PN ABX-899
98186

Biotechnology and Genetic Engineering in Egypt:

Principal Research Findings of the Science and Technology Project

September 1995

PREPARED BY

SCIENCE AND TECHNOLOGY COOPERATION PROJECT
ACADEMY OF SCIENTIFIC RESEARCH AND TECHNOLOGY

This report was prepared for USAID/Egypt with the assistance of the Project in Development and the Environment, USAID Contract Number: ANE 0178-Q-00-1047-00; USAID Project Number: 398-0365.

INTRODUCTION

This report covers the achievements of the Science and Technology Cooperation Project (STC) in the field of biotechnology and genetic engineering. It is one of a series of reports documenting the key areas of work undertaken by STC researchers in the fields of computer-aided manufacturing, biotechnology and genetic engineering, industrial waste management, and municipal waste management.

STC is a project funded by a grant from the U.S. Agency for International Development which is supplemented by in-cash and in-kind contributions from the Government of Egypt and the end-users. The goals of STC are improving the productivity and welfare of the Egyptian people by applying technology to specific development problems and building science and technology research capacity in selected fields. To this end, STC has funded 81 research subprojects.

Section I of this report represents a synthesis and analysis of key findings and conclusions from four STC research subprojects. This section was prepared by Dr. Atta Hazem, MD and revised by Dr. Talaat Abdel Aziz, MD. Section II of this report comprises summaries of the four research subprojects that formed the basis for section I. These summaries were prepared by a PRIDE team led by Dr. Walid Gamaledin. Members of the team included Ms. Naglaa El Bialy, Dr. Farida El Dars, and Dr. Dina El Naggar. This team was assisted by Ms. Nivene El Koshairi and Ms. Hala Rafik. The subproject summaries were based on reports issued by subproject research teams, interviews with Principal Investigators, and site visits. Dr. Todd Martensen graciously provided expert comments on an early draft of this report.

The views expressed in this report do not necessarily reflect those of the Project in Development and the Environment or USAID.

TABLE OF CONTENTS

		Page
SECTION I	APPLICATIONS OF GENETIC ENGINEERING IN RESEARCH AND ANALYSIS: SYNTHESIS AND ANALYSIS	I-1
	A. The Basics of Molecular Biology (MB)	I-1
	A1. Overview of Nucleic Acid Structure and Function	I-1
	A2. The Genetic Code	I-11
	A3. Hybridization of Nucleic Acid Strands	I-13
	A4. Detection of Probe/Target Complexes	I-16
	A5. Molecular Hybridization Techniques	I-16
	B. Recombinant DNA Technology (Genetic Engineering)	I-17
	B1. Introduction, Basic Features	I-17
	B2. Cloning Amplifies DNA	I-21
	B3. Probes Search Libraries for Specific Genes of DNA Molecules	I-25
	B4. Biomedical Importance of Recombinant DNA	I-26
	B5. Practical Applications of Recombinant DNA Technology	I-27
	C. Polymerase Chain Reaction (PCR)	I-32
	C1. Introduction	I-32
	C2. Terminology	I-32
	C3. The Theory	I-33
	C4. The Methodology	I-37
	C5. Detecting The Amplified Product	I-39
	C6. Optimization of PCR	I-41
	C7. Practical Considerations	I-44
	C8. In Situ Hybridization	I-46
	C9. PCR Applications	I-48
SECTION II	BIOTECHNOLOGY AND GENETIC ENGINEERING: RESEARCH SUBPROJECT SUMMARIES	II-1
A.	Subproject No. 6: Enzyme Production for Clinical and Industrial Applications Hybridization Technique in Egyptian Females and its Relation to Cervical Cancer	II-1
	A1. Subproject Description	II-1
	A2. Subproject Achievements	II-2
B.	Subproject No. 35: Introduction of Bacillus Thuringensis-Endotoxin Gene into Cotton Plants and Testing its Expression and Toxicity	II-3
	B1. Subproject Description	II-3
	B2. Subproject Achievements	II-5

TABLE OF CONTENTS
(continued)

C. Subproject No. 42: Molecular Genetic Techniques in the Diagnosis and Prognosis of Chronic Myeloid Leukaemia before and after Bone Marrow Transplantation Hybridization Technique in Egyptian Females and its Relation to Cervical Cancer II-6

C1. Subproject Description II-6
C2. Subproject Achievements II-7

D. Subproject No. 45: Screening for and Detection of Human Papilloma Virus by DNA Hybridization Technique in Egyptian Females and its Relation to Cervical Cancer II-7

D1. Subproject Description II-8
D2. Subproject Achievements II-9

SECTION I

SYNTHESIS AND ANALYSIS

SECTION I
APPLICATIONS OF GENETIC ENGINEERING IN RESEARCH AND ANALYSIS: SYNTHESIS AND ANALYSIS

A. The Basics of Molecular Biology (MB)

A1. Overview of Nucleic Acid Structure and Function

Unlike carbohydrates, lipids, and proteins, nucleic acids are not used primarily for either cellular energy or the structural integrity of the cell. Instead, they are used for the storage and expression of genetic information. There are two chemically distinct types of nucleic acids: deoxyribonucleic and (DNA) and ribonucleic acid (RNA). The DNA contained in a single fertilized egg encodes the information that directs the development of an organism. This development may ultimately require the production of billions of cells. Each of these cells is specialized, expressing only those functions that are required for it to perform its role in maintaining the organism. Therefore, the DNA must be able not only to replicate precisely each time a cell divides, but also to have the information that it contains be selectively expressed.

Members of the second class of nucleic acids, the RNAs, participate in the process of selective expression of the genetic information stored in the DNA. DNA is found not only in chromosomes in the nucleus of eukaryotic organisms, but also in mitochondria and in the chloroplast of plants. RNA synthesized in the nucleus performs its function in the cytoplasm, whereas RNA synthesized in the mitochondria or chloroplast remains within these organelles. Prokaryotic cells, which lack nuclei, have a single chromosome but may also contain nonchromosomal DNA in the form of plasmids.

A1a. Nucleotide Structure

Nucleotides are the building blocks of both DNA and RNA. They also serve as the energy currency of the cell and as structural components of a number of essential cofactors. Nucleotides are composed of a nitrogenous base, a monosaccharide, and one, two, or three phosphate groups.

- **Purines and Pyrimidines**

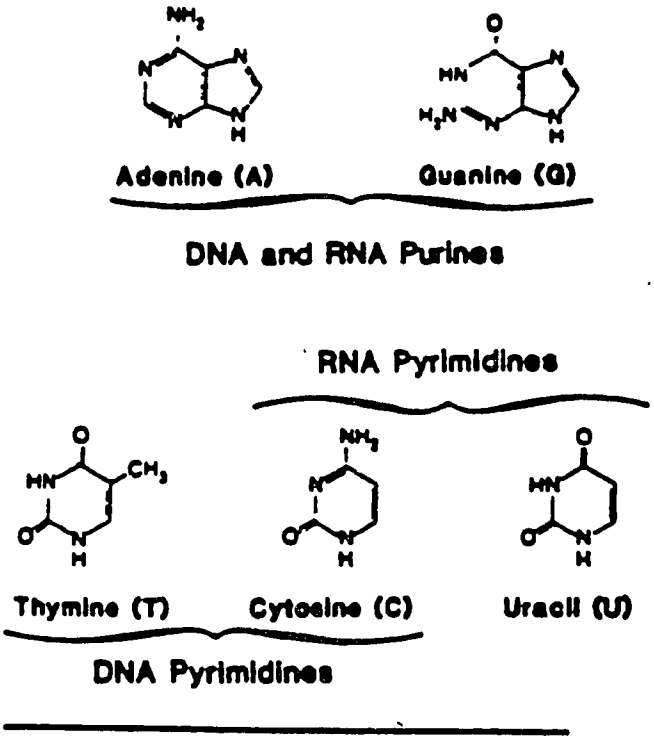
1. The bases belong to two families of compounds: the purines and the pyrimidines. (Figure 1.1)

2. Both DNA and RNA contain the same purine bases: adenine (A) and guanine (G).

3. Both DNA and RNA contain the pyrimidine cytosine (C), but they differ in their second pyrimidine bases: DNA contains thymine (T), whereas RNA contains uracil (U). (Note that T and U differ only by one methyl group present on T but absent on U (see Figure 1.1).

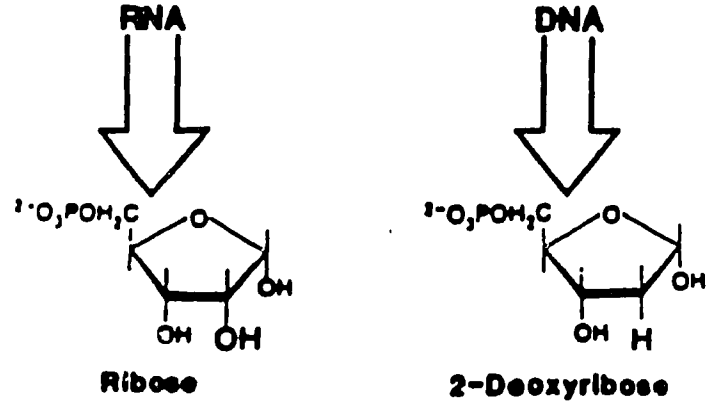
4. Unusual bases are found occasionally in some species of DNA and RNA, for example, in some viral DNA and in transfer RNA. These include the following:

Figure 1.2



Purines and pyrimidines commonly found in DNA and RNA.

Figure 1.2



Pentoses found in nucleic acids.

- 2 -

DNA: 5-methylcytidine, 5-hydroxymethylcytidine, glucosylated 5-hydroxymethylcytidine, hydroxymethyluracil.

RNA: Inosine, thymine, methylated purines and pyrimidines, thiouracil, dihydrouracil, isopentenyl adenine, acetyl cytosine, pseudouracil.

The presence of an unusual base in a nucleotide sequence may aid in its recognition by specific enzymes or protect it from being degraded by nucleases.

- **Nucleosides**

1. The addition of a pentose sugar to a base produces a nucleoside.

2. If the sugar is ribose, a ribonucleoside is produced. If the sugar is 2-deoxyribose, a deoxyribonucleoside is produced. (Figure 1.2)

3. The carbon and nitrogen atoms in the ring(s) of the base and the sugar ring are numbered separately. Note that the atoms in the rings of the bases are numbers 1-6 in pyrimidines, and 1-9 in purines, while the carbons in the pentose are numbers 1'-5'. Thus, when you refer to the 5'-carbon, you are specifying a carbon atom in the pentose rather than an atom in the base.

- **Nucleotides**

1. Nucleotides are phosphate esters of nucleosides.

2. The phosphate group is usually attached by an ester linkage to the 5'-OH of the pentose. Such a compound is called a nucleoside 5'- phosphate or a 5-nucleotide. (Note: The type of pentose is denoted by the prefix in the names "5'-ribonucleotide" and "5'-deoxyribonucleotide.")

3. If one phosphate group is attached to the 5'-carbon of the pentose, the structure is a nucleoside monophosphate (NMP) like AMP or CMP. If a second or third phosphate is added to the nucleotide, a nucleoside diphosphate or triphosphate results.

4. The dissociation of hydrogen from the phosphate groups at neutral pHs is responsible for the acidic nature of nucleotides and nucleic acids.

A1b. Structure of DNA

- **Building Blocks of DNA**

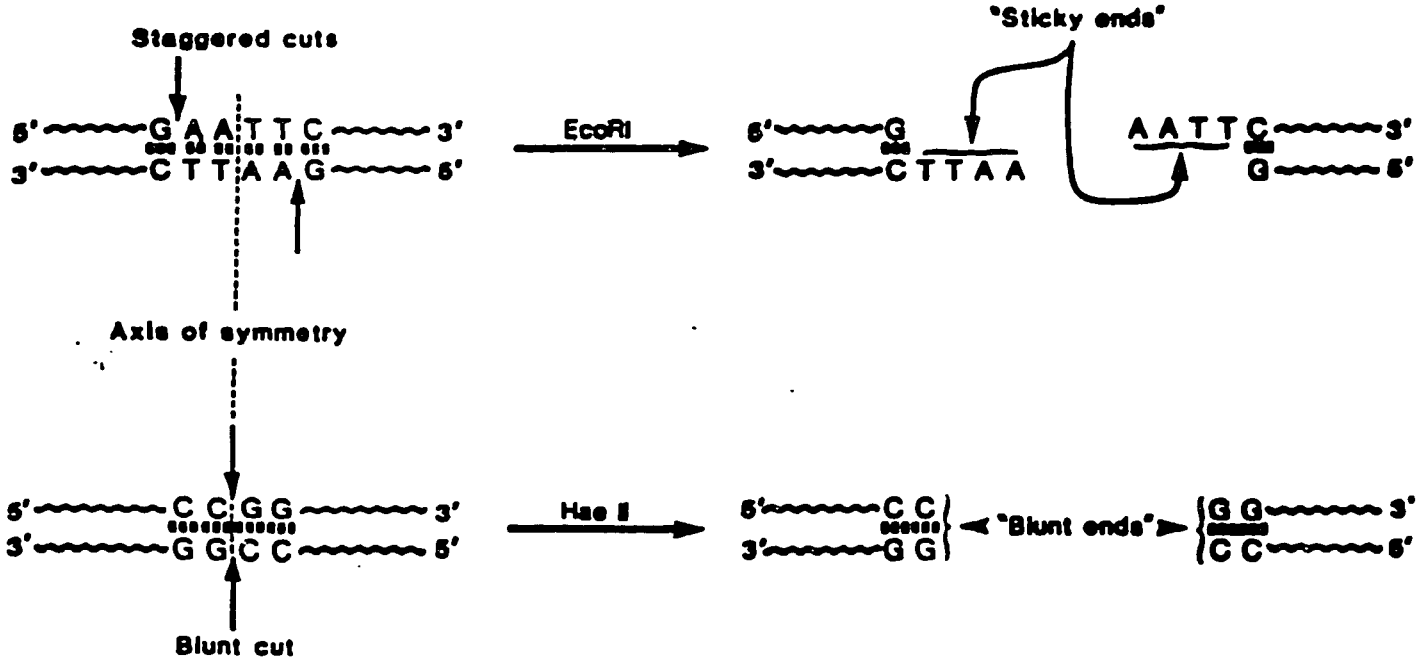
1. DNA is a polynucleotide that contains many mononucleotides covalently linked to each other by 3',5'-phosphodiester bonds.

2. These bonds join the 5'-hydroxyl group of the pentose of one nucleotide to the 3'-hydroxyl group of the pentose of another nucleotide through a phosphate group.

3'

Figure 1-3

III. Structure of DNA



Examples of *restriction endonuclease* cleavage sites on DNA showing twofold, rotational symmetry. (Cleavage sites shown are those of the restriction endonucleases *Eco R1* and *Hae II*). The cleavage sites are found at the arrows. The axes of symmetry are marked by dotted lines.

3. The resulting long, unbranched chain has polarity, with a 5'-end that is not attached to another nucleotide and a 3'-end that is also unattached. The bases located along the resulting ribose-phosphate backbone are, by convention, always written in sequence from the 5'-end of the chain to the 3'-end.

4. Phosphodster linkages between nucleotides (in both DNA and RNA) can be cleaved hydrolytically by chemicals or can be enzymatically hydrolyzed by a family of nucleases, for example, deoxyribonuclease for DNA and ribonucleases for RNA.

   a. Nucleases that cleave the nucleotide chain at positions in the interior of the chain are called endonucleases. Those that cleave the chain only by removing individual nucleotides from one of the two ends are called exonucleases.

   b. A special group of nucleases produced by microorganisms are called restriction endonucleases. These enzymes recognize short stretches of DNA (generally 4 or 6 base pairs) that contain specific base sequences. These sequences, which differ for each restriction endonuclease, have twofold rotational symmetry (Figure 1.3). This means that within a short region of the double helix, the nucleotide sequence on the "top" strand, read 5'---3' is identical to that of the "bottom" strand, also read in the 5'---3' direction. Therefore, if you turn the page upside down, that is, rotate is 180°, the structure remains the same. Because they cleave the DNA at specific positions, these enzymes are used experimentally to obtain precisely defined DNA segments and are among the most important tools in the cloning of genes.

- **Structure of DNA**

1. With the exception of a few viruses that contain single-stranded DNA, DNA exists as a double-strand molecule in which the two strands wind around each other forming a double helix.

2. The two-chains are coiled around a common axis. (Figure 1.4).

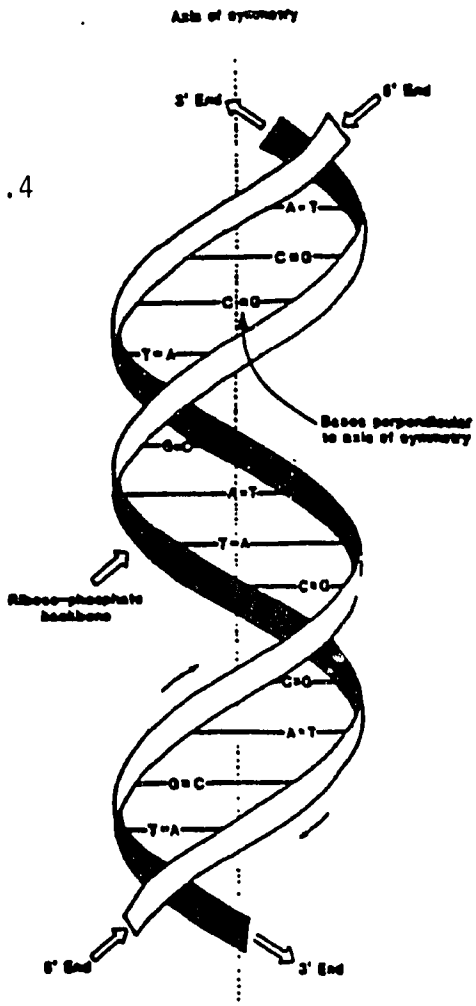
3. The chains are paired in an antiparallel manner, that is, the 5'- end of one strand is paired with the 3'-end of the other strand.

4. In the most common type of DNA helix, the classic "B" form, the hydrophilic ribose-phosphate backbones of each chain are on the outside of the molecule, whereas the hydrophobic bases are stacked inside, perpendicular to the axis of the helix. The overall structure resembles a twisted ladder. (Figure 1.4) The spatial relationship between the two strands in the helix creates a major (wide) groove and a minor (narrow) groove.

5. The bases of one strand are paired with the bases of the second strand so that an adenine is always paired with a thymine, while a cytosine is always paired with a guanine. Therefore one polynucleotide chain of the DNA double helix is always the complement of the other. Given the sequence of bases on one chain, the sequence of bases on the complementary chain is determined. (Figure 1.5)

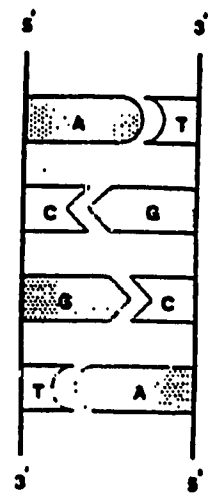
Figure 1.4 - Figure 1.7

Figure 1.4



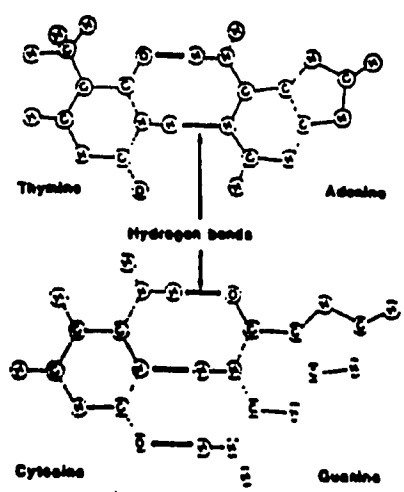
DNA double-helix illustrating some of its major structural features

Figure 1.5



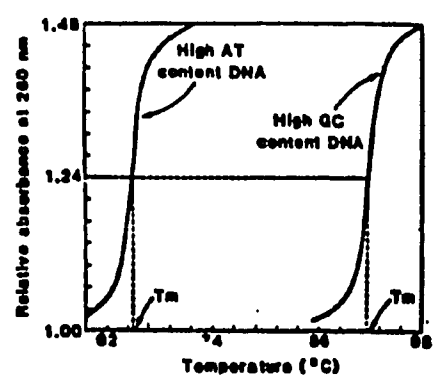
Two complementary DNA sequences.

Figure 1.6



Hydrogen bonds between complementary bases.

