Chapter 2

DETECTION AND ANALYSIS OF NUCLEIC ACIDS

A. Introduction

The year 2003 marked the 50th anniversary of the paper by James Watson and Francis Crick reporting the solution of the three dimensional structure of deoxyribonucleic acid (DNA) (9). This achievement earned them a Nobel Prize, which they shared in 1962 with Maurice Wilkins for his crystallographic studies on DNA structure (Table 2). It also set the stage for a new era which led to breakthroughs in our understanding of the genetic code and the development of techniques that allowed for the rapid sequencing of DNA. Eventually, the entire genome of a number of organisms would be sequenced, culminating in the complete sequencing of the human genome in 1991. An understanding of DNA structure also permitted the development of methods to clone and manipulate DNA (recombinant DNA technology), as well as alter the sequence of DNA using site directed mutagenesis to gain insights into the function of individual proteins. The advent of the molecular biology revolution also allowed us to begin to understand how genes are regulated. The regulation of gene expression is not only the key to our understanding of how genes are turned on and off, but to how genes are regulated in a tissue specific manner, how cellular differentiation occurs, and how tissues and organs develop. From the development of simple methods to clone, sequence, and manipulate DNA came a realization of its power to help solve the genetic basis of human disease. This, in turn, resulted in the development and evolution of the Biotechnology Industry, an industry that was founded to develop ways to synthesize hormones, enzymes and other proteins, rather than rely on their purification from natural sources. The production of hormones in the laboratory made it practical to treat patients on a large scale for the first time.

The description of the “double helix” by Watson and Crick, while a watershed discovery, was not the beginning of our understanding of the

Table 2. Selected Nobel Laureates (Molecular Biology)

George W. Beadle Edward L. Tatum Joshua Lederberg	genetic recombination and discoveries on the chemical basis of genes in bacteria	1958†
Severo Ochoa Arthur Kornberg	synthesis of RNA and DNA	1959†
Francis H. C. Crick James D. Watson Maurice H. F. Wilkins	discovery of the 3-dimensional structure of DNA as a double helix; base pairing	1962†
François Jacob André Lwoff Jacques Monod	analysis of how genes are regulated (bacteria and viruses); advent of "molecular biology" era	1965†
Robert W. Holley Har Gobind Khorana Marshall W. Nirenberg	solution of the genetic code; chemistry of protein synthesis	1968†
Max Delbrück Alfred D. Hershey Salvador E. Luria	mechanisms of replication and genetic structure of viruses (bacteriophages)	1969†
David Baltimore Renato Dulbecco Howard M. Temin	characterization and function of RNA tumor viruses; the discovery of reverse transcriptase	1975†
Werner Arber Daniel Nathans Hamilton O. Smith	discovery of restriction enzymes; use of restriction enzymes for genetic analysis	1978†
Paul Berg Walter Gilbert Frederick Sanger	first recombinant DNA (Berg); DNA sequencing methods (Gilbert, Sanger)	1980*
Richard J. Roberts Phillip A. Sharp	"split genes" (intron/exon structure of eukaryotic genes)	1993†
Kary B. Mullis Michael Smith	invention of the PCR (Mullis); site-directed mutagenesis (Smith)	1993*

†Nobel Prize in Physiology or Medicine
*Nobel Prize in Chemistry

composition of DNA or its role in the makeup of genes. The idea that genes are independent and heritable is usually traced to the studies of the Czech monk, Gregor Mendel, who demonstrated that phenotypic traits are inherited in his studies of peas in 1866. Soon afterwards (1868), Friedrich Miescher isolated a new substance that was neither protein nor sugar from the nucleus of cells which he called nucleic acid. However, the link between nucleic acid and genetic makeup was not made until the landmark publication by Oswald Avery and his colleagues, Colin MacLeod and Maclyn McCarty, in 1944 (10). In this paper, they showed for the first time that hereditable traits could be transferred using purified

DNA from one bacterium to another (and its daughter cells). As is often true of ground-breaking discoveries, the acceptance of this one was tempered by skepticism due to the prevailing wisdom of the time that proteins were responsible for heredity. Indeed, the composition of the nucleic acids that made up the DNA was considered to be too simple to be responsible for something as complex as a gene. However, the conclusion from Avery's research was unmistakable: nucleic acids, and thus DNA, was responsible for the genetic makeup of an organism. How DNA was responsible for the transmission of genetic traits could only be understood after Watson and Crick had solved the 3-dimensional structure of DNA and the genetic code had been broken. But these ground-breaking studies would not have been possible without the work of Avery. While the awarding of a Nobel Prize is sometimes controversial (we'll come back to that later), the failure to award one to Avery has been considered by some to be one of the most egregious oversights in Nobel history.

