Chapter 167
Molecular Medicine Demystified

Nikhil Moorchung

INTRODUCTION
The term “molecular medicine” is used to describe the role that the knowledge of deoxyribonucleic acid (DNA) is having on medical practice today. Molecular biology is the use of DNA-based knowledge in research.

The starting material for any form of molecular medicine is DNA. Deoxyribonucleic acid was discovered in 1869 by a Swiss physician called Friedrich Miescher who isolated the DNA from cell nuclei. The term “nucleic acid” was derived from Miescher’s “nuclein”. The double-helical structure of DNA was proposed by James Watson and Francis Crick, a hypothesis that has been confirmed several times since then. In 1956, Arthur Kornberg discovered an enzyme called DNA polymerase which enabled small segments of double-stranded DNA to be synthesized. Other significant discoveries include the discovery of messenger ribonucleic acid (mRNA) and the fact that each amino acid is encoded in DNA by a nucleotide triplet.

The polynucleotide chain reaction (PCR) was first described by Kary Mullis and this discovery made it possible to target segments of DNA with oligonucleotide primers and then amplify these segments. This method has several diagnostic and research applications. This method can be used to target not only DNA sequences but is capable of targeting RNA sequences as well.

CHROMOSOME
Deoxyribonucleic acid is packaged into chromosomes that can be essentially thought to be tightly coiled long chains of genes. Chromosomes can be visualized during cell division using a light microscope when they appear as thread-like structures or “colored bodies”. The word “chromosome” is derived from the Greek “Chroma” meaning color and “Soma” meaning body. Chromosomes are the structures that help distinguish one species from another and enable the transmission of genetic material from one generation to the next.

Structure of the Eukaryotic Chromosome
The chromosome is usually not visible in non-dividing cells. However, during mitosis and meiosis, they condense and become visible under the light microscope. The condensation of metaphase chromosomes probably results from folding of 30 nm chromatin fibers. Each metaphase chromosome consists of two sister chromatids, which are linked at a constricted region called the centromeres. Each centromere divides the chromosome into small and long arms designated as p for petit (French—small) and q for the next letter of the alphabet. The tips of the chromosome arms are called telomeres. Centromeres and telomeres will be elaborated later (Figure 1).

Morphologically, chromosomes are classified according to the position of the centromeres. If the centromeres are located centrally, it is called metacentric. If it is terminal, it is acrocentric. If the location is intermediate, it is called submetacentric.

Chromosomes are conventionally divided into regions and each region is divided into bands, numbering outward from the centromeres. A given point on a chromosome is designated by the chromosome number, the arm (p or q), the region and the band. In the case of longer autosomes, the q arm may be divided into three or four regions and the p arm into three regions. Within the major regions, the dark and light bands are numbered consecutively. To give an example of the methods used for assigning loci, the glucose-6-phosphate dehydrogenase (G6PD) gene is placed at q28, meaning it is in band 8 of region 2 of the q arm.

A shorthand notation system exists for the description of chromosomal abnormalities. Normal male and female karyotypes are designated as 46 XY and 46 XX respectively. A baby with Down’s syndrome would be designated as 47 XY + 21.

Figure 1: Structure of the eukaryotic chromosome
Classification of Chromosomes
Individual chromosomes differ not only in their positions but also in their overall length. Based on the three parameters of length, position of the centromeres and the presence or absence of satellites, most chromosomes can be identified. Chromosomes are classified into groups labeled as A to G on the basis of overall morphology. The classification of chromosomes is shown in Table 1.

In the normal human, the nucleus contains 46 chromosomes which are further broken up into 22 sets of autosomes and a single pair of sex chromosomes. The sex chromosomes are XX in females and XY in males. Members of a pair of chromosomes are known as homologs. Somatic cells have a diploid set of chromosomes whereas the gametes like the sperm and ovum have only one set with only a single sex chromosome. The number, sizes and shapes of metaphase chromosomes constitute the karyotype which is distinctive for every species.

