Molecular Basis of Diagnosis of Genetic Diseases

The term *molecular disease*, refers to disorders in which the primary disease-causing event is an alteration, either inherited or acquired, affecting a gene(s), its structure, and/or its expression.

A *genetic disorder* can be defined as any mutation in gene which changes its instructions to make protein or improper functioning of the protein or its complete absence, is termed as

A *genetic disease* occurs when an alteration in the DNA of an essential gene changes the amount or function, or both, of the gene products—typically messenger RNA (mRNA) and protein but occasionally specific noncoding RNAs (ncRNAs) with structural or regulatory functions. Although almost all known single-gene disorders result from mutations that affect the function of a protein, are now known. These disorders can be classified as:

- **Single-gene disorders**: where a mutation affects only one gene. Example, Sickle cell anemia.
- **Chromosomal disorders**: where chromosomes (or parts of chromosomes) are missing or changed. Example, Down syndrome.
- **Complex disorders**: where mutations affect two or more genes. Example, Colon cancer.
- **Mitochondrial disorders**: where mutations affect the non-chromosomal DNA of mitochondria. Example, an eye disease called Leber's hereditary optic atrophy; a type of epilepsy called Myoclonus Epilepsy with Ragged Red Fibers (MERRF).

Genetic disorders are a major health problem as they may develop in later life and are usually incurable. Therefore, molecular diagnosis of these disorders, especially at the prenatal stage becomes extremely important to prevent their transmission to next generation.

**Molecular Diagnosis of Genetic Diseases**

Early Diagnosis of any disease is key to efficient therapy to the patient. “Molecular diagnostics” is a holistic term defining various diagnostic tests that can evaluate a person’s health accurately at a molecular level, i.e., detecting and measuring specific genetic sequences (also called biological markers or biomarkers) in deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) or the proteins they express.

Molecular diagnostics is a reliable marker that helps determine whether a person is predisposed to get a disease, or actually has a disease, or whether a particular treatment option would be effective for a specific disease.
The last century depended on microscopy and histopathology or cellular pathology for identifying diseases. Clinicians classified cancer cells based on their pathology, that is, how they appeared under the microscope. Although they are the initial tools for rapid identification of the presence of infection, these methods have significant limitations. They are prone to individual errors at all stages of histopathological diagnosis, like processing of tissues, difficulty in identifying histological sections leading to misdiagnosis, etc. A more reliable genetic information from molecular diagnosis provides crucial information to enable correct diagnosis of a disease (Fig. 1).

**Figure 1: Old and New methods for molecular diagnostics**
Source: http://www.cancer.gov/cancertopics/understandingcancer/moleculardiagnostics/AllPages

**Molecular Diagnostic Techniques**

A number of techniques have been employed for molecular diagnostics of a particular disease. Some of the powerful tools are described briefly below:

- **Immobilization of DNA by Southern or dot blotting or in situ hybridization (ISH) or Fluorescent in situ hybridization (FISH):**—a labelled DNA or RNA strand that hybridizes with the target, complementary sequence and hence identifies and quantifies the target sequence in the sample (e.g., blood, tissue, saliva, etc.);

- **Amplification of DNA using the polymerase chain reaction (PCR):** the process of exponentially increasing the amount of a specific DNA or RNA sequence found in a sample until there are so many copies that they can be detected and measured;

- **Restriction endonuclease digestions:** identifying RFLPs (Restriction Fragment Length Polymorphisms) that are linked to the disease gene.

- **Microarrays:** which measure the expression of a large number of genes, or detect single nucleotide polymorphisms (SNPs), or genome regions;
Sequencing: a technique used to map out the sequence of nucleotides that comprise a strand of DNA.

The next section describes the use of the above listed techniques in some genetic diseases as test examples in order to appreciate the impact of molecular diagnostics.

Cystic Fibrosis or mucoviscidosis
It is the most common and life-threatening autosomal recessive disorder, meaning that the abnormal or mutated gene is located on one of the autosomes and two copies of the gene are necessary to have the trait or disease, one inherited from the mother, and one from the father (Fig. 2).