Figure 1.7



Melting temperature of DNA molecules with different nucleotide compositions. (At a wavelength of 260 nm, single-stranded DNA has a higher relative absorbance than does double-stranded DNA.)

a. The base pairs are held together by hydrogen bonds: two between A and T, and three between G and C. (Figure 1.6) These hydrogen bonds, plus the hydrophobic interactions between the stacked bases, stabilize the structure of the double helix.

b. The two strands of the double helix separate when the hydrogen bonds between the paired bases are disrupted. Disruption can occur if the PH of the DNA solution is altered so that the nucleotide bases ionize or if the solution is heated.

c. When DNA is heated, the temperature at which half of the helical structure is lost is defined as the melting temperature (T^m). The loss of helical structure in DNA is called denaturation and can be monitored by measuring its absorbance at 260 nm. Because there are three hydrogen bonds between G and C but only two between A and T, DNA that contains high concentrations of A and T will denature at a lower temperature than G-and C-rich DNA (Figure 1.7).

6. Three structural forms of DNA have been described: the B form, described by Watson and Crick in 1953, the A form and the Z form:

a. The B form is a right-handed helix with 10 residues per 360° turn of the helix, and with the planes of the bases perpendicular to the helixal axis. Chromosomal DNA is thought to consist primarily of B DNA.

b. The A form is produced by moderately dehydrating the B form. It is also a right-handed helix, but there are 11 base pairs per turn, and the planes of the base pairs per turn, and the planes of the base pairs are titled 20° away from the perpendicular to the helical axis. The conformation found in DNA-RNA hybrids or RNA-RNA double-stranded regions is probably very close to the A form.

c. Z DNA is a left-handed helix that contains about 12 base pairs per turn. It appears to occur naturally in specific regions of DNA and may play an important role in regulating gene expression.

- **Plasmids**

1. Prokaryotic organisms contain single large, circular chromosomes. In addition, most species of bacteria also contain small, circular, extrachromosomal DNA molecules called plasmids.

2. Plasmid DNA carries genetic information and undergoes replication that may or may not be synchronized to chromosomal division. Plasmids may carry genes that convey antibiotic resistance to the host bacterium and may facilitate the transfer of genetic information from one bacterium to another. A description of these processes can be found in a microbiology textbook.

3. Plasmids can be readily isolated from bacterial cells, their circular DNA cleaved at specific sites by restriction endonucleases, and foreign genes inserted into the circle. This hybrid plasmid can be reintroduced into a bacterium and large numbers of copies of the plasmid containing the foreign DNA produced. This DNA can then be further studied.

The method of amplification of specific genetic material by a microorganism is called gene cloning, and is another important tool of "genetic engineering."

4. When double-helical DNA is converted to a closed, circular molecule, it can itself be twisted to form a right- or left-handed supercoil (also called "supertwist" or "superhelix").

   a. Energy is required to introduce supertwists, since it introduces a strain in the DNA double helix.

   b. Supertwisting is of biological significance because superhelical DNA has a more compact shape than its relaxed precursor. Also, negative supertwists result in a partial unwinding of the double helix, which is important in DNA replication.

   c. Supertwists can also occur in linear DNA. They are caused by the unwinding of the double helix or can be enzymatically introduced.

A1c. Organization of Eukaroyotic DNA

- **Histones**

1. A typical human cell contains 46 chromosomes, each of which probably contains a single molecule of DNA. Eukaryotic DNA, however, is not "bare," but rather is associated with tightly bound proteins, including the histones, which serve to order the DNA into structural units called nucleosomes that resemble beads on a string.

2. The histones are small proteins and are positively charged at physiologic pH owing to their high content of lysine and arginine. Because of their positive charge, the histones form ionic bonds with the negatively charged DNA.

3. The five major classes of histones, designated H1, H2A, H2B, H3, and H4, fall into two main groups.

   a. The first group of histones form the structural core of the individual nucleosome "beads." Two molecules each of H2A, H2B, H3, and H4 associate to form the nucleosome core, around which a segment of the DNA double helix is wound, forming a negatively supertwisted helix.

   b. Histone H1, of which there are several related species that form the second group of histones, is not found in the nucleosome core but appears to bind to the DNA chain between the nucleosome beads (the linker DNA). H1 appears to aid the packing of nucleosomes into a more compact structure called the chromatin fiber. During cell division, the chromatin condenses further to form the chromosomes that are visible in the light microscope.

4. In addition to the histones, small nonhistone proteins called protamines bind to the DNA. Numerous enzymes and regulatory proteins also associate with the chromatin through sequence specific interactions that may facilitate or inhibit DNA replication or RNA transcription.

8

A1d. Structure of RNA

There are three major types of RNA: ribosomal RNA (rRNA), transfer RNA (tRNA), and messenger RNA (mRNA). Like DNA, all three types of RNA are long, unbranched molecules composed of mononucleotides joined together by phosphodiester bonds. They differ as a group from DNA in several ways; for example they are considerably smaller than DNA and, as previously discussed, contain ribose instead of deoxyribose and uracil instead of thymine. The three types of RNA also differ from each other in terms of size, function, and special structural modifications.

- **Ribosomal RNA**

1. There are three distinct size species of rRNA (23s, 16s, and 5S) in prokaryotic cells and eukaryotic mitochondria. In the eukaryotic cytosol, there are four rRNA size species (5S, 5.8S, 18S, and 28S). (Note: "S" is the Svedberg unit, which is related to the molecular weight of the compound.) Together, they make up about 80 percent of the RNA in the cell.

2. Ribosomal RNA is found in association with a number of different proteins as components of the ribosomes—the complex structures that serve as the sites for protein synthesis.

- **Transfer RNA**

1. The smallest of the RNA molecules (4S), transfer RNA has between 73 and 93 nucleotide residues. There is at least one specific type of tRNA molecule for each of the 20 amino acids commonly found in proteins. Together they make up about 15 percent of the RNA in the cell.

2. The tRNA molecules contain many unusual bases and have extensive intrachain base pairing, as shown in Figure 1.8. Additional features of this structure are also identified in Figure 1.8.

3. Each tRNA serves a an "adaptor" molecule that carries its specific amino acid to the site of protein synthesis. There, it recognizes the genetic code word that specifies the addition of its particular amino acid to the growing peptide chain.

- **Messenger RNA**

1. mRNA comprises only about 5 percent of the RNA in the cell yet is by far the most heterogeneous type of RNA in terms of size. The mRNA carries the genetic information from the DNA and is used as the template for protein synthesis.

2. Special structural characteristics of eukaryotic mRNA include a long sequence of adenines (a "poly-A tail") on the 3′-end of the RNA chain, plus a "cap" on the 5′-end consisting of a molecule of 7-methylguanosine attached "backward" through a triphosphate linkage, as shown in Figure 1.9.

Figure 1.8

V. STRUCTURE OF RNA

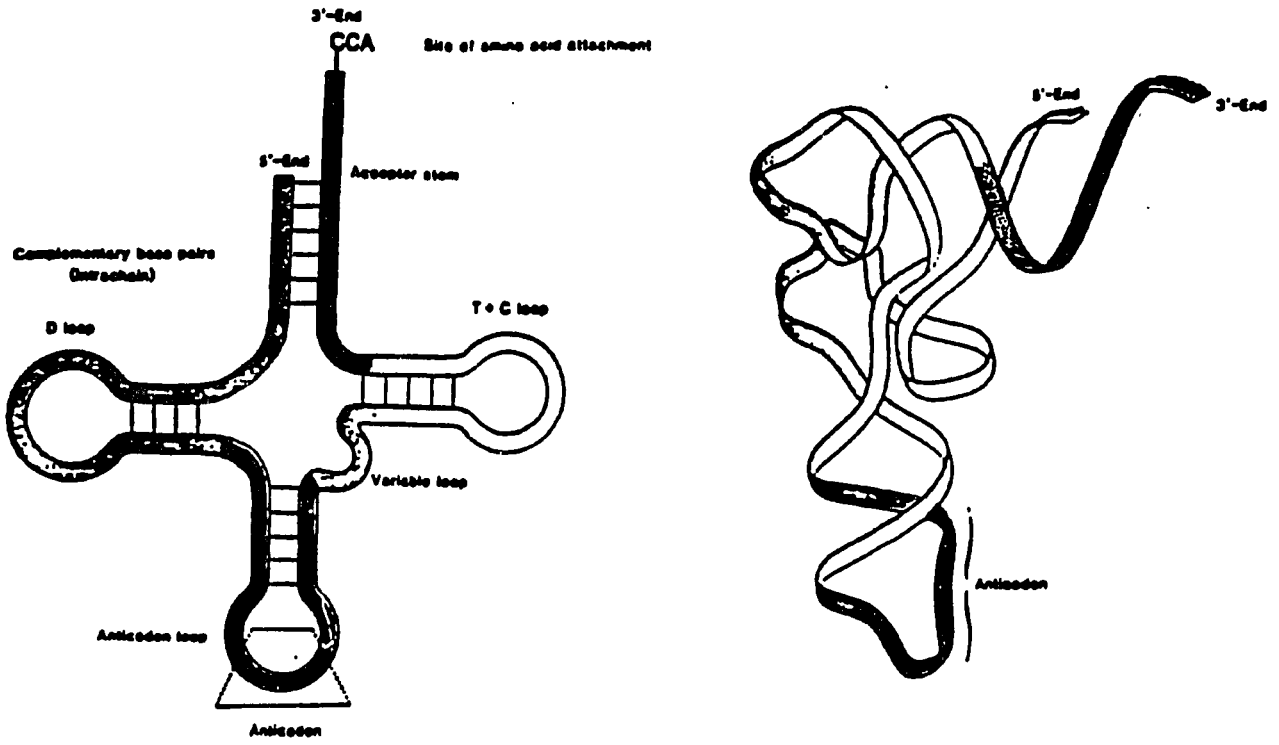
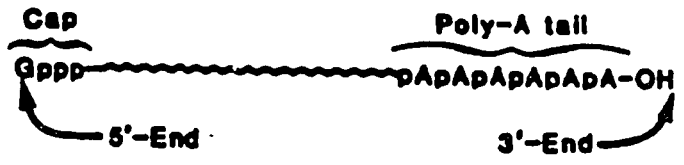


Figure 1.9



Structure of messenger RNA.

10

A2. The Genetic Code

Genetic information, stored in the chromosomes and transmitted to daughter cells through DNA replication, is expressed through its transcription to RNA and subsequent translation into a polypeptide chain. This flow of information from DNA to RNA to protein is termed the "Central Dogma" and is descriptive of all organisms (with the exception of some viruses that have RNA as the repository of their genetic information). The process of translation requires a genetic code, through which the information contained in the nucleic acid sequence is expressed to produce a specific sequence of amino acids. Any alteration in the nucleic acid sequence may result in the improper amino acid being inserted into the polypeptide chain, potentially causing a disease or even death to the organism.

A2a. Codons

1. The genetic code is a dictionary that gives the correspondence between a sequence of bases and a sequence of amino acids. The individual words in the code are each composed of three nucleotide bases. These genetic words are called codons.

2. The codons are found on the messenger RNA. The sequence of a codon is always read from its 5′-end to its 3′-end.

3. There are four nucleotide bases—adenine, guanine, cytosine, and uracil—that can be used to produce a three-base codon. There are therefore 64 different combinations of bases, taken three at a time. This (dictionary) can be used to translate any codon sequence and thus to determine which amino acids are coded for by the mRNA sequence. For example, the codon 5′-CAU-3′ codes for histidine, whereas 5′-AUG-3′ codes for methionine.

4. Three of the codons, UAG, UGA, and UAA, do not code for amino acids, but rather are termination codons. When one of these codons appears in an mRNA sequence, it signals that synthesis of the peptide chain coded for the that mRNA has been completed.

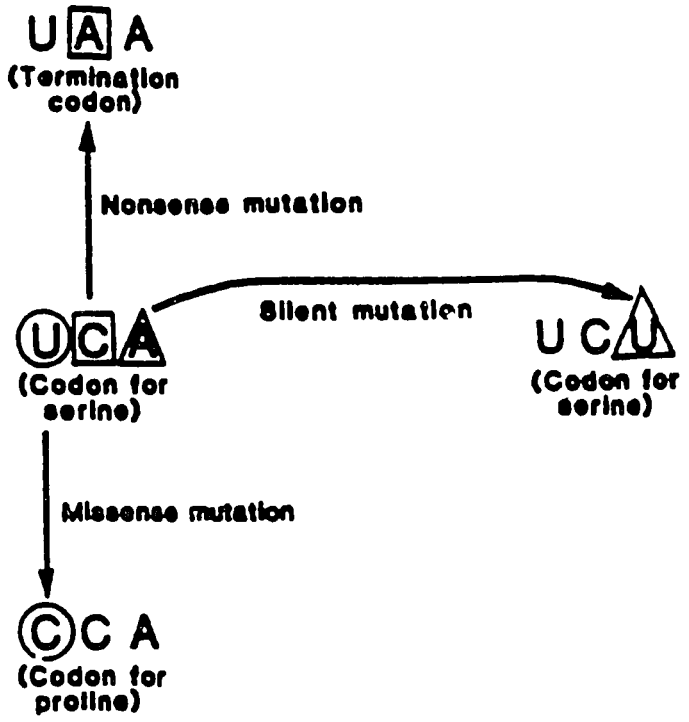
5. Changing a single nucleotide base on the mRNA chain can lead to any one of three possible results. (Figure 1.10)

   a. The codon containing the changed base may still code for the same amino acid. For example, if the serine codon UCA is given a different third base (say, UCU), it will still code for serine. Therefore, this is a silent mutation.

   b. The codon containing the changed base may code for a different amino acid. For example, if the serine codon UCA is given a different first base (say, CCA), the codon will code for a different amino acid, in this case, proline. This substitution of an amino acid is called a missense mutation.

   c. The codon containing the changed base may become a termination codon. For example, if the serine codon UCA is given a different second base (say, UAA), the new codon will cause the termination of translation at that point. The creation of a termination codon at an inappropriate place is called a nonsense mutation.

Figure 1.10



Possible effects of changing a single nucleotide base in the coding region of an mRNA chain.

12

A2b. Characteristics of the Genetic Code

1. The genetic code is specific: A specific codon always codes for a single amino acid.

2. The genetic code is usually stated to be universal (with a few minor exceptions); that is, the specificity of the genetic code has been conserved from very early stages of evolution, with only slight differences in the manner the code is translated.

3. The genetic code is redundant (sometimes called degenerate). Although each triplet corresponds to a single amino acid, a given amino acid may have more than one triplet coding for it. For example, arginine is specified by six different codons.

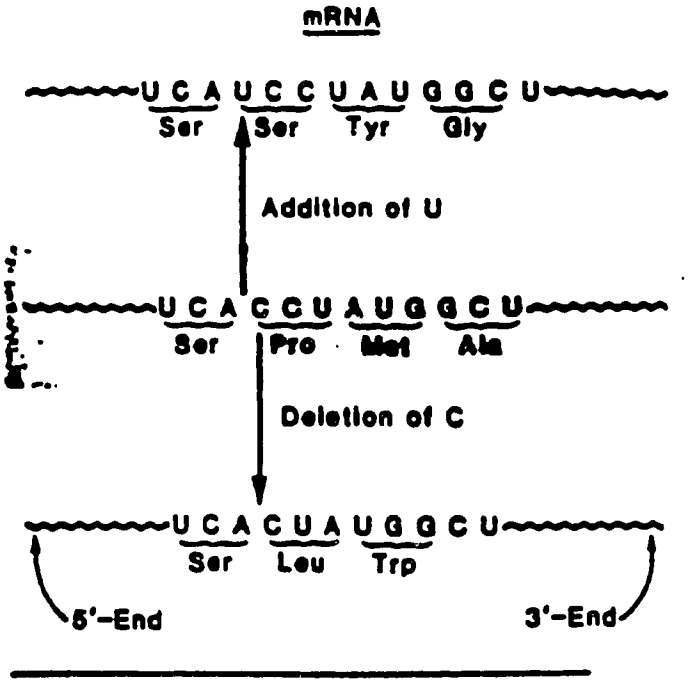
4. The genetic code is nonoverlapping and commandless: The code is read from a fixed starting point as a continuous sequence of bases, taken three at a time. For example, ABCDEFGHIJKL... is read as ABC/DEF/GHI/JKL...without any "punctuation" between the codons. Note that if one or two nucleotides are either deleted from or added to the interior of a message sequence, the reading frame will be altered, and the resulting amino acid sequence may become radically different from this point onward. (Figure 1.11)

A3. Hybridization of Nucleic Acid Strands

When two DNA strands meet, they ornient each other in opposite or antiparallel directions to allow base pair matching to occur. If no base matching is present, they go their separate ways. However, if there is sufficient base pair matching they will join together or hybridize. The specific term used to describe the degree of base pair matching that determines if the strands stay together is homology. How much homology is needed from two strands to stay together? Although it is true "the more, the better," another important variable is how close the base pair matches are to one another. Adjacent base pair matches in a sequence will hold together more strongly than the same number of base pair matches dispersed over the DNA sequence. (Figure 1.12) Dispersed base pair matches are typical of unrelated DNA strands, whereas clustered base pair matches are expected for related complementary DNA molecules. Clearly, if two DNA strands are completely homologous and have 100 percent base pair matching, then the strands would tend to remain hybridized under most conditions. Conversely, hybridized strands with poor homology (e.g., only 10 percent of base pairs matched) would tend to dissociate or denature readily under most conditions. Whether hybridized strands with intermediate homology—where, for example, 50 percent of the base pairs matched—would remain hybridized would depend greatly on the reaction conditions. Given that hydrogen bonds are the glue that keeps two hybridized strands together and that many chemicals and conditions can affect hydrogen bonding, a term is needed that describes if the hybridization reaction conditions relatively favor or disfavor hydrogen bonding. The term is stringency. Under low stringency conditions hybridized strands with intermediate homology would tend to remain hybridized whereas hybridized strands with poor homology would dissociate.

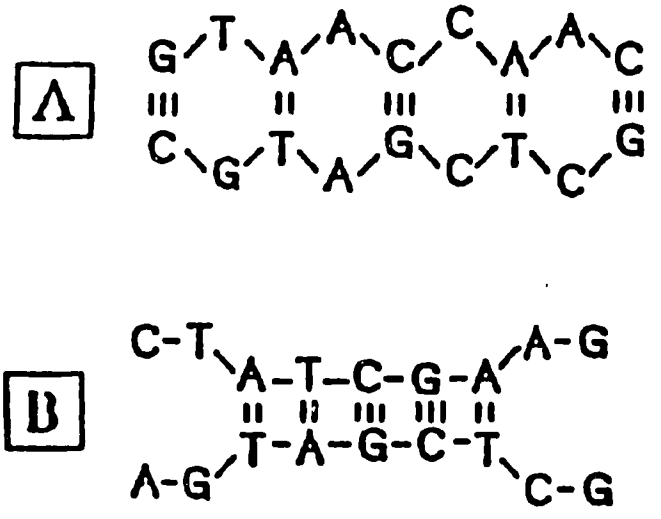
At high stringency conditions, only hybridized strands with strong homology would tend to remain hybridized. This key term is illustrated in the following example: three distinct hybridized DNA complexes, each 10 base pairs long, share variable homology—1 base pair match, 5 base pair matches, and finally complete homology of 10 base pair matches. The small number of base pairs being used for illustrative purposes; the relative percentage of base pair matching is the important concept. Under low stringency conditions, the strands with 10 percent homology will

Figure 1.11



Frameshift mutations can cause an alteration in the reading frame of mRNA.

Figure 1.12



Incomplete base pair matching between two DNA strands. For each double stranded molecule, 47% of the base pairs are complementary, which may be sufficient to keep the strands hybridized. However, the less dispersed arrangement of the base pair matches depicted in the complex B compared with complex A will allow it to remain hybridized under conditions where the other complex would denature.

14

separate. The hybridized strands with 50 percent and 100 percent homology remain together at low stringency despite disruption of some hydrogen bonds between matched base pairs because sufficient numbers of hydrogen bonds remain. However if one adds a chemical such as formamide, which may disrupt hydrogen bonds between matched base pairs, there may not be sufficient remaining hydrogen bonds to keep the strands with 50 percent homology together though the strands with complete homology will stay hybridized as sufficient hydrogen bonding still persists. This change in conditions, which disfavors hydrogen bonds and thus demands more homology for nucleic acid strands to remain hybridized, is referred to as increasing the stringency of the reaction or, more commonly, as high stringency.

