SECTION

1

LABORATORY TECHNIQUES

CHAPTER

1

BIOCHEMICAL BASIS OF APPLIED MOLECULAR GENETICS

Molecular genetics employs known principles of DNA structure and function to investigate the molecular basis for genotype-directed phenotypes under normal and pathological conditions. The DNA segments most often studied by molecular geneticists are those that encode genes, the smallest unit of genetic heredity. The term *applied molecular genetics* is used here to describe a rapidly growing set of laboratory-based research tools that exploit the information potential of organismal DNA. As an introduction to the field of applied molecular genetics, we begin with a brief review of information transfer from DNA to RNA to protein, highlighting the most relevant concepts. This is followed by a discussion of basic nucleic acid biochemistry and a short description of the most common molecular genetic laboratory reagents and techniques. Laboratory practicum 1 illustrates how to identify the transcriptional start site of a newly isolated gene.

FLOW OF GENETIC INFORMATION: DNA → RNA → PROTEIN

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) polymers consist of repeating units of deoxynucleotides and ribonucleotides, respectively. With the exception of some viruses, almost all organisms on this planet store their cellular blueprints for life in double-strand DNA molecules called chromosomes. In eukaryotic cells, chromosomes are copied during cell division, recombined and shuffled as a result of sexual reproduction, and transcribed into complementary RNA molecules through a process called gene expression. Figure 1.1 shows a schematic representation of how chemical information stored in the DNA coding sequences of a gene is transmitted to the protein synthesis machinery in the cell by mRNA "transcripts." This relationship between the DNA, RNA, and protein sequence information of a gene is sometimes referred to as the biochemical flow of genetic information. Although there are a few examples in nature where these simple hierarchical relationships do not hold true, the core principles of applied molecular genetics follow, for the most part, the classic paradigm of DNA → RNA → protein.

To understand fundamental molecular genetic principles, it is useful to think about how chemical information, stored in a simple nucleic acid polymer, could direct the development and maintenance of a living complex organism. First, remember that each

Figure 1.1 Schematic representation of genetic information processing in the cell. A gene is the fundamental unit of information storage and includes both coding sequence and transcriptional regulatory sequences. In eukaryotes, most gene coding sequences are contained within segments of DNA called exons (gray boxes) that are separated by noncoding sequences called introns. Transcriptional initiation begins at the 5′ end of the gene in the promoter region. Following elongation and transcriptional termination, post-transcriptional RNA processing mechanisms fuse exonic coding sequences by RNA splicing and add a 3′ polyadenylate tail (AAAA). Protein synthesis (mRNA translation) occurs in the cytoplasm of eukaryotes, whereas in prokaryotes, transcription and translation are coupled processes.

deoxyribonucleotide unit in DNA contains one of the four bases, guanine (G), adenine (A), cytosine (C), or thymine (T). Second, calculate the total number of sequence combinations that are possible for an oligonucleotide that is just 10 nucleotides long. Because each position in the oligonucleotide could be a G, A, C, or T, there would be 4^{10}, or 1,048,576, different sequence possibilities. Because the amount of DNA in a human cell contains more than 10^9 nucleotides, it is easy to imagine that there is more than enough storage space in the chromosomal DNA "hard drive" of a cell (4^{billion} sequence combinations) to encode someone even as unique as yourself! Nature's ability to encrypt the essence of life in long strings of DNA sequence is therefore not a problem of storage space, but rather how accurately the cell machinery is able to retrieve and interpret this vast amount of information. Over the past three decades, biochemists, geneticists, and cell biologists have been able to decipher the basic components of this

information processing in the cell, which together have laid the groundwork for modern molecular genetics. Indeed, applied molecular genetics is the exploitation of this knowledge to investigate and utilize the processes of DNA synthesis (replication), RNA synthesis (transcription), and protein synthesis (translation) not only to access, but also to manipulate the information potential of organismal DNA. Let's briefly review the important characteristics of these three information processing "algorithms" to understand better the key principles of applied molecular genetics.

DNA Synthesis

Cell division requires that DNA be duplicated to produce an exact chromosomal copy. The two major concepts to remember about DNA replication are the following:

1. DNA is a double-strand molecule that is stabilized by hydrogen bonding between complementary base pairs in two antiparallel strands: guanine bonds with cytosine (G-C base pair) and adenine bonds with thymine (A-T base pair).
2. Initiation of DNA synthesis requires a template primer with a free 3′ hydroxyl group and polymerization always proceeds in the 5′ to 3′ direction.

RNA Synthesis

RNA synthesis, also known as DNA transcription, requires RNA polymerase enzymes that use single-strand DNA as a template to make complementary copies of information stored in the DNA sequence. RNA polymerases synthesize RNA in the 5′ to 3′ direction just as DNA polymerases do. However, there are two important differences between RNA and DNA synthesis. First, uridine is the ribonucleotide base that pairs with adenine (rather than thymine), and second, RNA polymerases do not require a template primer.

It is important to keep in mind the following points about RNA synthesis:

1. RNA synthesis is required for the transcription of DNA information contained within the genetic unit called a gene. RNA synthesis begins at a specific initiation site on the 5′ end of a gene (upstream) and terminates at the 3′ end (downstream).
2. Multiple copies of short-lived RNA molecules are synthesized from a single DNA template; the number of RNA transcripts synthesized per unit time depends on the rate of transcriptional initiation by RNA polymerase.

Protein Synthesis

The primary cellular machine involved in protein synthesis, also called translation, is the ribosome. This very abundant macromolecule contains a well-characterized arrangement of large ribosomal RNA molecules and numerous ribosomal proteins. Three main concepts in protein synthesis must be emphasized:

1. The DNA sequence of a gene, as faithfully copied into mRNA, contains information for protein synthesis in the form of triplet codons. There are 64 possible

triplet codons, of which 61 can specify the 20 amino acids (with redundancy), and three correspond to termination codons. This is called the genetic code, a copy of which is printed on the inside cover of the book.

2. The 5′ end of mRNA directs ribosome binding and the subsequent initiation of protein synthesis. The first amino acid, usually the methionine codon AUG, corresponds to the amino terminus (*N*-terminus) of the encoded protein. The ribosome "reads" the mRNA in the 5′ to 3′ direction until reaching the penultimate codon in the mRNA, which specifies the carboxy terminal amino acid. The ribosome disengages at the subsequent termination codon.
3. Because the genetic code is based on triplets, and the 5′ end of mRNA contains nucleotides upstream of the initiator methionine codon to accommodate ribosome binding, the ribosome can theoretically begin translation in any of three possible protein coding registers called "reading frames."