Autosomal Recessive Inheritance

![Autosomal Recessive Inheritance Diagram](https://www.nfed.org/learn/genetics-inheritance)

**Figure 2: Autosomal recessive inheritance**
Source: [https://www.nfed.org/learn/genetics-inheritance](https://www.nfed.org/learn/genetics-inheritance)

Symptoms

Cystic Fibrosis is an inherited disease caused by a genetic mutation (defect) on chromosome 7. The defective gene results in abnormalities in the production and function of a protein called the cystic fibrosis transmembrane conductance regulator (CFTR). In healthy cells CFTR acts as a chloride channel and a regulator of sodium, chloride and bicarbonate transport. Symptoms are highly variable between patients, but the prominent symptom is production of thick, sticky
mucous that clogs their respiratory airways and lungs. As known, the airways of lung are lined by ciliated epithelium, which maintain sterile conditions in the respiratory passage by clearing out hydrated mucus secretions out of the airways. Now, in the pathological condition of cystic fibrosis (CF), there is reduced airway surface hydration, which further leads to decreased mucus clearance. Thus, the airways are clogged with mucus that is dehydrated and thicker than normal (Fig. 3). This inability to clear mucus from the lung increases the susceptibility of CF patients to lung infections, which results in lung damage. The widespread presence of CFTR throughout the body (lungs, salivary glands, pancreas, liver, kidneys, sweat ducts and reproductive tract) helps to explain why CF is a multisystem condition affecting many organs. The two major systems affected are the lungs and the gastrointestinal tract. Moreover, the mucous may cause gastrointestinal pseudo-obstructions, block pancreatic and bile ducts and interfere with exocrine function, thereby making it a multiple organ disease. Additionally, it is also one of the main causes of male infertility, because the mucus also clogs vas deferens (tubes that carry sperm from the testes).
Figure 3a and 3b. In CF, impairment of CFTR function causes reduced fluid production. Enhanced sodium absorption through epithelial Na+ channels (ENaC) and basolateral Na/K ATPase pumps results in increased fluid absorption leading to drier airways and impaired ciliary clearance.

Source: http://www.cfmedicine.com/htmldocs/cftext/basicproblem.htm

Pathogenesis

Cystic Fibrosis is caused by mutations in the CFTR (cystic fibrosis trans membrane conductance regulator) gene. It is located on chromosome 7 (Fig. 4). (Locus: 7q31.2 - The CFTR gene is found in region q31.2 on the long (q) arm of human chromosome 7). This gene provides instructions for making a protein called the cystic fibrosis trans membrane conductance regulator. It belongs to a family of genes called ABC (ATP-binding cassette transporters) and ATP (ATPase superfamily).
B.Sc. (Hons) Zoology
Biotechnology
Sem VI

- **Locus**: 7q31.2 - The CFTR gene is found in region q31.2 on the long (q) arm of human chromosome 7.
- **Gene Structure**: The normal allelic variant for this gene is about 250,000 bp long and contains 27 exons.
- **mRNA**: The intron-free mRNA transcript for the CFTR gene is 6129 bp long.
- **Coding Sequence (CDS)**: 4443 bp within the mRNA code for the amino acid sequence of the gene’s protein product.
- **Protein Size**: The CFTR protein is 1480 amino acids long and has a molecular weight of 168,173 Da.
- **Protein Function**: The normal CFTR protein product is a chloride channel protein found in membranes of cells that line passageways of the lungs, liver, pancreas, intestines, reproductive tract, and skin. CFTR is also involved in the regulation of other transport pathways.
- **Associated Disorders**: Defective versions of this protein, caused by CFTR gene mutations, can lead to the development of cystic fibrosis (CF) and congenital bilateral aplasia of the vas deferens (CBAVD).

