Molecular Pathology in Anatomic Pathology Practice
A Review of Basic Principles
Jennifer L. Hunt, MD

Molecular testing in pathology emerged shortly after polymerase chain reaction became a standard molecular biology assay.1 Testing efforts began in the clinical laboratories primarily with assays for genetically inherited diseases and assays for clonality in hematologic malignancies.1,2 Today, the field has evolved into “molecular diagnostics,” which encompasses testing in almost every area of anatomic pathology.3 Molecular testing is now even making its way definitively into both surgical pathology and cytopathology, although molecular anatomic pathology is still young with few standard tissue-based molecular assays. As more clinically valuable information is gained from molecular pathology testing of tissues, unique challenges are also becoming apparent at the intersection between tissue diagnosis and DNA diagnosis. This review focuses on basic molecular pathology concepts, with particular emphasis on the challenge of tissue-based testing in anatomic pathology.

(Arch Pathol Lab Med. 2007;132:248–260)

DNA AND RNA STRUCTURE AND FUNCTION

Discovery

The double-helix structure of DNA was determined by the landmark studies of James Watson and Francis Crick. The research of Watson and Crick was first published in March of 1953 (Nature. 171:737–738, 964–967).4 Although they often get credit for discovering DNA, many other scientists had crucial discoveries that were used by Watson and Crick for the development of their celebrated model of the structure of double-stranded DNA. Nucleic acids were discovered in 1869 by a German scientist named Johan Friedrich Miescher. He published his work about the material “nuclein” in a German journal in 1871 (Hoppe-Seyler’s medicinisch-chemische Untersuchungen. 4:441–460). Then, in 1943, an American scientist named Oswald Avery proposed that DNA was the carrier for the genetic code. The basic helical structure of 1 DNA molecule was suggested privately by Rosalind Franklin, but she did not initially publish her work. In 1950, a biochemist named Erwin Chargaff showed that bases had a one-to-one relationship and suggested base pairing in DNA (Chargaff’s rule: A=T, C=G). Although many other scientists contributed to the body of knowledge about DNA, Watson and Crick’s solution to the structure is still considered to be the foundation of molecular biology today.

The discovery of RNAs was also dependent on the basic work done by Miescher but was solidified in 1963 when the first sequence of RNA was uncovered in the laboratory of Robert W. Holley.5 He was the winner of the 1968 Nobel Prize for Physiology/Medicine for this discovery.

Structure

Nucleic acids are polymers that are made up of strings of mononucleotides, or nucleotide building blocks, held together by phosphodiesterase bonds. Each nucleotide is composed of a phosphate unit, a sugar unit, and a heterocyclic base. The sugar unit is a 5-carbon atom ring. It is either ribose (for RNA) or deoxyribose (for DNA); deoxyribose is missing the oxygen on the second carbon atom. The combination of this base and a sugar is called a nucleoside. The 4 bases are identified as adenine (A), guanine (G), cytosine (C), and thymine (T). Uracil (U) replaces thymine in the structure of RNA. The bases are ring structures that contain a nitrogen atom and are thus also known as nitrogenous bases. There is a single ring in pyrimidines (C, T, U) and a double ring in purines (A, G).

Most DNA in cells is double stranded, arranged in the famous double-helix structure. The 2 linear strands are complementary to one another. The major force promoting formation of this helix is complementary base pairing in which pyrimidines bind to purines as follows: A forms 2 hydrogen bonds with T and G forms 3 hydrogen bonds with C. One consequence of this binding disparity is that it takes more energy (eg, a higher temperature) to disrupt G–C bonds than A–T bonds.

The 2 strands of DNA that are wound together run antiparallel to one another, or in opposite directions. The 2 strands have been given many different names. Most commonly, the forward strand is referred to as the coding or sense strand, and the reverse strand is called the noncoding or antisense strand. By convention, the sequence of DNA is usually written in the 5’ to 3’ direction (with 5’ on the left and 3’ on the right); the numbers refer to the number of the carbon in the nucleotide carbon ring.

The 3 main types of RNA are messenger RNA (mRNA), transfer RNA, and ribosomal RNA, all of which are single stranded. Transcription is the process of transferring information from DNA to mRNA, and it occurs in the nu-
clesus. Posttranscriptional modifications also occur in the nucleus, including the splicing out of introns and the addition of a polyA tail. This modified mRNA then leaves the nucleus and is available as a template to make proteins in the cytoplasm. The transfer of information from mRNA to make protein is called translation. This process requires ribosomal RNA assembled into ribosomes and transfer RNA to decipher the codon sequence and deliver each amino acid. The short regulatory RNAs are new RNA molecules that have been recently described in the literature. Among the most well studied are microRNA and short interfering RNA.

**Function and Stability**

DNA has two very important roles in cellular function. The first is inheritance, or the transmission of information from parent to offspring or from stem cell to progeny cell. In other words, this molecule passes the code from one generation to the next. The second function is that DNA provides information for the translation of genetic messages into functional protein products, which carry out all cellular work.

In 2001, a major milestone was met with the publication of the results of the Human Genome Project, which was the sequencing of the entire human genome. The vast information obtained in this enormous project has enhanced our understanding of the genetic code. It was originally predicted that humans would have more than 60000 genes in DNA, but through the Human Genome Project it was shown that there are probably less than 30000 genes, making up only about 1% of genome.

DNA is hardy enough to withstand many environmental, physical, or chemical insults. In diagnostic pathology, we take advantage of the stable nature of DNA by extracting DNA from tissues and blood that have been processed even under the most severe conditions. DNA is naturally degraded by DNase, but only when it is loosely coiled or single stranded. In normal living cells, chromosomal regions that are transcriptionally active have looser DNA-protein structures and are more sensitive to DNase degradation. DNA from fresh or frozen tissue, or formalin-fixed and paraffin-embedded archival tissue, is fairly well protected from this natural degradation. However, chemicals used in typical tissue processing (formalin, alcohols, and xylene) affect the quality of DNA through chemical interactions (see “Sample Type”).

RNA functions transiently in cells to carry information for modulated gene expression and dynamic translation into proteins. Because cells must have tight control over the termination of a signaling event, RNA is not hardy and is particularly susceptible to degradation. RNase is the enzyme that breaks down RNA and it is ubiquitously present in all tissues. This enzyme is robust and can survive in the environment at ambient temperature. Because of the sensitivity of RNA, tissue for mRNA extraction requires special handling and storage to avoid this rapid and efficient natural degradation (see “Sample Preparation”).