NUCLEIC ACID BIOCHEMISTRY

To understand many of the principles underlying applied molecular genetics, it is important to be familiar with two chemical properties that affect the behavior of DNA and RNA in solution: (1) the molecular forces that affect the structure of nucleic acid polymers and (2) the kinetic parameters that determine rates of denaturation and renaturation of complementary heteroduplexes. In addition, because many of the starting points for molecular genetic methods require the design of custom oligonucleotides, we examine the basic steps required to produce nucleic acid polymers synthetically using solid support chemistry.

Structure of Nucleic Acid Polymers

The chemical structure of a DNA–RNA heteroduplex is shown in Figure 1.2. The key features to note are (1) the phosphodiester linkages between repeating nucleotide units; (2) the antiparallel polarity of the DNA–RNA heteroduplex, such that the 5′ to 3′ DNA strand is base paired with the 3′ to 5′ RNA strand; and (3) the complementary DNA and RNA strands joined by hydrogen bonding between T-A, G-C, and A-U base pairs. The chemical structure of the bases present in DNA (G, C, A, T) and RNA (G, C, A, U), allow for the formation of hydrogen bonds between opposing purine (G and A) and pyrimidine (C, T, and U) bases. Importantly, the number of hydrogen bonds formed between G-C base pairs is three, whereas only two hydrogen bonds are formed between A-T and A-U base pairs. This difference in hydrogen bonding capacity contributes directly to the thermal stability of double-strand DNA molecules.

Noncovalent interactions are responsible for the three-dimensional structure of the DNA double helix. The two primary sources of noncovalent interactions are the base to base hydrogen bonds formed between antiparallel strands, and the hydrophobic interactions that occur between adjacent bases on the same strand (van der Waals interactions) (Fig. 1.3). Although double-strand DNA appears rigid in molecular models, it is actually quite flexible in the presence of DNA binding proteins, which can bend DNA helices to angles greater than 100 degrees. The ability of DNA strands to dissociate and reassociate locally is critical to its function as a molecular database. DNA replication, recombination, and transcription require unwinding of the DNA double helix, which is

Figure 1.2 A DNA–RNA heteroduplex is formed during RNA synthesis and in the priming step of DNA replication. The bases have been rotated relative to the phosphate backbone to illustrate the hydrogen bond formation between base pairs. Note that the RNA strand has 2′-OHs on the ribose as well as uridine in place of thymine, and that purines (G, A) hydrogen bond with pyrimidines (C, T, U). Noncovalent interactions between G and C residues are more stable than A-T or A-U base pairs because of the extra hydrogen bond that is formed.

accomplished in vivo by specialized helix-destabilizing proteins. Pure double-strand DNA can be unwound in vitro using elevated temperatures or chemical denaturants as described in the next section.

Denaturation and Renaturation of Nucleic Acid Duplexes

The relative amount of single- or double-strand DNA in solution can be experimentally determined using spectrophotometery to measure ultraviolet light absorbance at a wavelength of 260 nanometers (optical density, OD_{260}). The aromatic bases in DNA are less accessible to ultraviolet light in the double-strand, compared to single-strand form, which creates a measurable difference in the observed OD_{260}. Using this empirical difference in absorbance, it is possible to observe the effect of temperature on DNA structure by monitoring OD_{260} over a temperature range of 55–90ºC. Figure 1.4 shows a melting curve of double-strand DNA that demonstrates that the amount of denatured DNA rises sharply over a narrow temperature range, indicating that denaturation is a

Figure 1.3 The molecular structure of a DNA double helix in the B form. The phosphodiester backbone is antiparallel and the bases from each strand are noncovalently bound through hydrogen bonding. Additional hydrophobic interactions between adjacent bases on the same strand also contribute to helix stability. DNA binding proteins associate with the double helix through noncovalent interactions with bases in portions of the DNA structure called the major and minor grooves.

cooperative process. This cooperativity indicates that once the DNA hybrid has been locally disrupted, it requires only a small amount of additional energy to separate the two strands completely. The temperature at which 50% of the DNA is denatured is called the T_m or melting temperature.

The T_m of a nucleic acid duplex is strongly affected by three factors: (1) base composition, (2) duplex length, and (3) ionic strength of the solution. Base composition is an important determinant of the T_m because G-C base pairs contain one more hydrogen bond than A-T base pairs. Therefore, duplex molecules that contain a high G-C content are more stable and have a higher T_m than do A-T-rich molecules. Duplex length of the hybrid affects the T_m because the overall stability of a double-strand molecule is directly proportional to the number of base pairs. This is especially evident for duplex molecules containing less than 150 consecutive base pairs. Two molecular genetic applications

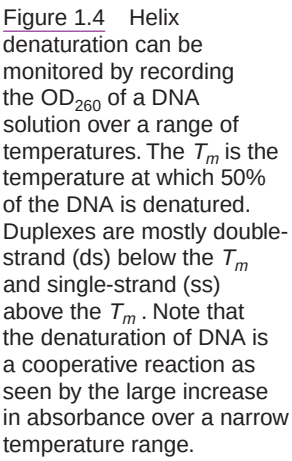

Figure 1.4 Helix denaturation can be monitored by recording the OD_{260} of a DNA solution over a range of temperatures. The T_m is the temperature at which 50% of the DNA is denatured. Duplexes are mostly double-strand (ds) below the T_m and single-strand (ss) above the T_m. Note that the denaturation of DNA is a cooperative reaction as seen by the large increase in absorbance over a narrow temperature range.

where extent of duplex formation is an important consideration are the use of short oligonucleotides in hybridization reactions and heteroduplex formations between molecules that are less than 100% complementary. The use of homologous, but not identical, DNA molecules in hybridization reactions is common when sequence divergence exists between two genes, for example, across species or among members of a related gene family. The T_m for heteroduplexes decreases by approximately 1°C for every 1% sequence mismatch. The third factor influencing the T_m of a given heteroduplex is the ionic strength of the solution. In high Na^+ concentrations (1*M*) the T_m is increased owing to electrostatic shielding of the negative phosphate charges in the DNA backbone, whereas in low Na^+ concentrations (0.1*M*) the T_m is decreased. It is possible to approximate the T_m of a short complementary oligonucleotide (10–20 bases) in a solution containing 1*M* Na^+ using the empirical formula:

$$T_m\,(^{o}\mathrm{C}) = 2(\text{number of A} + \text{T}) + 4(\text{number of G} + \text{C})$$

The reverse of nucleic acid denaturation is renaturation, also referred to as reassociation or hybridization. This bimolecular process is most affected by temperature, ionic strength, molar concentration of the two complementary strands, and reaction time. Two other factors that can be introduced experimentally are the effect of denaturing agents such as formamide or urea, both of which lower the T_m, and the inclusion of dextran sulfate in hybridization reactions, which increases the rate of reassociation by as much as 10-fold. Under optimal conditions of temperature and ionic strength, which is usually 10–15ºC below the T_m in a solution containing 0.2*M* Na^+, the concentration of nucleic acid becomes the rate-limiting step in the hybridization reaction. The term $C_o t$ is used to describe the kinetics of hybridization between two nucleic strands in solution and is defined by the product of [nucleic acid] × (time). Put simply, when the concentration of two complementary strands in a solution is high, it takes a shorter time for hybridization to occur than it does when one or both of the strands are present at a low concentration.