**Figure 4**: Location of CFTR gene on chromosome 7

**Discovery of Cystic fibrosis disorder**

Medieval folktale was superstitious and believed that infants having salty skin were ‘bewitched’ as they routinely died early. The saying goes that “Woe to that child which when kissed on the forehead tastes salty. He is bewitched and soon must die”. This saying, from is an early reference to the common genetic disease recognized today as cystic fibrosis. Salty skin is now recognized as a sign of CF. It was not until 1936, however, that Dr. Guido Fanconi named this condition “cystic fibrosis with bronchiectasis.” In 1949, Dr. Charles Upton Lowe established that CF was a genetic disorder, and in 1953, Dr. Paul A. di Sant’Agnese reported that children with CF secrete excessive salt in their sweat after observing dehydration in these children during a New York City heat wave.

**Structure of CFTR**

The CFTR gene is approximately 189 kb in length. CFTR is a glycoprotein with 1480 amino acids. The protein consists of five domains.

1. Two membrane-spanning domains (MSD1 and MSD2) that form the chloride ion channel, each with six spans of alpha helices.
2. Two nucleotide-binding domains or folds (NBD1 or NBF1 and NBD2 or NBF2) that bind and hydrolyze ATP (adenosine triphosphate), and

3. A regulatory (R) domain.

Delta F508, the most common CF-causing mutation, occurs in the DNA sequence that codes for the first nucleotide-binding domain (NBD1). NBD1 is connected to MSD2 by a regulatory "R" domain that is a unique feature of CFTR, not present in other ABC transporters (Fig. 5).

19% of the CFTR protein make up the twelve transmembrane domains (M1 - M12). These domains have been shown to be comprised of typical a-helical secondary structure. Many of the residues within these regions form the channel lining residues and have a major role in the regulation of pore function.

Figure 5: Structure of CFTR after proper folding in the Endoplasmic Reticulum (ER), R domain contains phosphorylation sites (black triangles), the branched structure on right half represents potential glycosylation sites.

Function of CFTR

CFTR functions as a channel across the membrane of cells that produce mucus, sweat, saliva, tears, and digestive enzymes. The CFTR protein is a cyclic-AMP dependent channel, thus with increasing levels of c-AMP inside a secretory epithelial cell, PKA (protein kinase A) is activated. PKA in turn phosphorylates the R-domain of CFTR protein, thus opening the channel. Thus, the ion channel only opens when its R-domain has been phosphorylated by PKA and ATP is bound at the NBDs.

In opened state, the channel transports negatively charged chloride ions into and out of cells, which in turn control the movement of water in tissues by osmosis. With chloride ion transport, the sodium ion balance is also maintained in and out of the cell. This water intake is responsible for the production of thin, freely flowing mucus, which can be easily cleared off from the airways of lung (Fig. 6).

![Figure 6: Normal CFTR channel and a mutant channel](http://cyfbin.weebly.com/cyfb.html) For educational purpose

So, in patients suffering from cystic fibrosis, the function of chloride channels is disrupted due to mutations in the CFTR gene, thereby preventing the usual flow of chloride ions and water into and out of cells. This in turn leads to thick, dehydrated mucus, which can not be cleared off easily by the ciliated epithelium, further leading to clogging or obstruction of air passages of lungs. Additionally, the inability of epithelial tissue (lining the ducts of sweat glands) to absorb chloride and the consequent impairment of sodium absorption from the duct lumen...
Common Mutations leading to Cystic Fibrosis

According to the UMD-CFTR (Universal Mutation Database-Cystic Fibrosis Mutation Database), there are 910 mutations known in the CFTR gene. Different CFTR mutations result in different disease phenotypes. The most common of these is a deletion of just three nucleotides in the DNA sequence that codes for the first nucleotide-binding domain (NBD1), which leads to the deletion ("Δ") of an amino acid (phenylalanine, F) at position 508 of the protein sequence (Fig. 7. This is denoted as ΔF508, and is found in around 90% of CF patients.

Due to ΔF508 in CFTR gene, NBD1 is not completely formed. Thus, after translation of this mutated CFTR gene, the CFTR protein is not able to fold correctly and hence the cell’s quality-control mechanism marks this defective protein for degradation. The end result is the absence of a properly folded and functional CFTR channel, thereby preventing the passage of ions and water into and out of the cell.