Within 20 years of the discovery of the 3-dimensional structure of DNA, its chemical structure was solved, DNA replication was understood, and the basic parameters of how DNA is precisely transcribed into RNA (ribonucleic acid), which is then translated into protein, became clear. The development of many new technologies was required to achieve these breakthroughs, and many of the techniques that were invented for these discoveries remain in use today. Some of these were so instrumental in their applicability to solve biological problems that they also garnered Nobel Prizes for their inventors (Table 2).

The following are brief descriptions of some of the common methods used to manipulate nucleic acids, especially DNA.

B. Basic methods for nucleic acid analysis

Gels and gradients for separation of nucleic acids

The size of a piece of DNA, usually produced as the result of restriction digestion (see below) can be determined by separation on gradients or gels. For gradient separation, sucrose density gradients were common at one time but were supplanted by gels once electrophoretic separation techniques were developed. Two types of gel matrices are commonly used for the separation of nucleic acids: those made from polyacrylamide, and those using agarose. When a mixture of DNA fragments is applied to one end of the gel and subjected to an electric current, the DNA fragments migrate toward the positive pole of the field, the anode, due to the negatively charged phosphate groups that make up the DNA backbone. Following electrophoretic separation, the DNA can be visualized by one or more methods as discussed below.

Polyacrylamide gel electrophoresis (PAGE)

PAGE gels, as previously mentioned, are used for high resolution separation of nucleic acids, such as for resolving DNA ladders in DNA sequencing reactions or for RNase protection assays. For nucleic acid separation, PAGE gels served as the first matrix used to demonstrate that restriction fragments of nucleic acids could be resolved in gels, and their sizes could be calculated from their migration patterns (11). In general, relatively low percentage gels are used for resolving nucleic acids.

Agarose gel electrophoresis

Agarose gel electrophoresis is one of the standard methods of molecular biology and is a method used to resolve and separate nucleic acid fragments of different sizes. The use of agarose as a matrix to separate DNA was quickly adapted once it was established that DNA fragments could be separated by gel electrophoresis.

Agarose is a sugar polymer isolated from seaweed; when boiled and cooled, it forms a "gel", much like jello. The agarose gel serves to impede the migration of the DNA, so that larger fragments of DNA migrate slower than smaller fragments. Because DNA migrates in an electric field as a consequence of the phosphates on the DNA backbone, the charge to mass ratio is identical for all DNA fragments, irrespective of their length. As a result, while the separation is due to the charge of the DNA, the migration of DNA fragments is inversely proportional to the log (base 10) of the molecular weight of the fragment. Thus, if fragments of unknown size are compared with a control "ladder" containing fragments of known sizes (these are commercially available), the sizes of unknown fragments can be readily calculated. Depending on the concentration of agarose in the gels, DNA fragments from under 100 base pairs (bp) to 30,000 base pairs (30 kilobases, or kb) can be resolved by standard agarose gel electrophoresis. Following separation of the DNA, the fragments can be visualized in the gel using fluorescent dyes that intercalate (*i.e.,* insert) between the bases of the DNA helix. The most common dye used is ethidium bromide, which appears red upon exposure to ultraviolet light, but other dyes, such as SYBER green, are becoming increasingly popular.

Pulsed field electrophoresis

This is a method that allows for the separation of much larger fragments of DNA than standard agarose gel electrophoresis. In this method,

the direction of the electrical current is altered periodically, allowing for the separation of fragments from 30 kb to several million base pairs.

