

Molecular Biology

n.

... The branch of biology that deals with the manipulation of DNA so that it can be sequenced or mutated. If mutated, the DNA is often inserted into the genome of an organism to study the biological effects of the mutation.

Source: The American Heritage® Dictionary of the English Language, Fourth Edition
Copyright © 2000 by Houghton Mifflin Company.
Published by Houghton Mifflin Company. All rights reserved

Molecular Biology I: Fundamentals of Recombinant DNA Construction

I. Learning Objectives.

- A. Understand how recombinant DNA technology is accelerating advances that are leading to improved health care.
- B. Describe the basic methods for recombining a DNA of interest with a self-replicating vector and amplifying it in a host cell.
- C. For restriction enzymes describe:
 1. the sequence characteristics of the recognition/cleavage sites.
 2. the three different types of DNA "ends" produced by restriction enzymes.
- D. Textbook reading for reference and clarification — Lehninger, Chapter 29; Gelehrter, Chapter 5.

- II. Introduction.** Molecular biology is a fundamental advance of modern biological sciences that is defining the present and future of medical science. Recombinant DNA technology provides an efficient way to elucidate the molecular mechanisms of life. This technology is yielding a deep understanding of health and disease that includes the molecular rationale for the development of risk predictors, diagnostics and therapeutics.

New York Times
July 25, 2002

Bioterror's New Frontier

To the Editor:

Your restrained response to the molecular genetic synthesis of the polio virus in "Synthetic Bioterror" (editorial, July 18) puts this research accomplishment in proper perspective. The synthesis of substantially larger genomes is more difficult and is unlikely to constitute a real bioterror threat in the immediate future.

It is more likely that the next generation of bioterror agents will be "mix and match" organisms produced by common molecular biology and genetic techniques involving the introduction of elements encoding drug resistance or even vaccine resistance into receptive organisms like staph and anthrax.

Hospital-acquired infections already kill 100,000 people a year, and a recent report by the Centers for Disease Control and Prevention documenting the first case of a vancomycin-resistant staph strain from a Michigan patient is much more troubling than the potential genetic reconstitution of viruses like smallpox.

DAVID PERLIN
Newark, July 19, 2002
The writer is scientific director, Public Health Research Institute.

- A. **Advances that are making major contributions toward understanding, diagnosing and treating diseases.**
1. **Restriction enzymes** – enable dissection of huge DNA molecules into defined fragments that can be managed.
 2. **DNA □Cloning via recombinant DNA methods** – making many copies of single DNA fragments in order to purify them out of complex mixtures of DNA such as the genome.
 3. **Inexpensive Production of Synthetic DNA** (currently ~50 cents per base) for use as hybridization probes. Hybridization of probes □to target DNA is a precise indicator of the DNA sequence of the target DNA.
 4. **Polymerase Chain Reaction** is another method for amplifying specific DNA fragments. The amplified fragments can be used for nucleotide sequence determination.
 5. **Automatic DNA Sequencing** enables high throughput nucleotide sequence determination.
 6. **Microarray Technology** – Enables high throughput determination of □the mRNA present in different cell types and under different conditions (i.e., gene expression). Enables high throughput determination of single nucleotide polymorphisms (SNPs).

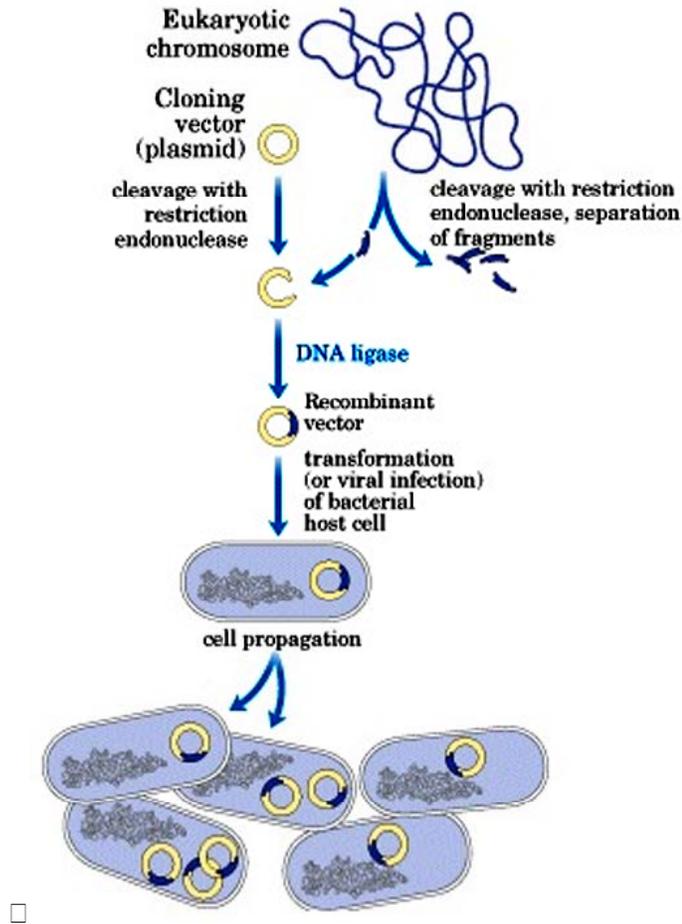
III. **Overview of basic recombinant DNA techniques.** DNA is easily cut (with restriction endonucleases) and pasted (using DNA ligase) to make chimeric/recombinant molecules (i.e., comprised of DNA from different sources). Typically, a DNA molecule of interest, called the *insert* is pasted together with a *vector*, a piece of DNA that enables the recombinant molecule to be replicated and harbored in the host organism. Recombinant DNA molecules are constructed for the purpose of cloning the DNA, which means making multiple copies of a single molecule. The recombinant molecule in one clone (i.e., one colony of the host organism) can be separated away from other DNA fragments by allowing a single cell to proliferate to form a colony or a large culture in liquid medium. Thus the DNA of the clone is *amplified* and *purified*.