In addition to ΔF508 mutation, the other types of mutation causing CF have been categorized into five classes:

I. Mutations that affect the biosynthesis of CFTR protein; mRNA processing is interrupted by the introduction of premature translation termination signals or by aberrant splicing. Example, W1282X mutation, where instead of inserting the amino acid tryptophan (W), the protein sequence is prematurely stopped (indicated by an X) at position 128.

II. Mutation that affect CFTR protein maturation; CFTR is produced but does not fold correctly, giving rise to improper maturation of the protein. Example, ΔF508 mutation.

III. Mutations that alter the channel regulation, CFTR is improperly activated as mutations affect binding and hydrolysis of ATP or phosphorylation of the R-domain. Example, G551D
missense mutation, where aspartate (D) is added instead of a glycine amino acid (G) at position 551.

IV. Mutations that does not allow proper chloride flux due to defective conduction through the pore, although some mutations cause lower chloride channel activity. Example, R117H missense mutation, where histidine is added instead of arginine at position 117.

V. Mutations that reduce the level of normally functioning CFTR at the cell membrane; Example, 3120+1G>A Splice-site mutation in gene intron 16, where glycine is replaced by alanine at the first nucleotide in the intron following nucleotide 3120 in the cDNA.

VI. Mutations that affect the regulation of other ion channels such as the sodium channel.

Additional information: Video

Cystic fibrosis disorder

Watch the video and learn more about the molecular mechanism of cystic fibrosis:
https://www.youtube.com/watch?v=_j99-xgOlaw

Source: YouTube

Possible Methods used in CFTR testing

A wide range of techniques is used to identify CFTR mutations. These can be divided into two groups:

1. Direct or Targeted methods (i.e, testing DNA samples for the presence or absence of specific known mutation(s)) and

2. Indirect or Scanning methods (i.e, screening samples for any deviation from the standard sequence).

Targeted Methods:

For known mutations like ΔF508, direct testing procedures include PCR and dot blot techniques.

(A) Detection using hybridization technique

The method includes hybridization to sequence-specific oligonucleotide (SSO) probes, which are of two kinds:

(1) SSO1, which is complementary to the DNA in the normal sequence (with no deletion of nucleotides);

(2) SSO2, which is complementary to the DNA in mutated sequence (with the deletion).
After CFTR DNA is amplified using PCR, it is placed onto a membrane in duplicate as “dot blots.” (Lemna et al., 1990). Each blot is then hybridized to one of the above two SSOs. As expected, DNA from individuals homozygous for the deletion hybridizes only to SSO2, while DNA from individuals homozygous for the normal sequence hybridizes only to SSO1. On the other hand, DNA from individuals heterozygous for the deletion hybridizes to both oligos, SSO1 & SSO2. In the Figure 8 below, five subjects are homozygous for the mutation (2, 3, 4, 5 and 8) and three are heterozygous for the mutation (1, 7 and 9), while one subject lacks delta mutation (no. 6).

Figure 8: Expected results of dot blot using SSO1 and SSO2 with CFTR DNA isolated from 9 individuals or subjects.

(B) Detection using PCR analysis

As already stated, ΔF508 is caused by a 3 bp deletion in CFTR gene, which leads to loss of amino acid phenylalanine, hence, the detection method using PCR includes designing of primers that can distinguish a normal gene from a mutant gene. As expected with these primers, two kinds of amplified products will be formed of different sizes: a 154 bp product (from a normal individual) and a 151 bp product (from an individual with the CF disease). On the other hand, the carrier individuals will be heterozygous for the mutation, hence both the bands will appear. These patterns are depicted below (Fig. 9):
(C) Detection using RFLP analysis

Restriction Fragment Length Polymorphism (RFLP) can be used to detect genetic variations as it is based on a simple principle, that even minor base changes in a gene or any DNA sequence can change the pattern of restriction enzyme digestion fragments obtained (Fig. 10). Since, the genetic diseases like cystic fibrosis are associated with such genetic alterations like deletions (ΔF508), which can create or delete a restriction enzyme recognition site, hence they can be easily detected using RFLP analysis.
Apart from these listed methods, there are still many more testing procedures for detection of mutations causing cystic fibrosis, but description of those methods is beyond the scope of this chapter. In conclusion, as we understand that cystic fibrosis is a chronic disease, with new advances in detection methods and treatment options available, the rate of illness of patients suffering from the disease has been decreased at an alarming rate. The video about Cystic fibrosis helps appreciate this.

