accessibility of the target tissue or organ. For instance, electroporation may be applied to the skin and skeletal muscle cells for gene transfer, but not to the heart and brain, because these organs are difficult to access and electroporation causes cell injury and death. In contrast, viral gene carriers and liposome can be used for gene transfer into target cells in most types of tissues and organs. Furthermore, various cell types may possess different capabilities of accepting foreign genes, and the cell state may also influence the rate of gene transfer for a given type of gene transfer mediation approach. For instance, retroviral gene carriers can infect dividing cells, but not nondividing cells, whereas adenoviral gene carriers can infect both dividing and nondividing cells. However, the retroviral approach may induce more stable gene expression than the adenoviral approach. Thus, all possible factors should be taken into account for the choice of a gene transfer approach. Here, several common gene transfer mediating approaches are briefly discussed.

Virus-Mediated Gene Transfer. Several types of virus, including retrovirus, adenovirus, adeno-associated virus, herpes simplex virus, have been used as gene carriers for mediating gene transfer into mammalian cells. The basis for virus-mediated gene transfer is that viruses are capable of infecting mammalian cells, carrying therapeutic genes, and integrating the carried genes into the host cell genome. Viruses can be modified or genetically engineered to remove harmful components and accommodate desired therapeutic gene fragments. Once transferred into the cell, viruses are able to integrate their genome into

the host genome and use the synthetic machineries of the host cells to produce viral DNA, RNA, and proteins. Thus, viruses are natural gene transfer carries. To be used for gene transfer mediation, a viral gene carrier ought to be replication-deficient, nonimmunogenic, and nontoxic, but are able to facilitate gene transfer to the genome of a target cell. The choice of a viral carrier for gene transfer is dependent on the nature of the virus and the state of the target cells, as we will see in the following sections.

RETROVIRUS-MEDIATED GENE TRANSFER. Retroviruses belong to a family of RNA viruses. A typical retrovirus contains two parts: the viral core and the envelope. The viral core is composed of two identical RNA strands and several enzymes, including reverse transcriptase, protease, and integrase. The envelope, composed of a membrane and glycoproteins, encloses the viral core. Common retroviruses used for gene transfer include Moloney virus and lentivirus. The Moloney virus is a murine leukemia virus that causes lymphoid leukemia in mice. The lentivirus is a retroviral strain that causes maedi (a chronic pulmonary disease found in Iceland) and visna (a disease affecting the central nervous system and causing paralysis) in sheep.

Retroviruses can interact with mammalian cells through cell membrane receptors, and enter the cell through receptor mediation. In the cytoplasm, the viral RNA genome can be converted into DNA by the reverse transcriptase. The converted DNA can be integrated into the host genome and replicate together with the host DNA during cell division. Thus, genetic information of the virus can be transmitted to the genome of the host cell and carried to the next generation. Any foreign genes that are inserted into the retroviral genome can be integrated into the host genome. The virus itself can produce viral proteins based on the viral mRNA and reproduce the same type of virus.

The RNA of the retrovirus can be engineered through a packaging process, generating viral vectors. Briefly, retroviral particles can be modulated to remove a selected RNA sequence known as the *packaging signal*, which is responsible for packaging the RNA core into the viral envelope. When transfected into a host cell line, the viral particles can express functional genes, such as env, gag, and pol. The env and pol genes encode viral envelope proteins, whereas the gag gene encodes the reverse transcriptase. While the virus is able to convert its RNA to DNA and produce necessary proteins for the construction of the viral envelope, it cannot pack the core RNA into the envelope. At the same time, another group of viral particles is selected and engineered to remove selectively the harmful sequences without modulating the packaging signal. A therapeutic gene fragment can be inserted into the viral genome to generate a recombinant viral gene carrier. When the recombinant viral gene carrier is transfected into the same host cell line containing the empty viral envelopes, the recombinant gene carrier can be packed into the empty envelopes under the action of the packaging signal, thus generating a complete viral structure that contains the therapeutic gene without the harmful gene sequences. This engineered viral structure is harmless and nonreplicable, and can be used to for gene transfer. Retrovirus-derived gene carriers have been used in experimental models and clinical trials.

There are several advantages for retrovirus-derived gene carriers. These carriers can induce a high rate of gene integration into the host cell genome with relatively stable and long-term expression (several months). However, retroviral gene carriers exhibit several disadvantages, including toxicity and limitation of effectiveness to only dividing cells. In particular, retroviral vectors may be potentially tumorigenic. The integration of a viral gene into the host cell genome may induce mutation of tumor suppressor genes or

oncogenes, potentially leading to tumorigenesis. Furthermore, viral genes can be expressed to generate viral proteins, which potentially cause host cell immune reactions, a potential factor that reduces the stability of the transfected gene and the duration of gene expression.

ADNOVIRUS-MEDIATED GENE TRANSFER. Adenoviruses are double-stranded DNA viruses and belong to a family composed of more than 30 serotypes. Many of the serotypes invade human cells and cause infectious disorders of the upper respiratory tract, known as *cold* or *flu*, and conjunctivae. Others infect the simian, bovine, canine, avian, and murine species. The mechanisms by which the adenovirus enters mammalian cells and causes infection remain poorly understood. It has been hypothesized that cell membrane receptors specific to adenoviral proteins may mediate the invasion of viruses (Fig. 10.8).

The genome of adenoviruses can be modified to remove replication control sequences and thus to prevent adenoviral replication and to eliminate toxic influence. The DNA genome of a typical adenovirus contains several regions, including E1 to E4. The E1 region is responsible for viral replication. The replacement of the E1 region with a foreign gene results in the generation of replication-defective viruses. Such engineered adenoviruses can be used as a gene transfer carrier. There are several advantages for the use of adenoviruses in gene transfer. First, the adenoviral gene carrier can induce a relatively high efficiency of gene transfection compared to a retroviral carrier. Second, an adenoviral carrier can infect both dividing and non-dividing cells. Third, an adenoviral vector can carry large DNA inserts [≤8 kb (kilobases)]. However, gene transfer mediated by the adenoviral vector often results in a relatively low rate of gene integration into the host cell genome and a short duration of gene expression (about several weeks). Although the toxicity of adenoviruses is lower than that of retroviruses, the viral particles can still induce infectious disorders in the host cells. Furthermore, the expressed viral proteins can induce host immune reactions, a potential factor that reduces the stability of the transfected gene and the duration of gene expression.