This discussion leads to another key term, the melting temperature, or as commonly abbreviated, Tm. If one takes two strands of DNA that share homology and hybridizes them, at any given time some of the strands will remain hybridized whereas others will have separated. This separation, as noted above, is referred to as denaturation. The ratio of hybridized/denatured DNA strands in the reaction will vary depending on the degree of homology as well as any condition that may affect hydrogen bonding between matched base pairs such as formamide concentration and temperature. The Tm or melting temperature is defined as that temperature under the specific reaction conditions where one-half of the hybridized strands are still hybridized and the other half are denatured. Tm will shift to lower temperatures if the conditions are adjusted to disfavor hydrogen bonding between matched base pairs. Two common and simple ways to favor denaturation of hybridize DNA and thus shift the Tm to lower temperatures are to add formamide and decrease the salt concentration, as is evident from the formula used to calculate the Tm:

Tm= 81.5°C
+ 16.6 log [Na+] (as molarity)
+ 0.41 (% G + C content) - 0.61 (% formamide)
- [500/# base pairs in the DNA/DNA hybrid]

For practical purposes, one usually deals with completely homologous, relatively long (>500 base pair) DNA strands under "standard conditions" (25 percent formamide and 100 mM Na+ with a GC content of 50 percent) where the Tm will be 69.9° C. Thus, one must bring the temperature substantially above 70°C in order to denature most of the DNA. This explains why denaturing is usually done at around 100°C and the temperature then brought to about 40° C. At 40° C, the hybridized probe and target will remain attached since they share strong homology. However, the probe will dissociate from unwanted segments of DNA at 40°C because the temperature will be much above Tm for such complexes (assuming 10/500 base pair matches between probe and non-target DNA, 25 percent formamide, 100 Mm Na+, and a GC content of 50 percent, the Tm would be 19.9° C). This is the basis of the superior specificity of molecular hybridization based analyses.

As an example to illustrate these important points, let us assume the probe is human papillomavirus (HPV 31) and the tissue contains both HPV 31 and HPV 52, which shares weak homology with HPV 31. Under low stringency conditions both HPVs would be evident, but at high stringency one would no longer see any evidence of the hybridization between HPV 31 and 52. This is to illustrate a simple but important point—the actual number of hybridized molecules between the probe and two different strands will be a function of the degree of homology and the relative number of the two distinct strands.

- **The Different Hybridization Assays**

Molecular biology is based on the concept of a known nucleic acid sequence hybridizing and thus identifying another nucleic acid strand in a given sample. The known strand is usually labeled and is called the probe. Standard labeling procedures include radioactive nucleotides such as ^{3}H, ^{35}S, ^{32}P, and nonisotopic tagged nucleotides that can be detected by a variety of methods including, to name just a few, fluorescence and colorimetic changes. The most common calorimetric reaction is based on biotin-labeled nucleotides that avidly bind streptavidin that, in turn, is conjugated to alkaline phosphatase, which can then catalyze the precipitation of a substrate that can easily be seen with light microscopy. The unlabeled strand in the sample being analyzed that is homologous to the probe is called the target.

A4. Detection of Probe/Target Complexes

Alkaline phosphatase-streptavidin is a detection system used for the localization of HPV 51 in a low-grade cervical squamous intraepithelial lesion (SIL). The signal is a blue precipitate that appears black in a black-and-white photograph, and marks the presence of alkaline phosphatase conjugated to streptavidin which, in turn, is bound to the biotin on the probe. It is to be noted that the virus is most abundant toward the surface of the lesion. This might be expected given that HPV is transmitted by sexual contact.

A5. Molecular Hybridization Techniques

Several techniques have evolved based on the ability of a labeled probe to bind to and thus permit the detection of the target nucleic acid sequence of interest. One approach is to extract the DNA, both target and non-target, from a sample and bind it to a filter where it can be hybridized with the labeled probe. This is called filter hybridization. Oftentimes the sample DNA is directly placed on the filter with the aid of a vacuum manifold, which has slot-like spaces for each sample, hence the term slot blot (or dot blot) hybridization. Alternatively, the sample DNA may first be separated according to size and configuration by electrophoresis on a gel and then transferred to a filter—this is termed Southern blot hybridization (Figure 1.13). In either of these techniques, the tissue must be destroyed thus precluding direct histological correlation. In the other major strategy based on hybridization of a target and the probe, the target DNA is not extracted but rather kept in the intact cell where it may bind to the probe. This, of course, is in situ hybridization.

As depicted in Table 1.1, an important difference between filter and in situ hybridization is their detection thresholds. Detection of a DNA sequence by in situ hydridization implies a selective increase in its numbers due to, for example, oncogene amplification or viral proliferation. However, only one virus need be present per every 100 cells for the Southern blot test to detect it, though the in situ test would be scored as negative. Why is there such a disparity in detection thresholds? There are probably several reasons. For one, the probe may find it more difficult to find the target if it has to transverse the labyrinth of nuclear proteins and nucleic acids in situ analysis relative to more "naked" DNA that has been attached to a filter in slot blot or Southern blot hybridization. Second, the extraction and purification of DNA characteristic of filter hybridization leads to a concentrating effect of rare nucleic acid sequences. The final assay based on molecular biological principles is, of course, the polymerase chain reaction (PCR). This is the subject of Section IC.

Table 1.1.Selected Features of Three Major Hybridization Assays

	In situ	Southern blot	Slot blot
Detection Threshold	20 copies/cell	1 copy/100	1 copy/200 cells
Samples	Fresh or fixed	Fresh	Fresh
Background	Low	Low	Low-high
Time	≤ day	2-7 days	1-5 days
Cell localization	Yes	No	No
Specialized equipment	None	Transfer unit Gel cast/ electrophoresis unit	Vacuum manifold
Samples/run	≤ 25	≤ 15	≤ 75
Detection latent infection	No	Yes	Yes

B. Recombinant DNA Technology (Genetic Engineering)

B1. Introduction and Basic Features

B1a. Introduction

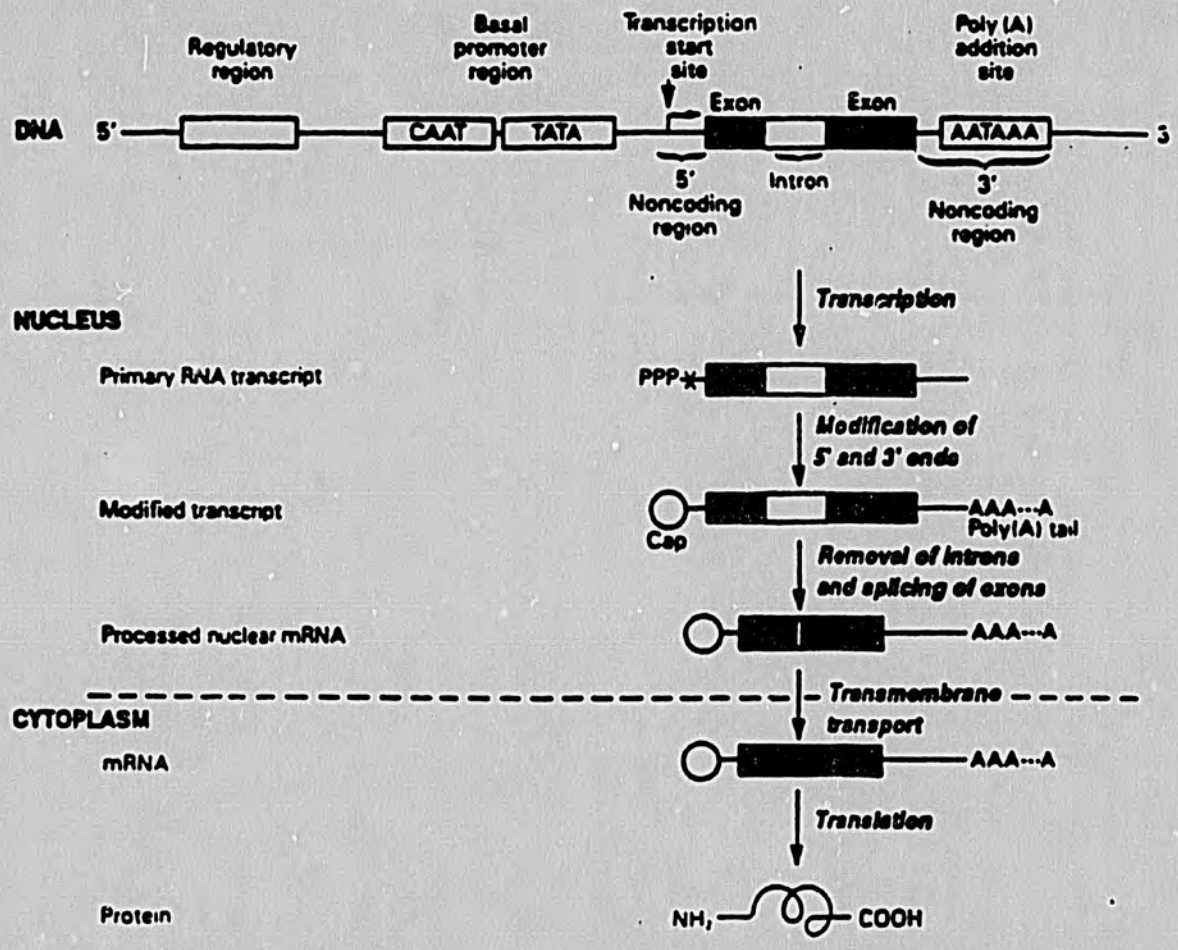
Recombinant DNA technology has revolutionized biology and is having an ever-increasing impact on clinical medicine. Much has been learned about human genetic disease from pedigree analysis and study of affected proteins, but in many cases where the specific genetic defect is unknown, these approaches cannot be used. The new technology circumvents these limitations by going directly to the DNA molecule for information. Manipulation of a DNA sequence and the construction of chimeric molecules, so-called genetic engineering, provides a means of studying how a specific segment of DNA works.

This section is aimed at clarifying this rather complex topic. It presents the basic concepts of recombinant DNA technology and its applications to clinical medicine. In order for this section to be complete, some repetition of subjects discussed in section IA will be found.

- **DNA Is Organized into Genes**

In general prokaryotic genes consist of a small regulatory region (100-500 bp) and a large protein coding segment (500-10,000 bp). Several genes are often controlled by a single regulatory unit. Most mammalian genes are more complicated, in that the coding regions are interrupted by noncoding regions that are eliminated when the primary RNA transcript is processed into mature messenger RNA (mRNA). The coding regions (those regions that appear in the mature RNA species) are called exons, and the non-coding regions, which interpose or intervene between the exons, are called introns (Figure 1.12). Introns are always removed from precursor RNA before transport into the cytoplasm occurs. The process by which introns are removed from precursor

Figure 1.12



Organization of a eukaryotic transcription unit and the pathway of eukaryotic gene expression. Eukaryotic genes have structural and regulatory regions. The structural region consists of the coding DNA and 5' and 3' noncoding DNA sequences. The coding regions are divided into 2 parts: (1) exons, which eventually are ligated together to become mature RNA, and (2) introns, which are processed out of the primary transcript. The structural region is bounded at its 5' end by the transcription initiation site and at its 3' end by the polyadenylate addition or termination site. The promoter region, which contains specific DNA sequences that interact with various protein factors to regulate transcription, is discussed in detail in Chapters 39 and 41. The primary transcript has a special structure, a cap, at the 5' end and a stretch of A's at the 3' end. This transcript is processed to remove the introns, and the mature mRNA is then transported to the cytoplasm, where it is translated into protein.

18

RNA and by which exons are ligated together is called RNA splicing. Incorrect processing of the primary transcript into the mature mRNA can result in disease in humans (see below); this underscores the importance of these post-transcriptional processing steps. Regulatory regions for specific eukaryotic genes are usually located in the DNA that flanks the transcription initiation site at its 5′ end (5′ flanking sequence DNA). Occasionally, such sequences are found within the gene itself or in the region that flanks the 3′ end of the gene. In mammalian cells, each gene has its own regulatory region. Many eukaryotic genes (and some viruses that replicate in mammalian cells) have special regions, called enhancers, that increase the rate of transcription. Some genes also have DNA sequences, known as silencers, that diminish transcription. Mammalian genes are obviously complicated, multicomponent structures.

- **Genes Are Transcribed into RNA**

Information generally flows from DNA to mRNA to protein, as illustrated in Figure 1.12. This is a rigidly controlled process involving a number of complex steps, each of which no doubt is regulated by one or more enzymes or factors; faulty function at any of these steps can cause diseases.

B1b. Recombinant DNA Technology Involves Isolation and Manipulation of DNA to Make Chimeric Molecules

Isolation and manipulation of DNA, including end-to-end joining of sequences from very different sources to make chimeric molecules (e.g., molecules containing both human and bacterial DNA sequences in a sequence- independent fastion), is the essence of recombinant DNA research. This involves several unique techniques and reagents.

- **Restriction Enzymes Cut DNA Chains at Specific Locations**

Certain endonucleases, enzymes that cut DNA at specific DNA sequences within the molecule (as opposed to exonucleases, which digest from the ends of DNA molecules), are a key tool in recombinant DNA research. These enzymes were originally called restriction enzymes because their presence in a given bacterium restricted the growth of certain bacterial viruses called bacteriophages. Restriction enzymes cut DNA into short pieces in a sequence specific manner, in contrast to most other enzymatic, chemical, or physical methods, which break DNA randomly. These defensive enzymes (hundreds have been discovered) protect the host bacterial DNA from DNA of foreign organisms (primarily infective phages). However, they are only present in cells that also have a companion enzyme that methylates the host DNA, rendering it an unsuitable substrate for digestion by the restriction enzyme. Thus, site-specific DNA methylases and restriction enzymes always exist in pairs in a bacterium.

Restriction enzymes are named after the bacterium from which they are isolated. For example Eco RI is from Escherichia coli, and Bam HI is from Bacilium amyloliquefaciens. The first three letters in the restriction enzyme name consist of the first letter of the genus (E) and the first two letters of the species (co). These may be followed by a strain designation (R) and a roman numeral (I) to indicate the order of discovery (e.g., Eco RI, RII). Each enzyme recognizes and cleaves a specific double-stranded DNA sequence that is 4-7 bp long. These DNA cuts result in blunt ends (HpaI) or overlapping (sticky) ends (Bam HI), depending on the mechanism used by the enzyme. Sticky ends are particularly useful in constructing hybrid or chimeric DNA molecules (see below). If the nucleotides are distributed randomly in a given DNA molecule, one can calculate

Figure 1.13

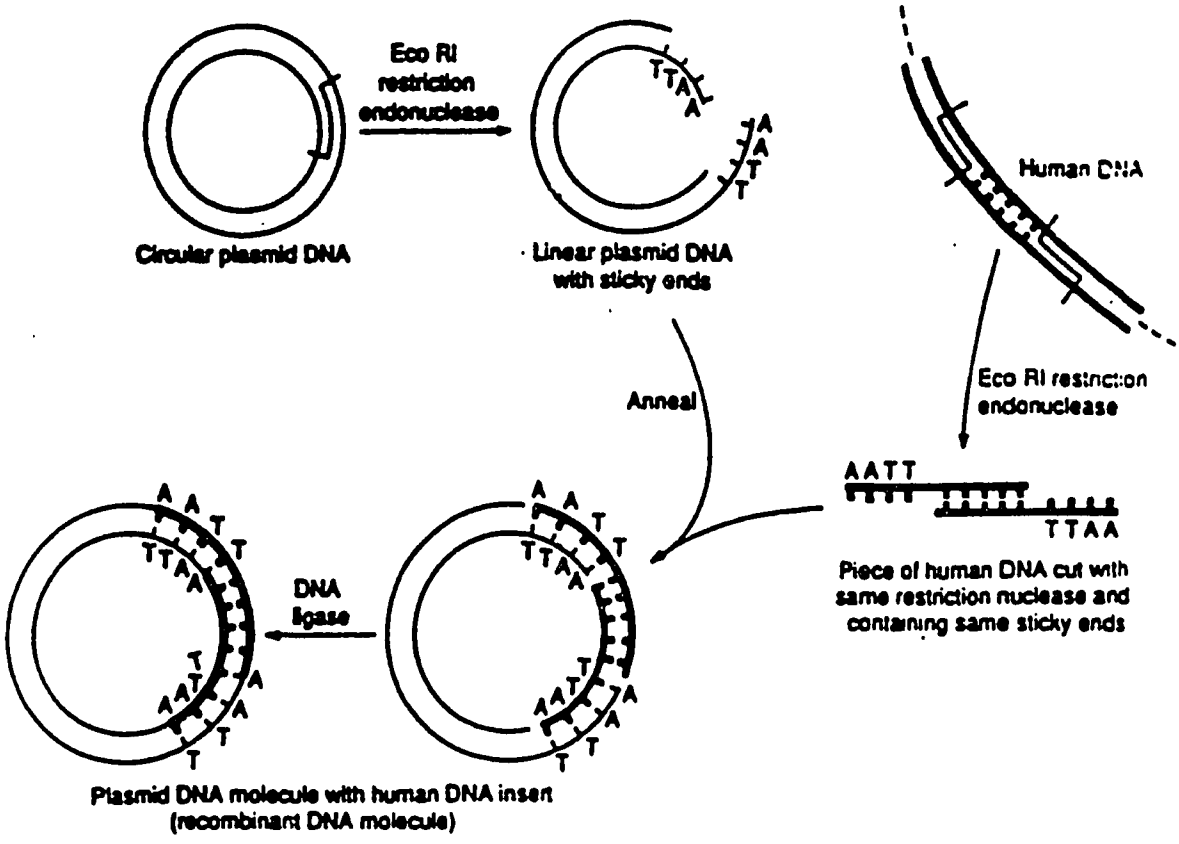


Figure 1.13 Use of restriction nucleases to make new recombinant or chimeric DNA molecules

When inserted back into a bacterial cell (by the process called transformation), the plasmid DNA replicates not only itself but also the physically lined new DNA insert. Since recombining the sticky ends, as indicated, regenerates the same DNA sequence recognized by the original restriction enzyme, the cloned DNA insert can be cleanly cut back out of the recombinant plasmid circle with this endonuclease. If a mixture of all of the DNA pieces created by treatment of total human DNA with a single restriction nuclease is used as the source of human DNA, a million or so different types of recombinant DNA molecules can be obtained, each pure in its own bacterial clone.

how frequently a given enzyme would cut a length of DNA. For each position in the DNA molecule there are 4 possibilities (A, C, G, or T); therefore, a restriction enzyme that recognizes a 4bp sequence will cut, on average, once every 256 bp(4^4), whereas another enzyme that recognizes a 6-bp sequence will cut once every 4096 bp (4^6). A given piece of DNA will have a characteristic linear arrange of sites for the various enzymes; hence, a restriction map can be constructed. When DNA is digested with a given enzyme, the ends of all the fragments will have the same DNA sequence. The fragments produced can be isolated by electrophoresis on agarose or polyacrylamide (see the discussion of blot transfer, below); this is an essential step in cloning and a major use of these enzymes. A number of other enzymes that act on DNA and RNA are an important part of recombinant DNA technology. Many of these are referred to in this and subsequent subsections.

- **Restriction Enzymes and DNA Ligase Are Used to Prepare Chimeric DNA Molecules**

Sticky-end ligation is technically easy, but some special techniques are often required to overcome problems inherent in this approach. Sticky ends of a vector may reconnect with themselves, with no net gain of DNA. Sticky ends of fragments can also anneal, so that tandem heterogeneous inserts form. Also sticky-end sites may not be available or in a convenient position. To circumvent these problems, an enzyme that generates blunt ends is used, and new ends are added using the enzyme terminal transferase, If poly d(G) is added to the 3′ ends of the vector and poly d(C) is added to the 3′ ends of the foreign DNA, the two molecules can only anneal to each other, thus circumventing the problems listed above. This procedure, called homopolymer tailing, also generates an Sma I restriction site, and so it is easy to retrieve the fragment. Sometimes, synthetic oligonucleotide linkers with a convenient restriction enzyme sequence are ligated to the blunt-ended DNA. Direct blunt-end ligation is accomplished using the enzyme bacteriophage T4 DNA ligase. This technique, though more difficult than sticky-end ligation, has the advantage of joining together any pairs of ends. The disadvantages are that there is no control over the orientation of insertion or the number of molecules annealed together and there is no easy way of retrieving the insert.