The study of chromosomes is called cytogenetics. The metaphase chromosome is usually studied. Chromosomal analysis is done to look for numerical abnormalities (monosomies, polysomies) or to look for structural abnormalities (translocations and deletions).

DEOXYRIBONUCLEIC ACID STRUCTURE
Deoxyribonucleic acid is composed of a double-stranded antiparallel helix composed of deoxyribonucleotides linked together in a linear polymer by phosphodiester bonds between neighboring sugar residues. The strands are held together by hydrogen bonds between laterally opposed base pairs of purines and pyrimidines in which adenine binds with thymidine (double hydrogen bonds) and guanine binds with cytosine (three hydrogen bonds) (Figure 2).

Therefore, the two nucleotide sequences are complementary. Each nucleotide consists of an organic nitrogen containing base linked to a five-carbon sugar that has a phosphate group attached to 5’ carbon. In the case of DNA, the sugar is deoxyribose and in the case of RNA, the sugar is ribose.

Sugar-phosphate backbone is on the outside of the helix and forms the backbone of the helix with the purine and pyrimidine bases extending as side groups.

There are two purines and three pyrimidines. Adenine and guanine are the purines and thymine, cytosine and uracil are the pyrimidines. Adenine can pair with either thymine (in DNA) or uracil (in RNA). Guanine pairs with cytosine. These associations between a larger purine and a smaller pyrimidine are called Watson-Crick base pairs.

POLYMERASE CHAIN REACTION
The purpose of a PCR is to make a huge number of copies of a segment of DNA. This could be a gene, a portion of a gene or an intronic region. There are three major steps in a PCR, which are repeated for 30 or 40 cycles. This is done on an automated cycle, which can heat and cool the tubes with the reaction mixture in a very short time.

Denaturation
During the denaturation, the double strand melts and opens up to form single-stranded DNA. The temperature of denaturation is not fixed but it usually is about 95°C.

Annealing
This is done at a lower temperature than the denaturation. The primers have to anneal to that very specific segment of DNA. After this, the polymerase can attach and starts copying the template.

Extension
This is done at 72°C. The nucleotide bases are added from the 5’ end to the 3’ end.

These cycles repeat again and again until a few million copies of the segment are generated. These copies need to be visualized. Visualization is done on a gel, an agarose gel or a polyacrylamide gel. The PCR products are loaded on the gel and run using electrophoresis. After running the gel, the PCR products may be visualized on the gel under ultraviolet (UV) light as a clear band.

It is important to realize that PCR may not be the end of all investigation. The PCR may be useful in cases of tuberculosis (TB), in the detection of TB DNA in a sample of human tissue. But, what happens when one wants to detect a substitution or a deletion in the human genome. That is when the downstream applications of the PCR are useful. The PCR product may be used for restriction fragment length polymorphism (RFLP) or for sequencing. These techniques are more likely to shed light on small mutations in the human genome.

MICROARRAYS
For understanding the complex interactions between a malignant growth and the human body or the human body and the multifaceted environment, it is necessary to study two basic features; firstly, it is necessary to study the genome and determine if some of the numerous polymorphic loci were instrumental in causing the disease. The second object that needs to be studied is the expression of the transcripts, i.e. the presence of specific sequences of RNA.

Prior to the advent of the microarray, studying either of the two factors noted above was tedious and laborious. Single nucleotide polymorphisms (SNPs) had to be mapped using the basic PCR followed by either a RFLP or sequencing. Expression analysis entailed the conversion of RNA to cDNA and performing a PCR for different genes. The number of genes that could be studied was limited and therefore the depth of understanding was equally limited.

Microarrays or biochips are technologies which have revolutionized our understanding of disease. The uses of microarrays for gene expression profiling, genotyping, mutation detection and gene discovery are leading to remarkable insights into the function of thousands of genes previously known only by their gene sequence.