Alkaline agarose gels

Agarose gel electrophoresis works well for separating double stranded DNA fragments and for the separation of RNA, but does not resolve single stranded DNA species. The incorporation of a base such as sodium hydroxide into the buffers used in the gels permits good resolution of single stranded DNAs.

Restriction enzymes and restriction fragments

Restriction endonucleases are enzymes that cleave DNA at specific sites. They are one of the most important tools in the analysis of DNA and for recombinant DNA technology. Restriction enzymes come from bacteria. They are a component of the "restriction-modification systems" of bacteria. These modification systems are thought to have evolved to help protect bacteria against foreign DNA, whether introduced by viral infection (by bacteriophages) or by DNA from other species of bacteria taken up by the host bacterium (see below). The fundamental work on restriction modification systems by Werner Arber, including the discovery of restriction enzymes, helped lead to the first insights into their sequence specificity by Kelly and Smith in 1970 (12) and to their use to characterize restriction fragments of viral DNA by Danna and Nathans in 1971 (11). These studies on restriction enzymes earned the Nobel Prize for Arber, Hamilton Smith and Daniel Nathans in 1978.

Hundreds of restriction enzymes (over 600 at this point) have been isolated. Restriction enzymes recognize short specific sequences in double stranded DNA. The sequences are generally palandromic, meaning they read the same (have the same nucleotides) in opposite orientations on the complementary DNA strands. Palandromic sequences can occur because of the base-pairing that is required in double stranded DNA (dsDNA). The 4 nucleotide bases adenine (A), cytosine (C), guanine (G) and thymine (T) exist in dsDNA as A-T and G-C complementary pairs. When the appropriate sequence is recognized, the restriction enzymes cut the DNA at specific sites within this recognition site, giving rise to a restriction fragment, which represents the piece(s) of DNA produced after cleaving the DNA with one or more restriction enzymes.

The actual cuts produced by a restriction enzyme can be "blunt" (that is, they cut at the same place on both strands of DNA) whereas others make symmetrical cuts with "overhangs". For example, the restriction

enzyme *Eco*RI (isolated from the bacterium *Escherichia coli*) recognizes the DNA base sequence GAATTC and cuts between the G and A (on the opposite strand the sequence is CTTAAG, and the cut occurs between the A and G on that strand). Restriction enzymes are often classified by the number of nucleotides they recognize. Because *Eco*RI recognizes a 6 base pair sequence, it is called a "6-cutter". Most restriction enzymes recognize 4, 6 or 8 bp sequences, although some have longer recognition sites. The longer and more specific the sequence recognized by a restriction enzyme, the fewer of these sites that are expected to be present in any given stretch of DNA. While the majority of restriction enzymes recognize DNA sequences irrespective of their methylation status, some restriction enzymes recognize a sequence motif in unmethylated but not methylated DNA.

Restriction enzymes are a basic tool in the analysis of DNA (restriction mapping) and recombinant DNA technology. Restriction enzymes are used to enzymatically digest genomic DNA into fragments before electrophoretic separation and Southern blotting, to generate fragments of DNA for cloning into vectors, and production of small fragments from a gene for making probes for Southern or Northern blots. DNA can be digested to completion using restriction enzymes or, for the production of genomic libraries where large stretches of DNA are desired, partial digests can be carried out. Partial digestion means that the enzyme is used in sub-optimal concentrations and that the DNA is incubated with the enzyme for a limited period of time. In this way, large pieces of DNA with internal sites that have not been cut by the restriction enzyme are generated and can be used for cloning purposes.

Restriction fragment length polymorphism (RFLP)

RFLPs are a type of length polymorphism in the genomes of homologous segments of DNA of different individuals that have been digested with a restriction enzyme. The digested DNA is resolved on agarose gels, and specific probes are used to identify regions of the DNA that exhibit length polymorphisms. The use of RFLPs to help map genes involved in human diseases was first suggested by David Botstein (13), and is now widely used not only to help map genes that contribute to diseases, but in forensics for the precise identification of DNA samples (see below).