**Huntington’s Disease**

Also known as Huntington's chorea, Huntington’s Disease is a progressive, and fatal autosomal dominant neurodegenerative disorder. Autosomal dominance indicates that there is one copy of disease causing or mutant gene and one normal gene on a pair of autosomes and the former is dominant over the latter.
Since the Huntington gene is not on a sex-determining chromosome, the disease is not sex-linked. This means that males and females have an equal chance of inheriting the disease. A person with Huntington’s disease has one non-HD allele and one HD allele. Hence, there is a 50% chance that the non-HD allele will be passed on and a 50% chance that the HD allele will be passed on. This means that each child of an individual with HD has a 50% chance of getting the HD allele (Fig. 11). Individuals with a chance of inheriting the disease are sometimes described as “at-risk.”

**Autosomal Dominant Inheritance**

- In autosomal dominant inheritance, only one copy of a disease allele is necessary for an individual to be susceptible to expressing the phenotype.
- With each pregnancy, there is a one in two (50%) chance the offspring will inherit the disease allele.
- Unless a new mutation has occurred, all affected individuals will have at least one parent who carries the disease allele.
- Autosomal dominant inheritance is often called vertical inheritance because of the transmission from parent to offspring.
- Across a population, the proportion of affected males should be equal to the proportion of affected females.
- Male-to-male transmission can be observed.
- Examples of diseases with autosomal dominant inheritance include **myotonic muscular dystrophy** and **Huntington disease**.

![Huntington's Disease Passed On Through Generations](https://edu.glogster.com/glog/huntingtons-disease/270gvkq4q1w)

**Figure 11: Pattern of autosomal dominant inheritance**

Source: [https://edu.glogster.com/glog/huntingtons-disease/270gvkq4q1w](https://edu.glogster.com/glog/huntingtons-disease/270gvkq4q1w)

There are three major types of symptoms observed in individuals suffering from Huntington’s disease (HD), commonly referred to as the ‘clinical triad’ of movement abnormality (referred
to as chorea), emotional disturbance, and cognitive impairment. The sequence in which symptoms develop varies from person to person. There are two forms of this disease.

- Early-onset Huntington disease, which is rare and begins in childhood or the teens.
- Adult-onset Huntington disease, which is the most common form. Persons with this form usually develop symptoms in their mid 30s and 40s.

**Symptoms and Causes**

The pathology of HD shows a general atrophy of brain and the principal neuropathological characteristic is selective loss of neurons within the striatum. The most noticeable atrophy is found in the caudate nucleus and putamen (which together comprise the corpus striatum portion of the basal ganglia, Fig. 12). In advanced cases, the total brain weight is reduced by 25-30%. Severe loss of medium spiny neurons, especially those synthesising neurotransmitters, enkephalin and gamma-Aminobutyric acid (GABA), is most prevalent in the striatum.

![Figure 12: HD affects the basal ganglia of brain](https://runkle-science.wikispaces.com/Huntington's+Disease+(HD))

Substantial neuronal loss is also evident in the globus pallidus and subthalamic nucleus. Additionally, degeneration in the cerebral cortex is also found with widespread loss of neurons. Thus, the major degeneration, which normally occurs first in the striatum, leads to chorea, and the subsequent loss in the cerebral cortex causes dementia. With such severe loss of neurons, the communication between the neural networks is affected drastically, leading to the obvious symptoms related to movement, cognitive, and psychiatric disorders.