B2. Cloning Amplifies DNA

A clone is a large population of identical molecules bacteria, or cells that arise from a common ancestor. Cloning allows for the production of a large number of identical DNA molecules which can then be characterized or used for other purposes. This technique is based on the fact that chimeric or hybrid DNA molecules can be constructed in cloning vectors, typically bacterial plasmids, phages, or cosmids, which then continue to replicate in a host cell under their own control systems. In this way, the chimeric DNA is amplified. The general procedure is illustrated in Figure 1.13.

Bacterial plasmids are small, circular, duplex DNA molecules whose natural function is to confer antibiotic resistance to the host cell. Plasmids have several properties that make them extremely useful as cloning vectors. They exist as single or multiple copies within the bacterium and replicate independently from the bacterial DNA. The complete DNA sequence of many plasmids is known; hence, the precise location of restriction enzyme cleavage sites for inserting the foreign DNA is available. Plasmids are smaller than the host chromosome and are therefore easily separated from the latter, and the desired DNA is readily removed by cutting the plasmid with the enzyme specific for the restriction site into which the original piece of DNA was inserted.

Figure 1.14

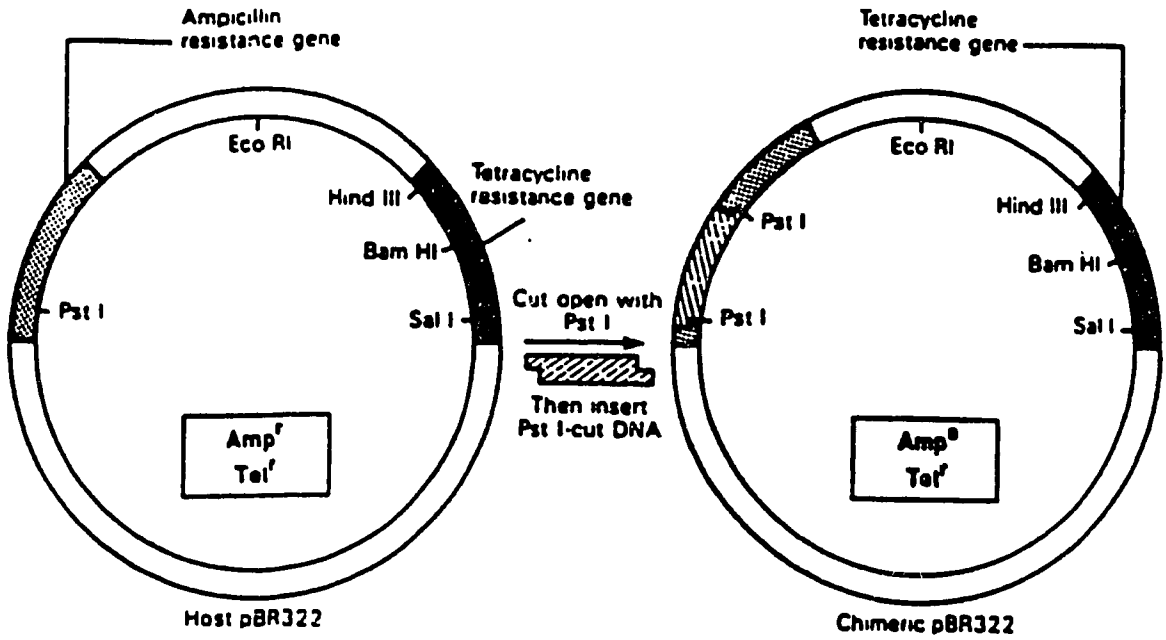


Figure 1.14 A method of screening recombinants for inserted DNA fragments

Using the plasmid pBR322, a piece of DNA is inserted into the unique PST 1 site. This insertion disrupts the gene coding for a protein that provides ampicillin resistance to the host bacterium. Hence, the chimeric plasmid will no longer survive when plated on a substrate medium that contains this antibiotic. The differential sensitivity to tetracycline and ampicillin can therefore be used to distinguish clones of plasmid that contain an insert.

22

Phages usually have linear DNA molecules into which foreign DNA can be inserted at several restriction enzyme sites. The chimeric DNA is collected after the phage proceeds through its lytic cycle and produces mature, infective phage particles. A major advantage of phage vectors is that while plasmids accept DNA pieces about 6-10 kb long, phages can accept DNA fragments 10- 20 kb long, a limitation imposed by the amount of DNA that can be packed into the phage head.

Even larger fragments of DNA can be cloned in cosmids, which combine the best features of plasmids and phages. Cosmids are plasmids the contain the DNA sequences, so-called cos sites, required for packaging lambda DNA into the phage particle. These vectors grow in the plasmid form in bacteria, but since much of the unnecessary lambda DNA has been removed, more chimeric DNA can be packaged into the particle head. It is not unusual for cosmids to carry inserts of chimeric DNA that are 35-50 kb long.

Insertion of DNA into a function region of the vector will interfere with the action of this region, and so care must be taken not to interrupt an essential function of the vector. This concept can be exploited, however, to provide a selection technique. The common plasmid vector pBR322 has both tetracycline (tet) and ampicillin (amp) resistance genes. A single Pst I site within the amp-resistance gene is commonly used as the insertion site for a piece of foreign DNA. In addition to having sticky ends), the DNA inserted at this site disrupts the amp-resistance gene and makes the bacterium carrying this plasmid amp-sensitive (Figure 1.14).

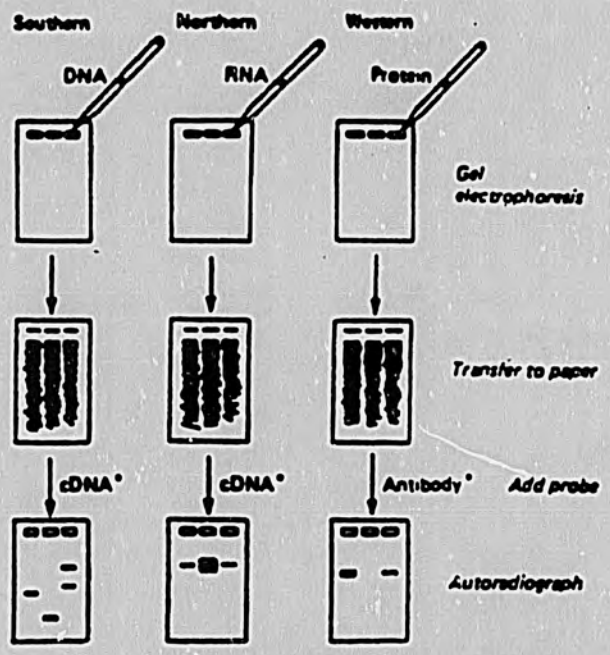
Thus, the parental plasmid, which provides resistance to both antibiotics, can be readily separated from the chimeric plasmid, which is resistant only to tetracycline. Additional confirmation that insertion has taken place comes from sizing the plasmid DNA obtained from the putative recombinant on an agarose gel, since the chimeric DNA molecule is now larger than the host vector DNA.

- **A Library Is a Collection of Recombinant Clones**

The combination of restriction enzymes and various cloning vectors allows the entire genome of an organism to be packed into a vector. A collection of these different recombinant clones is called a library. A genomic library is prepared from the total DNA of a cell line or tissue. A cDNA Library represents the population of mRNAs in a tissue. Genomic libraries are prepared by performing partial digestion of total DNA with a restriction enzyme that cuts DNA frequently (e.g., Sau IIIA). The idea is to generate rather large fragments, so that most genes will be left intact. Phage vectors are preferred for these libraries because they accept large pieces of DNA (up to 20 kb). The goal is to achieve a complete library. The number of fragments required to attain this objective is inversely related to fragment and directly related to genome size. human library that contains 106 recombinant fragments of large size has a 99 percent probability of being complete. Thus, the theoretical chances of finding any single-copy gene are excellent.

cDNA libraries are prepared by first isolating all the mRNAS in a tissue and then copying these mole-clues into double-stranded DNA, using (sequentially) the enzymes reverse transcriptase and DNA polymerase. For technical reasons, full-length cDNA copies are seldom obtained, and so smaller DNA fragments are cloned. Plasmids are often the favored vectors for cDNA libraries because they are much more convennt to work with than are phages or cosmids, although various lambda phage vectors have special advantages for cDNA cloning (see below).

Figure 1.15



The blot transfer procedure In a Southern, or DNA, blot transfer, DNA isolated from a cell line or tissue is digested with one or more restriction enzymes This mixture is pipetted into a well in an agarose or polyacrylamide gel and exposed to a direct electrical current DNA, being negatively charged, migrates toward the cathode the smaller fragments move the most rapidly After a suitable time, the DNA is denatured by exposure to mild alkali and transferred to nitrocellulose paper, in an exact replica of the pattern on the gel, by the blotting technique devised by Southern The DNA is annealed to the paper by exposure to heat, and the paper is then exposed to the labeled cDNA probe which hybridizes to complementary fragments on the filter After thorough washing, the paper is exposed to x-ray film, which is developed to reveal several specific bands corresponding to the DNA fragment that recognized the sequences in the cDNA probe. The RNA, or Northern blot is conceptually similar. RNA is subjected to electrophoresis before blot transfer This requires some different steps from those of DNA transfer, primarily to ensure that the RNA remains intact, and is generally somewhat more difficult The protein, or Western, blot, proteins are electrophoresed and transferred to nitrocellulose and then probed with a labeled antibody or other probe molecule

24

A vector in which the protein coded by the gene introduced by recombinant DNA technology is actually synthesized is known as an expression vector. Such vectors are now commonly used to detect specific cDNA molecules in libraries and to produce proteins by genetic engineering techniques. These vectors are specially constructed to contain very active inducible promoters, proper in-phase translation initiation codons, both transcription and translation termination signals, and appropriate protein processing signals, if needed. Some expression vectors even contain genes that code for protease inhibitors, so that the final yield of product is enhanced.

B3. Probes Search Libraries for Specific Genes of cDNA Molecules

A variety of molecules can be used to "probe" libraries in search of a specific gene of cDNA molecule or to define and quantitate DNA or RNA separated by electrophoresis through various gels. Probes are generally pieces of DNA or RNA labeled with a ^{32}P-containing nucleotide. The probe must recognize a complementary sequence to be effective. A cDNA synthesized from a specific mRNA can be used to screen either a cDNA library for longer cDNA or a genomic library for a complementary sequence in the coding region of a gene. A popular technique for finding specific genes entails taking a short amino acid sequence and, using the codon usage for that species making an oligonucleotide probe that will detect the corresponding DNA fragment in a genomic library. If the sequences match exactly, probes 15-20 nucleotides long will hybridize. cDNA probes are used to detect DNA fragments on Southern blot transfers and to detect and quantitate RNA on Northern blot transfers. Specific antibodies can also be used as probes, provided that the vector used synthesizes protein molecules that are recognized by them.

- **Blotting and Hybridization Techniques Allow Visualization of Specific Fragments**

Visualization of a specific DNA or RNA fragment among the many thousands of "contaminating" molecules requires the convergence of a number of techniques, which are collectively termed blot transfer. Figure 1.15 illustrates the Southern (DNA), Northern (RNA), and Western (Protein) blot transfer procedures. (The first is named for the person who devised the technique, and the other names began as laboratory jargon but are now accepted terms). These procedures are useful in determining how many copies of a gene are in a given tissue or whether there are any gross alternations in a gene (deletions, insertions, or rearrangements). Occasionally, if a specific base is changed and a restriction site is altered, these procedures can detect a point mutation. The Northern and Western blot transfer techniques are used to size and quantitate specific RNA and protein molecules, respectively.

Colony or plaque hybridization is the method by which specific clones are identified and purified. Bacteria are grown on colons on an agar plate and overlaid with a nitrocellulose filter paper. Cells from each colony stick to the filter and are permanently fixed thereto by heat, which with NaOH treatment also lyses the cells and denatures the DNA so that it will hybridize with the probe. A radioactive probe is added to the filter, and after washing, the hybrid complex is localized by exposing the filter to x-ray film. By matching the spot on the autoradiography to a colony, the latter can be picked from the plate. A similar strategy is used to identify fragments in phage libraries. Successive rounds of this procedure result in a clonal isolate (bacterial colony) or individual phage plaque.

All of the hybridization procedures discussed in this section depend on the specific base-pairing properties of complementary nucleic acid strands described above. Perfect matches

hybridize readily and withstand high temperatures in the hybridization and washing reactions. These complexes also form in the presence of low salt concentrations. Less than perfect matches do not tolerate these stringent conditions (i.e., elevated temperatures and low salt concentrations); thus hybridization either never occurs or is disrupted furing the washing step. Gene families, in which there is some degree of homology, can be detected by varying the stringency of the hybridization and washing steps. Cross species comparisons of a given gene can also be made using this approach.

B3a. Techniques Are Available to Determine the Sequence of DNA

The segments of specific DNA molecules obtained by recombinant DNA technology can be analyzed for their nucleotide sequence. This method depends upon having a large number of identical DNA molecules. This requirement can be satisfied by cloning the fragment of interest, using the techniques described above. The enzymatic (Sanger's) employs specific dideoxynucleotides that terminate DNA strand synthesis at specific nucleotides as the strand is synthesized on purified template nucleic acid. The reactions are adjusted so that a population of DNA fragments representing termination at every nucleotide is obtained. By having a radioactive label incorporated at the end opposite the termination site, one can separate the fragments according to size using polyacrylamide gel electrophoresis. An autoradiography is made, and each of the fragment produces an image (band) on an x-ray film. These are read in order to give the DNA sequence. Semi-automated DNA sequencing is now possible. Another method, that of Maxam and Gilbert, employs chemical methods to cleave the DNA molecules where they contain the specific nucleotides.

B3b. Oligonucleotide Synthesis Is Now Routine

The automated, chemical synthesis of moderately long oligonueleotides (100 nucleotides) of precise sequence is now a routine laboratory procedure. Each synthetic cycle takes a few minutes, so an entire molecule can be made by synthesizing relatively short segments that can then be ligated to one another. Olignucleotides are now indispensable for DNA sequencing, library screening, DNA mobility shift assays, the polymerase chain reaction (see below), and numerous other applications.

B4. Biomedical Importance of Recombinant DNA

Understanding recombinant DNA technology is important for several reasons: (1) the information explosion occurring in this area is truly staggering. To understand and keep up with this field, one must have an appreciation of the fundamental concepts involved; (2) it offers a rational approach to understanding the molecular basis of a number of diseases (e.g., familial hypercholesterolemia, sickle cell disease, the thalassemias, cystic fibrosis, muscular dystrophy); (3) using recombinant DNA technology, human proteins can be produced in abundance for therapy (e.g.. insulin, growth hormone, plasminogen activator); (4) proteins for vaccines (e.g., hepatitis B) and for diagnostic tests (e.g., AIDS test) can be obtained; (5) recombinant DNA technology is used to diagnose existing diseases and predict the risk of developing a given disease; (6) special techniques have led to remarkable advances in forensic medicine; and (7) gene therapy for sickle cell disease, the thalassemias, adenosine deaminase deficiency, and other diseases may be devised.

26

B5. Practical Applications of Recombinant DNA Technology

The isolation of a specific gene from an entire genome requires a technique that will discriminate one part in a million. The identification of a regulatory region that may be only 10 bp in length requires a sensitivity of one part in $3X10^8$, a disease such as sickle cell anemia is caused by a single base change, or one part in $3X10^9$. Recombinant DNA technology is powerful enough to accomplish all these things.

B5a. Gene Mapping Localizes Specific Genes to Distinct Chromosomes

Gene localizing thus can define a map of the human genome. This is already yielding useful information in the definition of human disease. Somatic cell hybridization and in situ hybridization are two techniques used to accomplish this. In situ hybridization, the simpler and more direct procedure, a radioactive probe is added to a metaphase spread of chromosomes on a glass slide. The exact area of hybridization is localized by layering photographic emulsion over the slide and, after exposure, lining up the grains with some histologic identification of the chromosome. This often places the gene at a location on a given band or region on the chromosome.

Hundreds of genes have been mapped. This human map will become more complete in ensuing years, and there are plans to sequence the entire human genome. The following conclusions can already be drawn: (1) genes that code for proteins with similar functions can be located on separate chromosomes (α- and β-globin). (2) Genes that form part of a family can also be on separate chromosomes (growth hormone and prolactin). (3) The genes involved in many hereditary disorders known to be due to specific protein deficiencies, including X chromosome-linked conditions, are indeed located at specific sites. Of most interest, perhaps, is the fact that because of the availability of defined and cloned restriction fragments, the chromosomal location for many disorders for which the protein deficiency is unknown is being defined, e.g., Huntington's chorea chromosome 4; cystic fibrosis chromosome 7; adult polycystic kidney disease, chromosome 16; and Duchenne-type muscular dystrophy, chromosome X. Once the defect is localized to a region of DNA that has the characteristic structure of a gene, a synthetic gene can be constructed and expressed in an appropriate vector and its function can be assessed, or the putative peptide, deduced from the open reading frame in the coding region, can be synthesized. Antibodies directed against this peptide can be used to assess whether this peptide is expressed in normal persons and whether it is absent in those with the genetic syndrome.

B5b. Proteins Can Be Produced for Research and Diagnosis

A practical goal of recombinant DNA research is the production of materials for biomedical application. This technology has two distinct merits: (1) it can supply large amounts of material that could not be obtained by conventional purification methods (e.g., interferon, plasminogen activating factor); and (2) tt can provide human material (e.g., insulin, growth hormone). The advantages in both cases are obvious. Although the primary aim is to supply products, generally proteins, for treatment (insulin) and diagnosis (AIDS test) of human and other animal diseases and for disease prevention (hepatitis B vaccine), there are other real and potential commercial applications, especially for agriculture. An example of the latter is the attempt to engineer plants that are more resistant to drought or temperature extremes or more efficient at fixing nitrogen.

27

B5c. Recombinant DNA Technology Is Used in the Molecular Analysis of Disease

Normal Gene Variations: There is a normal variation of DNA sequence just as there is with more obvious aspects of human structure. Variations of DNA sequence, polymorphism, occur approximately once in every 500 nucleotides, or about 107 times per genome. There are, no doubt, deletions and insertions of DNA as well as single-base substitutions. In healthy people, these alterations obviously occur in noncoding regions of DNA or at sites that cause no change in function of the encoded protein. This heritable polymorphism of DNA structure can be associated with certain diseases within a large kindred and can be used to search for the specific gene involved, as is illustrated below. It can also be used in a variety of applications in forensic medicine.

Gene Variations Causing Diseases: Classical genetics taught that most genetic diseases were due to point mutations which resulted in an impaired protein. This may still be true, but if on reading the initial subsections of this section one predicted that genetic disease could result from derangement of any of the steps illustrated in Figure 1.12, one would have made a proper assessment.

This point is nicely illustrated by an examination of the β-globin gene. Defective production of β-globin results in a variety of diseases and is due to many different lesions in and around the β-globin gene.

Point Mutations: The classic example is sickle cell disease, which is caused by mutation of a single base out of the $3X10^9$ in the genome, a T-to-A DNA substitution, which in turn results in an A-to-U change in the mRNA corresponding to the sixth codon of the β-globin gene. The altered codon specifies a different amino acid (valine rather than glutamic acid), and this causes a structural abnormality of the β-globin molecule. Other point mutations in and around the β-globin gene result in decreased or, in some instances, no production of β-thalassemia is the result of these mutations. (The thalassemias are characterized by defects in the synthesis of hemoglobin subunits, and so β-thalassemia results when there is insufficient production of β-globin.)

Deletions, Insertions, and Rearrangements of DNA: Studies of bacteria, viruses, yeasts, and fruit flies show that pieces of DNA can move from one place to another within a genome. The deletion of a critical piece of DNA, the rearrangement of DNA within a gene, or the insertion of a piece of DNA within a coding or regulatory region can all cause changes in gene expression resulting in disease. Again, a molecular analysis of β-thalassemia produces numerous examples of these processes, particularly deletions, as a cause of disease. The globin gene clusters seem particularly prone to this lesion. Deletions in the β-globin cluster, located on chromosome 16, cause β-thalassemia. There is a strong ethnic association for many of these deletions, so that northern Europeans, Filipinos, blacks, and Mediterranean peoples have different lesions all resulting in the absence of hemoglobin A and β-thalassemia.

A similar analysis could be made for a number of other diseases. Point mutations are usually defined by sequencing the gene in question, though occasionally, if the mutation destroys or creates a restriction enzyme site, the technique of restriction fragment analysis can be used to pinpoint the lesion. Deletions or insertions of DNA larger than 50 bp can often be detected by the Southern blotting procedure.