ADENO-ASSOCIATED VIRUS AS A GENE CARRIER. Adeno-associated virus belongs to a family of nonpathogenic human parvoviruses with a single-stranded DNA genome. This viral type is smaller than other types of viruses and cannot replicate unless associated with helper viruses, usually adenoviruses or herpesviruses. In the absence of helper viruses, the adeno-associated virus usually integrates its genome into the host chromosomal DNA and enters a latent state. Adeno-associated viruses can infect dividing as well as nondividing cells, and cause stable and long-term gene expression. Compared to the

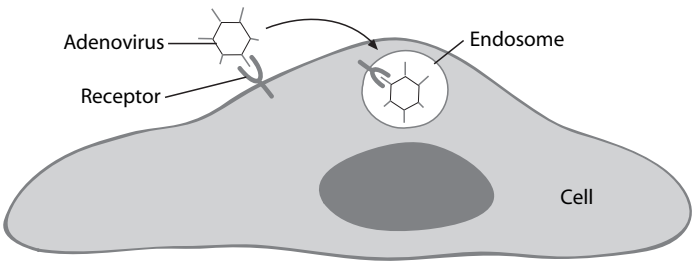


Figure 10.8. Schematic representation of adenovirus-mediated gene transfection.

adenovirus, the adeno-associated virus exhibits reduced immunogenicity and toxicity, since this virus consists of fewer genes that encode proteins harmful to the host cells. In a typical adeno-associated viral vector, the only wildtype sequences left are the inverted terminal repeats (ITRs) that flank the inserted gene cassette. These sequences are necessary for the packaging of the viral genome. Because of the lack of the replication capability, it is not necessary to modulate the genome of the adeno-associated virus for gene transfer. A desired therapeutic gene can be directly inserted into the viral genome and the viral vector can be used for gene transfer (Fig. 10.9). This type of virus has been extensively used for gene transfer in experimental models. However, adeno-associated viral vectors can accommodate relatively small gene inserts because of its limited genome size.

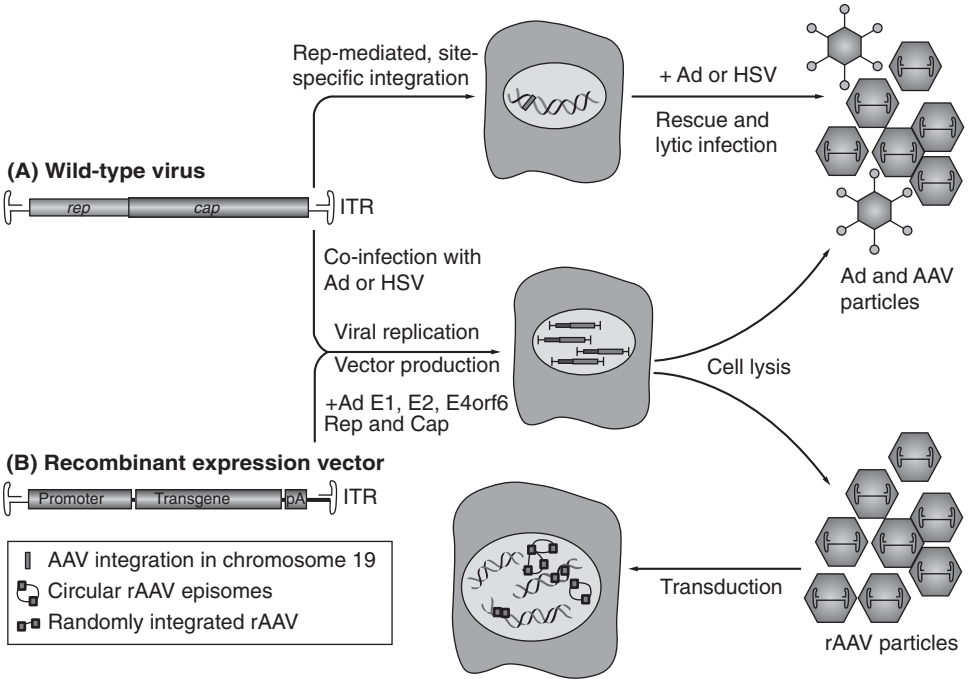


Figure 10.9. Adeno-associated virus as a vector for gene transfection. (A) The structure of the wildtype adeno-associated virus (AAV) is shown. A single-stranded DNA genome is encompassed by palindromic inverted terminal repeats (ITRs). The *rep* open reading frame (ORF) encodes proteins that are involved in viral replication, and the *cap* ORF encodes proteins that are necessary for viral packaging. AAV integrates into the human genome at a specific locus on chromosome 19 (red) and persists in a latent form. It can exit this stage only if the cell is co- or superinfected with helper virus such as adenovirus (Ad) or herpes simplex virus (HSV), which provide factors necessary for active AAV replication. (B) The generic genedelivery vector based on AAV is depicted. The viral genome is replaced by an expression cassette, which usually consists of a promoter, transgene, and polyA (pA) tail. For production of the recombinant virus (rAAV), Rep and Cap proteins as well as Ad or HSV elements (Ad E1, E2 and E4, or f6) have to be provided in *trans*. Examples of intracellular forms of the delivery vector that are responsible for transgene expression following transduction with rAAV (double-stranded circular episomes and randomly integrated vector genomes) are depicted in red. (Reprinted by permission from Macmillan Publishers Ltd.: Vasileva A, Jessberger R: *Nature Rev Microbiol* 3:837–47, copyright 2005.)

The efficiency of gene transfer mediated by adeno-associated viruses is usually lower than that mediated by adenoviruses.

HERPES SIMPLEX VIRUS-MEDIATED GENE TRANSFER. Herpes simplex virus belongs to a family of double-stranded DNA viruses that cause herpes simplex in humans. The virus can also be found in other mammalian species, such as canine, bovine, and simian. The virus can invade mammalian cells, integrate its DNA into the host genome, and mature in the host nucleus. This type of virus has been used for mediating gene transfer in experimental models. Herpes simplex virus-derived gene vectors can carry large gene fragments, infect dividing as well as non-dividing cells, and induce efficient gene transfection. However, the duration of gene expression mediated by herpes simplex viruses is often shorter (about one week) than that mediated by retroviral and adenoviral vectors. A possible reason is that herpes simplex viruses induce cytotoxic and/or immune reactions in the host cells, which shut down the expression of the transfected gene. A herpes simplex viral vector lacking the harmful genes, such as the immediately early genes, exhibits much reduced toxicity, improved gene transfection efficacy, and prolonged gene expression, as demonstrated in experimental models.

Receptor-Mediated Gene Transfer. Certain types of cell membrane receptors can be used to mediate gene transfer. A typical example is the transferrin receptor, which interacts with transferrin. The transferrin molecules can spontaneously link to a polymeric linker, such as polylysine, which can also link to DNA fragments at different sites, forming DNA–polylysine–transferrin complexes. These complexes can adhere to the cell membrane through the interaction of transferrin with the transferrin receptor (Fig. 10.10). Such an interaction activates the endocytosis mechanism of the target cells, which in turn take up the triplet DNA complexes. The transferred DNA complexes can be transported from the cytoplasm to the nucleus via endosomes.

Although the receptor-mediated gene transfer approach has been successfully used to transfer genes into mammalian cells, the transfer efficiency is low because most DNA molecules are trapped in the endosomes and cannot be released before degradation. To reduce DNA degradation and facilitate gene transport, a codelivery of replication-defective adenoviruses can improve the efficiency of receptor-mediated gene transfer. The mechanism for this enhancement is that endocytosed adenoviruses can disrupt the endosome membrane. Thus, the cotransferred triplet DNA complexes can be released. In this preparation, the DNA fragments are not inserted into the genome of the adenovirus. The virus is merely used as a gene transfer helper.

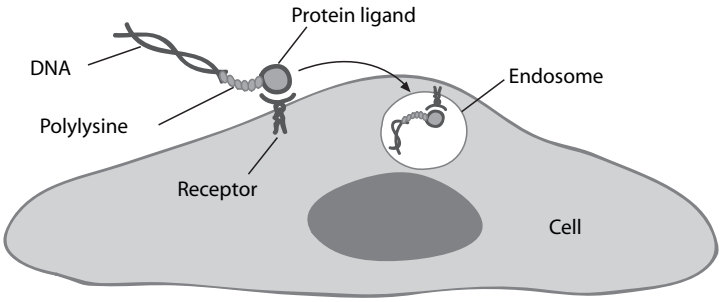


Figure 10.10. Schematic representation of receptor-mediated gene transfection.