$C_o t$ curves plot percent reassociation versus $C_o t$ (mole-seconds/liter) and are used to measure the sequence complexity of DNA samples (Fig 1.5). DNA from organisms with

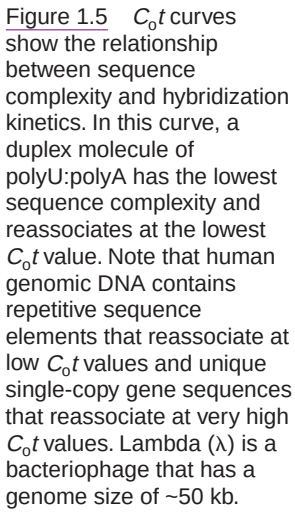

Figure 1.5 C_0t curves show the relationship between sequence complexity and hybridization kinetics. In this curve, a duplex molecule of polyU:polyA has the lowest sequence complexity and reassociates at the lowest C_0t value. Note that human genomic DNA contains repetitive sequence elements that reassociate at low C_0t values and unique single-copy gene sequences that reassociate at very high C_0t values. Lambda (λ) is a bacteriophage that has a genome size of ~50 kb.

small genomes have low sequence complexity and reassociate at much lower C_0t values than do denatured DNA samples from more complex organisms.

One way to think about how C_0t values affect hybridization kinetics, and how this relates to sequence complexity, is to imagine that at the same total DNA concentration, fragments of λ bacteriophage DNA require significantly less time to sort through all possible complementary strands than do similar sized fragments of *E. coli* DNA that are derived from a genome that is 1000 times larger. The C_0t curve obtained from the analysis of human DNA is a mixture of curves (Fig. 1.5). This irregularity is due to the abundance of repetitive DNA sequences that reassociate at low C_0t values, compared to the unique DNA sequences representing single-copy human genes that require much longer times to reassociate owing to their very low concentrations.

Table 1.1 summarizes factors that affect denaturation and renaturation of nucleic acid duplexes.

Chemical Synthesis of DNA and RNA

The ability to synthesize single strands of DNA or RNA using solid support chemistry has had a dramatic impact on the development of applied molecular genetic methods.

TABLE 1.1 Factors that affect the denaturation and renaturation of nucleic acid duplexes

Parameter	Effect on T_m	Effect on rate of renaturation
Base composition	↑ T_m with ↑% G-C	No effect
Hybrid length	↑T_m with ↑length >500 bp; no effect on T_m	↑Rate with ↑length
Ionic strength	↑T_m with ↑[Na^+]	Optimal at 1.5 M Na^+
% bp mismatch	↓T_m with ↑% mismatch	↓Rate with ↑% mismatch
DNA concentration	No effect	↑Rate with ↑[DNA]
Denaturing agents	↓T_m with ↑[formamide], [urea]	Optimal at 50% formamide
Temperature	Not applicable	Optimal at 20°C below T_m

Figure 1.6 Chemical synthesis of DNA using phosphoramidites. (a) Chemical structure of a phosphoramidite showing the dimethoxytrityl (DMT) blocking group on the 5′ carbon and the methylated 3′-phosphite and diisopropylamine groups attached to the 3′ carbon. (*b*) Five sequential steps are required for each cycle of nucleotide extension.

Custom-designed oligonucleotides are available commercially and are used routinely in numerous experimental procedures. For example, oligonucleotides are used as template primers in DNA sequencing and PCR reactions (Chapter 6) and for the incorporation of site-specific mutations in cloned genes (Chapter 3). In addition, chemically modified ribonucleotides can be used to synthesize large quantities of RNA for use as "antisense" inhibitors of RNA function (Chapter 7). Figure 1.6 outlines the basic steps required for in vitro DNA synthesis using the phosphoramidite method.

In vitro DNA synthesis reactions take place inside sealed columns that contain glass beads that serve as the solid support for the sequential chemical reactions. Single phosphoramidites for each of the four bases are added to a growing chain that is initiated at the 3′ end. The five chemical steps required for each nucleotide addition are (1) deblocking the 5′ end by DMT removal, (2) amidite activation of the incoming phosphoramidite, (3) coupling of the nucleotides through a 5′–3′ linkage, (4) capping of unreacted nucleosides to prevent extension of incomplete products, and (5) oxidizing the phosphate triester to stabilize the 5′-3′ linkage. After the required number of linkages are formed through repeated cycling, the oligonucleotide products are demethylated, released from the column, and chemically treated to produce a population of 5'-hydroxylated molecules. The efficiency of each coupling step in the reaction is critical and must be >98% to produce significant yields of a full-length product.

DNA METABOLIZING ENZYMES

Enzymes that modify and metabolize nucleic acids are essential tools for many applied molecular genetic methods. The commercial availability of these enzymes has led to the development of molecular genetic "enzyme kits," which can often be useful components of molecular genetic research strategies. However, as many researchers who use these molecular genetic kits will attest to, it can sometimes be difficult to troubleshoot a failed experiment when the protocol reads "combine equal volumes of solution A (red cap) with buffer B (yellow cap) and incubate for 30 minutes at room temperature in 1/10 volume of enzyme reaction mix C (blue cap)." Therefore, when using DNA metabolizing enzymes, it is important to have an understanding of the function (and limitation) of each enzyme and to follow standard biochemical laboratory practices to optimize enzyme activity. Three classes of enzymes are described here that represent the primary biological reagents for the most common molecular genetic applications: sequence-specific DNA restriction enzymes, ligases and kinases, and DNA and RNA polymerases. Nucleases are another important class of nucleic acid metabolizing enzymes that are described at the end of this chapter as reagents in laboratory practicum 1.