Pedigree Analysis: Sickle cell disease again provides an excellent example of how recombinant DNA technology can be applied to the study of human disease. The substitution of T

for A in the template strand of DNA in the β-globin gene changes the sequence in the region that corresponds to the sixth codon from

CCTGAGG	coding strand
GGACTCC	template strand

to

CCTGTGG	coding strand
GGACACC	template strand

This mutation destroys a recognition site for the restriction enzyme Mst II (CCTNAGG; denoted by the small vertical arrows). Other Mst II sites 5' and 3' from this site are not affected and so will be cut. Therefore, incubation of DNA from normal (AA), heterozygous (AS), and homozygoous (SS) individuals results in three different patterns on Southern blot transfer (Figure 1.16). This illustrates how a DNA pedigree can be established using the principles discussed in this section. Pedigree analysis has been applied to a number of genetic diseases and is most useful in those caused by deletions and insertions or the rare instances in which a restriction endonuclease cleavage site is affected , as in the example cited in this paragraph. The analysis is facilitated by the PCR reaction, which can provide sufficient DNA for analysis from just a few nucleated red blood cells.

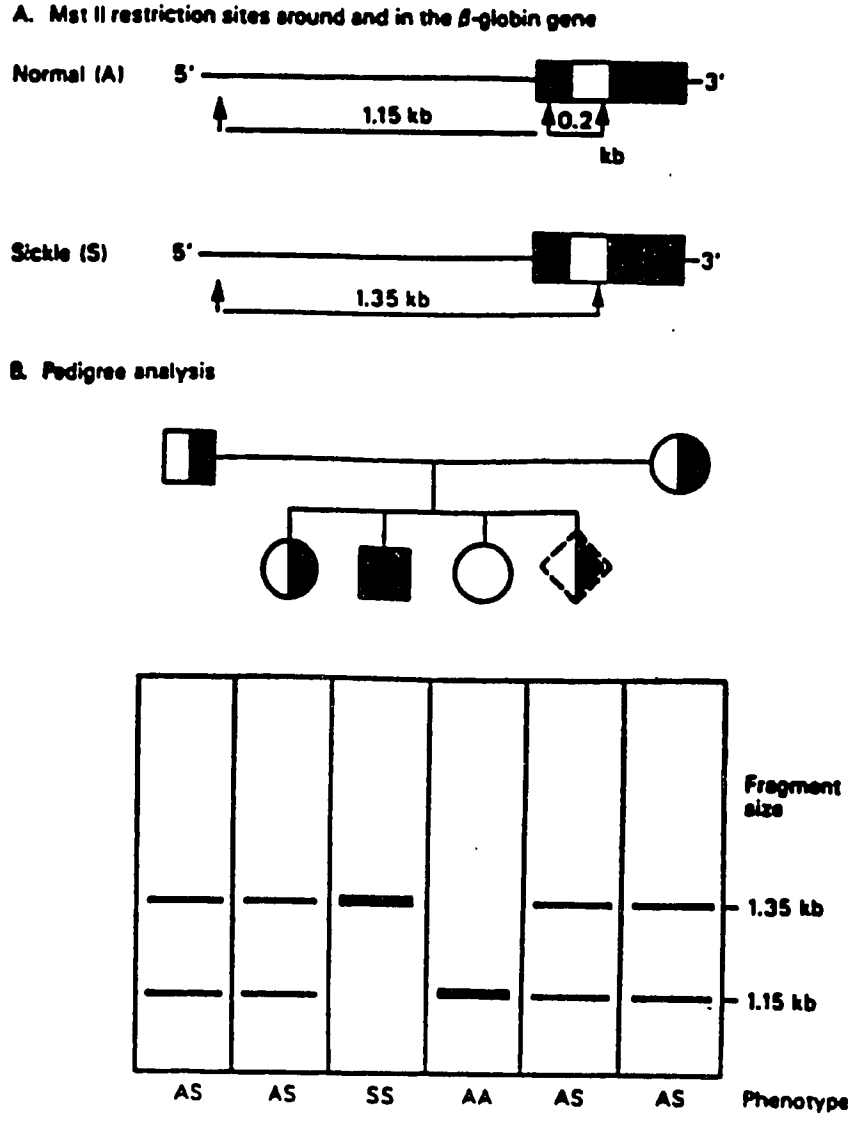
Prenatal Diagnosis: If the genetic lesion is understood and a specific probe is available, prenatal diagnosis is possible. DNA from cells collected from as little as 10 mL of amniotic fluid (or by chorionic villus biopsy) can be analyzed by Southern blot transfer. A fetus with the restriction pattern AA in Figure 1.16 does not have sickle cell disease, nor is it a carrier. A fetus with the SS pattern will develop the disease. Probes are now available for this type of analysis of many genetic diseases.

Restriction Fragment Length Polymorphism (RFLP): The differences in DNA sequence cited above can result in variations of restriction sites and thus in the length of restriction fragments. An inherited difference in the pattern of restriction (e.g., a DNA variation occurring in more than 1 percent of the general population) is known as a restriction fragment length polymorphism, or RFLP, RFLPs result from single-base changes (e.g., sickle cell disease) or from deletions or insertions of DNA into a restriction fragment (e.g., the thalassemias) and are proving to be a useful diagnostic tool. They have been found at known gene loci and in sequences that have known function: thus, RFLPs may disrupt the function of the gene or may have no biologic consequences.

RFLPs are inherited, and they segregate in a Mendelian fashion. A major use of RFLPs (thousands are now known) is in the definition of inherited diseases in which the functional deficit is unknown. RFLPs can be used to establish linkage groups, which in turn, by the process of chromosome walking, will eventually define the disease locus. In chromosome walking, a fragment representing one end of a long piece of DNA is used to isolate another that overlaps but extends the first. The direction of extension is determined by restriction mapping, and the procedure is repeated sequentially until the desired sequence is obtained.

29

Figure 1.16



Pedigree analysis of sickle cell disease. The top part of the figure (*A*) shows the first part of the β-globin gene and the Mst II restriction enzyme sites (T) in the normal (A) and sickle cell (S) β-globin genes. Digestion with the restriction enzyme Mst II results in DNA fragments 1.15 kb and 0.2 kb long in normal individuals. The T-to-A change in individuals with sickle cell disease abolishes one of the 3 Mst II sites around the β-globin gene, hence, a single restriction fragment 1.35 kb in length is generated in response to Mst II. This size difference is easily detected on a Southern blot (*B*). (The 0.2-kb fragment would run off the gel in this illustration.) Pedigree analysis shows 3 possibilities: AA = normal (○); AS = heterozygous (◐); SS = homozygous (■). This approach allows for prenatal diagnosis of sickle cell disease or trait (◐).

30

The X chromosome-linked disorders are particularly amenable to this approach, since only a single allele is expressed. Hence, 20 percent of the defined RFLPS are on the X chromosome, and a reasonably complete linkage map of this chromosome exists. The gene for the X-Linked disorder, Duchenne-type muscular dystrophy, has been found using RFLPs. Likewise, the defect in Huntington's chorea has been localized to the terminal region of the short arm of chromosome 4, and the defect that causes polycystic kidney disease is linked to the β-globin locus on chromosome 16.

Gene Therapy: Diseases caused by deficiency of a gene product are amenable to replacement therapy. The strategy is to clone a gene (e.g., the gene that codes for adenosine deaminase) into a vector that will readily be taken up and incorporated into the genome of a host cell. Bone marrow precursor cells are being investigated for this purpose because they presumably will resettle in the marrow and replicate there. The introduced gene would begin to direct the expression of its protein product, and this would correct the deficiency in the host cell.

This somatic cell gene replacement would obviously not be passed on to offspring. Other strategies to alter germ cell lines have been devised but have been tested only in experimental animals. A certain percentage of genes injected into a fertilized mouse ovum will be incorporated into the genome and found in both somatic and germ cells. These transgenic animals are proving to be useful for analysis of tissue-specific effects on gene expression and effects of overproduction of gene products (e.g., those from the growth hormone gene or oncongenes) and in discovering genes involved in development, a process that heretofore has been difficult to study. The transgenic approach has recently been used to correct a genetic deficiency in mice. Fertilized ova obtained from mice with genetic hypogonadism were injected with DNA containing the coding sequence for the gonadotropin-releasing hormone (GnRH) precursor protein. This gene was expressed and regulated normally in the hypothalamus of a certain number of the resultant mice, and these animals were in all respects normal. Their offspring also showed no evidence of GnRH deficiency. This is, therefore, evidence of somatic cell expression of the transgene and of its maintenance in germ cells.

B6. Applications of Genetic Engineering in Agriculture

- Genetic transformation of plants, e.g., cotton plants to make them resistant to cotton worms and other insects.
- Production of large amounts of enzymes, e.g., horseradish peroxidase to be used in commercial diagnostic kits.
- Genetic transformation of plants which cannot fix nitrogen, e.g., wheat, to allow them to fix nitrogen of air thus saving the use of fertilizers.
- Production of salt resistant plants which can grow in salty soil.
- Production of stress-resistant strains which can withstand unusual environmental conditions.
- Production of medical plants and plants with pharmacologic significance and modifications of their products.

- Changing the habitat of certain plants, e.g., allowing coca plants (tropical) to grow in the north pole.

C. Polymerase Chain Reaction (PCR)

C1. Introduction

In 1984, a team of scientists at Cetus Corporation (U.S.A.) developed a DNA amplification procedure based on an in vitro process; known as the Polymerase Chain Reaction (PCR). This method can produce large amounts of a specific DNA fragment from a complex DNA template in a simple enzymatic reaction.

PCR is characterized by the three Ss: selectivity, sensitivity, and speed. Virtually pure DNA fragments from complex genomes can be obtained in a matter of hours rather than the weeks or months traditional cloning requires. The method utilizes a DNA polymerase and two oligonucleotide primers to synthesize a specific DNA fragment from a single stranded template sequence. The amount of starting material needed for PCR can be as little as a single molecule rather than the usual millions of molecules required for standard cloning and molecular biological analysis. Although purified DNA is used in many applications, it is not required for PCR, and crude cell lysates also provide excellent templates. The DNA need not even be intact in contrast to the requirements of other standard molecular biological procedures, as long as some molecules exist that contain sequences complementary to both primers. The speed and sensitivity of PCR have been widely recognized by scientists in both medicine and basic biology, and the method has been applied to problems that a few years ago were thought to be inaccessible to molecular analysis.

C2. Terminology

Before beginning the discussion of how PCR works, certain terms need to be defined. Essential components of PCR are short segments of DNA called oligonucleotides (usually about 20 base pairs long). Because these short DNA sequences will serve as the "primers" or initiation points of DNA synthesis and amplification in PCR, they are commonly referred to as primers. The synthesis of oligonucleotides is a simple matter and several commercial laboratories will perform this service for a fee of about $100-$200 per 20 base pair sequence (unlabeled). In order to determine the sequence to choose to detect a given DNA molecule, one must know at least part of the sequence of that molecule. There is extensive literature detailing the partial and in many cases complete sequences of oncogenes, viruses, etc. Indeed, there is much discussion of an ambitious project to map out the entire sequence of the human genome.

Other important terms to discuss centers around the orientation of the native, double stranded DNA molecule. Each of the building blocks of DNA, the nucleotide, has two ends that participate in their joining. At one end is the triphosphate group; this end is referred to as the 5 prime ($5'$) end. The other end, referred to as the 3 prime ($3'$) end, has a hydroxyl (OH) group. When the nucleotides are joined together, the $5'$ triphosphate group of one joins with the $3'$ OH group of the other nucleotide, releasing pyrophosphate. The result is a dinucleotide that has a free $3'$ OH group on one side and a free $5'$ triphosphate group on the other side. The enzyme that catalyzes the synthesis of DNA, called DNA polymerase, can join only the free $5'$ triphosphate group of the single nucleotide to the $3'$ OH group of the growing chain. Because DNA polymerase can only function on the $3'$ OH group of the growing chain, DNA synthesis can only occur in a $5'$ to $3'$ direction. This is an important concept to understand in PCR. As mentioned

earlier, native DNA is composed of two complementary strands oriented in opposite directions in order to allow base pair matching to occur.

The final item to discuss is the DNA polymerase itself. In order for DNA polymerase to add a nucleotide to the 3' OH group of the growing chain, it must locate an area of transition from single stranded to double stranded DNA. The enzyme will "fill in the gap" and synthesize DNA using the single strand as its template.

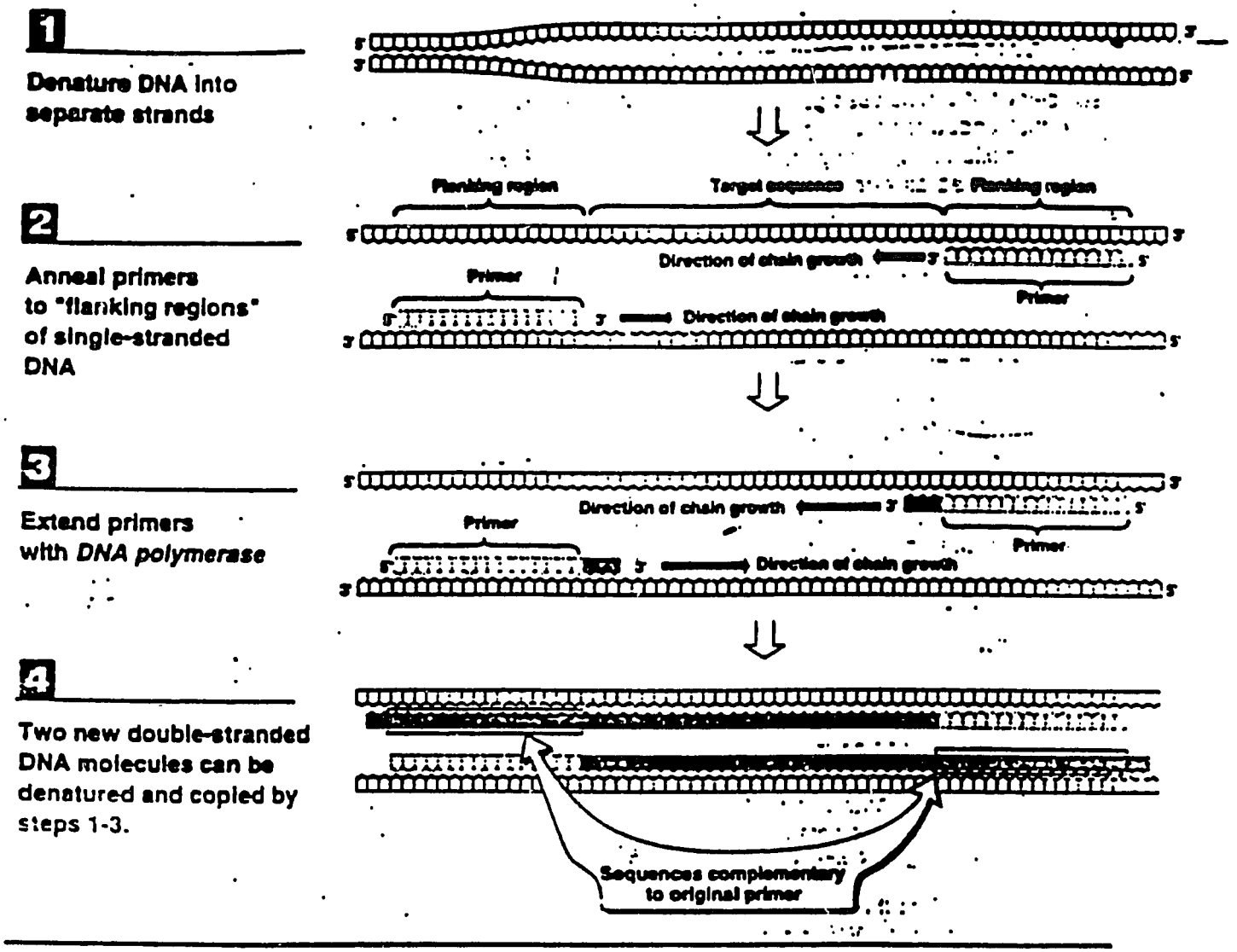
C3. PCR: The Theory

The essential reagents that will be needed for PCR include the DNA (double stranded) sample, the primers, the DNA polymerase with its essential cation Mg^2, a buffer (its exact composition will be discussed later), and the four nucleotide building blocks dATP, dCTP, dGTP, and dTTP, which we will collectively refer to as the dNTPs. You will recall that DNA polymerase cannot work on double stranded DNA. Thus, the first step is to denature the DNA. This can be done by increasing the temperature to 95°C. Even though there is typically no formamide in the reaction mixture, most of the DNA will denature because 95°C is sufficiently greater than the T_m of native double stranded DNA unless an excessive amount of salt is present in the sample. A problem becomes immediately apparent. How will the DNA polymerase withstand this high temperature? Most enzymes would be destroyed if exposed to such temperatures even for a very brief period of time. A major advance in PCR was the isolation of DNA polymerase that could withstand the high temperatures needed for denaturation. One source of this DNA polymerase was a bacterium that lives in hot springs called Thermus aquaticus. Hence, the term taq (DNA) polymerase (AmpliTaq, Perkin Elmer Corporation, Norwalk, Connecticut) came to be. This is not to say that taq polymerase cannot be destroyed by high temperatures, as will be discussed later. However, the enzyme can withstand the 95°C for the five minutes needed to denature the sample DNA and the repetitive 94°C used throughout the amplification process.

After the DNA is denatured, one more event must occur before DNA synthesis may be catalyzed by the enzyme. As mentioned earlier, in order for DNA polymerase to begin DNA synthesis, it must find an area of transition from single stranded to double stranded DNA. If the temperature remains at 95°C, most of the DNA will remain single stranded. However, if the reaction is brought to room temperature, then much of the DNA will renature, becoming double stranded again, which does not serve as a template for DNA polymerase. Clearly, it would be advantageous to reach an intermediate temperature where the competition of native DNA and primer DNA for the target favors the latter to allow the primers to initiate synthesis of the desired DNA sequence. At the proper temperature, which is typically around 55°C, sufficient primer-target DNA hybridization occurs, which leads to the synthesis of strands essential for the amplification step of PCR (Figure 1.17). The primers will seek out a region with similar homology. If a 20 base primer can find a region that is completely complementary, it will likely hybridize to it and remain hybridized at 55°C, this complex would most probably dissociate. But, what if an intermediate situation occurs where 15 of 20 base pairs are complementary. Will this complex remain hybridized? The answer is "maybe" and it will depend on factors such as the temperature and, as discussed earlier, the arrangement of the homologous base pairs (i.e., are they directly adjacent or dispersed throughout the sequence?).

Assume that the primer has found a single stranded region of the DNA and hybridized to it. DNA polymerase will find this region and rapidly begin to synthesize DNA and thus, extend the

Figure 1.17


1
Denature DNA into separate strands
2
Anneal primers to "flanking regions" of single-stranded DNA
Flanking region
Target sequence
Flanking region
Direction of chain growth
Primer
Primer
Direction of chain growth
3
Extend primers with DNA polymerase
Direction of chain growth
Primer
Primer
Direction of chain growth
4
Two new double-stranded DNA molecules can be denatured and copied by steps 1-3.
Sequences complementary to original primer

primer in a 5' to 3' direction (Figure 1.17, step 3). Of these two events the primer hybridization is the rate limiting one. The taq polymerase is very efficient and may add thousands of nucleotides in a complementary fashion to the template in a matter of a few minutes.