Liposome-Mediated Gene Transfer. Liposomes are lipid vesicles of ~200 nm in diameter. Lipid molecules can spontaneously form multilaminar vesicles, which can be converted to smaller uni-laminar vesicles by sonication. Each liposome contains about 2500 lipid molecules. Each lipid molecule is composed of a hydrophilic and hydrophobic end. When liposomes are mixed with a water-based fluid, the hydrophilic end of the lipid molecule interacts with the water molecules and the hydrophobic end is enclosed inside. When liposomes and DNA molecules are mixed, liposome–DNA complexes can form spontaneously. The interaction between positive and negative charges is responsible for the complex formation. Some types of liposome, such as lipofectin (*N*-[1-(2,3-dioleyloxy)propyl]-*N*,*N*,*N*-trimethylammonium chloride), derivatives of cholesterol and diacyl glycerol, and lipopolyamines, are positively charged at the hydrophilic heads on the external surface of the vesicle. DNA molecules are negatively charged at the phosphate groups. The positively charged liposomes can form complexes with the negatively charged DNA molecules. There exist also negatively charged liposomes. These liposomes may repel negatively charged DNA molecules and are not suitable for mediating gene transfer. However, some negatively charged liposomes may entrap DNA molecules within the lipid vesicles to form DNA–liposome complexes. The negatively charged liposomes are less efficient in gene transfer than positively charged liposomes.

Liposomes have long been used for mediating gene transfer into mammalian cells. Although the transfection mediated by liposomes is not as effective as that by viral vectors, the efficiency of gene transfection is acceptable. There are several mechanisms underlying liposome-mediated gene transfer. First, liposome–DNA complexes can bind to the lipid cell membrane. Cationic liposomes attach to the cell membrane more efficiently than anionic liposomes, because the cell membrane is more negatively charged. Attached liposome–DNA complexes can be either endocytosed or fused into cells (Fig. 10.11). The endocytosed liposome–DNA complexes are transported and released into cytoplasm by endosomes. Some liposomes may be degraded by lysosomal enzymes before being released into cytoplasm. Second, liposome-DNA complexes can diffuse directly through the nucleus membrane and enter the cell nucleus.

Calcium Phosphate-Mediated Gene Transfer. Calcium phosphate can bind to DNA to form complexes that facilitate precipitation of DNA molecules to the cell surface. The interaction of calcium phosphate–DNA complexes with the cell membrane triggers cell endocytosis, by which DNA molecules are taken up and transported into the cytoplasm. A fraction of the endocytosed DNA can reach the nucleus, while the remains are degraded in the cytoplasm. There are several advantages for the use of calcium phosphate: (1) calcium phosphate is a physiological ingredient and is harmless to the host cells—calcium

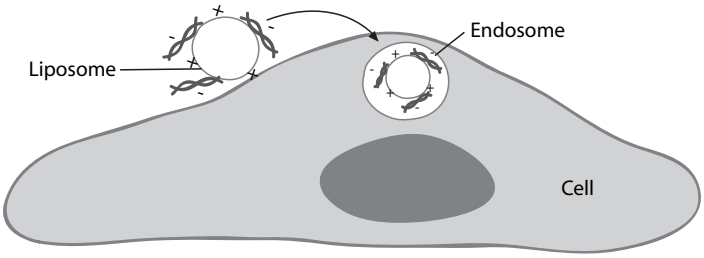


Figure 10.11. Schematic representation of liposome-mediated gene transfection.

phosphate is probably the safest substance among those used for mediating gene transfer and (2) calcium phosphate is easy to prepare and use. The only procedures involved in gene transfer are preparation of a calcium phosphate buffer supplemented with DNA and delivery of the DNA mix to target cells. However, the efficacy of calcium phosphate-mediated gene transfer is generally lower than that mediated by viral vectors and liposomes.

Electroporation-Mediated Gene Transfer. Electroporation is electrophysical process that induces the formation of hydrophilic pores in the cell membrane under the influence of an electric field. When cells are subject to an electric field between two electrodes, an electric pulse of appropriate voltage can induce the formation of pores in the cell lipid membrane. The rate of pore formation and the size of the pores are dependent on the maximal voltage across the cell membrane. The higher is the voltage, the higher is the rate of pore formation and the larger is the pore size. The duration of the pore opening is proportional to the duration of the electric pulse delivered to cells. The duration of electrical pulse may also affect the pore size. The longer is the duration of the electrical pulse, the larger is the pore size. Pores about 10 nm in diameter can be achieved in mammalian cell membranes by selecting an appropriate voltage. Short pulses usually result in a uniform population of small pores with a short duration of pore opening.

The process of electroporation is transient and reversible. After the termination of an electric pulse, the cell membrane pores shrink gradually and sealed eventually within several minutes. During the period of pore opening, DNA fragments present in the buffer can enter the cytoplasm and migrate into the cell nucleus. There are several possible mechanisms for the entrance of DNA fragments into the cytoplasm: (1) DNA fragments may directly move into the cytoplasm through the cell membrane pores, (2) DNA may be transported into the cytoplasm by osmotic forces established with respect to the concentration gradient of the DNA molecules, (3) electroporation may enhance endocytosis of the DNA fragments attached to the cell membrane, and electroporation may impose electrophoretic influence on the DNA fragments, enhancing DNA transport into the cytoplasm.

Electroporation is a highly efficient approach for gene transfer and has been extensively used in gene transfer for cultured cells. It has also been used for in vivo gene transfer into cells in accessible tissues and organs, such as the skin and skeletal muscle (Fig. 10.12). The efficacy of gene transfer mediated by electroporation is usually higher than that mediated by liposome and calcium phosphate, but lower than that mediated by viral vectors. A significant drawback for electroporation is cell injury induced by electrical pulses. In cell culture models, a standard electrical pulse (1000 V/cm in voltage), which causes effective gene transfer, may induce injury or death in about 30–50% of cells. Electroporation also causes cell injury and death for in vivo gene transfer. Thus, the voltage level for electroporation should be optimized to minimize cell injury through trial and error experiments for each specific type of cell, tissue, and organ.

Gene Gun-Mediated Gene Transfer. This is approach mediated by gene particle bombardment. Microparticles coated with the gene of interest can be accelerated by a force (e.g., a pressure) to penetrate through the cell membrane and thus to deliver the gene into the cytoplasm or the cell nucleus. The force used for gene transfer can be controlled to achieve an appropriate distance of penetration. This technique is easy to use and is efficient compared to other gene transfer-mediating approaches. However, bombardment induces cell injury and death.