Single nucleotide polymorphisms (SNP)

SNPs are single base pair variations in the DNA sequence of an individual gene. They occur in the human population at a relatively high

frequency and are very abundant, occurring at a rate of about 1 per 1000 base pairs. SNP genotypes are used in mapping disease susceptibility loci as well as in comparative genetics, evolutionary genetics, and forensic analysis.

Southern blotting

After resolving complex mixtures of DNA (such as restriction digested genomic DNA) on a gel, how can you determine which restriction fragment in a smear of DNA contains a particular gene? One method that was initially used was to manually cut slices out of the gel, isolate the DNA from the slices, and then hybridize the isolated DNA with a specific probe. This was not only highly laborious but also extremely inefficient. A second method, which revolutionized the ability to analyze DNA, was introduced by E.M. Southern in 1975 (14), and was rapidly adapted to the analysis of other macromolecules. Southern realized that because gels are porous, macromolecules within the gel could be transferred to another medium by a method called "blotting through". The mediums used are membranes, usually constructed of nitrocellulose or nylon. The DNA is denatured (with NaOH, since only ssDNA can transfer), the membrane is placed onto the gel, and then capillary action or a pump-based suction method is used to transfer the DNA from the gel onto the membrane using high salt solutions. While the capillary action method of transfer is "low tech", it is highly efficient and still used by many investigators since it takes longer and allows them the time to go home and sleep! The DNA is then immobilized on the membrane using heat (nitrocellulose) or ultraviolet light (nylon), and the entire membrane can now be probed.

Blotting methods work well for agarose gels but not for polyacrylamide gels, particularly the higher percentage gels used for proteins. For these, the transfer is accomplished by electroblotting (see Chapter 1).

Dot blotting and slot blotting

These are methods to immobilize bulk unfractionated DNA onto a membrane using a suction manifold. The manifolds have circular wells (dot blots) or slit wells (slot blots). They are frequently used to determine if a nucleic acid is present in a sample and to determine the relative abundance of the nucleic acid. They are most frequently used for RNA analysis. Since fractionation is not used, they do not rely on the integrity of the RNA. These blotting methods are now rarely used and have been supplanted almost entirely by polymerase chain reaction methods, which are more rapid, accurate, and quantitative.

Probes to identify nucleic acids on membranes

To identify a particular nucleic acid (whether DNA or RNA) on a membrane, probes are used. A probe is a general name for a substance that is used to specifically identify a macromolecule. It can be a chemical, an antibody, a protein, or a fragment of DNA.

For identifying nucleic acids on membranes, DNA probes are often used. These probes are labeled by one of several methods with either a radioactive isotope, a dye or fluorescent compound, or an enzyme. The probes are hybridized (annealed) to nucleic acids on the membrane. If a complementary nucleic acid is present, the probe anneals, and its presence can be detected by a method suited to the detection of the probe, depending on the labeling method. The probe must be single stranded in order to hybridize. As a result, labeled double stranded DNA fragments must be denatured prior to hybridization. Single stranded probes can also be synthesized using the appropriate vector (see below).

After hybridization reactions, membranes are washed in solutions containing formamide, SSC (a salt solution containing sodium chloride and sodium citrate) and SDS (sodium dodecyl sulfate). By reducing the concentration of SSC and increasing the temperature at which the membrane is washed, only highly complementary segments of nucleic acids will remain annealed to the probe. This is generally referred to as increasing the stringency between probe and target sequence.

Autoradiography (exposing the blot to x-ray film) is typically used to visualize the hybridization of a radioactive probe to the membrane. Alternatively, the hybridized probe can be identified and quantitated by phosphorimaging. For probes that have been labeled with fluorescent dyes or enzymes, colorimetric methods that rely on chemiluminescence can also be used (see Chapter 4).