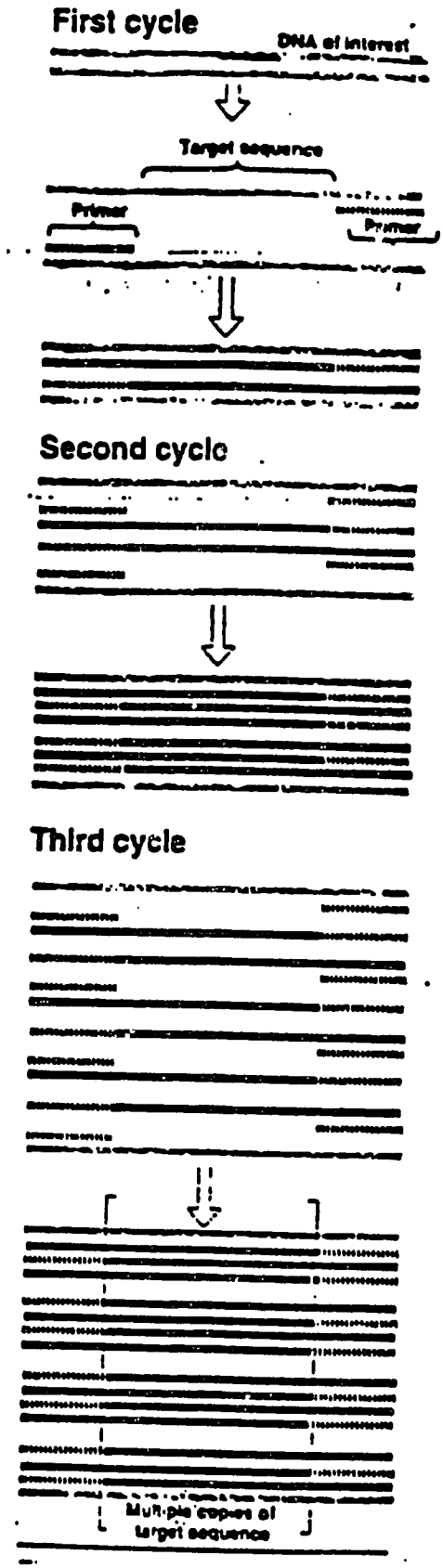
The next step is to separate the newly synthesized DNA from the template. This is easily accomplished by raising the temperature to 94°C for one minute. Note that there will be two newly synthesized segments of DNA (strands A' and B'), each having one of the two different primers at its 5' point of origin (Figure 1.17, step 4). If this process of DNA synthesis and denaturation continued we would get an arithmetic increase in the amount of these DNA segments. After 25 cycles, assuming only one DNA sequence that hybridized to the primers was available, we might be able to synthesize about 50 copies of the molecule. That is not good production and would be below the detection levels of conventional assays, such as Southern blot hybridization, typically used to detect the newly synthesized DNA. Furthermore, the newly synthesized fragment could include non-target sequences distal to the region of interest. What we need is a geometric and specific progression of DNA synthesis. This will be accomplished in part by carefully determining how far apart the primers are from each other in the DNA sequence we are attempting to amplify.

As seen in Figure 1.17, step 2, the primers are chosen such that each will be complementary to a region in either one or the other DNA strands (strands A' and B') made in the first cycle. Recall that strand A and B may each be over 1,000 base pairs long. When the temperature is raised to 94°C, the new strands separate from the template DNA. When the temperature is decreased to, for example, 55°C (called the annealing temperature), each of the primers may anneal to their complementary base pair sequences on strands A and B, respectively (Figure 1.17, step 2).

As seen in Figure 1.17, step 3, after the primers bind to strands A and B the taq polymerase will then catalyze the addition of nucleotides toward the other primer. Further, when the taq polymerase actually reaches the other primer DNA synthesis will stop. Notice what happens after these products are denatured. One has synthesized two new molecules (called strands C and D; Figure 1.18) that are the exact same size—the size is determined by the number of nucleotides between the two primers plus, of course, the number of nucleotides in the primers. When the temperature is brought back to the annealing range, the primers can find their complementary region at the 3' ends of strands C and D, respectively. The Taq polymerase will induce synthesis of copies of these two, strands. Thus, the two copies can yield four identical copies. In the next cycle these four copies can yield eight copies, and so on. The end result is the rapid amplification of the target DNA sequence located form mprimer 1 to primer 2.

Each denaturing/annealing sequence is referred to as a cycle. The absolute maximum of target-specific product that can be produced is 2^n where n is the number of cycles. Most experiments involve 20-30 cycles. Thus, at 33 cycles (cycle 3 is where application begins) one could theoretically produce 1,073,741,824 copies of just one DNA molecule! Although this number in itself is mind-boggling, the degree of amplification may appear even more impressive if one considers a typical PCR reaction that seeks to detect a virus in a human cell; for example, if one analyzes 100,000 cervical cells (small-sized biopsy) for human papillomavirus (HPV). Assume that only 100 viral particles are present in the entire tissue. Therefore, at the initiation of PCR there is roughly one viral segment of DNA for every 1 billion similarly sized fragment of human DNA! However, after about three hours and 40 cycles the amount of viral DNA may be greater than the amount of human DNA in the sample. Although the reaction does not work with 100 percent efficiency, it is routine to amplify as little as 0.01 pg of DNA to 100 ng (10^7 fold

Figure 1.18


First cycle
DNA of interest
Target sequence
Primer
Primer
Second cycle
Third cycle
Multiple copies of target sequence

amplification), which may then be visualized on an agarose or acrylamide gel stained with ethidium chloride.

A salient question at this stage is: do the primers find only the target DNA sequence that one wishes to amplify or do they sometimes "mistakenly" bind to non-target sequences? unfortunately, the primers may bind to non-target sequences. The question of PCR specificity brings up two issues that are illustrated in the following example. An attempt will be made to amplify HPV DNA from a sample of human cervical cells using 20 base pair primers that were selected because each is complementary to 20 base pair regions 500 nucleotides apart in the HPV 16 genome. The first issue is: how common is it that the human (non-target) DNA will contain a 20-base pair region with the exact complementary sequence? Actually, this is a very rare event. But even if it occurred there would be an extra protective device against nonspecific amplification. The resultant strand must also have a complementary region that is highly homologous to the other primer or geometric amplification would not occur. Second, can the 20 base pair HPV 16 primer bind to regions of the human genome where there is less than complete homology? The answer is yes, depending on the reaction conditions and, if the 3' end of the primer finds its base pair match, this bond may serve as the initiation point of DNA synthesis. We refer to such DNA synthesis as mis-priming. Another possible route for the primers is to bind themselves, which may seem like an unlikely event until one considers the relatively high concentration of primer present in the reaction mixture. Nonspecific amplification may then follow extension of primers onto one another's sequence to form primer oligomers; the latter process needs no additional DNA.

C4. The Methodology

Now that the theory behind PCR has been discussed, we will elaborate on the actual benchtop protocol. Many protocols have been published. This section will outline a common procedure with special reference to pitfalls and "helpful hints."

C4a. The Polymerase Chain Reaction

- **Tissue Preparation**

1. Cut one 4-10 μm paraffin-embedded section
2. Xylene 1 minute, centrifuge 5 minutes
3. 95 percent ETOH 1 minute, centrifuge 5 minutes
4. Dry, suspend in 180 l water at 60°C for 5 minutes
5. Add 20 μl of protease K (10 mg/ml) - 3 hr to 15 hr at 55°C
6. Inactivate protease K at 95°C for 10 minutes

- **PCR Reaction Mixture**

To each tube add:

1. 5 μl PCR buffer [as per the GeneAmp Kit (Perkin Cetus Elmer)]
2. 3 μl MgCI (stock solution 25 mM)
3. 6 μl dNTPs (stock solution 10 mM each)
4. 5 μl sample
5. 1 + 1 μl of primer 1 and primer 2 (stock solution 20 μM each)
6. 28.7 μl water
7. Overlay with 100 μl mineral oil
8. 0.3 μl taq polymerase (add taq after temperature reaches 55°C;)

- **Cycling Protocol**

1. Heating block to 94°C for 5 minutes
2. Cycling begins - 55°C for 2 minutes and 94°C for 1 minute
3. 25-40 cycles
4. 4°C until ready to process samples

- **Detection**

1. Electrophorese 25 l of reaction mixture
2. Southern blot hybridization analysis

The first step involves the extraction of the DNA from the test sample. For paraffin-embedded tissue, one 4- to 10-m section is cut into an Eppendorf tube. Actually, before this is done one needs to check the type of fixative used in the processing of the tissue. Fixative such as 10 percent neutral buffered formalin and 95 percent ethanol works well. However, fixatives that contain either picric acid (e.g., Bouin's solution) or heavy metals such as mercury (e.g., Zenker's solution) will have altered the DNA such that it may no longer be amplifiable. For tissues fixed in the desired manner, we recommend cutting just one section and using a disposable blade to reduce the possibility of one sample contaminating another. If multiple fragments are cut one runs the risk of not removing all the paraffin, which may inhibit the amplification. To remove the paraffin resuspend the tissue in xylene and vortex briefly, then centrifuge. The xylene may be decanted and residual xylene removed with 95 percent ethanol. The tissue, now free of paraffin, is then suspended in water.

Before continuing the discussion of the protocol, mention should be made of how to limit the problem of contamination when doing PCR. The extreme sensitivity of the technique can be its own worst enemy as contamination of a few target molecules into the sample may lead to a false- positive result. Our laboratory experience has taught us the utility of "dedicated" pipettemen, a separate laboratory work area restricted to PCR, and the practice of never working out of stock solutions. After making these changes, an initial rate of contamination in our laboratory of about 20 percent has been reduced to about 1 percent. We have not noted any utility in doing PCR in ventilation hoods or in using cotton-plugged pipette tips. This information is not presented as the final work on PCR contamination but rather just our own experience. Other laboratories may find utility in these latter procedural steps depending on their particular circumstances. Finally, brief mention should be made of Uranil-N-glycosidase (UNG). This is an enzyme that destroys any DNA into which has been incorporated dUTP. A common source of contamination in PCR is the target DNA synthesized in previous reactions (i.e., cross-over contamination). By including dUTP in the reaction mixture any cross-over DNA contaminant would be destroyed by the action of the enzyme UNG. Unfortunately, this system would not prevent a false-positive from native, non-dUTP containing DNA that might be inadvertently included in the reaction mixture. This may be a particular problem for DNA sequences commonly present such as when one analyzes lower genital tract tissues for HPV.

C5. Detecting the Amplified Product

The end product of PCR is, of course, newly synthesized DNA. Some (hopefully most) of this will be the target of interest but, under standard conditions, non-specific synthesis from mis-priming and primer oligomerization will also occur. Several methods are available for detecting DNA. The simplest way is to use electrophoresis and a gel, usually agarose. In this way, one can separate the DNA according to its size. This is well suited to PCR because one attempts to generate large numbers of DNA fragments of the same size. A common way to visualize the DNA on the gel directly is to use intercalating agents such as ethidium bromide, which fluoresce with ultraviolet light. One can also use known size standards and thus determine if the size of the amplified fragment is as expected from the known distance between the primers. However, even fragments of the expected size are not necessarily the amplified target DNA. Further, amplification may occur with less than optimal efficiency such that the final amount of synthesized DNA is below the detection threshold of direct visualization using ethidium bromide (about 50 ng of DNA). For these reasons, more sensitive techniques are often employed.

The two commonly used techniques are those discussed in the previous section under filter hybridization—slot blot and Southern blot hybridization analyses. As outlined in Section IB, one can detect as little as 0.1 pg of DNA with these methodologies. Typically, one chooses as the probe a labeled segment internal to the primers that will detect the amplified product but not other non-specific DNA synthesized as the result of mis-priming or primer oligomerization. The higher sensitivity of filter hybridization may detect amplified DNA not evident with ethidium staining given that the latter technique requires about 50 ng of DNA for detection. We prefer Southern blot hybridization (Figure 1.19) because it gives one the added assurance of a hybridization signal with the size of the amplified DNA. Slot blot hybridization is much more prone to false-positive results because there may be so much DNA concentrated on the filter that a signal may be evident even though the DNA may be derived from nonspecific amplification and share relatively poor homology with the probe. Protocols for slot blot or Southern blot hybridization are easily found and kits are commercially available to do either technique. In our experience we have found

Figure 1.19

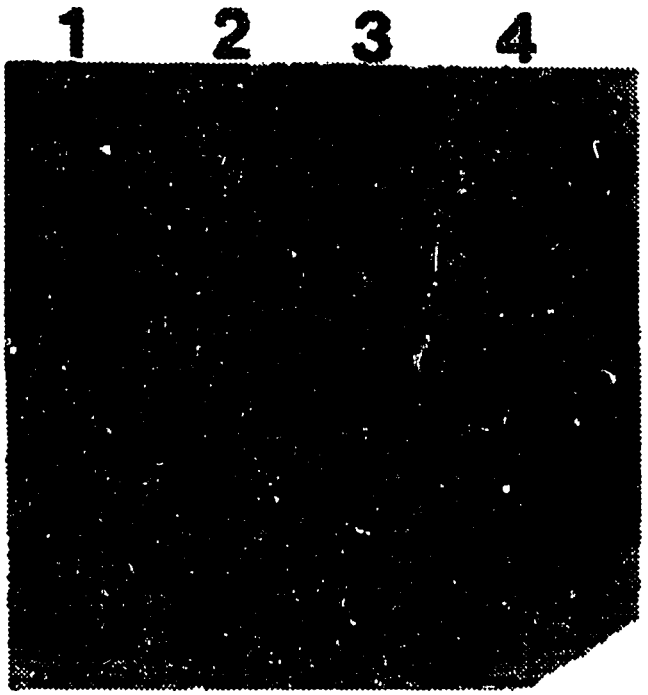


Figure 3. Southern blot analysis of initial PCR products. The PCR products from *Figure 1* were excised from the gel as four size fractions purified through Spin-X columns and were differentially amplified as shown in *Table 2*. 10 ng of each size selected class was separated on a 1% Nusieve gel; lane 1, <500 bp; lane 2, 500–1500 bp; lane 3, 1500–2000 bp; lane 4, >2000 bp. The gel was Southern blotted and the filter probed with hexaprime-labelled insert of clone 2, a highly abundant, constitutively expressed gene of transcript size 1150 bp. Strongest hybridization to the probe after stringent washing (65°C; 0.5 × SSC, 0.1% SDS) is seen in lane 2 corresponding to the PCR products of size 500–1500 bp.

commercially available kits to be easy to use and capable of giving highly reproducible results (e.g., Southern blot kit, ONCOR).

Another strategy for the detection of the DNA synthesized during PCR is direct incorporation of labeled nucleotides that can then be detected directly on a filter without the need for a probe-hybridization step. Because of their much longer shelf life and ease of disposability, much effort has been given to direct incorporation with nonisotopic nucleotides. Our efforts have focused on biotin- and digoxigenin-labeled dATP and dUTP, respectively (ONCOR and Boehringer Mannheim, Indianapolis, Indiana, respectively). With direct incorporation of either of these nucleotides, only one needs to transfer the PCR-synthesized DNA to a filter and then detect the biotin or digoxigenin by the use of alkaline phosphate-conjugated complexes of streptavidin or anti-digoxigenin, respectively. Pathologists especially like this procedure because detection of alkaline phosphatase-linked complexes is an everyday procedure for pathologists in immunohistochemistry laboratories. One may employ colorimetric detection by the use of a chromagen such as nitroblue tetrazolium (NBT) which, in the presence of 5-bromo-4-chloro-3-indolylphosphate (BCIP), yields a blue precipitate, to cite just one commonly used detection method. One may readily appreciate that direct detection of labeled nucleotides may give false-positive results unless actions have been taken to prevent or, at least greatly inhibit, the nonspecific pathways. Marked inhibition of nonspecific DNA synthesis is now possible by the use of hot start PCR and this modification has revealed fascinating and important practical and theoretical implications of the workings of PCR.

C6. Optimization of PCR

Polymerase chain reaction is an ingenious new tool for molecular biology that has had an effect on research similar to that of the discovery of restriction enzymes and the Southern blot. PCR is so sensitive that a single DNA molecule has been amplified and single-copy genes are routinely extracted out of complex mixtures of genomic sequences and visualized as distinct bands on agarose gels. PCR can be utilized for rapid screening and/or sequencing of inserts directly from aliquots of individual phage plaques or bacterial colons. Enhancements, such as the use of thermostable DNA polymerases and automation, of the method invented by Kary Mullis, have fostered the development of numerous and diverse PCR applications throughout the research community. Unquestionably, no single protocol will be appropriate to all situations. Consequently, each new PCR application is likely to require optimization. Some often encountered problems include: no detectable product or a low yield of the desired product; the presence of nonspecific background bands due to mispriming or misextension of the primers; the formation of "primer-dimer" that compete for amplification with the desired product; and mutations or heterogeneity due to misincorporation.

C6a. Enzyme Concentration

A recommended concentration range for Taq DNA polymerase is between 1 and 2.5 units (SA= 20 units/pmol.) per 100 μl reaction when other parameters are optimum. However, enzyme requirements may vary with respect to individual target templates or primers. When optimizing a PCR, we recommend testing enzyme concentration ranging from 0.5 to 5 units/100 μl and assaying the result by gel electophoresis. If the enzyme concentration is too high, non-specific background products may accumulate, and if too low, an insufficient amount of desired product is made. Note: Taq DNA polymerase from different suppliers may behave differently because of the different formulations, assay conditions, and/or unit definitions.

C6b. Deoxynucleotide Triphosphates

Stock dNTP solution should be neutralized to pH 7, and their concentrations should be determined spectrophotometrically. Primary stocks are diluted to 10 mM, aliquoted, and stored at -20°C. A working stock containing 1 mM of each dNTP is recommended. The stability of dNTP's during repeated cycles of PCR is such that approximately 50 percent remains as dNTP after 50 cycles.

Deoxynucleotide concentrations between 20 and 200 μM each result in the optimal balance among yield, specificity and fidelity. The four dNTP's should be used at equivalent concentrations to minimize misincorporation errors. Both the specificity and the fidelity of PCR are increased by using lower dNTP concentrations than those originally recommended for Klenow-mediated PCR (1.5 mM each). Low dNTP concentrations minimize mispriming at non-target sites and reduce the likelihood of extending misincorporated nucleotides.

One should decide on the lowest dNTP concentration appropriate for the length and composition of the target sequence, e.g. 20 μM each dNTP in a 100 μl reaction is theoretically sufficient to synthesize 2.6 μg of DNA or 10 pmol of a 400 bp sequence. Recently, the use of low uniform dNTP concentrations (2 μM each) enabled highly sensitive [1/107], allele specific amplification of ras point mutations.

C6c. Magnesium Concentration

It is beneficial to optimize the magnesium in concentration. The magnesium concentration may affect all of the following : primer annealing, strand dissociation temperature of both template and PCR product, product specificity, formation of primer-dimer artifacts, and enzyme activity and fidelity. Taq DNA polymerase requires free magnesium on top of that bound by template DNA, primers, and dNTPs. Accordingly, PCRs should contain 0.5 to 2.5 mM magnesium over the total dNTP concentration. Note that the presence of EDTA or other chelators in the primer stocks or template DNA may disturb the apparent magnesium optimum.

C6d. Other Reaction Components

A recommended buffer for PCRR is 10 to 50 mM Tris-HCl (between pH 8.3 and 8.8) when measured at 20°C. However, an extensive survey of other buffers has not been performed. Tris is a dipolar ionic buffer having a pKa of 8.3 at 20°C and a pKa of -0.021/°C. Thus, the pH of 20 mM Tris (pH 8.3) at 20°C varies between 7.8 and 6.8 during typical thermal cycling conditions. Up to 50 mM KCl can be included in the reaction mixture to facilitate primer annealing.

NaCl at 50 mM or KCl above 50 mM, inhibits Taq DNA polymerase activity. Gelatin or bovine serum albumin [100 μg /ml] and non ionic detergents such as tween 20 or Laureth 12 [0.05 to 0.1 percent; Mazer chemicals, Gurnee, Illinois] are added to help stabilizing the enzyme, although many protocols work well without added protein

C6e. Primers

Primer concentrations between 0.1 and 0.5 μM are generally optimal. Higher primer concentrations may promote mispriming and accumulation of nonspecific product and may increase the probability of generating a template-independent artifact termed a primer-dimer.

Nonspecific products and primer-dimer artifacts are themselves substrates for PCR and compete with the desired product for enzyme, dNTPs, and primers, resulting in a lower yield of the desired product.

Some simple rules aid in the design of efficient primers. Typical primers are 18 to 28 nucleotides in length having 50 to 60 percent G+C composition. The calculated strand-melting temperature (Tm[5]) for a given primer pair should be balanced. For this purpose, one can use the rule of thumb calculation of 2°C for A or T and 4°C for G or C. Depending on the application avoid complementarily at the 3' ends of primer pairs as this promotes the formation of primer-dimer artifacts and reduces the yield of the desired product. Also, runs (three or more) of C's or G's at the 3' ends of primers may promote mispriming at G+C rich sequences and should be avoided, when possible, as should palindromic sequences within primers. If all else fails, it usually helps to try a different primer pair. A less obvious reason for some primers failing to work is the presence of secondary structure in the template DNA. In this case, substitution of 7 deaza-2' deoxy GTP has been very useful.