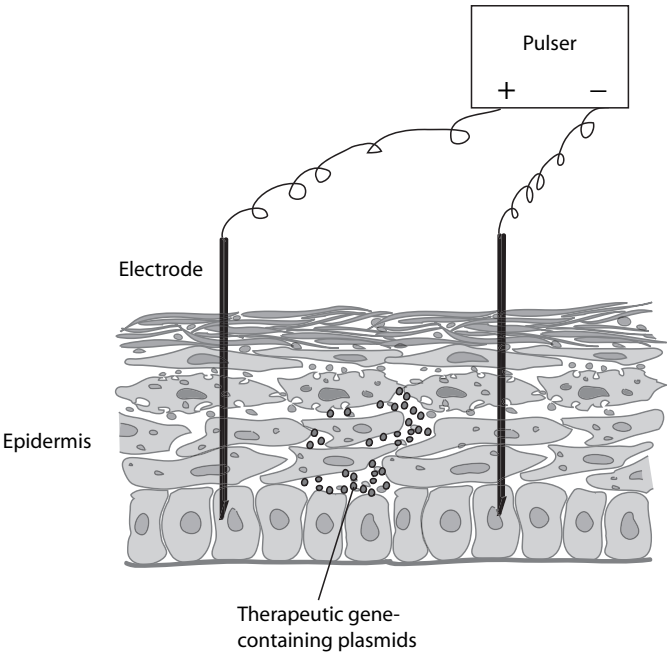


Figure 10.12. Schematic representation of electroporation-mediated gene transfection.

Assessing the Expression of the Transfected Gene. One of the most important steps in molecular engineering is to assess the expression of the transferred gene. Reporter genes are often used for such a purpose. Reporter genes are recombinant genes that encode proteins exhibiting special signs for visualization and thus can be used as markers for assessing the level of gene expression. Major criteria for the construction of a reporter gene are that the genes or expressed proteins are nontoxic, do not influence the function of the host cell, can be identified, and are not expressed in the target cell. Several types of reporter genes have been established and used for gene transfer. These include the β-galactosidase (gal) gene, luciferase gene, chloramphenicol acetyltransferase (CAT) gene, and fluorescent protein genes.

The β-*galactosidase gene* encodes a protein enzyme known as galactosidase and is found in *E. coli* cells. β-Galactosidase is a hydrolase that catalyzes the hydrolysis of the terminal residues of β-galactoside, forming d-galactose. d-galactose can react with x-gal, a chemical compound, to form a blue-colored substance. This feature has been used for detecting the efficiency of gene transfer. The presence of the blue color in cells transfected with the β-gal gene indicates the expression of this gene. β-galactosidase is not expressed in mammalian cells and nontoxic to the host cells. It is commonly used as a reporter gene in experiments of gene transfer. The β-galactosidase gene can be cotransferred with a therapeutic gene into target cells. It is assumed that the expression of the β-galactosidase gene indicates the expression of the cotransferred therapeutic gene. Alternatively, the β-galactosidase gene can be inserted into a recombinant vector that carries a therapeutic gene. The expression of the β-galactosidase gene is usually associated with the expression of the therapeutic gene. The second approach is preferable for detecting the efficiency of gene transfer.

The *luciferase gene* encodes a protein enzyme that oxidizes luciferin and is found in certain types of fish and insects, such as firefly bugs. The oxidization reduces luciferin to a compound that emits fluorescence. The emitted fluorescence can be observed by using a fluorescence microscope. Alternatively, the intensity of the emitted fluorescence can be detected by spectrophotometry. Since the luciferase gene is not expressed in mammalian cells, this gene can be used as a reporter gene for detecting the efficiency of gene transfer in terms of the fluorescence emitted by the luciferase substrate. In a cell sample transfected with the luciferase gene, positive fluorescent emission from the cell sample in the presence of luciferin indicates the expression of the transferred luciferase gene. In contrast, the absence of fluorescence in the presence of luciferin indicates the failure of gene transfer. Alternatively, an antiluciferase antibody developed by using luciferase as an antigen can be used to detect the expressed enzyme by immunohistochemistry. Luciferase is harmless to mammalian cells and is commonly used for the assessment of gene transfer efficiency.

The *chloramphenicol acetyltransferase gene* encodes chloramphenicol acetyltransferase (CAT), an enzyme that catalyzes the transfer of a catyl group from acetylcoenzyme A to the 3′-hydoxy position of chloramphenicol, $C_{11}H_{12}Cl_2N_2O_5$, which is a broad-spectrum antibiotic derived from *Streptomyces venezuelae* (fungus-like bacteria). The gene is not present in mammalian cells and can be used as a reporter gene for detecting the efficiency of gene transfer. The CAT catalytic activity can be monitored by a liquid scintillation enzyme assay and used for assessing gene expression. For the liquid scintillation enzyme assay, cell extracts can be incubated in a reaction mix containing ^{14}C- or ^{3}H-labeled chloramphenicol and *n*-butyryl coenzyme A. When the CAT gene is expressed, the CAT transfers the *n*-butyryl group of the *n*-butyryl coenzyme A molecule to chloramphenicol, forming *n*-butyryl chloramphenicol. The reaction products can be extracted with a small volume of xylene. The *n*-butyryl chloramphenicol compound can be partitioned into the xylene phase, while unmodified chloramphenicol remains predominantly in the aqueous phase. The xylene phase is mixed with scintillant and counted with a scintillation counter for the radioactivity of ^{14}C or ^{3}H, which is conjugated to chloramphenicol. The presence of radioactivity indicates the expression of the transferred CAT gene.

There are natural *fluorescent protein genes* that encode proteins capable of emitting fluorescence. These genes are found in certain types of florescent fish and insects. One of the most commonly seen natural fluorescent proteins is the green fluorescent protein, which is encoded by the green fluorescence protein gene. Recombinant fluorescent genes that encode red and cyanine proteins have been artificially constructed by modulating the gene sequences of natural fluorescence protein genes. These genes can be used as reporter genes for assessing the efficiency of gene transfer. Two methods can be used for such a purpose: (1) a fluorescence protein gene can be cotransferred with a therapeutic gene into target cells and used to assess the expression of the therapeutic gene or (2) a fluorescent protein gene can be inserted into a recombinant vector that contains a functional or therapeutic gene (Fig. 10.6). The fluorescent gene can be expressed together with the therapeutic gene when transfected into target cells, thus indicating the efficiency of gene transfer. It is important to note that the fluorescent protein gene should be inserted into an appropriate site that does not influence the transcription of the therapeutic gene. Such a site is usually located at the end of the therapeutic gene. However, trial-and-error experiments should be carried out to search for a correct insertion site. An advantage by using the fluorescent protein gene is that fluorescent signals can be examined in living cells, allowing the

observation of dynamic changes in molecular action. Thus, fluorescent protein gene has been extensively used in cell biology research.

Assessing the Effectiveness of Gene Transfer. An important aspect of gene therapy is to test the effectiveness and efficacy of gene therapy. One of the critical issues is whether gene transfer restores the physiological function of the target gene, cell, tissue, and organ, and corrects pathological alterations due to genetic disorders. The design of a test is largely dependent on the function of the target gene as well as the structure and function of the target cells. For instance, for a gene that encodes an enzyme, such as a matrix metalloproteinase, the level and activity of the enzyme can be detected before and after the gene transfection. The level of the enzyme can be assessed by immunoblotting analysis, and the function of the enzyme can be evaluated by detecting the level of the substrate modification. For a gene that encodes a protein responsible for regulating a specific cell activity, such as smooth muscle cell contractility, the level of the protein and the degree of cell activity can be assessed before and after the gene transfection. Similarly, the level of cell proliferation and apoptosis can be measured to assess the effectiveness of delivered genes that encode proteins for the regulation of cell mitogenic and apoptotic activities, respectively. Regardless of the gene type, a significant phenotype change in response to gene transfection suggests the effectiveness of the therapeutic gene.