IV. **The uses of recombinant DNA technology.**

- A. Obtaining single purified DNA species in the large quantities necessary for nucleotide sequence determination□. Without these techniques, current technology is not sensitive enough to determine, for example, the nucleotide sequence of a single DNA fragment (e.g., a region of a gene) within the vast sea of the three billion base pairs that comprise an individual's genome.
- B. Recombinant DNA technology has enabled the sequencing of the entire genomes of diverse species that include man, mouse, fly, worm, yeasts, plants, bacteria, and infectious organisms. Some of these organisms serve as models for human biology that enable experimentation and genetic tricks that are powerful for elucidating human gene functions. Others are models for foodstuff or are infectious organisms with healthcare significance. To determine the

nucleotide sequence of a genome, the genome is fragmented into many pieces. Each piece is recombined with a single type of *vector*. A *vector* is a piece of DNA that enables the DNA linked to it to be grown inside a *host* organism. Each genomic fragment/vector recombinant molecule is inserted individually into a different single cell of a host organism such as the bacterium *Escherichia coli* (see figure below). The **pool** of the single cells with each single cell containing a different genomic DNA fragment-vector is called a genomic DNA *library*. Each of the single cells can be expanded into a colony (also called a clone). [A culture grown from a single □ clone contains □ a single genomic DNA fragment (in the context of the vector) that is amplified by the growth expansion of the colony.] The purified DNA from the expanded culture is subjected to the methods of nucleic acid sequence determination. The entire nucleotide sequence of a genome can be obtained by determining the nucleotide sequences of all the inserts (genomic fragments) of the recombinant clones.

- C. Recombinant DNA and other biotechnology methods enable disease genes to be identified. DNA sequence, in conjunction with environmental factors, determines disease susceptibility. The gene variants that are currently known to cause diseases with certainty or to increase the likelihood of diseases are only the tip of the iceberg of those expected to be discovered. DNA sequence variants (*polymorphisms*) that cause disease are being identified through large clinical/laboratory research efforts that will continue into the foreseeable future..
- D. Recombinant DNA molecules are used for the synthesis of proteins in host organisms (e.g., *Escherichia coli*, yeast, mammalian cells). Recombinant DNA is engineered so that the host organism's DNA, RNA and protein synthesis machineries act on the recombinant molecule to produce the protein encoded by the cDNA. Purified recombinant proteins are used in biomedical research, diagnostics and as medicines.
- E. Recombinant molecules are used to create *transgenic* animals. The recombinant molecules replace the endogenous gene on the chromosome. Some recombinants are designed to *knock-out* genes to create mice (or other organisms) that lack specific genes. The developmental and biochemical abnormalities of the mice lacking specific genes provide important clues for inferring gene function. Recombinant molecules can also be engineered to □ replace chromosomal genes with mutated versions. *Transgenic mice with mutated genes* are powerful tools for learning about gene function.
- F. Recombinant molecules are used in gene therapy applications. The vectors enable DNA to be inserted into the cells and tissues of patients. Typically, the recombinant molecule is engineered to produce a functional version of a protein that may be inactive in the patient because of a mutation.



V. **Construction of a recombinant plasmids.** Plasmids are circular DNA molecules that replicate separately from the host genome. There are many naturally occurring bacterial plasmids.