Sequence-Specific DNA Restriction Enzymes

Biochemists discovered more than 30 years ago that most bacteria contain endonucleases that degrade the DNA genomes of infectious bacteriophages. They found that the host bacteria are "immune" to these nucleolytic enzymes because of site-specific DNA methylations that prevent endonuclease attack on the bacterial genome. This form of bacterial immunity requires two distinct enzymatic activities: (1) restriction endonucleases that cleave double-strand DNA at specific sites, and (2) DNA methylases that modify bases at these same sites in the host cell genome (Fig. 1.7). Different species of

Figure 1.7 Restriction enzymes are site-specific endonucleases that cleave double-strand DNA. (*a*) Restriction enzymes and their corresponding methylases function in bacteria to protect against bacteriophage infection. (*b*) Type II restriction enzymes bind to specific DNA sequences as homodimers and produce a double-strand break in the phosphodiester backbone.

bacteria contain their own sets of endonucleases and corresponding methylases. The term “restriction” refers to the function of these enzymes in restricting the host range of bacteriophage infection.

Based on mechanistic differences between several types of bacterial restriction systems, three classes of restriction enzymes have been described. Type II restriction enzymes are used in molecular genetic applications because they can be used in vitro to recognize and cleave within DNA sequences typically consisting of four to eight

nucleotides. By using DNA lacking the specific methylations for a given restriction enzyme (e.g., eukaryotic DNA or plasmid DNA obtained from a methylase-deficient bacterial strain), it is possible to cleave any DNA molecule that contains the recognition sequence for a particular Type II restriction enzyme.

A large number of restriction enzymes have been characterized and shown to bind as dimers with high affinity to specific DNA sequences (Fig. 1.7). The DNA recognition sites for most all restriction enzymes are palindromes and double strand cleavage produces a 5′ phosphate and 3′ hydroxyl at the DNA termini (Fig. 1.8). There are three types of restriction enzyme cleavage reactions, two of which result in the formation of staggered or "sticky" ends having a 5′ or 3′ extended terminus, and a third that results in flush or "blunt" DNA termini on both strands. Applications of restriction enzymes to recombinant DNA cloning methods are discussed in Chapter 2 and biochemical properties of the most common restriction enzymes are listed in Appendix D.

Ligases and Kinases

Ligases and kinases are another class of enzymes that play an important role in recombinant DNA methodologies. DNA ligases catalyze the formation of 5′–3′ phosphodiester bonds in double-strand DNA molecules. In vivo, ligases function in DNA replication to repair single-strand nicks resulting from DNA repair processes, and to join adjacent Okasaki DNA fragments produced in the lagging strand of a DNA replication fork. The T4 bacteriophage DNA ligase is an ATP-dependent ligase that is com-

Figure 1.8 Restriction enzymes cleave palindromic DNA sequences to produce double-strand breaks. Restriction enzyme cleavage results in the formation of 5′ PO_4^- and 3′ OH termini with (*a*) 5′ staggered ends, (*b*) blunt ends, or (*c*) 3′ staggered ends.

monly used in DNA cloning strategies to "ligate" two DNA fragments (Fig. 1.9). *E. coli* DNA ligase is also used for some applications, however, it differs from T4 DNA ligase in that it uses NAD for a co-factor and cannot efficiently ligate blunt-end fragments.

Kinases are enzymes that phosphorylate specific substrates by covalently attaching the γ phosphate from ATP to a reactive group on the target molecule. T4 polynucleotide kinase is an enzyme that phosphorylates 5′ hydroxyl termini on DNA and RNA (Fig. 1.10). T4 polynucleotide kinase is often used to label DNA radioactively using [γ-^{32}P]ATP for the purpose of making high specific activity radioactive probes. This

Figure 1.9 DNA ligases catalyze the formation of a phosphodiester bond in nicked double-strand DNA. A ligase–AMP complex forms a transient intermediate with the 5′ phosphate initiating a nucleophilic attack on the 3′ hydroxyl group. The reaction shown here illustrates the two-step ligation of heterologous EcoRI DNA fragments producing covalently closed double-strand DNA.

Figure 1.10 T4 polynucleotide kinase can be used to label radioactively the 5′ ends of DNA in a reaction using [γ-^{32}P]ATP. DNA termini with 5′ phosphates can be dephosphorylated by the enzyme bacterial alkaline phosphatase to produce a 5′ hydroxyl that is the optimal substrate for polynucleotide kinase. In addition to the forward reaction, phosphorylated DNA can also be labeled with T4 DNA kinase in a two-step exchange reaction utilizing unlabeled ADP as an intermediate.

enzyme can also be used to phosphorylate synthetic DNA for certain DNA cloning strategies that utilize oligonucleotides. Bacterial alkaline phosphatase is an enzyme that removes the 5’ phosphate from DNA termini. Treatment of DNA with alkaline phosphatase is used to dephosphorylate DNA termini prior to labeling with T4 polynucleotide kinase and [γ-^{32}P]ATP or as a strategy to prevent re-ligation of vector DNA (Chapter 2). Polynucleotide kinase can also be used in a two-step exchange reaction to label 5′ phosphorylated DNA with [γ-^{32}P]ATP in the presence of excess ADP (Fig. 1.10).

DNA and RNA Polymerases

DNA and RNA polymerases direct the synthesis of complementary nucleic acids using single-strand DNA as the template. DNA synthesis requires a preexisting DNA or RNA primer with a 3′ hydroxyl, whereas RNA synthesis can initiate synthesis de novo. The

in vitro enzymatic synthesis of DNA and RNA has become a central component in a variety of molecular genetic applications, for example, the amplification of DNA sequences using the polymerase chain reaction (PCR). The polymerase chain reaction requires the use of a thermostable DNA polymerase called *Taq* DNA polymerase, which has optimal activity at 75°C (Chapter 6). One very important use of DNA polymerases in molecular genetic applications is for determining the nucleotide sequence of cloned DNA. This is done using a modified bacteriophage T7 DNA polymerase, called Sequenase™, in an in vitro reaction that contains dideoxynucleotides that serve as chain terminators. Other uses of DNA polymerases are for the production of complementary DNA (cDNA) using reverse transcriptase (Chapter 5) and for radioactively labeling DNA fragments with the Klenow fragment of *E. coli* DNA polymerase I.

RNA polymerases are used to synthesize single strand radioactive RNA probes using cloned DNA as a template and to produce mRNA suitable for in vitro protein synthesis or microinjection into cells. As with the DNA polymerases, commercially available bacteriophage RNA polymerases are especially useful in molecular genetic methods because of their high specific activity.