**Chromosomes, Genes, and Epigenetics**

In the nucleus of a living cell, DNA is wound around histones and organized in a consistent and patterned manner into chromosomes. Chromosomes can be differentiated from one another visually in classical cytogenetics preparations by their length and their banding pattern. Each diploid cell has 2 copies of the 22 somatic chromosomes, 1 copy inherited from the mother and 1 from the father. The final 2 chromosomes are the sex chromosomes, X and Y. The full complement of chromosomes in a normal human diploid cell is 46 chromosomes (44 somatic and 2 sex chromosomes).

The smallest functional unit of the genetic code is the gene. The many different types of genes are often classified by the function of the protein product for the gene. Genes have regulatory regions (eg, promoters), coding regions (exons), and noncoding regions (introns). The 5' promoter region will often contain specific well-recognized sequences and also a CG repeat unit. The CG repeat unit is also known as a “CpG island,” where the “p” refers to the diphosphate bond between the C and G nucleotides.

Much of molecular testing in pathology focuses on assessing the alterations in either the sequence or the amount of DNA or RNA. However, there are also alterations in gene transcription that do not alter the sequence of the gene. These are referred to as epigenetic alterations, and they are becoming important in tissue-based molecular testing.

There are 3 major epigenetic mechanisms of transcription regulation: methylation, histone modification, and RNA interference.

Methylation refers to enzymatic addition of a methyl group to the carbon 5 position of cytosines typically at CpG sites in DNA. In normal cells, methylation is usually in the coding regions and the CpG islands in promoters are unmethylated. Methylation is also the mechanism of inactivation of one of the X chromosomes in XX cells. In tumor cells, the balance of methylation is altered and the methylation state may be important for tumor genes.

Histone modifications are also involved in gene regulation. Posttranscriptional modification of the histones are responsible for the chromatin structure as either open chromatin that allows interaction with transcription factors or closed chromatin that does not allow permeation of transcription factors and therefore halts gene expression. Finally, microRNA and short interfering RNA are known to have an essential role in transcription. MicroRNAs are single-stranded RNA molecules of about 21 to 23 nucleotides long. They are thought to function in gene regulation. The sequence of microRNA is complementary to one or more mRNAs and annealing is thought to inhibit translation or promote degradation. Short interfering RNAs are also 20 to 25 nucleotides in length but are double stranded. Synthetic short interfering RNA is used experimentally for gene silencing. New evidence is linking the short regulatory RNAs to cancer.

**Variability in the Genome**

Much of the genome is nearly identical between individuals, with up to 99.7% the same from one person to the next. The variable areas are called polymorphisms. Although polymorphisms represent a small percentage of the overall genome, they are critical for several types of diagnostic molecular analyses. A polymorphism of some type is predicted to occur every 1000 base pairs in the genome. In molecular pathology, we capitalize on variable regions for many different applications of DNA technology, from human identification to measuring and assessing DNA alterations in tumors. Polymorphisms can range from a single base pair (single nucleotide polymorphisms), to alterations of a short sequence (short tandem repeats, microsatellites, and minisatellites), or alterations...
Table 1. Different Types of Polymorphisms Used in DNA Testing

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>Alternate Names</th>
<th>No. of Base Pairs</th>
<th>No. of Repeat Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single nucleotide polymorphism</td>
<td>SNP</td>
<td>1</td>
<td>None</td>
</tr>
<tr>
<td>Microsatellite</td>
<td>Short tandem repeat</td>
<td>1</td>
<td>Usually less than 45</td>
</tr>
<tr>
<td>Minisatellite</td>
<td>Variable nucleotide tandem repeats (VNTRs)</td>
<td>8–80 base-pair repeats</td>
<td>Usually less than 40</td>
</tr>
</tbody>
</table>

Polymorphisms are more common in the noncoding regions of genes. Mutations in polymorphisms are not common, except in a unique group of diseases called *trinucleotide repeat disorders*. These diseases result from expansion of triplet repeats in coding areas (Huntington disease) or untranslated regions (fragile X and myotonic dystrophy). There are also several examples of single nucleotide polymorphisms that are associated with diseases, such as a specific genotype of the *APO-E* gene (*APOE-e4*) that has an association with Alzheimer disease.

Using polymorphisms to create linkage maps was first described in a landmark paper in which restriction fragment length polymorphism analysis was used. Because polymorphisms are inherited, many hereditary diseases were originally mapped to specific chromosomal loci using restriction fragment length polymorphism analysis. The assays, however, were somewhat laborious and included restriction enzyme digestion of genomic DNA, electrophoresis, and Southern blotting and hybridization with radioactive-labeled probes. Incorporation of a polymerase chain reaction (PCR)–based assay for short tandem repeat polymorphisms greatly simplified testing for inheritance. Polymorphism analysis has expanded to include loss of heterozygosity analysis, identity tests, and microsatellite instability testing.

**EXTRACTION AND PURIFICATION OF DNA**

**Sample Type**

In anatomic pathology, the most common samples are tissues, cellular aspirates, and body fluids. All of these samples can be fresh or may be preserved in a variety of different ways. All types of fixatives, preservatives, and processing protocols do have an effect on the quality of nucleic acids that can be obtained from these samples.