A. **Restriction endonucleases**

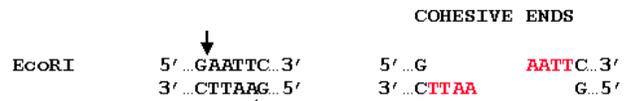
1. **Each DNA to be ligated** must be cut with enzymes that make *compatible ends* (generated by the restriction enzymes).
 - a) Any blunt end can be joined to any other blunt end.
 - b) Ends with overhanging, free, unpaired bases (protruding ends) can only be ligated to ends that have the same overhanging bases (i.e., *are complementary*).
2. **Restriction Endonucleases - background**
 - a) Cleave DNA at specific short DNA sequences (usually 4-7 bp) to generate specific DNA fragments ("restriction fragments").
 - b) Over 220 different DNA sequences are known to be recognized by the currently known restriction enzymes.
 - c) Enable huge DNA molecules to be cut into manageable fragments with defined ends.
 - d) Originate from bacteria. Restriction enzymes cleave invading viral genes "restricting" them from their genome. The bacteria modify (methylate) the restriction enzyme cleavage recognition/cleavage sites

of their own genomic DNA so it is not cleaved by their own restriction enzymes.

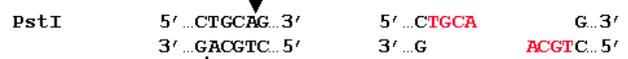
3. Restriction enzyme recognition sequences:

- a) *Usually palindromic* – the recognition/cleavage sequence when read in the 5' to 3' direction is the same on each of the two strands of duplex DNA.
- b) Restriction enzymes cut duplex DNA in one of three ways to produce different types of DNA ends at the cleavage site:

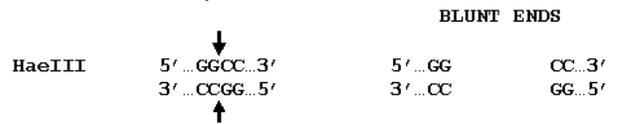
- (1) *5' protruding ends (cohesive)* – the cuts are staggered leaving protruding unpaired bases on the 5' end.



- (2) *3' protruding ends (cohesive)* – the cuts are staggered leaving protruding unpaired bases on the 3' end.



- (3) *blunt ends* – each strand of duplex DNA is cut directly across from each other.

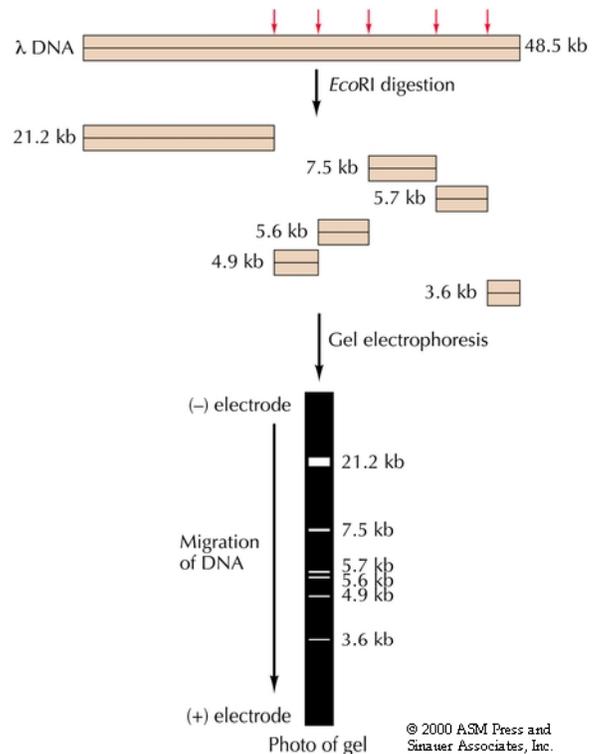


4. **Compatibility of ends** – Only DNA ends that are compatible can be ligated to each other using the enzyme *DNA ligase*.
 - a) *protruding ends (cohesive or also known as “sticky”)* – ends with unpaired bases can only ligate to ends with complementary overhanging bases. Two different DNA fragments that were cut with the same enzyme can be ligated together.
 - b) *blunt ends* – any blunt end can be ligated to any other blunt end.
5. **Frequency of Cutting** – Statistically, longer restriction enzyme recognition sequences are present less frequently because on average the recognition site will be present every 4^n base pairs, where n is the number of bases in the recognition sequence.
 - a) 4-base Cutter: 256-bp fragments, on average.
 - a) 5-base Cutter: 1024-bp, on average.
 - b) 6-base Cutter: 4096-bp, on average.
 - b) 7-base Cutter: 16384-bp, on average.