BIOCHEMICAL METHODS TO STUDY DNA AND RNA

Purification of Nucleic Acids

The most common step in any nucleic acid procedure is precipitation of DNA and RNA with ethanol in the presence of monovalent cations. If contaminating proteins must be removed prior to nucleic acid precipitation, then the sample is first extracted with an organic solvent such as phenol. Ethanol precipitation of nucleic acids from aqueous solutions yields a white or clear pellet that can easily be dissolved in an appropriate buffered solution such as a Tris-EDTA (pH 7). EDTA chelates Mg^{2+}, which functions as a co-factor for DNA nucleases (DNases) that may be present at very low levels in the sample. DNA and RNA can be separated using cesium chloride buoyant density gradient centrifugation in the presence of the DNA intercalating dye ethidium bromide. This type of density gradient is used routinely to separate different forms of DNA molecules based on the amount of ethidium bromide absorbed. Ethidium bromide intercalation causes local unwinding of the DNA helix, which reduces the molecular density of DNA. As shown in Figure 1.11, large linear genomic DNA molecules can bind more ethidium bromide than supercoiled plasmid DNA owing to the difference in topology of these two molecules. As described in Chapter 2, when plasmids are propagated in bacteria, topoisomerases introduce supercoils into circular DNA molecules. Because supercoiled plasmid DNA cannot be easily unwound, much less ethidium bromide can intercalate. In contrast, linear and relaxed circular DNA molecules are not as topologically constrained and can therefore absorb more ethidium bromide molecules, resulting in an overall decrease in molecular density.

It is also possible to separate DNA and RNA of different molecular weights using direct physical methods, such as size exclusion column chromatography and gel electrophoresis. Affinity matrices have also been developed using silica gels or anion exchange resins that preferentially bind nucleic acids under appropriate conditions and allow for the removal of proteins and polysaccharides from DNA preparations using crude cell lysates.

Figure 1.11 Cesium chloride density gradients can be used to separate genomic DNA and plasmid DNA based on differential buoyant densities resulting from the amount of ethidium bromide absorbed. The chemical structure of ethidium bromide is shown.

Gel Electrophoresis

The separation of nucleic acids by electrophoretic mobility is used for both analytical and preparative methods. DNA and RNA molecules are negatively charged owing to the phosphate backbone and the polymer structure results in a constant charge to mass ratio. Therefore, in a uniform electric field, nucleic acids migrate through a solid support matrix toward the positively charged anode at a rate that is inversely proportional to the $\log_{10}$ of the molecular weight. Physical measurements indicate that nucleic acids move through gel matrices as rods and that electrophoretic mobility of circular double-strand DNA can be affected by molecular topology and by the amount of superhelicity (supercoiled molecules migrate faster than relaxed circles). Two types of matrices, agarose and acrylamide, are used to separate DNA and RNA by gel electrophoresis. Because of the differences in pore size of these two gel matrices, standard agarose gels are used to separate nucleic acid molecules of 0.2–10 kb, and acrylamide gels are best suited for resolving nucleic acids less than 500 nucleotides long (see Appendix F).

Agarose is a purified linear polysaccharide polymer derived from a red seaweed that is commonly harvested for commercial applications. Liquid agarose gel is made by adding powdered agarose to a solution of electrophoresis buffer and heating to the boiling point to produce a homogeneous mixture. This solution is then poured into a Plexiglas gel support system and allowed to cool, resulting in the formation of a horizontal

slab gel containing slots at one end for sample loading (Fig. 1.12). Platinum electrodes are connected to a power supply and an electric field is established (~5 V/cm) in the presence of a Tris-acetate (or Tris-borate) electrophoresis buffer. Tracking dyes such as bromophenol blue or xylene cyanol are loaded with the samples to monitor electrophoretic mobility. Agarose concentration (mass/volume) determines the pore size and thus differentially affects the migration of small and large molecules. Applied voltage across the gel determines the current (field strength), which directly affects the velocity of electrophoretic mobility against a constant resistance. Nucleic acids are visualized by staining the gel with ethidium bromide, which fluoresces when exposed to ultraviolet light.

The other commonly used gel matrix for separating nucleic acids is polyacrylamide. Polyacrylamide, $[CH_2 = CHCONH_2]_n$, is formed by cross-linking chains of acrylamide with methylenebis-acrylamide $(CH_2 = CHCONH_2)_2CH_2$ in the presence of ammonium persulfate to generate free radicals, and TEMED (*N*,*N*,*N*′,*N*′-tetramethylethylenediamine), which stabilizes the free radicals and sustains the chemical reaction. An acrylamide matrix is formed with a porosity that is determined by both the concentration of acrylamide in the gel and the ratio of acrylamide to cross-linking agent (see Appendix F).

DNA Sequencing

A fundamental component of most applied molecular genetic strategies is a working knowledge of the DNA sequences being utilized in the approach. The nucleotide sequences of all cloned DNA fragments that have ever been published are archived in a large database managed by the U.S. government called GenBank. This database can be readily accessed through the Internet using the World Wide Web (Chapter 9). Up until

Figure 1.12 Horizontal agarose gel electrophoresis can be used to separate DNA molecules in the range of 0.2–10 kb; it is performed using a simple Plexiglas apparatus and an alternating current power supply. Nucleic acids can be visualized with ultraviolet light in agarose gels that have been stained with ethidium bromide.

the late 1980s, scientific journals still printed the actual sequence of cloned genes, and this information was then scanned into the database by National Institutes of Health (NIH) staff members. However, since about 1990, all newly characterized DNA sequences have been electronically deposited directly into GenBank through the Internet. These GenBank sequences are referred to by a specific database file number included in the published article.

There are two basic reasons why a researcher would need to determine the DNA sequence of a cloned DNA segment. First, if the experimental approach is aimed at characterizing genes or contiguous regions of genomic DNA that have not been studied before, the sequence of the unknown DNA would have to be determined so that it could be deposited into GenBank, which also allows other researchers to access the sequence information using a variety of computer algorithms (Chapter 9). Second, recombinant DNA fragments sometimes must be sequenced to (1) confirm the arrangement of cloned segments in a plasmid vector, (2) screen for sequence alterations introduced by mutagenesis, or (3) identify a gene product by comparing the sequence information with the GenBank database. Regardless of the objective, most DNA sequencing is now performed by automated DNA sequencing instruments that are run as an out-service, much like oligonucleotide synthesis services. The sequence information is returned to the user by E-mail or on a computer disk. Some DNA sequencing is still performed manually in the lab, however, it is much more cost-effective and labor-saving to have the sequence determined by a centralized facility using high-throughput instrumentation (Chapter 9).