**Structure of Huntingtin gene**

After an intensive research, the causative gene for HD was identified by in 1989 and was named ‘huntingtin’. The symbol for this gene is HTT. The HTT gene provides instructions for making a protein called huntingtin (originally known as IT-15). It is located on the short (p) arm of chromosome 4 at position 16.3 (Fig. 13).
It consists of 67 exons spanning 180 kb of DNA. The \textit{HTT} gene belongs to a family of genes called endogenous ligands (those that are produced in the body and not those introduced into the body, such as certain drugs). Although wild-type huntingtin is expressed throughout the human body, it is most concentrated in the male testes as well as the neurons of the central nervous system. Also, not much is known about the normal cellular function of HTT but some in vitro studies have identified a wide range of HTT-interacting proteins and suggest that it may be involved in different biological processes as protein trafficking, vesicle transport, anchoring to the cytoskeleton, clathrin-mediated endocytosis, postsynaptic signalling, transcriptional regulation and anti-apoptotic functions.

Huntingtin is a large protein (347 kilo-Da) that is made up of 3,144 amino acids. Near the N-terminal, the protein contains a repeating region of 11 to 34 glutamine residues that are encoded from a region of CAG repeats on DNA.

**Mutation that causes Huntington’s disease**

CAG repeats in normal HD gene range from 10 - 29 repeats but some may have upto 30-35 repeats (Fig. 14).
Thus, the CAG repeat is highly polymorphic in nature. When this number exceeds from this normal range, Huntington’s disease can occur. The inherited mutation that causes HD is known as a **CAG trinucleotide repeat expansion** (Fig. 15).

![Figure 15: No. of CAG repeats on chromosome 4, associated with the resulting tendency in an individual to get affected by HD](http://smithlhhsb122.wikispaces.com/Alison+P)

The (CAG)n repeat lies immediately upstream from a moderately polymorphic polyproline-encoding (CCG) repeat. (CCG)n tract may vary in size between 7 and 12 repeats in both normal and affected individuals, with no correlation being observed between the number of the repeats and the disease (Fig. 16). However, in a study reported by Pramanik et al., in 2000, there is an exciting relationship across major human morphological groups, between the CAG repeat numbers and CCG repeat locus in the huntingtin gene. For example, in Caucasian populations, CAG repeats in the disease range, are strongly associated with (CCG)7, while in Mongoloid populations, this association is with (CCG)10.

![Figure 16a: Huntington’s gene with CAG and CCG repeats](http://smithlhhsb122.wikispaces.com/Alison+P)
B.Sc. (Hons) Zoology
Biotechnology
Sem VI
Additional Information: 
Discovery of Huntington’s disease

The disorder of Huntington’s disease is named after George Huntington, an American physician who first described it in the late 1800s. The interest in the disorder began with his classic description in the article titled ‘On Chorea’ (meaning "mad dance") in 1872, in which he clearly described both the familial nature of the disorder and its symptoms (Fig. 17). He wrote about the illness that he called "an heirloom from generations away back in the dim past." Before the description of this disease, some HD sufferers were thought to be possessed by spirits or victimized as witches, and were shunned or exiled by the society. While others who exhibited symptoms of HD were accommodated by George Huntington and his family. Both father and grandfather of George Huntington were physicians and practiced medicine on Long Island, New York. As a boy, Huntington often accompanied his father on medical rounds and witnessed with the devastation of this disorder. From combined observations of his own, his father and grandfather, with chorea and their families on Long Island from almost eighty years, he was able to produce his groundbreaking paper at the age of 22, about a year after completing his medical degree at Columbia University in New York City.

Figure 17: Front page of George Huntington's communication, "On Chorea"; photo of George Huntington (on right).
The number of CAG repeats have been approximately correlated with the proportion of individuals having the mutation and who exhibit clinical symptoms of HD (i.e. penetrance) and it is summarized in the table below:

<table>
<thead>
<tr>
<th>Repeat length in triplets</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;27</td>
<td>Normal</td>
</tr>
</tbody>
</table>

**Figure 16b: Huntington’s gene with CAG repeats**

When CAG repeats are too many, the implication obviously will be an overabundance of amino acid glutamine (codon CAG codes for glutamine) in the translated Huntingtin gene. Because of this, Huntington’s disease is also called as ‘polyglutamine disease’ or a ‘polyQ disease’ (glutamine is represented by single amino acid code Q).