B. Separation of nucleic acid fragments by gel electrophoresis. In an electric field, nucleic acids, which are polyanions, migrate toward the positive pole. Gel electrophoresis separates DNA, independent of nucleotide sequence, according to size due to a sieving effect. Longer DNA molecules are slower to wind through the series of holes in the sieve. DNA fragments of different sizes resulting from restriction enzyme digestion can be separated by electrophoresis. The separated DNA molecules can be detected using a stain that stains all DNA regardless of nucleotide sequence, or by using a labeled nucleotide probe that hybridizes to a specific DNA fragment.

1. **Gels with different pore sizes are used to separate different size ranges of nucleic acids.** Big pores are used to separate big fragments and small pores are used to separate small fragments.

- a) **Polyacrylamide gels:** can resolve small DNA fragments that differ in length by only a single nucleotide in the range of a few base pairs to about 1,000 bp.
- b) **Agarose gels:** Are more porous than polyacrylamide. Can resolve fragments of 200 bp to 10,000 bp.
- c) **Pulse field gel electrophoresis:** Can resolve DNA molecules larger than 20 kb (kilobase), such as DNA molecules from an entire chromosome containing millions of nucleotides. The angle of the electrical field is oscillated in order to help very large DNA molecules reorient in order to wiggle through the pores of agarose gel.



© 2000 ASM Press and Sinauer Associates, Inc.

2. **Methods for visualizing electrophoresed DNA.**

- a) **Fluorescent dyes** such as ethidium bromide fluoresce intensely upon excitation with ultraviolet light when bound to DNA. Ethidium bromide is a non-specific stain, because it binds to DNA regardless of nucleotide sequence.
- b) **Probes** (fluorescent- or radioactive-labeled) – Labeled *probes* complementary to the DNA of interest are hybridized to the nucleic acid after it is electrophoresed (Southern or Northern blotting).

VI. Ligation of the vector with the DNA being inserted (*insert*) using the enzyme *DNA ligase*. *DNA ligase* requires that one of the DNA fragments to be joined have a 3' hydroxyl end and the other fragment a 5' phosphoryl end in order for it to catalyze a phosphodiester bond between the two molecules. The ligation reaction generates the desired recombinant molecule and can also generate side products that are not desired.

VII. Putting recombinant plasmid DNA into host cells.

A. Experimentally, plasmids can be put into bacterial cells by a process called *transformation*. Only a small fraction of cells are *transformed*, so a method is needed to select for *transformed* cells. The usual strategy is to ensure that the plasmid includes a gene that the host cell requires for growth under specific conditions, such as a gene for resistance to an antibiotic. Only cells that have been transformed by the recombinant plasmid can grow in the presence of that antibiotic, making any cell that contains the plasmid "selectable" under those conditions. Such a gene is sometimes called a selectable marker. The products of the ligation reaction with a plasmid vector are introduced into bacteria by a process called *transformation*, in which bacterial cells (e.g., *E. coli*) are permeabilized by different methods and are mixed with the products of the ligation reaction. Only a small fraction of the permeabilized cells actually take up the plasmid DNA and those cells take up only a single plasmid. A bacterium with a single plasmid (i.e., a clone) grows into a colony or culture that contains only a single recombinant product molecule.

B. Properties of Plasmid Vectors.

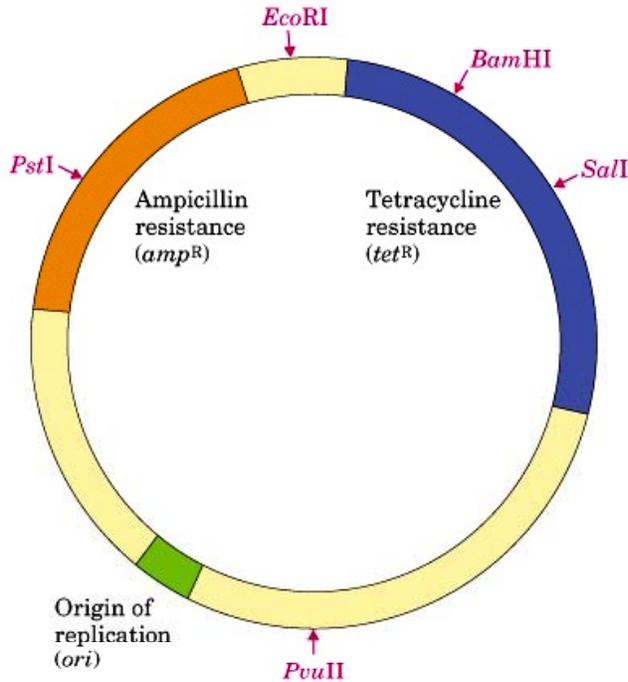
1. **Contain origins of replication.** Origins of replication are DNA sequences that the bacterial DNA replication machinery recognize and are the site for initiation of DNA replication. The origin of replication is required for the vector to be propagated in the host organism.
2. **Contain antibiotic resistance gene(s)** –The plasmid, by virtue of an antibiotic resistance gene, confers to host cells the ability to grow on media containing antibiotic while host cells without the vector (and antibiotic resistance gene) die (i.e., the antibiotic resistance gene enables selection of host cells that are transformed from those that are not).
3. **Contain unique restriction enzyme cleavage sites for insertion of foreign DNA.**