Biochemical methods for DNA sequencing were developed in the 1970s by two groups. The chemical cleavage reaction was developed by Allan Maxam and Walter Gilbert, and the chain termination method was described by Fred Sanger, the same biochemist who worked out *N*-terminal peptide sequencing in the 1950s. Because the Sanger sequencing strategy is more amenable to automation and can provide more sequencing information per reaction than the Maxam and Gilbert method, this enzyme-based method has become the standard procedure. Both methods are based on producing a pool of single-strand DNA molecules that all have the same 5′ end, but differ in length by one nucleotide owing to random 3′ ends that have been generated in vitro. Because base-specific reactions are used, it is possible to determine the DNA sequence of the starting material by separating the fragments on an acrylamide gel using electrophoresis. Radioactive- or fluorescent-labeling methods are utilized to identify DNA molecules that share the same 5′ end. Figure 1.13 illustrates how a set of chain termination reactions utilizing dideoxynucleotides (ddNTPs) lacking the necessary 3′ OH group required for elongation are used to determine the complementary sequence of a DNA template. Chapter 9 describes the principle of automated DNA sequencing, which is based on the use of fluorescently labeled dideoxynucleotides to produce a pool of truncated DNA molecules that can be detected by their emission spectra following laser excitation.

Membrane Blotting and Hybridization of Nucleic Acids

Perhaps one of the most important and universal molecular genetic techniques to be developed over the past 25 years has been the use of solid support membranes to analyze DNA sequence similarity by nucleic acid hybridization. In 1975, Ed Southern published a paper in the *Journal of Molecular Biology* describing a technique to analyze DNA sequences bound to nitrocellulose membranes. This method became known as the

Figure 1.13 Sanger dideoxynucleotide DNA sequencing produces a pool of extended oligonucleotides that have random 3′ ends but the same 5′ terminus. DNA synthesis is initiated at room temperature by adding a modified T7 bacteriophage DNA polymerase, called Sequenase™, in the presence of all four dNTPs, one of which is radioactively labeled, most commonly [α-^{33}P]dATP or [α-^{35}S]dATP. The labeling reaction is then split into four tubes, each of which contains a different ddNTP, and incubated for a few minutes at 37^{o}C. Chain termination in the ddGTP reaction produces a pool of products that all end with ddGMP. Because the ddNTP chain terminators are present at a ~100-fold lower concentration than the dNTPs, chain elongation is able to proceed until a ddNTP is incorporated. The chain termination products are separated by denaturing polyacrylamide gel electrophoresis using tracking dyes (xylene cyanol and bromophenol blue), and the composite DNA sequence is determined by identifying which reaction contains the predominant truncated product at each position in the gel using autoradiography.

"Southern blot" and has been instrumental not only in understanding genome organization, but also in inspiring the development of powerful gene isolation techniques based on the principles of nucleic acid membrane blotting and hybridization.

The basic principle of the Southern blot, as it is done currently, is illustrated in Figure 1.14. Agarose gel electrophoresis is used to separate DNA fragments generated by restriction enzyme digestion, and then the gel is photographed to record the migration of DNA molecular weight markers. The gel is soaked in an alkaline solution (1*N* NaOH) for 30 minutes to denature the DNA, and then the gel is neutralized in Tris buffer. The gel is mounted onto a simple DNA transfer system that is assembled by stacking various materials in the following order: a glass plate placed over a reservoir containing a high salt solution called 20X SSC (3*M* NaCl/0.3*M* sodium citrate), a paper wick, the agarose gel containing denatured DNA, and a piece of nylon membrane. When absorbent paper such as a stack of paper towels is placed on top of the nylon membrane, it allows the 20X SSC solution to be drawn from the reservoir through the gel and into the absorbent paper. This process causes directional diffusion of the DNA out of the gel and transfers it directly onto the nylon membrane maintaining the DNA separation pattern seen on the gel. Following DNA transfer (~12 hours), the membrane is exposed to ultraviolet light to attach the single-strand DNA covalently to the nylon membrane by cross-linking. Detection of specific DNA fragments is accomplished by hybridizing a single-strand radioactive probe to the nylon filter under conditions that promote DNA reassociation. After the excess unhybridized probe is removed by washing, the nylon membrane is analyzed by autoradiography.

Other applications of DNA–DNA hybridizations are the DNA-based screening of genomic libraries (Chapter 3) and restriction-fragment-length polymorphism analysis (Chapter 4). Several useful variations of the original Southern blotting technique have subsequently been developed. The most closely related technique is called the "Northern blot." Northern blots use RNA gels and are processed in essentially the same way as Southern blots (Fig. 1.14) with the exception that organic denaturants (formaldehyde or glyoxal) are used to denature the RNA fully, rather than NaOH, which hydrolyzes RNA. Not to be outdone, immunologists and biochemists developed a protein-based blotting system called the "Western blot," which involves the electrophoretic transfer of proteins from a polyacrylamide gel onto a nylon membrane. These membranes are incubated (not hybridized!) with antibodies under conditions that allow the detection of specific proteins. Two somewhat esoteric variations on the antibody-based Western blot are the "Southwestern blot," which uses double-strand radioactive DNA probes to identify putative DNA binding proteins, and the "Farwestern blot" which relies on high affinity protein–protein interactions, similar to those found in yeast two-hybrid screens (Chapter 4), to identify candidate binding proteins.

Laboratory Practicum 1. *Identifying the transcriptional start site of a gene transcript*

Research Objective

A molecular endocrinologist is interested in estrogen-regulated gene expression in mammary epithelial cells. She has recently isolated a nearly full-length cDNA clone and the corresponding 5′ genomic DNA sequences for a gene that is induced 10-fold by estrogen treatment of human mammary epithelial cells in culture. Her research objective is to identify the 5′ end of the gene transcript in order to facilitate future studies aimed at investigating estrogen-regulated expression of this gene in normal and tumorigenic mammary epithelial cells.

Figure 1.14 Membrane blotting and hybridization using the Southern blot technique. The use of nylon membranes, rather than nitrocellulose as originally done by Southern, has increased the utility of Southern blotting by allowing the more durable nylon membranes to be rehybridized multiple times with different DNA probes. The four DNA samples (S), and molecular weight markers (Mkr), are visualized by staining the gel with ethidium bromide.

Available Information and Reagents

1. Based on the cDNA sequence, and the estimated size of the gene transcript from Northern blots, she predicts that the 5′ end of the transcript is 50–150 nucleotides farther upstream of the sequence in her longest cDNA clone.
2. A plasmid subclone of genomic DNA has been constructed that corresponds to a 2 kb region that overlaps with the most 5′ cDNA sequence and therefore is likely to contain the gene promoter.
3. An antisense oligonucleotide has been designed that is 24 nucleotides long and has a 3′ end that is located 10 nucleotides downstream of the 5′ terminus of the cloned cDNA fragment.
4. RNA from untreated and estrogen-treated mammary epithelial cells has been isolated and shown by Northern blots to contain a 10-fold difference in steady-state levels of the new gene transcript.