Potential Negative Effects of Gene Transfer. Gene transfer may not be always beneficial and may potentially induce harmful effects:

1. Gene transfer requires the mediation of gene carriers, which are often toxic (liposomes), traumatic (electroporation and bombard), or infectious (adenovirus and retrovirus). These mediating approaches may negatively influence the function of host cells.
2. It is difficult to control the gene insertion sites in the target genome. Gene transfer by any mediating means may potentially cause gene insertion into incorrect sites in the genome, resulting in host or guest gene mutation, upregulation, or downregulation.
3. A therapeutic gene may be suitable for certain physiological functions, but may negatively influence other functions. A typical example is the genetic manipulation of the *ras* gene, which encodes the Ras protein, a critical molecule for the transduction of mitogenic signals and for the regulation of cell survival and proliferation. Because of its proproliferative effect, the *ras* gene has been considered a potential gene target for the treatment of atherosclerosis and cardiovascular hypertrophy. However, the suppression of the *ras* gene, while potentially reduces the proliferative activity, inevitably causes a negative effect on other functions mediated by this protein such as cell survival. Thus, the potential negative effects of gene transfer should be taken into account in the design of a molecular engineering approach.

HOMOLOGOUS RECOMBINATION [10.3]

Homologous recombination is a natural process by which DNA damages, such as doublestrand breaks (DSBs) and interstrand crosslinks, are repaired by generating a functional DNA copy with the homologous DNA sequence of the intact sister chromatin as a template

and replacing precisely the damaged DNA sequence. This process can completely restore the structure and function of the damaged DNA. Thus, homologous recombination is an ideal genetic approach for the correction of a mutant gene and can be utilized for therapeutic purposes.

Homologous recombination is originally discovered in yeast and *E. coli*. In mammalian cells, homologous recombination is rare, but is significantly increased in frequency when DNA double-strand breaks occur. Thus, one strategy to induce homologous recombination in mammalian cells is to introduce double strand breaks to desired target genes, which are responsible for the disorder of interest. The natural homologous recombination machinery in the cell is capable of recognizing the double-strand breaks on the target gene and initiating homologous gene recombination or site-specific gene replacement, resulting in the restoration of the structure and function of the mutant gene.

An effective approach for introducing DNA double strand breaks to a target gene is to deliver specific zinc finger nucleases to cells or tissues. Zinc finger nucleases are proteins that can recognize and cut target DNA at specific sites, induce double-strand breaks, and enhance site-specific homologous recombination. A typical zinc finger nuclease is composed of characteristic zinc finger domains. Each zinc finger domain consists of about 30 amino acids and can bind a zinc ion, which is critical to the stability and function of the zinc finger protein. A zinc finger protein can recognize and bind to a specific target gene by interacting with the major groove of the DNA double helix. The catalytic domain of the zinc finger nuclease can then cut the double strand of the bound DNA, resulting in double strand breaks.

Zinc finger nucleases can be designed and constructed with high specificity to desired target DNA sequences. To date, several types of three-finger zinc finger nucleases have been constructed with specificity to unique triplet DNA sequences, including 5′-ANN-3′, 5′-CNN-3′, 5′-GNN-3′, and 5′-TNN-3′, where N represents any nucleotide. Determination of the specificity of a zinc finger nuclease is based on the structure of the protein enzyme and the target triplet nucleotides.

For therapeutic purposes, a triplet DNA sequence, which is unique to a mutant gene and serves as a specific binding site for a zinc finger nuclease, can be identified from a cell type. A zinc finger nuclease can be designed, constructed, and delivered to the target cells. The zinc finger nuclease can recognize the specific binding site and induce double-strand breaks in the target mutant gene. The double strand breaks can initiate site-specific homologous recombination that replaces the mutant gene with a corresponding functional gene. Thus, the induction of the artificial double strand breaks and homologous recombination represent potential therapeutic approaches for the molecular treatment of genetic disorders.

ANTISENSE OLIGONUCLEOTIDE-BASED THERAPY [10.4]

Antisense oligonucleotide therapy is to design and use a DNA or RNA oligonucleotide sequence to hybridize complementary target mRNA, which is considered the “sense” sequence, and thus to block the translation of a specific protein, which is involved in the initiation and development of the disorder of interest. An example is the use of antisense oligonucleotides to downregulate the expression of angiotensinogen, thus to reduce the level of angiotensin I and II, for the treatment of hypertension (see page 699 for mechanisms). Furthermore, the hybridization of the target mRNA with an antisense

oligonucleotide may induce the activation of an enzyme known as *ribonuclease* H (RNase H), which cleaves the hybridized mRNA. These approaches can be used to effectively downregulate protein translation, providing a therapeutic means for the treatment of disorders. An antisense DNA or RNA sequence specific to a selected mRNA molecule can be designed and constructed. The constructed antisense oligonucleotides can be delivered to target cells on the basis of endocytosis, a natural process that takes up small particles on the cell surface. Several gene transfer-mediating methods, such as liposome- and electroporation-mediated transfer, can be used to enhance the delivery of the antisense oligonucleotides. The antisense approach can be potentially used for therapeutic purposes by repressing the translation of specific proteins, which are involved in the pathogenesis of disorders.

SMALL INTERFERING RNA-BASED THERAPY [10.5]

Small interfering RNA (siRNA), also known as *short interfering RNA*, is a short RNA sequence, which is specific to and can hybridize to a target mRNA, induce the degradation of the target mRNA, and thus knock down the expression of the encoded protein. The process of siRNA-induced mRNA degradation is referred to as RNA interference, posttranscriptional gene silencing, or transgene silencing. This mechanism was originally discovered in petunia plant cells and *Caenorhabditis elegans*. Further investigations have demonstrated that mammalian cells also exhibit RNA interference. Since RNA interference can be effectively used to suppress the translation of specific proteins, siRNA can be potentially used as therapeutic agents for pathological disorders that involve abnormal expression of specific proteins.

A siRNA molecule is generated under the action of an enzyme, known as "dicer," which is an RNAse III molecule composed of a helicase domain, two RNAse III domains, a double-stranded RNA (dsRNA)-binding domain, and a PAZ domain (Fig. 10.13). The dicer enzyme can splice dsRNA into short siRNA. A typical siRNA is a double-stranded RNA about 21 nucleotide in length, and consists of a 5′ phosphate overhang at one end and a 3′ hydroxyl overhang at the other end. A siRNA fragment can form a complex with a multiprotein complex known as the *RNA-induced silencing complex* (RISC), which binds to a specific sequence of the siRNA and unwind the siRNA into single strands. The single-stranded siRNA can hybridize to complementary mRNA in the cytoplasm. The RISC complex is activated and can degrade the substrate mRNA bound by the siRNA, thus suppressing protein translation.