Two basic types of fixatives are commonly used in anatomic pathology laboratories: cross-linking fixative (formaldehyde, glutaraldehyde, and paraformaldehyde) or precipitation fixative (alcohol, methanol, and acetone). One of the most popular fixatives is 10% buffered neutral formalin, which has been extensively studied in reference to molecular testing. All fixatives damage nucleic acids to some degree through chemical interactions. Some types of fixatives, such as picric acid, mercury-containing solutions, and acid decalcifiers, may not be compatible with molecular testing. Formalin- and alcohol-based fixatives preserve DNA reasonably well, but there is high variability introduced by several different factors. Optimal time of fixation is between 12 and 24 hours, after which time, the DNA that is obtained is of much lower quality.
Table 2. Fixatives Used in Anatomic Pathology

<table>
<thead>
<tr>
<th>Fixative</th>
<th>Components</th>
<th>DNA Quality</th>
<th>RNA Quality</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unbuffered formalin</td>
<td>Formaldehyde</td>
<td>Poor</td>
<td>Poor</td>
<td>16</td>
</tr>
<tr>
<td>Buffered formalin</td>
<td>Formaldehyde, Phosphate buffers</td>
<td>Fair</td>
<td>Fair</td>
<td>16, 29</td>
</tr>
<tr>
<td>Methacarn</td>
<td>Methanol, Chloroform, Glacial acetic acid</td>
<td>Good</td>
<td>Good</td>
<td>92, 93</td>
</tr>
<tr>
<td>Ethanol</td>
<td>70%–100% Methanol, Glacial acetic acid</td>
<td>Good</td>
<td>Good</td>
<td>94–96</td>
</tr>
<tr>
<td>Carnoy</td>
<td>Ethyl alcohol, Chloroform</td>
<td>Good</td>
<td>Good</td>
<td>97–99</td>
</tr>
<tr>
<td>Glutaraldehyde (Karnovsky)</td>
<td>Parafomaldehyde, Sodium hydroxide, Glutaraldehyde, Buffer</td>
<td>Good</td>
<td>Unknown</td>
<td>98, 100</td>
</tr>
<tr>
<td>Histochoice</td>
<td>Proprietary, Amresco Inc, Solon, Ohio</td>
<td>Good</td>
<td>Unknown</td>
<td>101</td>
</tr>
<tr>
<td>HOPE</td>
<td>Proprietary, DCS Innovative, Hamburg, Germany</td>
<td>Good</td>
<td>Good</td>
<td>102</td>
</tr>
<tr>
<td>UMFIX (universal molecular fixative)</td>
<td>Proprietary, Sakura Finetek USA, Inc, Torrance, Calif</td>
<td>Good</td>
<td>Good</td>
<td>35, 36</td>
</tr>
<tr>
<td>Zinc buffered formalin</td>
<td>Formaldehyde and zinc</td>
<td>Good</td>
<td>Good</td>
<td>103–105</td>
</tr>
<tr>
<td>Prefer</td>
<td>Proprietary, Anatech, LTD, Battle Creek, Mich</td>
<td>Poor</td>
<td>Poor</td>
<td>26</td>
</tr>
<tr>
<td>Decalcifying acids</td>
<td>Various formulations</td>
<td>Poor</td>
<td>Poor</td>
<td>106–108</td>
</tr>
<tr>
<td>Bouin</td>
<td>Picric acid, Formaldehyde</td>
<td>Poor</td>
<td>Poor</td>
<td>81, 104</td>
</tr>
<tr>
<td>Mercury-containing solutions (B-5)</td>
<td>Mercuric chloride, Sodium acetate</td>
<td>Poor</td>
<td>Poor</td>
<td>100, 104</td>
</tr>
<tr>
<td>Hollande</td>
<td>Copper acetate, Distilled water, Picric acid, Formaldehyde</td>
<td>Fair</td>
<td>Fair</td>
<td>104</td>
</tr>
<tr>
<td>Zenker</td>
<td>Glacial acetic acid, Mercuric chloride, Potassium dichromate, Sodium sulfate</td>
<td>Poor</td>
<td>Poor</td>
<td>100</td>
</tr>
</tbody>
</table>

Storage of DNA in formalin for more than a week will almost irreparably damage nucleic acids. The cross-linking action of formalin gives the excellent cellular preservation essential for optimal light microscopy. Formalin penetrates the tissue and cross-links the proteins, which preserves architecture. Cross-linking will also affect histones, which are the proteins around which DNA is wound in the double-helix structure. During cross-linking, methylene bridges appear to be formed between the histones and the amino groups in DNA. Cross-linking creates 2 problems for extraction of DNA. First, the DNA will be less easily released, and, second, it is more susceptible to structural damage and fragmentation during extraction. The result is DNA of lower molecular weight and in lower quantities. The average fragment length of DNA obtained from formalin-fixed, paraffin-embedded tissues is often only between 300 and 400 base pairs. In contrast, DNA from fresh tissue or blood will often range up to thousands of base pairs in length. When the DNA is in large preserved fragments, it is referred to as high-molecular-weight DNA. Cross-linking may be reversible using certain pretreatment protocols, but some studies have suggested that complete removal of proteins is enough to overcome the negative effects of cross-linking.

New fixatives that do not contain formalin have recently become commercially available. These include several proprietary commercial fixatives such as Prefer, HOPE, and Histochoice (Table 2). Very few studies have explored the potential for molecular tests with these fixatives. One study did suggest that Prefer is not compatible with molecular testing, but other fixatives appear to perform reasonably well in molecular testing. Additional long-term studies of different fixatives are needed.

RNA is substantially degraded during formalin fixation and paraffin processing. Furthermore, comparison studies have shown that the quantitation of RNA from differentially expressed genes may be greatly altered in fixed tissue as compared with parallel fresh tissue. Despite these caveats, RNA can be obtained and used from paraffin-embedded tissue in certain circumstances. RNA fragments obtained from paraffin-embedded tissue samples are often small, ranging from 100 to 500 base pairs in length. On the other hand, the length of RNA fragments from fresh tissue can be thousands of base pairs in length. Optimal fixation protocols can aid in obtaining higher quality nucleic acids. Optimization includes avoiding over-fixation and under-fixation (from 12 to 24 hours is ideal). Fixation at 4°C is preferred to room temperature, although this may not be feasible in clinical anatomic pathology laboratories. Length of storage time of paraffin-embedded tissue blocks does not appear to affect the ability to perform molecular testing.

Several alternative tissue processing systems have been developed that allow for rapid fixation and rapid impregnation of tissues with paraffin. These include microwave systems and ultrasound-assisted processing. Initial studies using these processors have demonstrated that RNA and DNA of high quality can be obtained from these specimens.
Sample Preparation

Because DNA is fairly robust and does not degrade rapidly, fresh tissue or fluid can be stored in the refrigerator. RNA, on the other hand, is more labile because of RNase in tissue. Therefore, tissue or fluid samples intended for DNA testing should be handled carefully and quickly. Recent studies have suggested that the quality of RNA is not adversely affected by delays in processing of up to 6 hours, with proper storage. For prolonged storage, preservation of RNA is best when the samples are frozen, although an alternative is to use an RNA preservative, such as RNA Later (Ambion The RNA Company, Austin, Tex). RNA Later is not a cytologic preservative and morphology will be adversely affected with this type of preservation.