C6f. Primer Annealing

The temperature and length of time required for primer annealing depend upon the base composition, length, and concentration of the amplification primers. An applicable annealing temperature is 5°C below the true Tm of the amplification primers. Because Taq DNA polymerase is active over a broad range of temperatures, primer extension will occur at low temperatures, including the annealing step. The range of enzyme activity varies by two orders of magnitude between 20 and 85°C. Annealing temperature in the range of 55 to 72°C generally yield the best results. At typical primer concentrations (0.2 μM), annealing will require only a few seconds. Increasing the annealing temperature enhances discrimination against incorrectly annealed primers and reduces misextension of incorrect nucleotides at the 3' ends of primers. Therefore, stringent annealing temperatures, especially during the first several cycles will help to increase specificity. For maximum specificity in the initial cycle, Taq DNA polymerase can be added after the first denaturation step during primer annealing. Low extension temperature together with high dNTP concentrations favours misextension of primers and extension of misincorporated nuleotides. For these two reasons, some investigators have argued that PCRs should perform better using longer primers and only two temperatures; e.g., from 55 to 75°C for annealing and extension, and 94 to 97°C for denaturation and strand separation.

C6g. Primer Extension

Extension time depends upon the length and concentration of the target sequence and upon temperature. Primer extensions are traditionally performed at 72°C because this temperature was near optimal for extending primers on an MB-based model template. Estimates for the rate of nucleotide incorporation at 72°C vary from 35 to 100 nucleotides second^{-1} depending upon the buffer, pH, salt concentration, and the nature of the DNA template. An extension time of one minute at 72°C is considered sufficient for products up to 2 Kb in length. However, longer extension times may be helpful in early cycles if the substrate concentration is very low, and at late cycles when product exceeds enzyme concentration (approximately 1nM).

C6h. Denaturation Time and Temperature

The most likely cause for failure of a PCR is incomplete denaturation of target templet and/or the PCR product. Typical denaturation conditions are 95°C for 30 seconds, or 97°C for

15 seconds, however, higher temperatures may be appropriate, especially for G+C-rich targets. It only takes a few seconds to denature DNA at its strand separation temperature (T^{ss}); however, there may be lag time involved in reaching T^{ss} inside the reaction tube.

Incomplete denaturation allows the DNA strands to snap back and, thus, reduces product yield. In contrast, denaturation steps that are too high and/or too long lead to unnecessary loss of enzyme activity. The half-life of Taq DNA polymerase activity is > 2 hours, 40 minutes, and 5 minutes at 92.5, 95, and 97.5°C, respectively.

C6i. Plateau Effect

The term "plateau effect" is used to describe the attenuation in the exponential rate of product accumulation of 0.3 to 1 pmol of the intended product. Depending on reaction conditions and thermal cycling, one or more of the following may influence plateau: (1) utilization of substrates (dNTPs or primers); (2) Stability of reactants (dNTPs or enzyme); (3) end product inhibition (pyrophosphate, duplex DNA); (4) Competition for reactants by nonspecific products or primer-dimer; (5) reannealing of specific product at concentrations above 10-8M (may decrease the extension rate or processivity of Taq DNA polymerase or cause branch-migration of product starnds and displacement of primers); and (6) incomplete denaturation/starnd separation of product at high product concentration. An important consequence of reaching plateau is that an initially low concentration of nonspecific products resulting from mispriming events may continue to amplify preferentially. Optimizing the number of PCR cycles is the best way to avoid amplifying background products.

C7. Practical Considerations

While PCR provides a very powerful tool for molecular analysis, the investigator should be aware of some aspects of the procedure that should be considered in the design and interpretation of experiments.

C71. Misincorporation

The rate of nucleotide misincorporation during the PCR has been estimated in various studies by determining the frequency of misincorporated nucleotides in the sequence of cloned PCR products and calculating an average rate per cycle. The misincorporation rate would be expected to reflect not only the properties of the polymerase itself (e.g., presence or absence of 3'-5' exonuclease proofreading activity), but also the reaction conditions. Biochemical fidelity measurements of "nonproofreading polymerase" (e.g., AMV reverse transcriptase and Drosophila melanogaster DNA polymerase α) revealed that the nucleotide misincorporation is dependant upon the dNTP concentration. The initial estimate for nucleotide misincorporation (1.7×10^{-4} nucleotide polymerized per cycle) by Taq polymerase during PCR used 1.5 mM of each dNTP, 10 mM Mgcl2, and a 37°C annealing temperature to amplify a 272-base pair fragment from the HLA-DPB1 gene. More recent studies gave an average nucleotide misincorporation rate of 5×10^{-6} per cycle. These studies used lower dNTP AND $MgCl_2$) as well as higher annealing temperatures (54-55°C). In general, the rate of misincorporation as well as that of extension from a 3' mispaired primer may be reduced by minimizing the dNTP and $Mgcl_2$ concentrations. However, reaction conditions that maximize fidlty may reduce PCR efficiency. Also, although the calculations yield an "average" misincorporation rate per cycle, the fidelity of thermostable polymerases may vary during the PCR. The misincorporation rate of thermostable polymerases with 3'-5' exonuclease proofreading activity is likely to be lower.

For most PCR applications, such as direct sequencing or characterization probes or restriction enzyme digestions, it is the population of amplified products that are analyzed and therefore, rare misincorporated nucleotides do not pose any problem in the interpretation of experiments. The actual fraction of amplified molecules that contain a mutation is related to the number of bases in each target molecule, the rate of misincorporation per base per cycle, and the number of cycles. A mathematically precise theory for estimating the frequency of molecules without any misincorporation has recently been put forward. Using the most recent average estimate of misincorporation ($8.5 \times 10_{-6}$), only 3 percent of the molecules of a 200 base-pair PCR product will contain a misincorporated base somewhere in the sequence after 30 cycles. Of course most of the molecules with misincorporated nucleotides would not contain the same mutation.

C72. Hybrid PCR Products

When the sample being amplified is heterozygous (two different alleles for the target locus) or when many different but related sequences are amplified with the same primers (e.g., a multigene family), in vitro recombination or template-strand switching is theoretically possible. Although a very rare event, PCR hybrid sequences can be formed when a primer that has been partially extended on one of the templates is annealed and extended on a similar template in a subsequent cycle. For this so-called "jumping" to occur, the concentration of partially extended PCR primers must be quite high.

The model experiments that studied this phenomenon with plasmid DNA, however, used high concentrations of template, thereby increasing the likelihood that the level of partially extended primers would be sufficient to promote jumping. The likelihood of partial primer extension is also dependent on the distance between the primers, secondary structure, the time allowed for polymerase extension, the processivity of polymerase and the extent to which the original target DNA sequences contained single strand breaks. Again, most methods of analysis characterize the population of PCR products, not rare individual hybrid products. In general, the formation of hybrid molecules can be minimized if the minimum of PCR cycles required for product detection are carried out. Under some circumstances jumping PCR may be beneficial. Assume a sample of DNA is highly degraded and no completely intact target molecules exist. If the DNA contained a set of partially overlapping fragments from the target, then repeated jumping could eventually produce a full-length PCR product.

C73. Preferential Amplification

Under some circumstances preferential amplification of one allele relative to the other can occur in heterogenous samples. One possible mechanism for differential priming is that one of the two alleles may have a rare DNA sequence polymorphism in the primer-binding site. Another potential mechanism is that one allele could have a higher GC content, and therefore require a higher denaturation temperature to generate functional PCR template than does the other allele. In the case of length polymorphism detected as PCR products if different size, the smaller allele can, under some circumstances, be amplified preferentially. In addition, one allele in a heterozygous sample can be amplified preferentially due to stochastic fluctuation when sampling a very low number of target molecules.

C7d. Contamination

The main drawback of the PCR is its sensitivity, as clinical or laboratory generated contamination by environmental DNA, cloned plasmid and PCR products may produce false-positive results. However, amplification products are the most serious source of contamination. Amplified DNA from a single sample can contain several picomoles of PCR product. Even one millionth of that product could have disastrous consequences if it contained a 1-microgram sample of genomic DNA containing 300,000 copies of unique sequence. When carrying out PCR experiments, physically separating the areas in the laboratory where PCR reactions are set up from those where the PCR products are analyzed can be of significant help.

It is strongly suggested that an area such as a bacterial hood with a UV light be dedicated to the preparation of PCR reagents and amplification of samples. If a fume or laminar flow hood is used, the fan should be turned off, since the air flow is a potential source of sample contamination.

All pipetters, pipettes, bulbs, tubes, racks, etc., must be designated for PCR preparation only and kept in this hood or clean area at all times. These instruments should never be used to pipette PCR product DNA, plasmid DNA or other potentially contaminating sources. Reagent solutions should be aliquoted. Micropipette tips with built-in filters can also be useful. Surfaces and instruments should be cleaned frequently with bleach or hypochlorite. Minuscule amounts of amplified target sequences can cause carry-over contamination and invalidate an entire study.

These "PCR clean" techniques should also be used at the clinic or hospital where samples are collected. Controls are essential for monitoring contamination, reproductibility and sensitivity of the assays. Negative controls that mimic specimens (that is, containing human cells or human DNA without target DNA) should be interspersed throughout sample manipulation. Positive controls with low amounts of target DNA should be proceeded to determine the analytical sensitivity. Recently, several protocols for eliminating contaminating DNA sequences carried over from previous reactions have been developed. U.V. irradiation of the reaction tube contents can damage any contaminating sequences before the DNA template and polymerase are added. Before adding the sample template into a PCR mixture, the contents can be exposed to short-wave (254 and 300 nm together or just 254 nm) UV radiation. Such an attack should sufficiently nick and cross link any contaminating sequences rendering them unamplifiable. The disadvantage of this procedure is that the primers and Taq polymerase are adversely affected by long exposure UV, resulting in a subsequent loss of sensitivity the reaction. If primers can be ruled out as a source of contamination then they need not be exposed for long periods. Following 80 min of UV radiation exposure, up to 25ng of plasmid DNA cannot be detected by 40 cycles of PCR amplification. Shorter exposure times will deal with lesser amounts of contamination. Higher intensity UV sources will require less exposure.

C8. PCR in Situ Hybridization

The polymerase chain reaction (PCR) has been extensively used to amplify specific DNA sequences for use in the molecular analysis of many diseases.

Despite the widespread use of both PCR and in situ hybridization in the last several years, it has proved difficult to combine the two. If this combination could be accomplished DNA would be amplified in intact cells then visualized with in situ hybridization. The technique of PCR in situ hybridization was described in 1990 by Hasse, et al., who amplified visna virus DNA in

46

infected sheep choroid plexus cells and then detected the amplified DNA using in situ hybridization. In situ hybridization employed radioactive probes, multiple primer pairs and a fixed cell suspension. The multiple oligonucleotide primers were used to amplify a 1200 base pair DNA segment of the visna virus gag gene. The amplified DNA was identified within cells using in situ hybridization with a short 150 base pair probe and the sensitivity of in situ PCR was at least several fold greater than in situ hybridization alone, with a single copy gag gene being readily detected by in situ PCR. Although a useful start, the need to use cell suspensions limited the utility of this particular methodology. More importantly, the apparent need for multiple primer pairs severely limited the applicability of this PCR in situ technique because of the difficulty of generating multiple sequence specific primers for highly polymorphic targets, not to mention the high cost.

Nuovo and coworkers have further modified and developed the technique of in situ PCR for the identification of different types of human papillomavinis (HPV) in formlin-fixed paraffin embedded tissue samples. Compared with PCR in solution, the amplification of DNA from tissue sections or isolated cells is less efficient and this may be partly due to relatively poor access of primers, DNA polymerase and/or cofactor to target DNA and due to loss of tissue adherence and tissue drying. Accordingly optimal concentrations of the essential reagents such as primers, Mg^{2+}, and the DNA must be determined.

The "hot start" modification of the PCR technique, in which the oligonucleotide primers and DNA polymerase are only added at high temperature, produces the most consistence in situ PCR results, as this appeared to improve the efficiency of DNA amplification.

The success of in situ PCR depends on the amplified DNA remaining localized. It is not clear how the amplified DNA does not diffuse, although it is probable that network formation of the amplified DNA results in the development of an insoluble high molecular weight DNA complex. It was hypothesized that an amplified larger DNA segment would be less likely to diffuse than a smaller fragment; however, other studies of in situ PCR have found that the length of the DNA fragment to be amplified did not affect the final localization of DNA after in situ hybridization. The use of complementary tailed oligonucleotide promoters which enhance DNA complex formation does not appear necessary, although one study found that using primers tailed with complementary sequences minimized diffusion of the PCR product and this resulted in a stronger and better localized in situ hybridization signal.

Discussion of the PCR in situ hybridization technique involves two different methodologies. In one, called PCR in situ hybridization, non-labelled amplified DNA is detected by a labelled probe. The other called in situ PCR, refers to the direct in situ detection of the labelled nucleotide that has been incorporated into the amplified DNA. The labelled nucleotide, is then detected by the use of an antibody against it that is conjugated to alkaline phosphalase.

Nuovo et al 1991, described the detection of (HPV) type 16 DNA in paraffin embedded, formalin-fixed tissues of cervical squamous intrepethetial lesions (SILs) by in situ hybridization after amplification by the PCR. Using conventional in situ hybridization and biotin-labeled probe; variable numbers of superficial cells and none of the cells in the SILs showed detectable HPV 16 DNA. When the in situ assay was done after amplification, increased numbers of superficial cells had detectable HPV DNA, and the hybridization signal was much more intense. HPV DNA was also detected in basal and parabasal cells at the site of the lesion whereas it was not detectable in directly adjacent, normal squamous epithelium. Amplified HPV DNA was demonstrated in

formalin-fixed sitta cells using a biotin-labeled probe, demonstrating the ability to detect one copy of HPV 16 DNA.

In situ PCR has been used to detect low copy HIV-1 in peripheral blood mononuclear cells when HIV infection could not be detected by in situ hybridization. In addition, the utility of in situ PCR has been demonstrated by the identification of single-copy immunoglobulin gene rearrangements in human B-lymphocytes and the identification of different HLA_DQ haplotypes in human peripheral blood memonuclear cells .

Messenger RNA has also been detected using in situ PCR. Immunoglobulin heavy and light chain mRNA's were identified within single mouse hybridoma cells by sequentially performing in situ reverse transcription to produce DNA followed by in situ PCR. Rhodamine-labelled oligonucleotide were used during PCR and the amplified DNA was identified directly by fluorescence microscopy.

In situ PCR or in situ nucleic acid amplification and detection is a rapid, sensitive, and specific technique which combines the sensitivity of PCR with the cellular localization and resolution of in situ hybridization and is likely to have a variety of applications in histopathology.

C9. PCR Applications

C9a. Genetic Disorders

- **Characterization of a Gene Defect**

PCR can be used to identify sequence polymorphisms (e.g., single base mutations) for linkage analysis, even if no restriction sites are created or abolished, using either allele-specific primers or additional analytical techniques such as hybridization, restriction mapping or sequencing. Often more than one polymorphism can be simultaneously analyzed.

Length polymorphisms of hypervariable microsatellite regions can be detected purely by size, without the need for further analysis. This method is cheaper and faster than traditional Southern transfer. Once the gene responsible for a genetic disease is identified, PCR techniques can be used to rapidly characterize the individual defects in patients, in conjunction with sequencing and other techniques.

- **Parental Diagnosis, Neonatal Screening, and Carrier Testing**

Chorionic villus biopsy may be carried out from six weeks of gestation, and this is an established method of obtaining fetal tissue, with secondary fetal loss rates below 2 percent in many centers. PCR needs less material, so multiple DNA assays can be done on the same sample.

An important problem with the interpretation of PCR data from chorionic villi is the risk of material of material contamination. This does not apply to neonatal screening and PCR has potential for multiple genetic analysis of Guthr spots. PCR carrier testing can be done on hair roots and buccal washing. These non-invasive rapid tests make feasible programs of mass screening for carriers of genetic disease, as well as testing potential carriers within a known pedigree.

48

Several techniques have been developed using PCR to establish sex, by specifically amplifying parts of the Y chromosome. These techniques have been applied to single cells from blastomeres. and to chorionic villus biopss and have demonstrated the presence of a male fetus from maternal blood samples.

- **HLA Typing and Disease Susceptibility Genes**

PCR with direct sequencing of HLA loci has advantages over traditional HLA haptotyping methods. No expression of the genes is required for the analysis; there is considerable improvement in fine resolution, and very similar haplotypes can be distinguished down to the resolution level of a single base difference.

Restriction endonuclease mapping and allele specific hybridization can also be used for HLA typing, with or without prior PCR amplification, but DNA sequencing in more accurate and reliable (particularly if both complementary stands are sequenced) and can cover a greater range of polymorphisms. The rapid analysis of polymorphisms is fueling an expansion of research into HLA and other disease susceptibility genes.

- **Retrospective and Necropsy Studies**

PCR technology makes feasible the examination of pathological archival material in past cases of sudden infant death syndrome to look for inherited defects. Primers for many genetic diseases such as thalassemia, sickle cell anaemia, Haemophilia and phenyl ketonuiria have already been evaluated. In the next few years, it is expected that all single gene defects will be characterized, their genes sequenced, and PCR primers evaluated for the detection of individual cases. Since inherited mutations are carried in all cells, retrospective surveys can be done using paraffin embedded tissue from any organ. Also individual analyses can be arranged by a paediatric pathologists as part of a necropsy, using either paraffin-embedded or unprocessed material.

C9b. Identifying Micro Organisms

A vast body of work is accumulating on the use of PCR to identify pathogenic micro-organisms. The principle is that primers are constructed which will amplify a known sequence of a particular size from the pathogen, but will not amplify DNA from the host or from any other micro-organism. Primer sequences can distinguish between minor differences of strain and type, and as few as one molecule per million cells be detected. Thus, PCR can be a very sensitive and specific test for pathogens which are resistant or slow to culture like mycobacteria.

So, how can the pathologist evaluate published reports of microbial PCR assays?

1. The amplified bands must be the predicted size, to exclude non-specific amplification.

2. An initial confirmation of specificity is required either by Southern blotting and DNA hybridization with an internal probe of different sequence to the primers (since the primers will be incorporated into any amplified product) or by sequencing.

3. A control band is needed using host primers amplifying a different-sized DNA region to act as a reaction control for false negatives.

4. The major problem with PCR assays for micro-organisms is contamination. (Some assays and tissues are particularly prone to this; for example human papilloma virus assays on cervical smears taken in genito-urinary and family planning clinics.) Negative controls must be run with every assay including samples containing no DNA.

5. The extreme sensitivity of the technique may result in clinically insignificant doses of an organism being detected. (Often hybridization techniques might be more appropriate.) Sensitivity can be reduced by running fewer cycles, and the cycle number should always be given in any report of an assay. Ideally the initial evaluation should include an attempt to assess clinical significance.

Viral PCR assays have been used in aetiological studies, using both fresh clinical samples and archive material. PCR tests have found Epstein-Barr virus (EBV) in blood and salivary gland biopss from patients with Sjogren's syndrome.

Conversely, aetiological PCR studies have shown the absence of association of cytomegalovirus (CMV) with kaposi's sarcoma, and HPV with ovarian carcinoma.

C9c. Cancer

- **Activated Oncogenes**

PCR using primers with conserved sequences has been used to extend the numbers of potential oncogenes with related sequences, such as the tyrosine kinase family. However, the main use of PCR has been to amplify regions of known oncogenes or anti-oncogenes for further analysis, mainly sequencing, restriction endonuclease digestion, and allele-specific hybridization, in order to detect somatic mutations in tumor tissue or cells. It requires smaller amounts of starting material than Southern blotting.

In genes where known activating single base mutations have been found, panels of oligonucleotide probes matching the likely mutations have been used for allele specific hybridization after PCR amplification.

This has been done extensively for the ras oncogenes, and ras mutations have been found in many human cancers and also in many benign neoplasms such as in the colon and thyroid.

Where the mutation could not be predicted, for example with the retinoblastoma and p53 genes, PCR has been followed by direct sequencing.

- **Translations**

PCR is a very sensitive technique for identifying translocations (100,000 times, more sensitive than cytogenetics and 10,000 more sensitive than flow cytometry or Southern blotting). One primer is selected from each side of the putative translation site. Exponential amplification of the enclosed region will occur only if a translocation has brought the two primers into proximity.