In mammalian cells, there is a RNA interference machinery, which has been evolved as a defense mechanism against the invasion of retroviruses. The RNA interference machinery can recognize and separate double-stranded RNA molecules into two single strands to form siRNAs. The siRNA molecules can recognize and recruit RNases, which degrade mRNA transcripts complementary to the siRNA molecules. On the invasion of retroviruses, siRNA sequences are generated and utilized for the destruction of the invaded viruses. This mechanism is also used for posttranscriptional gene regulation in mammalian cells. During gene transcription, certain RNA transcripts can fold to form hairpin-like double-stranded RNA structures, sometimes referred to as *microRNA*, which cannot be translated into proteins. The RNA interference machinery can detect and destroy these microRNA structures by generating siRNAs. Since the RNA interference machinery may not be able to completely remove target mRNAs, siRNA-mediated mRNA degradation is also referred to as gene knockdown.

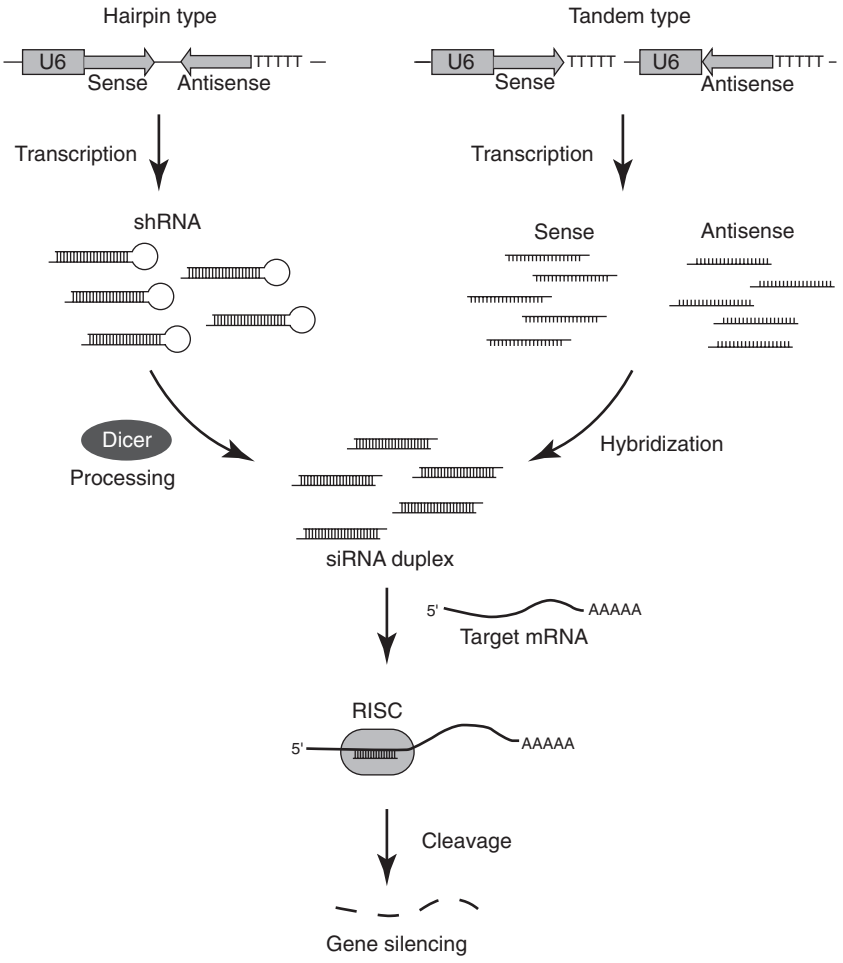


Figure 10.13. Schematic representation of a transcription system for production of siRNA. (Reprinted by permission of the Federation of the European Biochemical Societies from Itoa M et al: *FEBS Lett* 579:5988–95, copyright 2005.)

For therapeutic purposes, specific siRNA molecules can be synthesized in vitro and applied to target cells or tissues for inducing degradation of target mRNA. To synthesize a siRNA sequence, it is necessary to determine the target mRNA sequences, on which the siRNA acts, based on the sequence of a specific gene. The effective target mRNA sequences can be selected by using siRNA design programs provided by commercial carriers, such as Promega. Desired siRNA sequences can be synthesized by using an oligonucleotide synthesizer. The synthesized siRNA can be used for cell transfection. A liposome-mediated transfection method can be used to facilitate siRNA transfection. For experimental purposes, it is necessary to construct and use a siRNA control, known as "scrambled" siRNA. Such a siRNA is usually a 21-nucleotide sequence, which is nonspecific to any known mRNA. The control siRNA should be transfected to control cells simultaneously with the specific siRNA transfection. While siRNA transfection is

effective in knocking down gene expression, the effect is usually transient. It is necessary to conduct multiple transfection if long-term effect is required for experimental or therapeutic purposes.

To induce a long-term effect of RNA interference, a siRNA cloning vector can be constructed and used to transfect cells (Fig. 10.14). The transfected siRNA cloning vector contains a nucleotide sequence, which encodes a desired specific siRNA, and a gene expression vector, which induces gene expression when transfected into a cell. To construct a siRNA cloning vector, it is necessary to establish siRNA encoding gene sequence, which can be done by using a siRNA design program as described above. The siRNA-encoding gene sequence can be inserted into a cloning vector, such as the neomycin-resistant gene-containing psiSTRIKE cloning vector from Promega. The established siRNA cloning vector can be amplified by transfecting and growing E. coli cells. The amplified siRNA cloning vector can be purified and used for mammalian cell transfection. The following is an example of the siRNA cloning vector for the mRNA of the protein tyrosine phosphatase SH2 domain-containing protein tyrosine phosphatase 1 (SHP1).

The sequences for the SHP1 siRNA are as follows: 5′-ACC**GAAAGGCCGGAACA AAT GTGT***TTCAAGAGA**ACACATTTGTTCC GGCCTTTC***TTTTTC-3′ and 5′-TGCAGA AAA***AGAAAGGCCGGAACAAATGTGT** *TCTCTTGAA***ACACATTTGT TCCGGCC TTT**-3′. In these sequences, the boldfaced fragments represent the target