VIII. Strategies to isolate clones with the correct recombinant plasmids from incorrect recombinant plasmids.

Host cells containing any vector are able to grow and form colonies (clones) by virtue of the antibiotic resistance gene, but some of these clones do not contain the desired recombinant DNA (the most common incorrect ones are □ the vector that ligates without an insert). Several strategies are used to identify the desired recombinant molecules.

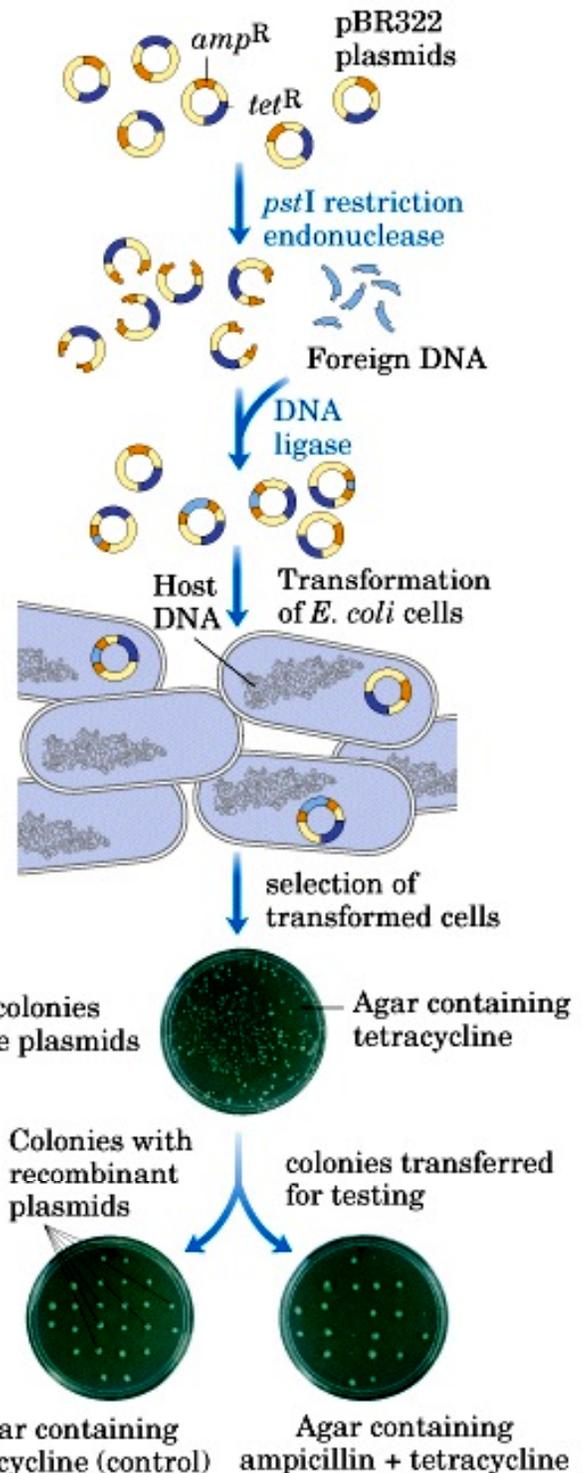
A. **Disruption** of a marker gene on the vector by insertion of DNA into it.

1. **Antibiotic resistance gene:** The plasmid vector pBR322 carries genes for resistance to ampicillin and tetracycline (amp^r and tet^r , see above left). If the foreign DNA is inserted into the amp^r gene, the amp^r gene is disrupted and can no longer encode the ampicillin resistance protein; cells harboring such plasmids cannot grow on “plates” containing ampicillin but



can grow on “plates” containing tetracycline. In contrast, cells transformed by the plasmid vector lacking the inserted foreign DNA (incorrect recombinants) grow on both ampicillin and tetracycline plates, while cells that do not take up the plasmid (not transformed) cannot grow on either antibiotic (see figure above right).