Basic Strategy

There are two methods commonly used to map the 5′ end of mRNA transcripts when both the cDNA and genomic DNA corresponding to the 5′ region of the gene have been cloned. The first method is shown in Figure 1.15 and is called RNase mapping. This method uses in vitro RNA synthesis to produce radioactively labeled complementary RNA that includes sequences upstream and downstream of the predicted 5′ terminus of the mRNA. Following solution hybridization between the complementary RNA probe and total cellular RNA, the reassociated heteroduplexes are treated with RNases that degrade single-strand (unhybridized) RNA. The products of RNase digestion are separated on a polyacrylamide gel and the sizes of the undigested and digested RNA probe are determined. The second method is called primer extension, which utilizes an end-labeled oligonucleotide that serves as a primer for cDNA synthesis using the enzyme reverse transcriptase (Fig. 1.16). In the presence of this gene-specific primer, dNTPs, and cellular RNA, reverse transcriptase synthesizes cDNA from any primer that is annealed to template RNA. The length of the longest end-labeled cDNA products should correspond to the total number of nucleotides between the 5′ end of the primer on the antisense strand and the 5′ terminus of the mRNA template. For both the RNase mapping and primer extension methods, the pattern of product formation from parallel reactions using RNA from either untreated or estrogen-treated cells would be used to confirm specific mapping of the estrogen-induced gene.

Comments

The availability of in vitro transcription systems using bacteriophage-specific promoters led to the development of the RNase mapping method by facilitating the synthesis of single-strand high specific activity probes. Both T7 and SP6 bacteriophage RNA polymerase-dependent in vitro transcription systems have been developed. Titration of the RNase digestion conditions is done as an initial experiment to avoid under- or overdigestion of the heteroduplex substrate (Fig. 1.15). In addition to mapping 5′ ends of gene transcripts, RNase mapping can be used to identify RNA splice sites and to measure quantitatively steady-state levels of RNA under different physiological conditions. The primer extension technique (Fig. 1.16) is a reliable method to confirm RNase mapping studies because it relies on product synthesis rather than substrate degradation. A third approach, not shown here, is called S1 nuclease mapping, which is similar to RNase mapping except that a single-strand end-labeled DNA probe is used.

Figure 1.15 Identification of the 5′ end of a gene transcript using RNase mapping. (*a*) The radioactive antisense RNA probe is synthesized in vitro using T7 bacteriophage RNA polymerase and T7 promoter sequences that flank the multiple cloning site. By titrating the RNase digestion time, it is possible to identify the major protected fragment. Discrete anomalous bands arise from preferential digestion of the RNA probe near regions of secondary structure. (*b*) Results from the autoradiograph would be used to predict that the transcriptional start site is located 128 nucleotides upstream (5') of the EcoRI site in exon 1.

Figure 1.16 Primer extension can be used to identify the 5' end of a gene transcript. (*a*) A 24 nucleotide (24-mer) antisense oligonucleotide was designed based on the cDNA sequence. This 24-mer includes two nucleotides from the EcoRI site at its 3′ end and is radiolabeled with T4 polynucleotide kinase and [γ-^{32}P]ATP at the 5′ end. Following primer hybridization to RNA from either estrogen-free (–E) or estrogen-treated (+E) cells, reverse transcriptase and dNTPs are added to initiate cDNA synthesis. A product of 150 nucleotides would corroborate the RNase mapping studies. Note that there is a ~10-fold difference in the amount of 150 nucleotide product, depending on whether the RNA was isolated from –E or +E cells (compare lanes 3 and 4). Moreover, cDNA synthesis should depend on the presence of reverse transcriptase (lane 1) and inclusion of the antisense 24-mer oligonucleotide in the reaction (lane 2). (*b*) The observed 150 nucleotide long extended cDNA product predicts that the 5′ mRNA terminus is 126 nucleotides upstream of the primer 3′ end.

Prospective

Once the 5′ end of a gene transcript is localized within the context of a genomic sequence, it becomes possible to test functionally for gene promoter activity using sequences within the first ~200 nucleotides upstream of the transcriptional start site. In this example, the researcher could construct a reporter plasmid (Chapter 4) that contains the putative estrogen-regulated promoter and test its activity in normal and tumorigenic mammary cells that have been treated with estrogens. Subsequent promoter mapping experiments could then be done to determine if this gene is a primary target of estrogen action, which may be important to understanding its regulation in normal and neoplastic mammary cells. In vitro transcription studies could also be performed to map the transcriptional start site. This would be done using truncated versions of the cloned genomic DNA as a template in reactions containing nuclear cell extracts and [α-^{32}P]UTP.

REFERENCES

Flow of genetic information

Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. Essential cell biology: An introduction to molecular biology of the cell. Garland Publishing, New York, 630 pp., 1998.

*Avery, O., MacLeod, C., MacCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J. Exp Med. 79:137–158, 1944.

*Crick, F.H.C. Codon-anticodon pairing: The wobble hypothesis. J. Mol. Biol. 19:548–555, 1966.

*Meselson, M., Stahl, F. The replication of DNA in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 44:671–682, 1958.

*Watson, J.D. The involvement of RNA in the synthesis of proteins. Science 140:17–26, 1963.

Nucleic acid biochemistry

*Britten, R., Kohne, D. Repeated sequences in DNA. Science 161:529–540, 1968.

Mattencci, M., Caruthers, M. Synthesis of deoxyoligonucleotides on a polymer support. J. Am Chem. Soc. 103:3185–3191, 1980.

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

*Watson, J.D., Crick, F.H.C. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. Nature 171:737–738, 1953.

Wetmur, J.G., Davidson, N. Kinetics of renaturation of DNA. J. Mol. Biol. 31:349–370, 1986.

Wing, R., Drew, H., Takano, T., Brodka, C., Tanaka, S., Itakura, K., Dickerson, R.E. Crystal structure analysis of a complete turn of B-DNA. Nature 287:755–758, 1980.

DNA metabolizing enzymes

Chaconas, G., van de Sande, J.H. 5′-^{32}P labeling of RNA and DNA restriction fragments. Methods Enzymol. 65:75–88, 1980.

Challberg, M.D., Englund, P.T. Specific labeling of 3′ termini with T4 DNA polymerase. Methods Enzymol. 65:39–43, 1980.

Doublie, S., Tabor, S., Long, A., Richardson, C., Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. Nature 391:251–258, 1998.

Engler, M.J., Richardson, D.C. DNA ligases. In The Enzymes, Vol. 15B (P.D. Boyer, ed.). Academic Press, New York, pp. 3–30, 1982.

* Landmark papers in applied molecular genetics.