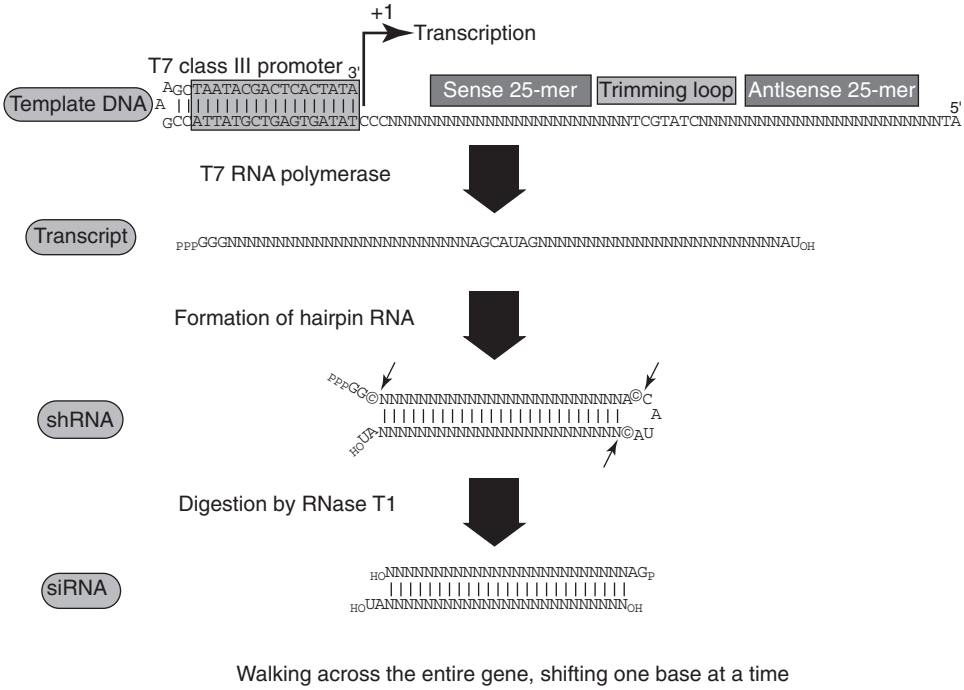


Figure 10.14. Schematic representation of gene silencing by a shRNA expression vector. The shRNA is transcribed from a shRNA cloning vector and processed by Dicer to produce siRNA. The processed siRNA enters the RNA-induced silencing complex (RISC), where it targets mRNA for degradation. (Reprinted by permission of the Federation of the European Biochemical Societies from Itoa M et al: *FEBS Lett* 579:5988–95, copyright 2005.)

sequence, the italic boldfaced fragments represent the target reverse complement, the underlined fragments are for the mRNA loop, and the remainder are the overhang fragment (5′-ACC) and the U6 termination sequence (TTTTC-3′) for cloning purpose. The sequences for the control scrambled siRNA for SHP-1 are as follows: 5′-ACC**GAAGAT-GCGAAGGGATAC TAC***TTCAAGAGA**GTAGTATCCC T TCGCATCTTC***TTTTTC-3′ and 5′-TGCAGAAAAA***AGAAGATGCGAAGGGATACTA** *TCTCTTGAA* **GTAGTATCCC TTCGCATCTT**-3′. The SHP1 specific and scrambled siRNA sequences can be synthesized by a commercial carrier such as Proligo.

BIBLIOGRAPHY

10.1. Disorders Due to Gene Mutation

Kopp P, Jameson JL: Transmission of human genetic disease, in *Principles of Molecular Medicine*, Jameson JL, Collins FS, eds, Humana Press, Totowa, NJ, 1998, pp 43–58.

Schneider AS, Szanto PA: *Pathology*, 3rd ed, Lippincott Williams & Wilkins, Philadelphia, 2006.

Pasternak JJ: *An Introduction to Human Molecular Genetics*: *Mechanisms of Inherited Diseases*, 2nd ed, Wiley-Liss, Hoboken, NJ, 2005.

Killeen AA: *Principles of Molecular Pathology*, Humana Press, Totowa, NJ, 2004.

Watson JD, Berry A: *DNA: The Secret of Life*, Knopf, New York, 2003.

Pasternak JJ: *An Introduction to Human Molecular Genetics: Mechanisms of Inherited Diseases*, Fitzgerald Science Press, Bethesda, MD, 1999.

10.2. Principles of DNA Engineering

Schena M: *Microarray Analysis*, Wiley-Liss, Hoboken, NJ, 2003.

Grompe M, Johnson W, Jameson JL: Recombinant DNA and genetic techniques, in *Principles of Molecular Medicine*, Jameson JL, Collins FS, eds., Humana Press, Totowa, NJ, 1998, pp 9–24.

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

Kmiec EB: Gene targeting protocols, in *Methods in Molecular Biology*, Vol 133, Humana Press, Totowa, NJ, 2000.

Speicher DW, Reim DF: Microsequence analysis of electroblotted proteins, *Anal Biochem* 207:19–23, 1992.

Jackson PJ: N-terminal protein sequencing for special application, in *Methods in Molecular Biology*, Vol 211: *Protein Sequencing Protocols*, 2nd ed, Smith BJ, ed, Humana Press, Totowa, NJ, 2003, pp 287–300.

Steinke L, Cook RG: Identification of phosphorylation sites by Edman degradation, in *Methods in Molecular Biology*, Vol 211: *Protein Sequencing Protocols*, 2nd ed, Smith BJ, ed, Humana Press, Totowa, NJ, 2003, pp 301–8.

Jackson PJ et al: Validation of protein sequencing in a regulated laboratory, in *Methods in Molecular Biology*, Vol 211: *Protein Sequencing Protocols*, 2nd ed, Smith BJ, ed, Humana Press, Totowa, NJ, 2003, pp 309–18.

Dupont DR, Yuen SW, Graham KS: Automated C-terminal protein sequence analysis using the alkylated-thiohydantoin method, in *Methods in Molecular Biology*, Vol 211: *Protein Sequencing Protocols*, 2nd ed, Smith BJ, ed, Humana Press, Totowa, NJ, 2003, pp 319–31.

Harvey DJ: Identification of sites of glycosylation, in *Methods in Molecular Biology*, Vol 211: *Protein Sequencing Protocols*, 2nd ed, Smith B.J. ed, Humana Press, Totowa, NJ, 2003, pp 371–83.

Aitken A, Learmonth M: Analysis of sites of protein phosphorylation, in *Methods in Molecular Biology*, Vol 211: *Protein Sequencing Protocols*, 2nd ed, Smith BJ, ed, Humana Press, Totowa, NJ, 2003, pp 385–98.

Copley RR, Russell RB: Getting the most from your protein sequence, in *Methods in Molecular Biology*, Vol 211: *Protein Sequencing Protocols*, 2nd ed, Smith BJ, ed, Humana Press, Totowa, NJ, 2003, pp 411–30.

Shannon JD, Fox JW: Identification of phosphorylation sites by edman degradation, *Tech Protein Chem* VI:117–23, 1995.

Edman P: Method for determination of the amino acid sequence in peptides, *Acta Chem Scand* 4:283–93, 1950.

Burke DT, Carle GF, Olson MV: Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors, *Science* 236:806–12, 1987.

Cohen SN, Chang ACY, Boyer HW, Helling RB: Construction of biologically functional bacterial plasmids in vitro, *Proc Natl Acad Sci USA* 70:3240–4, 1973.

Nathans D, Smith HO: Restriction endonucleases in the analysis and restructuring of DNA molecules, *Annu Rev Biochem* 44:273–93, 1975.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT et al: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science* 239:487–91, 1988.

Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors, *Proc Natl Acad Sci USA* 74:5463–7, 1977.

Benton WD, Davis RW: Screening lambdagt recombinant clones by hybridization to single plaques in situ, *Science* 196:180–2, 1977.

Broome S, Gilbert W: Immunological screening method to detect specific translation products. *Proc Natl Acad Sci USA* 75:2746–9, 1978.

Brown PO, Botstein D: Exploring the new world of the genome with DNA microarrays, *Nat Genet* 21(Suppl 1):33–7, 1999.

Gerhold DL, Jensen RV, Gullans SR: Better therapeutics through microarrays, *Nat Genet* 32(Suppl):547–51, 2002.

Grunstein M, Hogness DS: Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene, *Proc Natl Acad Sci USA* 72:3961–5, 1975.

Maniatis T, Hardison RC, Lacy E, Lauer J, O'Connell C, et al: The isolation of structural genes from libraries of eucaryotic DNA, *Cell* 15:687–701, 1978.

Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis, *J Mol Biol* 98:503–17, 1975.