C. DNA sequencing

Introduction

DNA sequencing is, as the name implies, a method that reveals the linear sequence of nucleic acids within a stretch of DNA. DNA is composed of four nucleotide bases: adenine (A), cytosine (C), guanine (G), and thymine (T). The base present at a particular position in a DNA sequence is identified by the presence of a product of a chemical or enzymatic reaction in a gel. The ability to sequence DNA was a major advance for a number of reasons. First, it allows for the prediction of a potential open reading frame (ORF) for genes within a genomic DNA sequence. An ORF is the sequence that is transcribed and translated

into protein, and in eukaryotic DNA is not contiguous but interrupted by intronic sequences. DNA sequencing is far more rapid and significantly cheaper than protein sequencing methods, and thus is the method of choice for most sequencing projects. The sequence of a protein can be used to predict its function by identifying potential protein interaction sites, DNA interaction sites, or motifs critical to enzymatic function or consensus enzyme recognition motifs. DNA sequencing also allows for the rapid identification of potential regulatory elements within genes (both upstream promoter sequences as well as intronic and 3′ enhancers) by comparing DNA sequence information with known transcription factor binding sites, or evaluating non-coding sequences in the introns around a gene that are conserved across species. These elements often predict sequences with regulatory functions. Sequencing therefore provides basic information about protein structure that can then be confirmed (or rejected) by directly testing these domains for functional activities.

Two major methods have been developed for the sequencing of DNA, one based on chemical methods (the "Maxam-Gilbert" method, named after its developers) and an enzymatic chain termination method often referred to as "dideoxy" sequencing or "Sanger sequencing" after its developer. Gilbert and Sanger shared with Nobel prize for these technical advances in 1980 with Paul Berg, who developed recombinant DNA techniques.

Both DNA sequencing techniques rely on reactions that generate oligonucleotide ladders, which terminate in definable nucleotides (A T G or C), although the methods used to achieve these are quite different. In both methods, a piece of DNA that can be manipulated in the sequencing reactions (often generated using restriction enzymes) is inserted into a cloning vector. Cloning vectors contain unique sites that flank both sides the cloned DNA fragment, the insert, to be sequenced. These sites are complementary to oligonucleotide primers that initiate the reactions. DNA sequencing generally relies on the use of high resolution denaturing polyacrylamide gels which can resolve single stranded DNA ladders over several hundred base pairs.

Maxam-Gilbert sequencing

This method (15) relies on the use of chemical reagents that modify specific bases, followed by the cleavage of the DNA at one or two specific nucleotides. Labeled DNA (usually radioactively labeled at the 3′ or 5′ ends of the strands) is subjected to one of 3 chemical reactions which modify bases in a specific way. Depending on the modification, the

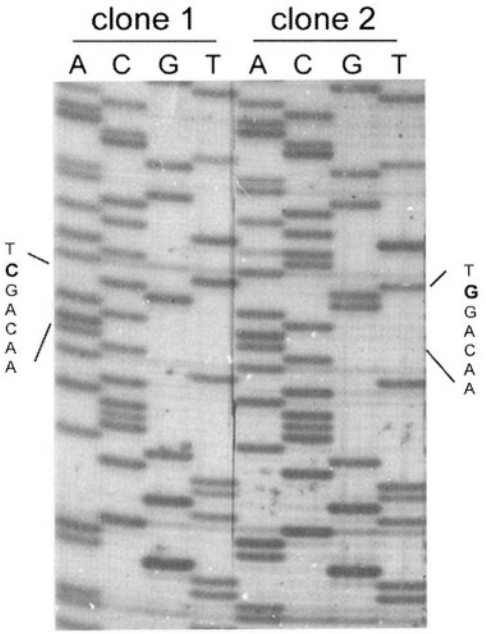
chemical piperidine then catalyzes the breakage of DNA at one or two predictable bases. Application of each reaction to adjacent lanes then gives "ladders" which can be deciphered to reveal the DNA sequence. Maxam-Gilbert sequencing, because it uses toxic chemicals, is not the method of choice for large scale sequencing, but it is still used for DNA footprinting where contact sites on DNA are identified.

Dideoxy or chain-termination sequencing (also known as Sanger sequencing)

This is an enzymatic method (16) that relies on the fact that the incorporation of dideoxynucleotides (ddNTPs) rather than deoxynucleotides (dNTPs) results in the termination of elongation of DNA catalyzed by DNA polymerase. DNA polymerases works by elongation of a DNA strand complementary to a template DNA, and requires a hydroxyl (OH) group at the 3′ position for DNA polymerase to add the complementary nucleotides. Since dideoxynucleotides lack a OH group present in deoxynucleotides, the incorporation of a ddNTP terminates elongation.