2. **Gene encoding a color marker:** The vector contains the gene encoding β -galactosidase (encoded by the β -*lacZ* gene). The “insert” DNA is inserted into this coding sequence. Clones with vector not having inserted DNA have an intact β -galactosidase coding



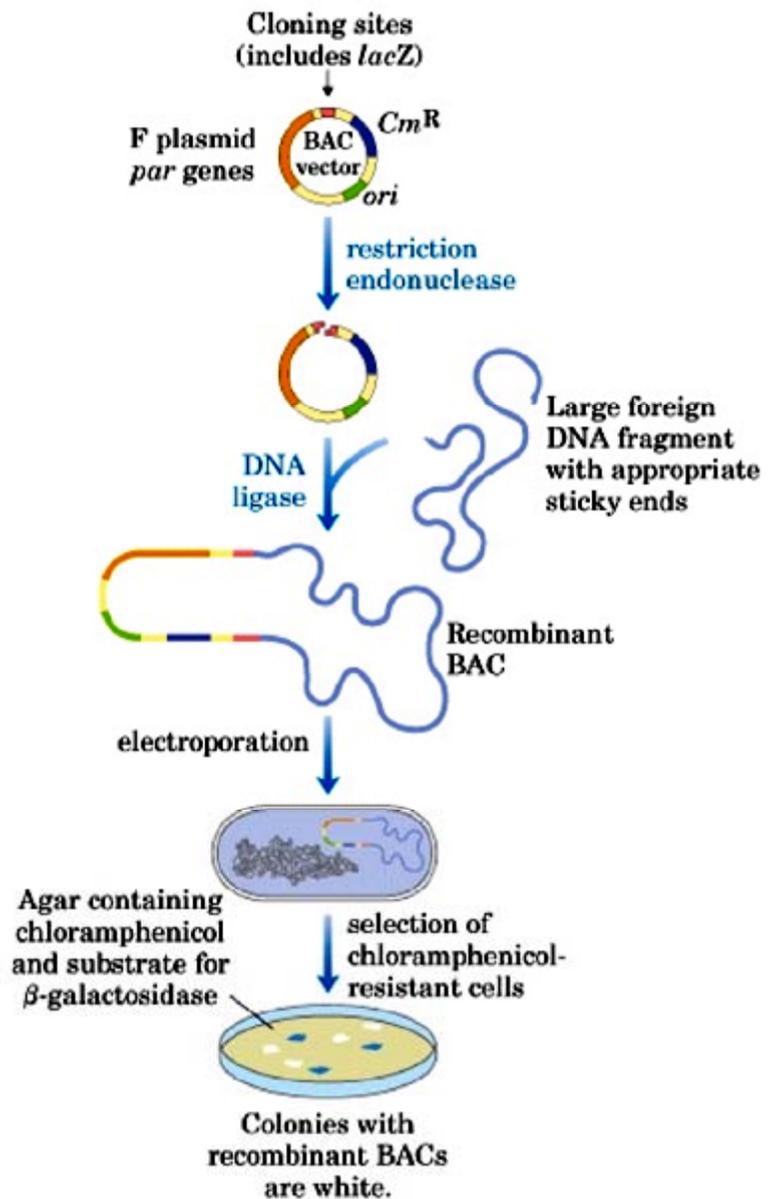
sequence. The β -galactosidase converts a galactose analog to a blue product making these colonies (clones) blue when grown on plates containing the galactose analog. Clones that have an insert disrupting β -galactosidase are white. The white colonies are chosen (See figure \rightarrow).

3. **Restriction mapping:** If the restriction sites of the insert DNA and vector DNA are known, then the pattern of restriction fragments of the correct recombinant can be predicted. The candidate plasmid clones are cut with restriction enzymes and those that have the predicted restriction fragments are the correct recombinant plasmids.

IX. Production of clinically useful proteins with biotechnology.

Vectors can be constructed for the production of large amounts of pharmaceutically valuable proteins that are difficult to obtain from human tissues. When a cDNA of interest is cloned into an *expression vector* and put into a host, the host cells serve as factories for the production of a wide variety of recombinant proteins.

Protein expression in bacteria. An *expression vector* contains the proper promoter elements for efficient transcription of the cDNA in the host. Thus, cDNA inserted into the expression vector is transcribed into mRNA, which is translated into protein in the host. Eukaryotic mRNA and the cDNA made from it do not contain introns. Since introns cannot be processed in bacteria, cDNAs, rather than genomic DNAs, are used for expression cloning in bacteria. The figure below demonstrates how a cDNA in an expression vector can be used to express a large quantity of recombinant human protein in bacteria.