Schildkraut CL, Marmur J, Doty P: The formation of hybrid DNA molecules and their use in studies of DNA homologies, *J Mol Biol* 3:595–617, 1961.

Ruitenberg MJ, Eggers R, Boer GJ, Verhaagen J: Adeno-associated viral vectors as agents for gene delivery: application in disorders and trauma of the central nervous system, *Methods* 28:182–194, 2002.

Chan L, Fujimiya M, Kojima H: In vivo gene therapy for diabetes mellitus, *Trends Mol Med* 9:430–5, 2003.

Zhdanov R, Bogdanenko E, Moskovtsev A, Podobed O, Duzgunes N: Liposome-mediated gene delivery: Dependence on lipid structure, glycolipid-mediated targeting, and immunological properties, *Meth Enzymol* 373:433–65, 2003.

Bessis N, Doucet C, Cottard V, Douar AM, Firat H et al: Gene therapy for rheumatoid arthritis, *J Gene Med* 4:581–91, 2002.

Botstein D, Shortle D: Strategies and applications of in vitro mutagenesis, *Science* 229:1193–201, 1985.

Herskowitz I: Functional inactivation of genes by dominant negative mutations, *Nature* 329:219–22, 1987.

Robertson E, Bradley A, Kuehn M, Evans M: Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector, *Nature* 323:445–8, 1986.

Jaenisch R: Transgenic animals, *Science* 240:1468–74, 1988.

10.3. Homologous Recombination

Bronson SK, Smithies O: Altering mice by homologous recombination using embryonic stem cells, *J Biol Chem* 269:27155–8, 1994.

Capecchi MR: Altering the genome by homologous recombination, *Science* 244:1288–92, 1989.

Gordon JW, Ruddle FH: Integration and stable germ line transmission of genes injected into mouse pronuclei, *Science* 214:1244–6, 1981.

Durai S, Mani M, Kandavelou K, Wu J, Porteus MH et al: Zinc finger nucleases: Custom-designed molecular scissors for genome engineering of plant and mammalian cells, *Nucleic Acids Res* 33:5978–90, 2005.

Gommans WM, Haisma HJ, Rots MG: Engineering zinc finger protein transcription factors: The therapeutic relevance of switching endogenous gene expression on or off at command, *J Mol Biol* 354:507–19, 2005.

Durai S, Mani M, Kandavelou K, Wu J, Porteus MH et al: Zinc finger nucleases: Custom-designed molecular scissors for genome engineering of plant and mammalian cells, *Nucleic Acids Res* 33:5978–90, 2005.

Vasileva A, Jessberger R. Precise hit: Adeno-associated virus in gene targeting, *Nat Rev Microbiol* 3:837–47, 2005.

Aplan PD: Causes of oncogenic chromosomal translocation, *Trends Genet* 22:46–55, 2006.

Glaser S, Anastassiadis K, Stewart AF: Current issues in mouse genome engineering, *Nat Genet* 37:1187–93, 2005.

Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L: Chromosome evolution in eukaryotes: A multi-kingdom perspective, *Trends Genet* 21:673–82, 2005.

Porteus MH, Carroll D: Gene targeting using zinc finger nucleases, *Nat Biotechnol* 23:967–73, 2005.

Gerton JL, Hawley RS: Homologous chromosome interactions in meiosis: Diversity amidst conservation, *Nat Rev Genet* 6:477–87, 2005.

Klug A: Towards therapeutic applications of engineered zinc finger proteins, *FEBS Lett* 579:892–4, 2005.

Park Y, Gerson SL: DNA repair defects in stem cell function and aging, *Annu Rev Med* 56:495–508, 2005.

Pellegrini L, Venkitaraman A: Emerging functions of BRCA2 in DNA recombination, *Trends Biochem Sci* 29:310–6, 2004.

10.4. Antisense Oligonucleotide-Based Therapies

Izant JG, Weintraub H: Constitutive and conditional suppression of exogenous and endogenous genes by anti-sense RNA, *Science* 229:345–52, 1985.

Izant JG, Weintraub H: Inhibition of thymidine kinase gene expression by anti-sense RNA: A molecular approach to genetic analysis, *Cell* 36:1007–15, 1984.

Wagner RW: Gene inhibition using antisense oligodeoxynucleotides, *Nature* 372:333–5, 1994.

Poulaki V, Mitsiades CS, Kotoula V, Tseleni-Balafouta S, Ashkenazi A et al: Regulation of Apo2L/tumor necrosis factor-related apoptosis-inducing ligand-induced apoptosis in thyroid carcinoma cells, *Am J Pathol* 161:643–54, 2002.

Kawai H, Sato W, Yuzawa Y, Kosugi T, Matsuo S et al: Lack of the growth factor midkine enhances survival against cisplatin-induced renal damage, *Am J Pathol* 165(5):1603–12, Nov 2004.

10.5. Small Interfering RNA-Based Therapies

Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T et al: A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA, *Science* 293:834–8, 2001.

Hutvagner G, Zamore PD: RNAi: Nature abhors a double-strand, *Curr Opin Genet Dev* 12:225–32, 2002.

Novina CD, Sharp PA: The RNAi revolution, *Nature* 430:161–4, 2004.

Sharp PA: RNA interference—2001, *Genes Dev* 15:485–90, 2001.

Hutvagner G, Zamore PD: A microRNA in a multiple-turnover RNAi enzyme complex, *Science* 297:2056–60, 2002.

Napoli C, Lemieux C, Jorgensen R: Introduction of a chalcone synthase gene into Petunia results in reversible co-suppression of homologous genes in trans, *Plant Cell* 2:279–89, 1990.

Dehio C, Schell J: Identification of plant genetic loci involved in a post transcriptional mechanism for meiotically reversible transgene silencing, *Proc Acad Sci USA* 91:5538–42, 1994.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: Potent and specific genetic interference by double–stranded RNA in Caenorhabditis elegans, *Nature* 391:806–11, 1998.

Aronin N: Target selectivity in mRNA silencing, *Gene Ther* 13:509–16, 2006.

Rodriguez-Lebron E, Paulson HL: Allele-specific RNA interference for neurological disease, *Gene Ther* 13:576–81, 2006.

Bonini NM, La Spada AR: Silencing polyglutamine degeneration with RNAi, *Neuron* 48:715–8, 2005.

Uprichard SL: The therapeutic potential of RNA interference, *FEBS Lett* 579:5996–6007, 2005.

Stecca B, Mas C, Ruiz i Altaba A: Interference with HH-GLI signaling inhibits prostate cancer, *Trends Mol Med* 11199–203, 2005.

Bennink JR, Palmore TN: The promise of siRNAs for the treatment of influenza, *Trends Mol Med* 10:571–4, 2004.

Ryther RC, Flynt AS, Phillips JA 3rd, Patton JG: siRNA therapeutics: Big potential from small RNAs, *Gene Ther* 12:5–11, 2005.

Hannon GJ, Rossi JJ: Unlocking the potential of the human genome with RNA interference, *Nature* 431:371–8, 2004.

Caplen NJ: Gene therapy progress and prospects. Downregulating gene expression: The impact of RNA interference, *Gene Ther* 11:1241–8, 2004.

Stevenson M: Dissecting HIV-1 through RNA interference, *Nat Rev Immunol* 3:851–8, 2003.