![Figure 17: Tri-nucleotide CAG repeat causes addition glutamine residues in Huntington protein.](http://smithlhhsb122.wikispaces.com/Josie+P. CC)

Since glutamine is polar, or a “charged” molecule, a large number of glutamine residues causes links (hydrogen bonds) to form within and between proteins. These Huntingtin aggregates (also known as neuronal inclusions, or inclusion bodies) accumulate in the nuclei of cells, where they interfere with the process of transcription, and they also affect other cellular processes such as axonal transport in neurons. These aggregates of mutant huntingtin have been found to
bind with a protein named Huntingtin-associated protein-1 (HAP-1), which is encoded by the HAP1 gene (Fig. 17). The binding is a result of interactions between polar residues (repeat of glutamines), which are aligned in the form of a 'zip fastener', termed as ‘polar zipper effect’.

**Molecular Methods for testing Huntington’s Disease**

Diagnostic and predictive testing for Huntington disease (HD) requires an accurate determination of the number of CAG repeats in the Huntington (HTT) gene.

**(A) Detecting (CAG)n repeats by PCR: Triplet repeat primed PCR**

The most obvious method of detecting HD is analysis of CAG repeats by simple PCR assay. Two primers are constructed which span the region of CAG trinucleotide repeat region and the resulting PCR product size can be estimated from polyacrylamide gel electrophoresis. For further confirmation, a primer set that amplifies a region encompassing both the CAG repeat and the adjacent (CCG)n is used (Fig. 18).

![Diagram of PCR primers](image)

**Figure 18:** shows the primer set (forward and reverse) to amplify the CAG repeat region.

Though, this method is simple and cost-effective, it suffers from few limitations, like:

a) If one of the CAG repeats is very large (more than about 150 repeats), it can be too large for the standard PCR method to detect it, so it can appear like there is only one, normal-sized gene.

b) If a mutation in the region of primer binding precludes amplification of one allele. Thus, the end result will be that one copy of the gene will remain undetected, giving a false diagnosis that the patient carries two identical copies with equal CAG repeat lengths.

c) If the HD locus has high C+G content and highly repetitive, it may not be responsive to routine PCR.

d) The radioactive analysis or silver staining of polyacrylamide gels produces several shadow bands around the main allele band due to the nature of the repeat amplifications. So, it is often
very difficult to distinguish between the main and shadow bands when they have similar intensities.

e) It is difficult to size the alleles for two reasons. First, if the product yield is low, it produces a very faint signal, and second, as the migration rate across the gel is different in different lanes, the calibrator size values may be rendered inaccurate. Therefore, conventional PCR and the usual detection of products are generally not suitable for appropriate diagnosis of individuals at risk for HD.

(B) Detecting (CAG)n repeats by Fluorescent PCR method:
As discussed previously, that the conventional PCR detection methods to diagnose HD are not completely accurate and suffer from some significant limitations, thus, an improvement of this technique in the form of fluorescent PCR has been designed. In this method (given by Tóth et al., 1997), primers are labeled with fluorescent markers and a fluorescent DNA sequencer (Genescanner) is used to separate, detect and analyse the fluorescently labelled PCR products.
It has particular advantages as compared with the conventional method:
1. It is more sensitive,
2. It can be used for quantitative measurement of bands,
3. Few PCR cycles are required for the same level of detection, resulting in cleaner band pictures and easier interpretation.
4. Even weak signals can be easily visualized, allowing the detection of alleles with high numbers of CAG repeats.

(C) Detecting (CAG)n repeats by Southern Blot Hybridization:
As we know that the number of CAG repeats can reach >40, thus such large amplifications may be difficult to obtain with the routine PCR method. Thus, in such cases, southern blot hybridization may prove useful for presymptomatic diagnosis of patients with high risk of HD. It is also useful for the confirmation of “homozygous normal” genotypes.
According to the method given by Guida et al., 1996, a 360 bp PCR-derived labeled probe (obtained from amplification of CAG repeats) is used for hybridization against the genomic DNA of individuals being tested for HD. This is a direct measurement of the CAG repeat in the HD gene and has been found to be a sensitive and highly accurate molecular diagnostic test.
B.Sc. (Hons) Zoology
Biotechnology
Sem VI
Detection of Huntingtonin Gene CAG Expansion Mutations by PCR