Protein production in yeast and mammalian cells. Bacteria lack the full repertoire of post-translational modification pathways that may be required to generate a fully active recombinant protein. In cases where post-translational modifications are required for the biological activity of the protein, the protein is produced in yeast or mammalian cells (see below right for Hepatitis B vaccine production).

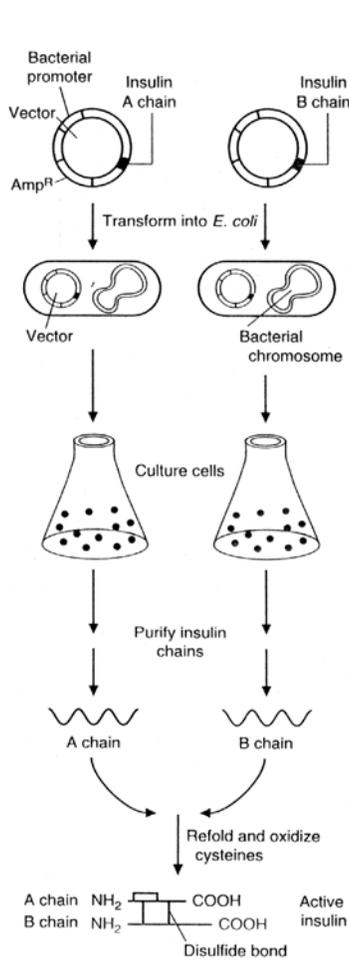
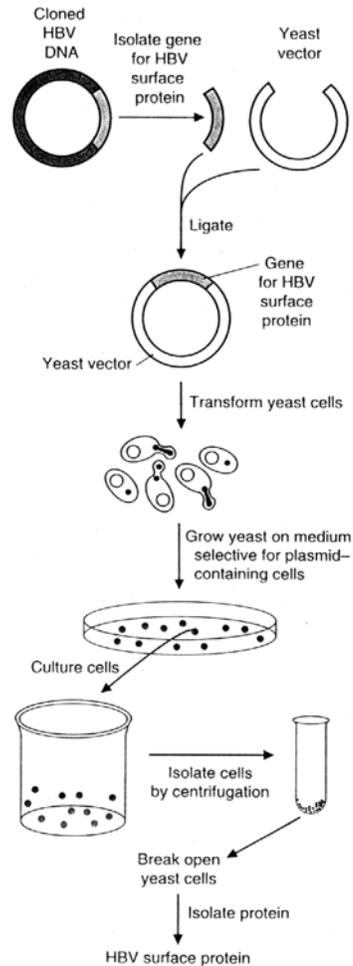


Fig. 16.13. Production of human insulin in *E. coli*. Amp^R = the gene for ampicillin resistance.



16.12. Production of a hepatitis B vaccine by recombinant DNA techniques.

X. Clinically used proteins from recombinant DNA technology

A. Hormones

1. **human insulin** – produced in *E. coli*. □ Used to treat diabetes.
2. **human growth hormone** - GENOTROPIN AQ® (UpJohn) is approved for the treatment of growth hormone deficiency (GHD), growth failure associated with chronic renal insufficiency (CRI) prior to kidney transplantation, and short stature associated with Turner syndrome. GENOTROPIN is synthesized in a strain of *E. coli* that has been modified by the addition of the gene for human growth hormone. Patients with growth hormone deficiency do not produce enough of their own growth hormone, a protein essential for growth. They simply do not grow or grow

extremely slowly. Patients with CRI or Turner syndrome also grow extremely slowly. Growth hormone can help patients with these conditions potentially grow at a normal rate.

B. Enzymes

1. **PULMOZYME® INHALATION SOLUTION** is a sterile, clear, colorless, highly purified solution of recombinant human deoxyribonuclease I (rhDNase), an enzyme which selectively cleaves DNA. The protein is produced by *genetically engineered Chinese Hamster Ovary (CHO)* cells containing DNA encoding for the native human protein, deoxyribonuclease I (DNase). By cleaving DNA found in the lungs of patients with cystic fibrosis, the viscosity of pulmonary mucous is reduced.