SEQUENCING
Sequencing is a process by which the sequence of nucleotides in a particular portion of DNA or RNA is obtained. This method has several obvious advantages to the pathologist. Firstly, it helps us to determine a sequence of nucleotides in a PCR product in order to determine if there is a mutation in the sequence or not. A classical example of this is in evaluating rat sarcoma (Ras) gene mutations where without sequencing, the determination of a mutation is impossible. It is also useful to confirm the presence of a SNP or a point mutation in cases where the RFLP is equivocal. Scientists use

<p>| TABLE 1 | The classification of chromosomes |</p>
<table>
<thead>
<tr>
<th>Group</th>
<th>Chromosome numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1–3</td>
</tr>
<tr>
<td>B</td>
<td>4–5</td>
</tr>
<tr>
<td>C</td>
<td>6–12 + X</td>
</tr>
<tr>
<td>D</td>
<td>13–15</td>
</tr>
<tr>
<td>E</td>
<td>16–18</td>
</tr>
<tr>
<td>F</td>
<td>19–20</td>
</tr>
<tr>
<td>G</td>
<td>21–22 + Y</td>
</tr>
</tbody>
</table>
it to characterize newly cloned cDNAs and to check the fidelity of a newly created mutation.

EPIGENETICS AND EPIGENOME

Epigenetics involves the heritable patterns of gene expression that do not involve changes in the sequence of the genome. There are several processes involved in epigenetics including DNA methylation, histone modifications, miRNAs and others. Epigenetic factors contribute to human disease.

Genomic Imprinting

Genomic imprinting is an epigenetic phenomenon by which epigenetic chromosomal modifications drive differential gene expression according to the parent of origin. This means that the expression of the gene is entirely according to the parent of origin. Expression is due to an allele inherited from the mother (as in H19 and CDKN1C genes) or it is because of an allele inherited from the father [such as the insulin-like growth factor (IGF)-2 gene]. This inheritance is independent of the classical Mendelian genetics. Usually imprinted genes are involved in a particular stage of development.

Deoxyribonucleic Acid Methylation

The most widely studied epigenetic modification is the cytosine methylation of DNA within the C-phosphate-G (CpG) dinucleotide. The CpG dinucleotide is a sequence of 5'–CG–3'. During evolution, the dinucleotide CpG has been progressively eliminated from the genome of higher eukaryotes and is present at only 5–10% of its predicted frequency. In the genome, there are smaller regions of DNA, called CpG islands ranging from 0.5 kb to 5 kb and occurring on average every 100 kb. CpG islands are usually found in the promoter region of genes. Chromatin containing CpG islands is generally heavily acetylated, lacks histone H1 and includes a nucleosome-free region. This is an open chromatin configuration and it allows for interaction of transcription factors with gene promoters.

Approximately half of all genes in mouse and humans (i.e. 40,000–50,000 genes) contain CpG islands. These are mainly housekeeping genes that have a broad tissue pattern of expression, but approximately 40% of genes with a tissue-restricted pattern of expression are also represented. Usually methylation is inversely correlated with the transcriptional status of the genes.

Histones and Epigenetic Regulation of Gene Expression

Histones, the protein backbone of chromatin, are also important in epigenetics. Today, they are recognized as being important translators between genotypes and phenotypes, having a dynamic function in the regulation of chromatin structure and gene activity. Understanding the importance of histones in the normal cell and how this change in neoplasia is still in its infancy compared with that of DNA methylation.

MICRORIBONUCLEIC ACIDS

Micrornucleic acids (miRNAs) were discovered in the early 1990s by Victor Ambros and colleagues. They found that miRNAs act as gene regulators. MicroRNAs had perhaps escaped detection because of their size as gene hunters were mainly interested in long mRNAs and disregarded very short RNAs.

MicroRNAs are ~22-nucleotide single-stranded RNAs that inhibit the expression of specific miRNA targets through Watson-Crick base pairing.