Stored tissues will yield the best DNA or RNA if they are snap frozen in liquid nitrogen. Degradation will occur if the frozen tissue is subjected to multiple freeze-thaw cycles. Because extraction is more efficient with smaller fragments, dicing larger tissues before storage may be helpful. It is usually recommended to further pulverize frozen tissue before extraction. This can be done mechanically using a variety of commercially available mortar and pestle implements. Frozen sections prepared in Tissue-Tek Optimal Cutting Temperature Compound (OCT; Sakura Finetek USA, Torrance, Calif) can be used for molecular analysis. However, OCT, which includes polyvinyl alcohol and polyethylene glycol, can interfere with molecular testing. It is recommended that samples for molecular testing not be stored in OCT or that excess OCT be cut away from the tissue before extraction.

When starting with paraffin-embedded material, there are several options for sample preparation. Tissue can be cored out of paraffin blocks or tissue sections can be cut from the blocks at 10 to 20 µm thick and made into tissue scrolls (rolled sections), which are stored in tubes. A number of scrolls can be placed into 1 tube for digestion; however, if too many tissue scrolls are digested together, the sample may paradoxically yield less DNA. Lastly, routine unstained sections can be prepared on glass slides for microdissection. Common practice is to thoroughly remove the paraffin wax with a series of xylene alcohol washes. Several studies, however, have recently suggested that paraffin does not interfere with assays and deparaffinization may not always be necessary. Either deparaffinization can be performed in the tubes or slides can be deparaffinized before microdissection. An alternate method of deparaffinization using microwave treatment has also been described.

Microdissection enriches samples for the cellular component of interest and excludes undesirable elements, such as necrotic areas, blood, or contaminating normal cells. The 2 most commonly used microdissection procedures include manual microdissection and laser capture microdissection. Both of these techniques have been described in detail elsewhere. Counterstaining is used for microdissected slides, the best stains to preserve nucleic acid quality are methyl green, Evans blue, and nuclear fast red stains.

In special circumstances, it may be necessary to use DNA obtained from an already stained glass slide, rather than from newly cut material from a paraffin block. This can include the use of diagnostic cytology samples, hematoxylin-eosin–stained slides, or even slides that have been stained for immunohistochemistry. In general, cytology material stained with Papanicolaou stain is quite amenable to extraction of DNA. Cells stained by this method often have been fixed in alcohol-based fixatives, with excellent preservation of nucleic acids. Typically, the Papanicolaou stain does not interfere with extraction or PCR. On the other hand, hematoxylin-eosin–stained slides produce much more variable PCR results. Some commercial preparations of hematoxylin appear to be better than others. One reason the hematoxylin stain may decrease DNA yield appears to be the fact that the nuclear stain interacts with histones, increasing resistance to digestion by proteinase K. Feasibility of molecular testing of immunohistochemistry slides or negative controls will depend on the antibody and the type of antigen retrieval used.

Tissue Digestion

A number of studies have shown that tissue digestion can improve the quality of DNA obtained for molecular testing. Tissue fragments are digested in a lysis buffer that can be purchased or can be prepared in the laboratory. The basic principles of digestion include breaking up tissue by enzymatic disruption or combinations of chemicals and heat. The most common digestion enzyme used is proteinase K, which effectively disrupts tissue components. The inclusion of chemical additives, EDTA, the detergent Tween 20, or sodium dodecyl sulfate has also been shown to improve digestion.

The time of digestion depends on several variables, including the type of tissue, the temperature, and the pH of the digestion buffer. Digestion in proteinase K overnight at 42°C has generally been shown to be effective. For shorter digestion times, samples may need to be heated to 55°C to 60°C; proteinase K functions optimally at 56°C. Although some studies have indicated a need for prolonged digestion times, others have indicated that long digestion times at increased temperature may actually result in lower DNA yield. Digestion can be visually assessed for completeness by the presence of few remaining intact fragments. The final step in digestion is a high heat incubation to inactivate the proteinase K. Methods that eliminate enzymatic digestion have also been shown to produce quality DNA; most of these methods involve heating samples in either alkaline solution or with chemical treatment.

Extraction Techniques

Most laboratories use an extraction technique to purify DNA and to eliminate contaminants. Some studies have argued that it is unnecessary to extract DNA and that unpurified digested lysate can be used directly in PCR. Using direct lysis without purification is not suitable for assays that require restriction enzyme digestion, however. Advantages of avoiding a purification step include a higher absolute DNA concentration, decreased processing time, and ease of use in the laboratory. Disadvantages of not extracting DNA are that quality or quantity of DNA cannot be measured by conventional methods and that contaminants may interfere with PCR.

Several techniques for purifying DNA exist, all of which can effectively decrease protein and chemical contaminants that may be present. Typical techniques include phenol-chloroform extraction, salting out, guanidium thiocyanate method, extraction via metal chelating agents (chelation), or purification using commercially available extraction kits.
Testing of DNA Quantity and Quality

Two common methods of analyzing purity and concentration of nucleic acid are used: spectrophotometry and fluorometry. Spectrophotometers measure absorbance of light in the UV spectrum. Nucleic acids absorb light maximally at 260 nm, whereas protein contaminants absorb at 280 nm. The ratio of absorbance at 260 to 280 nm is a measure of the purity of DNA. The ideal ratio is generally greater than 1.8. Diluted DNA is placed into a specially designed glass cuvette through which UV light is passed, after calibration. Fluorometers use a fluorescent dye that incorporates into double-stranded DNA. This method gives a high signal, is less sensitive to impurities that may be present, and is specific for double-stranded DNA. Because both methods use a significant amount of sample for measurement, they may not be ideal for small precious samples. In addition, neither method is very accurate at low nucleic acid concentration. Two newer machines, the Nanodrop 1000 spectrophotometer and 3300 fluorospectrophotometer (Nanodrop Technologies, Wilmington, Del), provide nucleic acid quantification using only 1 μL of sample, which may be useful for DNA isolated from small microdissected samples.