In dideoxy sequencing, denatured dsDNA or ssDNA is copied (sequenced) using a DNA polymerase. To copy DNA, a DNA polymerase needs a primer and a supply of the 4 bases (dNTPs). A primer is a short oligonucleotide sequence of DNA complementary to one strand that "primes' the process (for dideoxy sequencing, vectors that have so-called "universal priming sites" are generally used). To make the sequence readable, 4 different reactions are used. In each of the 4 different reactions a ddNTP (ddATP, ddCTP, *etc.*) is added in limiting concentrations. Included in each reaction is a radioactive nucleotide (usually α-^{32}P-dATP is used). The DNA polymerase initiates chain elongation in a 5′ → 3′ direction, starting at 3′ end of the oligonucleotide primer that has been annealed to the DNA template. The DNA polymerase adds deoxynucleotides, which are selected by base-pair matching, to the elongating DNA chain. DNA polymerases can incorporate both dNTPs and ddNTPs as substrates. As a result, chain elongation is terminated whenever a ddNTP is incorporated since they have no 3′ hydroxyl groups. After incubation, the individual reactions containing one of the 4 ddNTPs are run on separate lanes of an acrylamide DNA sequencing gel and then exposed to film. The banding pattern reveals the sequence of the DNA. Figure 2 shows an example of a dideoxy sequencing reaction, with part of the corresponding sequences of two independent clones of a given stretch of DNA. Notice that one of the sequenced clones has a C → G substitution, otherwise known as a mutation.

Figure 2. DNA Sequencing



Automated DNA sequencing

Automated sequencing uses the dideoxy sequencing method, but instead of using a radioactive label fluorescent ddNTPs are used, each labeled with a different fluorophore. Once the reactions are completed, they are resolved on a PAGE gel. Because the laser in the automated reader can distinguish the 4 different fluorescent labels, the sequencing can be accomplished by running all 4 reactions in a single lane. Automated sequencing is not only more efficient (more reactions can be loaded on each gel), but generally provides more sequence information per reaction, generating 600–1000 bp or more of readable sequence versus the 300–400 possible with manual sequencing methods.

D. Detection and quantitation of RNA

A number of methods have been developed to identify RNA species and quantify a specific RNA transcript, specifically messenger RNA (mRNA) within total cellular RNA, or total RNA isolated from tissues. mRNA is the "+" or "sense" strand of RNA that is transcribed and processed from DNA encoding a gene that is subsequently translated into protein. The following represent the more commonly employed methods.

Northern blotting

This is a method designed to identify individual RNA that has been fractionated on agarose gels and transferred to a membrane for probing. It is therefore analogous to Southern blotting for DNA. The only major difference is that because RNA can form secondary structures, it must be resolved on gels under denaturing conditions. Common denaturants include formaldehyde and glyoxal.

Northern blotting got its name from its developers, Alwine and Stark (17), who called it "Northern" blotting as a joke (after "Southern" blotting), but the name has stuck, and other blotting methods have since used compass directions as descriptions as well. As for Southern blotting, the probes used are generally DNA probes (cDNA probes, or double stranded genomic DNA fragments that are denatured before use). Northern blotting is commonly used to identify a mRNA transcript in either total cellular RNA or in enriched mRNA. mRNA can be purified using oligo dT cellulose and then fractionated on agarose gels in order to detect rarer transcripts, since most cellular RNA is ribosomal and transfer RNA (tRNA). Northern blots can provide quantitative measures of steady state RNA levels, especially comparative measurements of the relative abundance of RNA in different cell types. However, Northern blotting is not as quantitative as ribonuclease protection assays.