Figure 2: Structure of deoxyribonucleic acid (DNA)
pairing between the miRNA “seed region” and sequences commonly located in the 3’ untranslated regions (UTRs). The human genome is estimated to encode up to 1,000 miRNAs, which are either transcribed as stand-alone transcripts, frequently encoding several miRNAs, or generated by the processing of introns of protein-coding genes. The integration of miRNAs into introns of protein-coding genes serves to coordinate the expression of the miRNA with the mRNA encoded by that gene, within the necessity for a separate set of cis-regulatory elements to drive expression of the miRNA. It is not uncommon for intronic miRNAs to modulate the same biological processes as the protein encoded by the host gene miRNAs.

Bioinformatics and cloning studies have estimated that miRNAs may regulate 30% of all human genes and each miRNA can control hundreds of gene targets. MicroRNAs are highly conserved in sequence between distantly related organisms, indicating their participation in essential biological processes. It is well known that miRNAs have very important regulatory functions in such basic biological processes as development, cellular differentiation, proliferation and apoptosis.

**FLUORESCENT IN SITU HYBRIDIZATION**

Standard cytogenetic analysis of prenatal specimens detects chromosome aneuploidies and rearrangements with 99.5% accuracy. Prenatal cytogenetic analysis requires the isolation of metaphase chromosomes and takes 7–10 days for a final result. This wait which is imposed on both the clinician and the patient can increase the emotional burden on the patient and/or physician. Rapid analysis using fluorescence in situ hybridization (FISH) probes on direct amniocytes offers an opportunity to reduce anxiety through earlier decision-making.

Fluorescence in situ hybridization probes for enumerating chromosomes 13, 18, 21, X and Y can potentially identify 69–80% of all abnormalities found at amniocentesis, depending on the patients’ a priori risk and indication for the invasive procedure. In recent years, the possibilities for visualizing several chromosomal targets simultaneously have meant that FISH analysis has an increasing role to play in the study of patient samples. Fluorescence in situ hybridization can provide a rapid diagnostic answer in certain classes of cancer and a variety of human genetic diseases.

There has been a great increase in the diversity of DNA probes now available for the study of specific chromosomal changes in human cells, largely as a direct result of the success of the human genome project. Secondly, there has been an improvement in the labeling and imaging techniques which has made FISH a procedure which can be easily performed in many laboratories. Since FISH is performed on interphase nuclei, it is no longer always necessary to go through the lengthy process of preparing and analyzing metaphase cells derived from patient samples. Fluorescence in situ hybridization is also more sensitive than conventional cytogenetics and many new specific FISH clinical tests can assay for subtle chromosomal changes that would have been impossible to detect by conventional G-banded analysis. The problem with FISH vis-à-vis conventional in situ hybridization is that many tumors have very important regulatory functions in such basic biological processes as development, cellular differentiation, proliferation and apoptosis.

**COMPARATIVE GENOMIC HYBRIDIZATION**

Comparative genomic hybridization (CGH) or chromosomal microarray analysis (CMA) is a molecular cytogenetic method for the analysis of copy number changes (gains/losses) in the DNA content of a given subject’s DNA and often in tumor cells. It allows for the detection of chromosomal copy number changes without the need for cell culturing. It gives a global overview of chromosomal gains and losses throughout the whole genome of a tumor. Thus, CGH is a relatively fast screening technique that can point at specific chromosomal regions that might play a role in the pathogenesis or progression of tumors. Guided by CGH results, more specific molecular biological techniques (such as FISH, loss of heterozygosity analysis and sequencing) can be used to identify oncogenes and/or tumor suppressor genes in these regions.

**BIBLIOGRAPHY**

**General**


**Specific**

**Structure and Function of Deoxyribonucleic Acid**


**Chromosomal Structure and Function**


**Epigenetics**


**Polymerase Chain Reaction**


**Microarrays in Pathology**


**Sequencing and its Relevance to Pathology**


Chapter 167  Molecular Medicine Demystified