Quantitation of DNA by either method described previously does not assess for DNA fragmentation. Several methods to assess quality can be used. First, the simple and basic method of electrophoresing an aliquot of purified DNA in an agarose gel can be used (see “PCR Product Analysis”). DNA fragments in the sample separate according to their size, with high-quality, high-molecular-weight DNA being retained in the loading well and degraded DNA producing a broad smear that extends to the end of the gel (Figure 2). The presence of a smear indicates that the DNA is highly fragmented. DNA bands are normally detected by staining with ethidium bromide stain. This assay requires a reasonably large amount of DNA, and, therefore, if the starting amounts are small, it may not be practical to waste precious material in performing this quality assessment.

A newer platform that can measure DNA concentration and quality in 1 assay is the Agilent Bioanalyzer (Agilent Technologies, Santa Clara, Calif). This machine uses the principles of capillary electrophoresis but with microfluidics and “lab-on-a-chip” technology. The output shows the sizes of fragments as well as the concentration of DNA. The final test of quality of DNA, however, is whether PCR is successful, because the DNA has to be sufficiently preserved for the assay to work. Thus, simultaneous amplification of a known marker, such as the globin gene, is an excellent experimental control. If the quality of the DNA is poor, the control gene will not amplify. One option is to avoid direct assessments of DNA quality and to rely on appropriate results from known controls.

Storage

Extracted and purified DNA can be stored for prolonged periods, especially when stored optimally at −80°C. Storage at typical freezer or refrigerator temperatures (−20°C or +4°C) will cause degradation of the DNA over time. For prolonged storage, the DNA should be resuspended in a storage medium, either ethanol or some other solution such as trehalose, to improve preservation. RNA should be stored frozen, either as an aqueous solution at −70°C or as an ethanol precipitate at −20°C. If repeated sample use is anticipated, it is best to store smaller aliquots and to avoid repeat freeze-thaw cycles to maintain nucleic acid quality.

REVERSE TRANSCRIPTION FUNDAMENTALS

DNA polymerases used in PCR recognize the bases A, T, C, and G but do not recognize U, which is incorporated in the RNA equivalent of T. To use RNA as a template for PCR, it must first be converted into a DNA-like sequence called complementary DNA (cDNA). The process of conversion of RNA to cDNA is called reverse transcription (RT), and it is accomplished using the enzyme reverse transcriptase. The RT reaction requires addition of enzyme, template, and primers. Three different types of primers can be used in RT. Random hexamer primers are random variable nucleotide sequences of 6 base pairs in length. These primers ideally will amplify all types of RNA present in tissue samples. The oligo-DT primer consists of repeats of the base thymine, which anneals directly to the polyA tail of mRNA specifically. In reactions using oligo-DT primers, theoretically only the mRNA is amplified because it has been modified with the addition of the polyA tail. Finally, sequence-specific primers for individual genes can be used in the RT. This can eliminate the use of different primers for RT and PCR, allows for both reactions to be performed in 1 tube, and is generally considered to be quite specific. However, under certain conditions, nonspecific binding can enhance the possibility of false-positive results. The cDNA generated from RT reaction is used as a template for PCR, just as genomic DNA is used.

PCR FUNDAMENTALS

Discovery of PCR

Polymerase chain reaction was first conceived in 1983 by Dr Kary B. Mullis, who was a chemist working at Cetus Corporation in Emeryville, Calif. Dr Mullis won the Nobel Prize for chemistry in 1993 for his discovery of PCR. The most important discovery that led to PCR was uncovering the properties of a DNA polymerase (Taq) from Thermus aquaticus, an organism that lives at very high temperatures. Polymerase chain reaction was originally described in reference to detecting sickle cell anemia in a seminal paper in Science in 1985. Polymerase chain reaction was initially labor intensive and tedious because it was performed manually. The first thermal cycling machines were made by Cetus, partnering with Perkin Elmer. The patent for PCR was sold to Roche Molecular Systems, and since that time, a licensing fee is applied to all commercially performed assays that use PCR in clinical tests.

The discovery of PCR marked the beginning of an explosion of molecular applications to samples that could never have been analyzed before. Before PCR, the main assays that were performed on DNA and RNA used Southern blots and Northern blots. These procedures were time consuming, required large amounts of high-quality nucleic acids, and needed highly trained personnel.
Figure 4. Reaction phases of polymerase chain reaction. A, The reaction is begun at room temperature. The template DNA is double stranded, and the small pink and blue boxes represent the primers that have been added to the reaction mixture. The Taq enzyme and nucleic acids are not illustrated but would also be added at the beginning of polymerase chain reaction. B, The reaction mixture is then subjected to high heat in the denaturation phase of polymerase chain reaction. At this temperature, the double-stranded DNA separates into single-stranded DNA. C, The temperature is then taken down to the temperature that has been determined to be optimal for the particular primer pair used. This is usually between 50°C and 60°C. This is the annealing phase, when the primers can bind to the template DNA. D, The temperature is then taken back up to 72°C. This is the extension phase of polymerase chain reaction, when the Taq enzyme can function optimally. The primers are extended, creating a copy of the original template DNA strands. E, The previous 3 phases represent 1 polymerase chain reaction cycle. The process is repeated for 30 to 40 cycles. This figure shows the denaturing phase of the reaction for cycle 2.

The discovery of PCR solved 2 major problems in molecular biology: the lack of abundance of the starting nucleic acid and the lack of specificity of whole DNA preparations. In PCR, the nucleic acid area of interest is selectively amplified up to 1 million-fold, resulting in production of abundant and specific amplicons. For very small DNA samples, such as the starting material for preimplantation genetic testing, PCR-based methods to amplify the entire genome have been described, termed as whole genome amplification. Most whole genome amplification studies have described good reproducibility, although allelic dropout and biased amplification remain a concern.77–80

PCR Reagents and Mechanisms

In PCR, either genomic DNA or cDNA is used as a template for amplification of a specific region. Before a PCR assay can be designed, it is essential to obtain the nucleotide sequence of the target DNA. With the Human Genome Project, the sequence for nearly the entire genome is now available and can be obtained easily from Internet databases (www.ncbi.nlm.nih.gov/Genbank/index.html; accessed August 20, 2007). Once the sequence is known, PCR primers can be designed. Primers are short oligonucleotide sequences of between 15 and 22 base pairs in length that are designed to flank the sequence of interest. Two primers are needed, a forward primer (also called left, sense, or 5’ primer) and a reverse primer (also called right, antisense, or 3’ primer) (Figure 3). Computer programs are available to optimize primer design, both as commercial software packages or free online Web-based services. These programs are excellent for obtaining prim-
er sequences that avoid negative interactions such as primer-dimers (primers that bind to each other) and hairpin loops (primers that bind back on themselves). The programs also allow for setting the desired product length and for estimating the melting temperature of the primer. The melting temperature is the point at which 50% of the primer is stable in double-stranded complexes and the other 50% is free in the reaction mixture.