C. Cytokines

1. **Betaseron®** (Interferon β -1b) is a purified, sterile, lyophilized protein product produced by recombinant DNA techniques and formulated for use by injection. Interferon beta-1b is *manufactured by bacterial* fermentation of a strain of *Escherichia coli* that bears a genetically engineered plasmid containing the gene for human interferon betaser17. The native gene was obtained from human fibroblasts and altered in a way that substitutes serine for the cysteine residue found at position 17. Interferon beta-1b is a highly purified protein that has 165 amino acids and an approximate molecular weight of 18,500 Dalton. It does not include the carbohydrate side chains found in the natural material. Interferon beta-1b is used to treat the relapsing-remitting form of multiple sclerosis (MS). This medicine will not cure MS, but may decrease the number of relapses of the disease.
2. **Colony stimulating factors** help the bone marrow to make new blood cells. For example, Sargramostim (GM-CSF Leukine®) is a human granulocyte-macrophage colony-stimulating factor (rhuGM-CSF) produced by a yeast (*Saccharomyces cerevisiae*) expression system. It is a glycoprotein of 127 amino acids. □The amino acid sequence of Sargramostim differs from native human GM-CSF by a substitution at the leucine position 23, and the carbohydrate moiety may be different from the native protein. Clinically, Sargramostim is used to reduce the duration of neutropenia and incidence of infection in patients receiving myelosuppressive □chemotherapy or bone marrow transplantation, for mobilization of peripheral blood progenitor cells for collection, and for bone marrow graft failure or engraftment delay.

D. Antibodies

1. **Herceptin®** – the first humanized monoclonal antibody for the treatment of breast cancer. Herceptin® (Trastuzumab) as a single agent is indicated for the treatment of patients with metastatic breast cancer whose tumors overexpress the HER2 (human epidermal growth factor receptor2) protein. The antibody blocks the growth factor receptor. It's a recombinant humanized monoclonal antibody to HER-2. A monoclonal antibody is made by *mouse B-cells which have reacted with the HER-2 antigen and have been fused with a cell line that immortalizes it*□. The antibody was a mouse antibody to start with, and it was humanized because if mouse antibodies are injected into the human the immune system would attack them and the treatment would not be effective. Most of the molecule is now

the same as a human antibody; not only is there no immune reaction against it, but it can also bind to components of our own human immune system and have some immune effect in the body.

2. **Synagis®** (palivizumab) is a humanized monoclonal antibody (IgG1k) directed to an epitope in the A antigenic site of the F protein of respiratory syncytial virus (RSV). Palivizumab is a composite of human (95%) and murine (5%) antibody sequences.

E. Blood clotting factors

TNKase™ (Tenecteplase), a thrombolytic (clot-busting) agent is a recombinant version of tissue plasminogen activator. *It is produced using genetically modified Chinese hamster ovary cells.* It has been approved by the U.S. Food and Drug Administration for the reduction of mortality associated with acute myocardial infarction (AMI - heart attack). It is the first thrombolytic drug to date that can be administered in only five seconds and in just one dose (based on the weight of the patient). TNKase works by stimulating the body's own clot-dissolving mechanism by activating plasminogen, a naturally occurring substance secreted by endothelial cells in response to injury of the artery walls. Injury to artery walls would normally lead to clot formation. However, when TNKase activates plasminogen, the plasminogen converts into plasmin, which breaks down the fibrin mesh that binds the clot together. The clot is then dissolved, restoring blood flow to the heart.

table 29–3

Some Recombinant DNA Products in Medicine	
Product category	Examples/uses
Anticoagulants	Tissue plasminogen activator (TPA) activates plasmin, an enzyme involved in dissolving clots; effective in treating heart attack victims.
Blood factors	Factor VIII promotes clotting and is deficient in hemophiliacs; use of factor VIII produced by recombinant DNA technology eliminates infection risks associated with blood transfusions.
Colony stimulating factors	Immune system growth factors that stimulate leukocyte production; used to treat immune deficiencies and to fight infections.
Erythropoietin	Stimulates erythrocyte production; used to treat anemia in patients with kidney disease.
Growth factors	Stimulate differentiation and growth of various cell types; used to promote wound healing.
Human growth hormone	Used to treat dwarfism.
Human insulin	Used to treat diabetes.
Interferons	Interfere with viral reproduction; used to treat some cancers.
Interleukins	Activate and stimulate different classes of leukocytes; possible uses in treating wounds, HIV infection, cancer, and immune deficiencies.
Monoclonal antibodies	Extraordinary binding specificity is used in: diagnostic tests; targeted transport (of drugs, toxins, or radioactive compounds to tumors as a cancer therapy); many other applications.
Superoxide dismutase	Prevents tissue damage from reactive oxygen species when tissues briefly deprived of O ₂ during surgery suddenly have blood flow restored.
Vaccines	Proteins derived from viral coats are as effective in "priming" an immune system as the killed virus more traditionally used for vaccines, but are safer; first developed was the vaccine for hepatitis B.