Translations cn be detected by PCR in cytogenetically normal neoplasms such as Philadelphia negative chronic myeloid leukaemias. Tumor tissue (which can be fresh, frozen, formalin fixed, or paraffin embedded), blood, bone marrow, peritoneal washing and any body fluid or tissue can be assayed for trace amounts of translated DNA. Archive material can be

51.

studied to establish the incidence of particular translocations in various tumors. There are several additional factors to bear in mind in evaluating these studies. Contamination is a particular risk (e.g., peritoneal washing may be contaminated by the tumor resection specimen if taken during or immediately after surgery). Silent constitutional translocations have an estimated incidence of up to 1 percent these may be excluded with a control tissue from another part of the patient such as a hair root, buccal washing, or peripheral blood. False negatives can arise if the primer is too far away from the breakpoint. As a positive control, two primers on the same side of the breakpoint are used. To avoid false positive it is vital to ensure that each primer does not anneal to a non-specific site on the other chromosome—this can be confirmed by hybridization. Studies are in progress to assess the prognostic relevance of trace translocations after treatment, since there is doubt as to whether all cells carrying the translation are malignant.

C9d. Other Uses of PCR

- **Anthropology**

PCR combined with direct sequencing has been used to study polymorphisms in mitochondrial DNA which has a maternal lineage, enabling anthropological studies of world population movements. Thirteen thousand-year-old DNA has been amplified from extinct animals including the marsupial wolf and giant ground sloth.

- **Forensic Use**

DNA can be amplified from traces of material such as blood spots on clothing and hairs. Human DNA contains hypervariable regions called minisatellites which can give an individual "finger "finger print." Up to six of these minisatellites can be co-amplified in one PCR reaction.

SECTION II
BIOTECHNOLOGY AND GENETIC ENGINEERING: RESEARCH SUBPROJECT SUMMARIES

A. Subproject No. 6: Enzyme Production for Clinical and Industrial Applications

Awardee:	Molecular Biology Department, National Research Center (NRC)	
End-User:	El Nasr Company for Pharmaceutical Chemicals	
Principal Investigator:	Prof. Dr. Abd El Hady Ghazi	
Period:	01/02/ 1990 - 31/ 01/ 1993	
Budget Breakdown:		**LE**
USAID/STC	$ 109,510 and LE 215,862	**588,196**
In Cash	GOE: None End-user: None	
In Kind	GOE: LE 722,750 End-user: None	722,750
Total		1,310,946

A1. Subproject Description

This subproject aims to establish and develop guidelines for local production of peroxidase and urease enzymes utilizing indigenous resources and raw materials. The subproject is envisaged as a primary step toward the advancement of biotechnology and its applications in Egypt.

Enzymes are protein molecules that exist in all living organisms. Some enzymes contain haem or iron centers that play an important role in the catalytic activity of the molecule. These haem contained in enzymes have selective activities (oxidation/reduction), with certain types of organic compounds (substrates). This noted specificity in activity is generally controlled by the iron center and the protein structure of the enzyme itself.

Certain plant enzymes have been studied extensively because of their availability and high stability when isolated.

In general, plant enzymes have application in biotechnology. Early applications have been reported within the detergent industry, analytical and clinical diagnostic fields, and chemotherapeutic activities. More recent applications have been reported in the field of environmental pollution control.

In Egypt, the development and application of enzyme biotechnology is still at its inception. The country relies on imported enzymes such as peroxidase and urease for certain clinical and diagnostic activities. The absence of a local technical base for the advancement of biotechnology has been further aggravated by the unavailability of raw material required for enzyme preparation.

Peroxidase and urease enzymes are particularly useful tools in preparing diagnostic reagents and determining the prognosis of a number of diseases. This is done through testing of body fluids. Peroxidase is used mainly as an indicator in glucose and cholesterol diagnostic kits as well as in immunoassay kits. Urease is used mainly for detection of urea in medical uses, for example in kidney dialysis.

For purposes of their local manufacture, certain plant tissues and roots were screened within this subproject to identify potential sources of enzymes. Criteria for selection of screened plants included their availability at an appropriate cost within the local market. The selected plants were turnip and radish for peroxidase and melon and pumpkin seeds for urease. Following plant screening, tests for enzyme extraction took place.

Research results indicated that turnip roots and water melon seeds provided the best material for peroxides and urease extraction, respectively.

Several techniques have been employed by the subproject team for protein separation and purification in order to obtain the optimum enzyme concentration and homogeneity required for the various enzyme applications. Furthermore, the subproject team prepared powder and liquid forms of urease in commercial kits for diagnostic purposes.

Following this phase, the efficiency of the purified enzymes was tested against those imported under similar conditions. Sample diagnostic kits were manufactured utilizing local enzyme preparations. This activity was undertaken in collaboration with a publicly owned pharmaceutical company, El Nasr Company for Pharmaceutical Chemicals. Locally manufactured kits were distributed to members of an identified users group for further performance evaluation using regular control testing. Evaluation results indicated that local kits were as efficient as imported kits.

A2. Subproject Achievements

A2a. Potential for Widespread Application of Enzyme Technology within the Medical Sector

Having identified local substitutes for imported enzyme preparations, the subproject team has created a venue for biotechnological applications in Egypt. Particularly within the health sector, enzyme technology has been proven worldwide to have diagnostic and clinical applications that significantly contribute to more efficient prevention, diagnosis, and treatment of diseases.

A2b. Possibilities for Cleaner Food Production

A significant potential has been identified by the subproject team for the application of enzyme technology to the Egyptian food industry. More specifically, chemical and artificial food preservatives—presently used widely in food manufacturing operations—could possibly be substituted by natural enzymes that would have the same preserving effect. This is likely to contribute to the reduction of chronic health hazards associated with prolonged consumption of foods that contains chemical or artificial preservatives and will render food products safer for human consumption.

A2c. Identification of Potential for the Application of Biotechnology within the Environmental Sector

The successful application of bio- or enzyme technology within the medical field has revealed a significant potential for its further application within the environmental sector. Initially, this has been demonstrated in the substitution of surfactants used for the manufacturing of detergents by enzyme preparations. There is a growing trend toward expanding the application of these natural catalysts to other industrial operations. It may be possible to improve the quality of industrial wastewater through reducing its toxic/organic content with the subsequent likelihood of reducing costs of pollution abatement and wastewater treatment.

A2d. Development of Fundamentals for Local Enzyme Manufacture

The subproject team has identified sources and raw material for enzyme preparations that are readily available locally at an appropriate cost. This achievement will likely contribute to the advancement of this technology in Egypt, at both the research and applied levels. Members of the subproject team have gained more experience in this relatively new field that will enable them to pursue further research activities and advancement within this internationally growing technological sector. Creating opportunities for the local pharmaceutical-manufacturing sector is yet another achievement of this subproject.

B. Subproject No. 35: Introduction of Bacillus Thuringensis-Endotoxin Gene into Cotton Plants and Testing its Expression and Toxicity

Awardee:	Genetic Engineering and Tissue Culture Center, Menoufiya University	
End-User:	Ministry of Agriculture and Land Reclamation	
Principal Investigator:	Dr. Mahmoud Imam Nasr	
Period:	01/02 /1990 - 31/ 01/1993	
Budget Breakdown:		**LE**
USAID/STC	$ 165,225 and LE 277,047	**838,812**
In Cash	GOE None End-user LE 1,500	**1,500**
In Kind	GOE LE 165,000 End-user None	**165,000**
Total		**1,005,312**

B1. Subproject Description

This subproject aims at the identification and testing of natural and environmentally safe alternative defense mechanisms for the control of cotton pests, in order to maintain and possibly improve crop quality.

Cotton is one of Egypt's most important crops. Economically, a large portion of the cotton yield is used to develop and sustain the local industrial base to provide prime export commodities, hence the importance of the cotton crop as a national economic resource.

To ensure a high quality yield, cotton cultivation is a process that involves the efficient and timely application of certain control measures to prevent and reduce infection with the cotton leaf and boll worms (insects).

Some measures that were previously applied for pest removal were simple manual operations. These measures were effective in limiting the extent of damage at early stages of crop growth. Unfortunately, changes in agricultural practices over the past 50 years have led to detrimental pest attacks on plantation crops at their early phases of growth resulting in a reported decline in productivity within the cotton growing sector. Manual operations are insufficient to control such infestations, and there is a growing reliance on chemicals and pesticides to reduce and control worm populations.

Initially, pesticides were effective during the life cycle of the leaf worm, and reduced worm population densities within cotton plantations. However, pest populations developed resistance against commonly used pesticides, and have rendered their application ineffective.

According to the Egyptian Ministry of Agriculture and Land Reclamation, 70 percent of imported pesticides are used to control the cotton pests. However, growing environmental awareness both internationally and locally is expected to influence the chemical pesticide industry and end market consumer regarding the pesticide products.

The Ministry of Agriculture is a key player in defining and implementing nationwide cotton pest control strategies. This subproject has been designed to support the Ministry in enhancing natural resistance of the cotton plant through the introduction of Bacillus Thuringensis (BT) d-endotoxin gene into selected cotton strains. The production of these transgenic plants will effectively lead to a reduction of pesticide use in cotton cultivation.

BT strains were tested by the subproject team to identify suitable gene clones using gene expression tests. Plant regeneration and transformation was carried out using a tissue culture technique. The process yielded a complete transgenic plant.

Using antibiotic and gene-containing media, test cells were planted in large computer-controlled, illuminated incubators. The shoot apex from planted cotton strains were selected as the most suitable cell type for tissue culture regeneration. The process was monitored at different stages of plant growth. With the development of the shoot and root systems, selected transgenic plants were replanted within the experimental field area. Necessary precautions were undertaken until the plant reached maturity and produced seeds.

Gene expression techniques were used to detect the presence of endotoxin genes in transgenic tissues and seeds of selected cotton strains. Further testing against target insects (leaf and boll worms) and non-target insects (harmful and beneficial) was done with tissue extracts from these strains. This series of chemical and biological testing allowed for determining and monitoring the potential tolerance and/or resistance against BT endotoxins. Test results may also be useful to project effects of BT endotoxins on other species.

55

An assessment of the biochemical and neurological toxicity of newly prepared transgenic plants has been initiated within this subproject using specific immunoassays. Comparative mutagenicity testing of these plants versus commercially available endotoxin gene products was also initiated to evaluate relevant genetic risks, if any.

Finally the subproject team prepared an economic evaluation in cooperation with the Ministry of Agriculture as a first step toward larger-scale application of the research.

B2. Subproject Achievements

B2a. Economic Opportunities through Reduced Expenditure on Imported Pesticides

As stated earlier, 70 percent of imported pesticides are used for the control of cotton pests. Local production of these genetically engineered or transgenic plants is expected to result in a revolutionary change in agricultural practice and expenditure in Egypt. Savings in cotton crop cultivation should increase yield of this primary national product and ensure its high quality. This will likely result in improvements in the overall national economy.

B2b. Potential for Reducing Environmental Pollution Occurring As a Result of Extensive Use of Chemical Pesticides

Reduced use and reliance on chemical pesticides through their substitution with natural defensive mechanisms against cotton pests will have significant environmental impacts. This will result in improved quality of agricultural drainage water and reduced deterioration of soil quality by chemical contamination.

B2c. Expected Health Improvements for Pesticide Applicators

Persons who apply pesticides are currently exposed to acute and chronic poisoning. These hazards may be further aggravated by inadequate health infrastructure in some rural areas. The partial or total substitution of chemical pesticides by natural defense mechanisms is likely to reduce the adverse health impacts of these chemicals.

C. Subproject No. 42: Molecular Genetic Techniques in the Diagnosis and Prognosis of Chronic Myeloid Leukaemia before and after Bone Marrow Transplantation

Awardee:	Faculty of Medicine, Cairo University		
End-User:	National Cancer Institute		
Principal Investigator:	Prof. Dr. Mohamed Talaat Abd El Aziz		
Period:	20/02/1993 - 26/01/1994		
Budget Breakdown:			**LE**
USAID/STC	$ 133,900 and LE 206,225		**661,485**
In Cash	GOE End-user	None None	
In Kind	GOE End-user	LE 254,160 None	254,160

Total		915,645

C1. Subproject Description

This subproject aims to apply modern techniques in molecular biology for the detection of minimal residual disease in chronic myeloid leukaemia patients who have undergone bone marrow transplantation.

Leukaemias are a group of diseases affecting lymphatic and blood-forming systems usually arising in the bone marrow. Leukaemia results in overproduction of abnormal leukocytic elements at their site of origin. This may or may not be associated with an increase of these cells in the blood or internal organs.

In chronic myeloid leukaemia (CML), there is rarely any enlargement of the lymph nodes in the early stages. The disease may actually be present for 2-5 years before it is clinically recognized. At this point the treatment is usually bone marrow transplantation (BMT), with or without chemotherapy.

Laboratory diagnosis of CML will help establish a prognosis and regular treatment regime. Because of the late appearance of clinical symptoms of the disease and the failure to diagnose it at an early stage using classical laboratory techniques (karyotype and cytological techniques), the disease is usually diagnosed at the stage requiring BMT. It is worth nothing that 75 percent of patients who have undergone BMT will likely develop symptoms of recurrent CML within 5-10 years at which time they would require a second BMT.

The incidence of CML in Egypt is 1:10,000. Most of the recorded cases are diagnosed and treated at the National Cancer Institute (NCI), which is a public health facility based in Cairo. The subproject team applied their new diagnostic technique to CML patients received at the NCI over two years.

The new technique, referred to as the polymerase chain reaction (PCR) technique, was used by the subproject team to detect early cellular changes in patients who have undergone BMT as treatment for CML. Early changes were detected in two members of the group of patients. These were given a bufficote transfusion, which is an alternative method of treatment that is only successful if the disease is diagnosed in its very early stages.

Following up those two patients using PCR indicated a negative state of the disease. As such, those patients have been spared the need for a second BMT, which is a major operation with likely post-operative complications and a cost of LE 75,000.

Early in the subproject, the team set up a molecular biology laboratory at the Faculty of Medicine of Cairo University. Standardization of the laboratory techniques was carried out and training was provided to team members. Screening and testing of follow-up samples of CML patients has been conducted in the new laboratory.

C2. Subproject Achievements

Statistical evaluation and analysis of data collected throughout the subproject period will serve as the basis for defining the epidemiology of CML in Egypt. This will in turn facilitate further identification and screening of high risk groups for early detection of the disease.

C2a. Identification of New Types of Chromosomal Disorders

In the course of conducting the research, the subproject team identified new types of chromosomal disorders. These disorders have to date not been mentioned in international publications.

C2b. Early and Sensitive Detection of CML

The application of PCR in the diagnosis of recurrent CML patients will reduce the need for these patients to undergo a second BMT. The difficulties of this operation include a search for a matching donor, followed by a long list of post-operative complications. These usually occur as a result of the heavy doses of immunotherapy that is administered postoperatively and the weak state of the patient in general.

C2c. Reduced Health Care Costs

The cost of the bufficote transfusion given to patients with recurrent CML and positive PCR results is LE 1,000. BMT, however, which is the treatment of choice for these patients when diagnosed at a later stage of the disease using less-sensitive techniques, is LE 75,000. When undertaken at the NCI, BMT costs are subsidized creating a burden on the national health care budget.

C2d. Application for the Diagnosis of Endemic Diseases

The newly applied technique PCR can be used to diagnose early symptoms of Hepatitis B and C viruses. These viruses, currently a main cause of morbidity and mortality in Egypt, cause a deterioration in liver function that often leads to liver failure in infected patients. Early diagnosis of these liver changes may allow for better treatment of patients.

D. Subproject No. 45: Screening for and Detection of Human Papilloma Virus by DNA Hybridization Technique in Egyptian Females and its Relation to Cervical Cancer

Awardee:	Faculty of Medicine, Cairo University	
End-User:	National Cancer Institute	
Principal Investigator:	Prof. Dr. Mohamed Talaat Abd El Aziz	
Period:	20/12/1992 - 20/01/1996	
Budget Breakdown:		**LE**
USAID/STC	$ 99,000 and LE 241,680	**578,280**
In Cash	GOE None End-user None	

In Kind	GOE LE 272,520 End-user None	**272,520**
Total		**850,800**

D1. Subproject Description

The subproject aims to apply modern molecular biology techniques to detect human papilloma virus in biological tissues and in smeared cervico-vaginal cells of Egyptian females.

Cervical cancer is the second most common cancer affecting Egyptian females, after breast cancer. This is despite the fact no other form of cancer offers the opportunity of early detection such as cancer of the cervix does.

This high incidence rate persists because of inadequate screening of high risk groups and the absence of a sensitive and accurate diagnostic and prognostic (follow up) tool to date.

Adequate treatment and follow up of patients during the early stages of the disease should reduce considerably the occurrence of invasive cancer and minimize its complications.

The development of cervical cancer is often unassociated with alarming symptoms. As a result, the disease may progress into a moderately advanced stage before it is investigated. The value of screening using techniques that diagnose the disease in its early stages would allow for detection of the disease in sub-clinical forms that would dramatically respond to simple curative interventions.

The most commonly used technique is cytological examination of vaginal smears. Cytology has been shown to be of low sensitivity and low accuracy in detecting cases of pervasive legions. A high false negative smear rate is a problem that affects the reliability of test results, especially when applied to members of high risk groups, particularly women infected with the Human Papilloma Virus (HPV). For these cases, colposcopic examinations is helpful in improving the reliability of detection. Consequently, clinicians increasingly refer patients to gynaecologic clinics for colposcopic examinations and directed biopsies for final evaluation. The frequency of testing depends on the age and the presence of other predisposing factors. As such, the procedure may be both time consuming and costly when applied to a large segment of the population. By narrowing down high risk groups, the procedure may be rendered more cost-effective.

HPV has been identified as a primary aetiological factor in the occurrence of cervical cancer. HPV is a sexually transmitted virus that could be detected by cytological techniques. False negative detection rates using these techniques range between 20-45 percent. The need thus arose for a more sensitive diagnostic technique to be applied to females comprising the high risk group.

The subproject team selected a group of 145 females. These were partners of male patients visiting a sexually transmitted disease (STD) clinic for treatment of genital warts caused by HPV.

Applying the cytological techniques to members of the test group, an 18 percent positive test rate was achieved. Further application of the polymerase chain reaction (PCR) technique to

the samples that were proven negative yielded an additional 24 percent positive rate of infection with HPV.

The subproject team compiled a health profile of members of the test group in order to define better the epidemiological characteristics of this disease in Egypt. The team was able to identify new forms of the virus whose link to cancer occurrence has not yet been established.

To facilitate the clinical application of PCR diagnostics on a large scale, the subproject team established a screening clinic in the Obstetrics and Gynecology Department at the Faculty of Medicine. The clinic is comprised of a physical examination room and a cytology laboratory, and a moleculor biology laboratory within the Department of Biochemistry of Cairo University.

D2. Subproject Achievements

D2a. Earlier Detection of Cervical Cancer

The application of PCR to the diagnosis and treatment of cervical cancer could enhance cure rates and reduce treatment procedures as shown in the following table, assuming PCR will provide early diagnosis.

Stage at Diagnosis	Treatment	Duration	Cure Rate	Cost in LE
Carcinoma in Situ	Surgical	3 weeks	95%	1000
Invasive Cancer	Surgical/irradiation /chemotherapy	6 months	40%	10,000

D2b. More Sensitive Diagnosis of Cervical Cancer

Compared with other techniques used for diagnosis, PCR yields an accuracy rate of 100 percent as opposed to a 30 percent accuracy rate using cytological techniques.

D2c. Established Screening Procedures in Egypt

The screening clinic and laboratory established through this subproject are expected to serve a large segment of the population that uses public health facilities. The facility is the first molecular biology laboratory in the public health sector in Egypt.

D2d. Reducing National Health Care Costs

Early diagnosis of cervical cancer will reduce costs of treatment, depending on the stage at which it is detected. Particularly among groups benefiting from the public health sector, this will reduce national expenditure on treatment of late stages of cervical cancer.

D2e. Upgrading Technical Skills within the Health Care Sector

The training provided within this subproject to doctors and laboratory technicians will allow them to pursue their professional development in more fields of specialization.

D2f. Identification of Research Needs

The new forms of the virus detected within this subproject have been recommended by the subproject team as a new topic for research in order to verify its linkage to the occurrence of cancer. Presumably, this research will lead to improvements in cancer detection and treatment techniques in Egypt.

D2g. Reduction in Morbidity and Mortality Caused by Cervical Cancer

The epidemiological data and analysis compiled under this subproject will serve as a basis for a more in-depth study of one of the two most common types of cancer affecting the Egyptian female population.

Moreover, improved testing procedures will undoubtedly lead to earlier detection of the disease and thus enable earlier intervention, thereby reducing morbidity and mortality.