In designing primers for PCR assays on nucleic acid isolated from paraffin-embedded tissue, it is critical to consider the size of the expected PCR product. Although PCR from fresh tissue can successfully amplify large fragments (even up to 700–1000 base pairs in length), the degradation of DNA in fixed tissues will not allow for such large products to be generated. Optimal PCR product size for paraffin-embedded tissues appears to be 100 to 200 base pairs. In PCR using DNA from paraffin tissues, fragments greater than 300 base pairs can be amplified but not in a reliable and consistent manner.

Nucleotide building blocks (deoxyribonucleotide triphosphates [dNTPs]) are added to the reaction mixture for PCR. These are incorporated into the PCR product as amplification progresses. Additional substances are added, including PCR buffers that usually contain Mg2+ and detergent. A standard dNTP mixture can be purchased ready to use.

The most reliable and reproducible PCR assays contain a single set of primers to amplify the sequence of interest. Multiplex PCR has been described as a method to decrease cost and time to generate and analyze several different products in a single experiment. In this technique, more than one set of primers is used to generate 2 or more products in the same tube. Although the benefits for efficiency can be significant, it should be recognized that multiplex PCR requires special optimization and that it may be less reliable for DNA from paraffin-embedded samples.

There are different phases in PCR that take place at different reaction temperatures. Because these temperatures must be achieved rapidly, a machine called a thermocycler machine (also known as a thermal cycler) is used. Thermocyclers have individual wells to hold the thin-walled PCR tubes. The temperature in the wells can be rapidly increased or decreased for each phase of PCR (Figure 4, A through E).

The initial phase of PCR is denaturation, in which the double-stranded template DNA separates into single-stranded DNA. Denaturation occurs at high heat, usually 94°C or greater. At this high temperature, the hydrogen bonds that hold the coding strand and template strand together are overcome. The second phase is annealing, the optimal annealing temperature is usually about 5°C below the specified melting temperature of the primers. If the annealing temperature is too close to the melting temperature, annealing will be very slow and inefficient. If it is too low, however, nonspecific primer binding can occur, with generation of nonspecific PCR product. During annealing, the primers bind using complementary base pairing with the forward primer binding to the sense strand and the reverse primer binding to the antisense strand. Extension, the final phase of a PCR cycle, is performed at 72°C, the temperature at which Taq has the highest activity. Taq is a DNA polymerase and is thus able to sequentially add complementary bases, starting at the 3’ end of the annealed primer and processing along the template DNA in the 5’ to 3’ direction.

The 3 phases described previously are performed in sequence and constitute 1 cycle of a PCR. The cycles are repeated between 30 and 40 times. In each subsequent cycle, PCR products that have extended from the primer and the DNA template strand are denatured into single-stranded DNA fragments, each of which is then available as template. Each cycle sees the doubling of PCR product, which is the basis for exponential amplification. In the beginning of PCR, the reagents are not rate limiting and the exponential nature of amplification can be realized. However, as amplicon is generated, primers are exhausted, and the reaction eventually plateaus (plateau phase) (Figure 5).

**PCR Product Analysis**

An average 30-cycle PCR will result in an approximately 1 billion-fold amplification of the target sequence. The resultant amplification fragments are also called amplicon or PCR product. Several standard ways of visualizing this PCR product exist, many of which rely on electrophoresis.

In gel electrophoresis, PCR products are mixed with marker dyes and are loaded into wells at the top of an agarose gel. The entire gel is immersed in buffer and exposed to an electric current. Polymerase chain reaction products are essentially sequences of DNA, and because of the negative charge of the phosphate backbone, they migrate toward the positive charge. The rate of migration through the gel depends on several factors, the most important of which is the size of the fragments. Large molecules migrate at a slower speed than smaller molecules, resulting in separation of PCR products by size. A size ladder that contains DNA fragments of known size is usually loaded into the first lane of the gel so that the size of adjacent migrating PCR product can be estimated through comparison. A gel is read by staining the DNA, usually with ethidium bromide that intercalates between the bases in DNA. It needs to be visualized with UV light and a photograph can be taken of the gel for documentation.

Capillary gel electrophoresis is another powerful way to detect PCR products. This method is performed on a machine in a capillary tube filled with gel that is exposed to an electric current. For this detection system, one of the PCR primers in the set is initially synthesized with a fluorochrome tag at the 5’ end when it is synthesized. Each fragment of the synthesized PCR product will then contain a fluorescent label. Polymerase chain reaction product is loaded onto the machine, and as it passes through the capillary detection system the fluorochrome is excited by a laser. The light emitted of a specific wavelength is detected by the instrument. Like in agarose gels, the speed of migration is highly dependent on the size of the amplicons. A computerized printout shows the PCR products, with size of the PCR product on the horizontal axis and relative fluorescence of emitted light on the vertical axis (Figure 6). This assay is considered to be semiquantitative because the relative fluorescence (peak height) of different PCR products is proportional to the amount of PCR product generated, which in turn is proportional to the relative amount of genomic DNA in the original sample.
Quantitative PCR

Quantitative or real-time PCR has become an important assay in the clinical laboratory. Multiple instruments are commercially available that combine both PCR and the process of product detection. This is unlike traditional product detection, which is done after the completion of PCR. These instruments have the ability to analyze PCR in real time with quantitative results.