Fuchs, R., Blakesley, R. Guide to the use of type II restriction endonucleases. Methods Enzymol. 100:3–38, 1983.

*Jacobsen, H., Klenow, H., Overgaard-Hansen, K. The N-terminal amino-acid sequences of DNA polymerase I from *Escherichia coli* and of the large and the small fragments obtained by a limited proteolysis. Eur. J. Biochem. 45:623–627, 1974.

Kovall, R., Matthews, B. Structural, functional, and evolutionary relationships between λ-exonuclease and the type II restriction endonucleases. Proc. Natl. Acad. Sci. USA 95:7893–7897, 1998.

*Lawyer, F.C., Stoffel, S., Saiki, R.K., Myambo, K., Drummond, R., Gelfand, D.H. Isolation, characterization, and expression of Escherichia coli of the DNA polymerase gene from *Thermus aquaticus*. J. Biol. Chem. 264:6427–6437, 1989.

Lehman, I.R. DNA polymerase I of Escherichia coli. In The Enzymes, Vol. 14A (P.D. Boyer, ed.). Academic Press, New York, pp. 16–38, 1981.

*Linn, S, Arber, W. Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. Proc. Natl. Acad. Sci. USA 59:1300–1306, 1968.

*Modrich, P., Lehman, I.R. Enzymatic joining of polynucleotides. J. Biol. Chem. 245: 3626–3631, 1975.

Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I., Aggarwal, A. Structure of restriction endonuclease BamHI and its relationship to EcoRI. Nature 368:660–664, 1994.

Reid, T.W., Wilson, I.B. Escherichia coli alkaline phosphatase. In The Enzymes, Vol. 4 (P.D. Boyer, ed.). Academic Press, New York, pp. 373–415, 1971.

Richardson, C.C. Bacteriophage T4 polynucleotide kinase. In The Enzymes, Vol. 14A (P.D. Boyer, ed.). Academic Press, New York, pp. 299–314, 1981.

Roberts, R. Restriction and modification enzymes and their recognition sequences. Nucl. Acids Res. 11:r135–r167, 1983.

Rong, M., He, B., McAllister, W., Durbin, R. Promoter specificity determinants of T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 95:515–519, 1998.

*Tabor, S., Richardson, C.C. DNA sequencing analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84:4767–4771, 1987.

Verma, I.M. Reverse transcriptase. In The Enzymes, Vol. 14A (P.D. Boyer, ed.). Academic Press, New York, pp. 87–104, 1977.

Biochemical methods to study DNA and RNA

Arnheim, N., Southern, E.M. Heterogeneity of the ribosomal genes in mice and men. Cell 11:363–370, 1977.

Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., Smith, J., Struhl, K. Current protocols in molecular biology. John Wiley & Sons, New York, 1996.

Birnboim, H.C. A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymol. 100:243–255, 1983.

Chrambach, A., Rodbard, D. Polyacrylamide gel electrophoresis. Science 172:440–451, 1971.

*Denhardt, D.T. A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. 23:641–646, 1966.

Doel, M.T., Houghton, M., Cook, E.A., Carey, N.H. The presence of ovalbumin mRNA coding sequences in multiple restriction fragments of chicken DNA. Nucl. Acids Res. 4:3701–3713, 1977.

Helling, R.B., Goodman, H.M., Boyer, H.W. Analysis of R. EcoRI fragments of DNA from lambdoid bacteriophages and other viruses by agarose gel electrophoresis. J. Virol. 14: 1235–1244, 1974.

Holmes, D.S., Quigley, M. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114:193–197, 1981.

Khandjian, E.W. Optimized hybridization of DNA blotted and fixed to nitrocellulose and nylon membranes. Bio/ Technology 5:165–167, 1987.

Lis, J.T., Schleif, R. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. Nucl. Acids Res. 2:383–389, 1975.

*Marmur, J. A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol. 3:208–218, 1961.

*Maxam, A., Gilbert., W. A new method of sequencing DNA. Proc. Natl. Acad. Sci. USA 74: 560–564, 1977.

*McDonell, M.W., Simon, M.N., Studier, F.W. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. 110:119–146, 1977.

Nygaard, A.P., Hall, B.D. A method for the detection of RNA-DNA complexes. Biochem. Biophys. Res. Commun. 12:98–104, 1963.

*Radloff, R., Bauer, W., Vinograd, J. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: The closed circular DNA in HeLa cells. Proc. Natl. Acad. Sci. U.S.A. 57:1514–1521, 1967.

Sambrook, J., Fritsch, E.F., Maniatis, T. Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989.

*Sanger, F., Nicklen, S., Coulson, A. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467, 1977.

*Sharp, P., Sugden, B., Sambrook, J. Detection of endonuclease activities in *Hemophilus parainfluenza* using analytical agarose-ethidium bromide electrophoresis. Biochemistry 12: 3055–3062, 1973.

*Smith, L.M., Sanders, J., Kaiser, R., Hughes, P., Dodd, C., Connel, C., Heriner, C., Kent, S., Hood, L.E. Fluorescence detection in automated DNA sequence analysis. Nature 321: 674–679, 1986.

*Southern, E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517, 1975.

*Wahl, G.M., Stern, M., and Stark, G.R. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA 76:3683–3687, 1979.

Identifying the transcriptional start site of a gene transcript

*Berk, A.J., Sharp, P.A. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12:721–732, 1977.

Graves, B.J., Eisenberg, S.P., Coen, D.M., McKnight, S.L. Alternate utilization of two regulatory domains within the Moloney murine sarcoma virus long terminal repeat. Mol. Cell. Biol. 5:1959–1968, 1985.

Jones, K.A., Yamamoto, K.R., Tjian, R. Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell 42:559–572, 1985.

*McKnight, S.L., Kingsbury, R. Transcription control signals of a eukaryotic protein-encoding gene. Science 217:316–324, 1982.

*Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., Green, M.R. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. 12:7035–7056, 1984.

Miesfeld, R., Arnheim, N. Identification of the in vivo and in vitro origin of transcription of human ribosomal DNA. Nucl. Acids Res. 10:3933–3949, 1982.

Tremea, F., Batistuzzo de Medeiros, S.R., ten Heggeler-Bordier, B., Germond, J.-E., Seiler-Tuyns, A., Wahli, W. Identification of two steroid-responsive promoters of different strength controlled by the same estrogen-responsive element in the 5′-end region of the *Xenopus laevis* vitellogenin gene A1. Mol. Endocrinol. 3:1596–1609, 1989.

*Zinn, K., DiMaio, D., Maniatis, T. Identification of two distinct regulatory regions adjacent to the human β-interferon gene. Cell 34:865–879, 1983.