Ribonuclease (RNase) protection assay (RPA)

RPA is used to quantitate steady state levels of RNA, to identify the 5′ or 3′ ends of mRNA, and to characterize the splice junctions of the primary RNA transcripts that are processed into mRNA. The probes used for RPA are RNA probes that are generated by RNA polymerase from bacteriophage promoters, including SP6 (from *Salmonella typhimurium*) and T3 or T7 (from *E. coli*). The DNA to be used for the probe is cloned into a vector downstream of a bacteriophage promoter, and the RNA polymerase is used to produce an RNA probe (usually containing radioactive precursors) complementary to mRNA. Once the RNA probe is hybridized to RNA, RNase is used to remove the free probe and any single stranded segments of hybridizing RNA. The probe will be digested at any point where the probe and RNA do not hybridize. This can be the 5′ end of the RNA, or regions marking exon-intron boundaries. The remaining RNA which was "protected" by annealing to the RNA of interest is then fractionated on a sequencing gel, and the exact size and abundance of the probe remaining can therefore be determined. The size of the protected fragment reveals the start site or domain boundary,

while the abundance of the protected fragment reveals the abundance of the RNA transcript.

S1 analysis

S1 analysis is a method used to map exon-intron boundaries and determine the 5′ ends of RNA. S1 analysis usually relies on the use of a single stranded DNA probe that has been labeled on its 5′ end by a kinase. The probe is added to RNA, and the non-hybridizing portion of the probe is digested by S1 nuclease, fractionated by PAGE and analyzed by means similar to RNase protection.

Primer extension

This method is designed to quantitate the levels of a given RNA in a cell as well as to map the 5′ end of the RNA. Primer extension is used to confirm RPA and S1 mapping of the 5′ ends of RNA. Primer extension uses reverse transcriptase (RT; see below) and a primer (a short single stranded DNA) complementary to a sequence in the RNA. After hybridization, the reverse transcriptase (together with dNTP precursors) is used to synthesize ("extend") a cDNA copy of the RNA. Because RT is a polymerase, it synthesizes the cDNA copy in the 5′ → 3′ direction, which is directed toward the 5end of the RNA. Consequently, when the RNA template terminates, extension will end. The length of the cDNA synthesized identifies the length from the primer to the 5′ end of the RNA, and the amount of product synthesized is a direct reflection of the abundance of the RNA.

Other methods to quantitate RNA use modifications of the polymerase chain reaction, and are described in Chapter 3.

Reverse transcriptase (RNA-dependent DNA polymerase)

Reverse transcriptase (RT) is an enzyme isolated from retroviruses (Human Immunodeficiency Virus, or HIV, is one example of a retrovirus). Retroviruses contain a RNA genome, but their life cycle is dependent on the production of a DNA intermediary that integrates into the DNA of the host genome. RT is responsible for transcribing the RNA genome into the first strand of this DNA. A double stranded copy is then synthesized and the proviral DNA is integrated into the host cell genome.

The isolation and characterization of reverse transcriptase was a major breakthrough in our understanding of the life cycle of retroviruses. Its

discovery was all the more remarkable since the prevailing wisdom of the time was that DNA was transcribed into RNA, and that the reverse could not (and did not) occur. The identification of RT (18) won the Nobel Prize for its discoverer, David Baltimore, in 1975, which he shared with Renato Dulbecco and Howard Temin for their work on retroviruses and the discovery of tumor viruses.

RT has become an essential tool in biomedical research. In addition to its use in primer extension, RT is used in modified polymerase chain reaction (RT-PCR) studies for quantitative analysis of RNA, for the construction of cDNA libraries, and for differential gene expression studies, to name but a few of its many uses. While the RT used originally was derived from the avian myeloblastosis virus (AMV), a tumor-inducing retrovirus of chickens, most RT used today is genetically modified (one of the many uses of recombinant DNA), removing endogenous RNase activity and enhancing its polymerase activity. These engineered enzymes not only yield more cDNA product, but the cDNAs generated usually are significantly longer and therefore more likely to represent the entire mRNA than the cDNAs generated using native RT.

Like DNA polymerases, RT requires a primer. Oligo dT is frequently used because it hybridizes to the polyA tail characteristic of mRNA. Alternatively, random hexamers (6 bp long random sequences) are used, especially when there is a need to increase the representation of the 5′ ends of mRNAs, such as in the production of representative cDNA libraries. No matter how good the RT, many mRNAs cannot be synthesized into intact cDNAs representing the entire mRNA from the 3′ end to the 5′ end because they are too large.