The development of quantitative PCR techniques was based on 2 important discoveries. First, it was observed that Taq enzyme has 5’→3’ exonuclease activity. This means that Taq enzyme can hydrolyze or cleave an oligonucleotide that is bound to the DNA sequence that is being replicated during PCR. Not all commercial Taq enzymes, however, have equal exonuclease activity. Second, it was shown that oligonucleotides designed with a specific 5’ and 3’ fluorescent label do not emit significant signal when they are intact. This phenomenon is because of the fluorescent energy resonance energy transfer principal (FRET); 1 probe will quench the signal of the other probe when they are physically close in proximity to one another. These unique dual-labeled probes are sometimes called FRET probes. When the 2 fluorescent dyes are separated from each other physically, as they are when the probe is degraded or hydrolyzed, fluorescence is emitted.

The TaqMan assay, one popular quantitative PCR assay, uses dual-labeled probes, called the TaqMan probes. The TaqMan probe is designed to anneal to the middle of the DNA sequence that is being amplified. It has 2 fluorescent dyes attached to it—the 5’ end probe is called the reporter dye (FAM, TET, JOE, HEX, and VIC are names of fluorescent dyes that can be used), and the 3’ end is called the quencher dye (TAMRA). As long as these 2 dyes are next to each other at the 2 ends of the intact oligonucleotide, fluorescence is quenched and no signal emitted is above baseline. When the TaqMan probe is added to a typical PCR, it will bind to the template DNA. When Taq encounters the bound probe, its exonuclease activity hydrolyzes the probe, resulting in separation of reporter dye from quencher dye. Once released, the reporter dye emits fluorescence that is proportional to the amount of PCR product that has been produced. Newer types of probes have also now been developed, including molecular beacons, scorpions, and others. An alternative detection model that is less expensive is the use of a double-stranded DNA binding dye (SYBR green). It does not require a sequence-specific dual-labeled TaqMan probe to be purchased. SYBR green does not emit fluorescence when it is unbound but does emit when it is bound to double-stranded DNA.

The read-out for quantitative PCR is the cycle threshold.
This is a measurement of the number of PCR cycles required for the fluorescent emission to be above a pre-established baseline. Cycle threshold is inversely proportional to the amount of starting template. In other words, with a high starting template amount, the threshold will be reached sooner (lower cycle threshold). Most assays require a standard curve to be generated, using a DNA sample of known concentration. Additionally, normalization to a housekeeping gene can be used to account for minor variations in the PCR efficiency.

**Controls**

Controls are essential for the analysis of the result of PCR or RT-PCR; careful consideration of controls should be part of assay design (Figure 7). A water control without DNA is used to establish that the laboratory and reagents are free of DNA or PCR product contamination. Positive and negative controls demonstrate that the reaction is working with expected appropriate results for a known positive and negative sample. In many assays, it is also important to include PCR for a housekeeping gene (one that is uniformly present or expressed); this control will demonstrate adequate quality of the DNA and RNA template. Finally, for semiquantitative assays, including a sensitivity control will help to establish the lower limits of detection for the assay. Interpretation of any assay results should begin with an assessment of the controls. If the controls are not appropriately positive or negative, the sample results should not be interpreted and the assay should be repeated.

**PCR Inhibition**

Polymerase chain reaction can be inhibited by both endogenous and exogenous substances and this inhibition can be the cause of failed reactions. Inhibition should be suspected when reactions fail despite apparently adequate quality and quantity of DNA template and with good results from control assays. Typically, in reactions using DNA from fresh tissue, the effects of inhibition can be overcome by heating and/or diluting the starting sample. In paraffin-embedded samples, a paradoxical situation has been described. Increasing the amount of starting DNA template may decrease efficiency of PCR.

The inhibitor was originally thought to be an unknown endogenous substance in paraffin-embedded tissues. Experimentally, the inhibitor was described as being “resistant to boiling, not destroyed by proteinases, not soluble in organic solvents, and to have a molecular weight of less than 10 kilodaltons.” Additional studies proved that inhibition in these reactions was actually caused by the presence of low-molecular-weight, fragmented DNA. Small fragments retain some primer binding in the PCR but fail to fully amplify when the end of the fragment is reached. One study elegantly showed that DNA purification leads to increased concentrations of higher-molecular-weight DNA fragments from paraffin-embedded tissue samples.

Other endogenous or exogenous inhibitors can be more easily avoided. Contaminants from extraction procedures, including chemicals and proteins, can inhibit PCR. Good nucleic acid purification procedures and even filtration if necessary will usually completely remove these contaminants. Heme can inhibit PCR, and, therefore, it is optimal to use microdissection to avoid areas of tissue with extensive necrosis and hemorrhage. Finally, sections that have been counterstained, particularly with nuclear stains, are not optimal for molecular testing. These stains may not directly inhibit PCR but interfere with obtaining high-quality DNA.

**Contamination**

In any diagnostic molecular laboratory, it is critical to avoid contamination. Because molecular tests are highly sensitive, the presence of even the smallest amount of contaminants in PCR can lead to false-positive results. It is not necessary to maintain sterility, but because of the greatly reduced risk of contamination, molecular laboratories may opt for a near-sterile setup. Several areas are particularly prone to contribute to preanalytical contamination in molecular anatomic pathology testing. These include the grossing stations, cryostats, and the water bath used for histology tissue sectioning. Usually, this contamination is at the tissue levels through floaters and carry-over artifacts. Contamination in the molecular laboratory is more likely to be from nucleic acids. This is of particular concern, given the highly sensitive nature of PCR.

Even a very small amount of amplified product can contain thousands of copies of the target. Specific procedures are followed in the molecular laboratory to avoid contamination. Use of standard techniques for good laboratory practice should be supplemented by additional precautions at the bench, including at the technique level and at the equipment level. Techniques that minimize risk of contamination include carefully aliquoting reagents to avoid constant opening of tubes, avoiding popping off of tube caps that can aerosolize contents, avoiding talking during PCR setup, and using regular UV treatment of equipment. Special equipment and reagents can be purchased to help minimize contamination as well. This includes RNase/DNase-free tubes and pipette tips, aerosol-resistant barrier pipette tips, and high-quality reagents. Finally, dUTP substituted for dTTP in the dNTP mixture during PCR has been used to decrease contamination issues.

**DIAGNOSTIC TESTING IN ANATOMIC PATHOLOGY**

**General Background**

Molecular diagnostics has evolved significantly since the early days when assays were only beginning to be described. One universal theme has been the need for high-quality validation procedures. Without a solid foundation of robust validation, tests are much more susceptible to errors. Of course, validation does not eliminate the potential for error, but it does highlight areas that are prone to error and ensures that the interpretation of results takes into account the potential for false-negative and false-positive results.

For most accreditation programs, a validation process is mandatory for new test development. In molecular testing, validation tends to be rigorous, exhaustive, and time consuming. Validation strategies are not specified by accrediting agencies, and there is no standardization across assays, platforms, or institutions. As a summation for the validation process, a summary document should be prepared to describe the samples tested, clinical indications for testing, results from validation experiments, and test interpretation guidelines. This document is the laboratory director’s assurance that the test meets institutional or laboratory criteria for new test implementation. The general
principles to guide new assay validation are discussed briefly in the following.

**Principles of Validation**

In new test development, whether in reference to a molecular test or to an immunohistochemical stain, several guiding principles will help to assess the quality and validity of the assay. These principles have been nicely specified by the Centers for Disease Control in the ACCE approach, which was developed in a project carried out by the Foundation of Blood Research (http://www.cdc.gov/genomics/gtesting/ACCE.htm; accessed August 20, 2007). The ACCE model involves an assessment of 4 different validation variables: analytical accuracy (A), clinical validity (C), clinical utility (U), and ethical, legal, and social implications (E).

The first component of assessing a new test is ensuring analytical accuracy. Analytical validity of an assay refers to how precisely and how reliably the test performs in detecting the product of interest. An assessment of false positives and false negatives that are inherent in the technique used is essential for determining analytical accuracy. The second component of validation is to assess clinical validity. Clinical validity of an assay determines how well the test predicts the disease of interest. This component of validation begins with analysis of literature on the relationship of the test results to the disease of interest and other confounding diseases. Additionally, it is generally recommended that a laboratory validating a new assay should test internal populations with and without the disease to determine whether their population results are concordant with those reported in the literature. Clinical utility of an assay is an assessment of the risks and benefits of introducing the new test into clinical practice. In other words, what populations will benefit from the test. The final category, ethical, legal, and social implications should identify any potential pitfalls in testing that should be addressed before the test is offered.

One good validation strategy is to use the ACCE model as a template to guide validation of new tests. The ACCE review includes a series of 44 questions that will help laboratories better understand the validation components needed for emerging genetic tests.

**Special Concerns Regarding Consent**

For the most part, molecular testing on tissue samples is used for assessing somatic genetic changes, which are tumor-specific alterations that do not usually have hereditary implications. Therefore, there are no specific guidelines or suggestions for obtaining patient consent or authorization for testing. Only when there is a question of hereditary disease testing is there the need to consider patient consent before testing.

**REGULATORY ISSUES**

The US Food and Drug Administration currently regulates diagnostic testing in pathology under the Office of In Vitro Diagnostics. All diagnostic testing regulations fall under the regulations for devices, as opposed to the other familiar arm of the US Food and Drug Administration, which regulates drug development. There are 2 basic options for bringing in new diagnostic tests: A laboratory can purchase a US Food and Drug Administration-approved commercially available kit, or the laboratory can purchase analyte-specific reagents (ASRs), design the assay, and use internal validation procedures. The latter method accounts for a large percentage of molecular diagnostic tests that are called laboratory developed or homebrew tests. Laboratory-developed tests are not currently regulated directly by the US Food and Drug Administration.

Companies that are certified to sell ASRs are under some regulatory control. This includes the mandated use of “good manufacturing practices.” Companies are not allowed to sell ASRs to laboratories that are not Clinical Laboratory Improvement Amendments of 1988–certified to perform high complexity testing. Other regulations pertain to the packaging and language used for ASRs regarding clinical applicability. Companies selling ASRs cannot direct the user about clinical use of reagents in any way. They also cannot give specific protocols or any other advice relating to clinical use. These ASRs are meant to be high-quality “building blocks” to create assays in the laboratory, and no direction regarding their use can be dispensed by the manufacturer.

**LABORATORY SETUP**

Setting up a molecular anatomic pathology laboratory requires investments in space, equipment, and personnel. As with any laboratory for clinical testing, the space requirements will depend on the anticipated volume of assays, the specific equipment that will be used, and the number of technicans/technologists that will be performing the assays. There is one general important principle that relates to space that is specific to molecular pathology testing. To reduce the risk of contamination, it is crucial to have a physical separation of the “clean” and “dirty” spaces whenever amplification procedures are used. There are many different configurations for laboratory space, but in the molecular anatomic pathology laboratory, separate space for microdissection should also be included (Figure 8). Clean space refers to all pre-PCR procedures, including microdissection, DNA or RNA extraction, and PCR setup. Dirty space is where post-PCR products (amplicon) are handled. Personnel will need to be trained to keep all reagents, equipment, and protective gear separate for these 2 spaces and not to transport anything from the dirty space back into the clean space. This physical separation of space greatly reduces the risk of contamination (see “Contamination”).

Equipment needed for molecular anatomic pathology varies based on the type of assays that will be performed. Laboratories may want to invest in a microdissecting platform: either a manual microscope or a laser microdissection system. All standard equipment for PCR and RT-PCR are required, along with some specialized product analysis equipment, depending on assays being performed. Common platforms may include machines for quantitative PCR, a capillary electrophoresis machine for fragment analysis, and a unit for direct sequencing. When budgeting for these items, it is also important to budget for service contracts for maintenance and repair, which can be relatively costly.

Trained personnel for a molecular laboratory are in high demand. Very few specific training programs for technologists in molecular techniques exist. Even in low-volume molecular laboratories, a minimum number of qualified technologists are required to provide adequate turnaround time and coverage for vacation. Cross training of technologists from other areas has been used successfully in some laboratories. Many institutions do not have a sep-
Molecular testing will almost certainly expand dramatically in anatomic pathology during the next decade. More and more clinical laboratories will begin to perform testing using the principles and techniques described in this review. As this occurs, practicing pathologists will need more familiarity with the underlying principles and will also need to be comfortable with the testing applications.

I thank Sucheta Kulkarni, PhD, for critical review of this work.

References


Cawkwell I, Quirke P. Direct multiplex amplification of DNA from a formalin fixed, paraffin wax embedded tissue section. Mol Pathol. 2000;53:51–52.