![Detection of CAG Expansion Mutations by PCR and Southern Blot](https://slideplayer.com/slide/5987380/)

**Figure 19:** Detection of CAG Expansion Mutations by PCR and Southern Blot
Labeled PCR primer
Huntingtin
80–170 bp
>40 repeats
Huntington Disease
10–29 repeats (normal)

Autoradiogram of polyacrylamide gel

**Conclusion**

It can be concluded that accurate and error-free diagnosis of Huntington’s disease is extremely important, with its evident critical symptoms and clinical variability. A single test should not be used to confirm the possibility of HD, instead, it should be supplemented with additional testing procedures with full confidence. There is no ground for misdiagnosis, which may prove an extremely fatal and life threatening situation for individuals who shall not receive proper medical care on time. This also implies another situation of an unnecessary emotional turmoil for those who might not be at risk, but are somehow wrongly diagnosed for HD due to some technical laboratory error.

**Additional Information:**

**Discovery of Huntington’s disease**

Learn about the latest research in Huntington’s Disease by watching the video:

https://www.youtube.com/watch?v=UIsOCPVAV4Y

Source: YouTube
B.Sc. (Hons) Zoology
Biotechnology
Sem VI
Sickle Cell Anaemia

Sickle cell anemia, also known as Sickle cell disease was the first genetic disease to be characterized at the molecular level. It is the most common inherited blood disorder in the United States, especially, African Americans. It affects hemoglobin, a molecule that carries oxygen throughout the body. As the name suggests, the red blood cells of individuals suffering from this disease become sickle-shaped. This is due to a point mutation that causes the normal hemoglobin (designated as HbA) to become sickle-shaped (designated as HbS).

Sickle cell disease is an autosomal recessive disorder. When both parents are carriers of sickle cell disease, there is a 25% chance of each child having the disease. Moreover, there is a difference between sickle cell trait and sickle cell disease. In the former, the individual is only a carrier of HbS mutation with no disease symptoms, while in the latter, the individual suffers from the disease with all the symptoms.

Table: Genotype of HbA and HbS and resulting phenotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Description</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA HbA</td>
<td>Normal</td>
<td>Normal hemoglobin</td>
</tr>
<tr>
<td>HbA HbS</td>
<td>Sickle Cell Trait/Carrier</td>
<td>No disease symptoms</td>
</tr>
<tr>
<td>HbS HbS</td>
<td>Sickle Cell Disease</td>
<td>Sickle hemoglobin</td>
</tr>
</tbody>
</table>

Sickle haemoglobin

![Figure 20: Shape of normal (HbA) and sickle cell haemoglobin (HbS)](image-url)
Symptoms and Causes

In people with Sickle cell disease, abnormal hemoglobin molecules, called hemoglobin S, cause their red blood cells to form crescent shapes also known as sickling. These abnormal red blood cells can then block sections of blood vessels leading to severe pain (also known as a vaso-occlusive crisis). On the contrary, normal red blood cells are disc-shaped (like doughnuts) and hence can move easily through the blood vessels. Additionally, the abnormal sickle cells usually die after only about 10 to 20 days (in contrast to normal RBCs, which have a life span of 120 days). With such rapid destruction of RBCs, the bone marrow cannot make new red blood cells fast enough to replace the dying ones. The end result is severe loss of hemoglobin, leading to anemia and resulting lowered supply of oxygen to the entire body (Fig. 21).

Figure 21: (a) shows normal RBCs flowing freely through veins. The inset shows a cross section of a normal red blood cell with normal hemoglobin. (b) shows abnormal, sickled RBCs blocking and accumulating at the branching point in a vein. The inset image shows a cross-section of a sickle cell with long polymerized HbS strands extending and altering the cell.


Severe anaemia can cause a person to be pale, short of breath, easily tired, and poor growth. Other symptoms of sickle cell disease include recurrent infections, extreme pain and severe complications such as blindness and stroke. Anaemia can cause shortness of breath, fatigue,
and delayed growth and development in children. It may also cause yellowing of the eyes and skin, which are signs of jaundice. It also causes damage to the spleen, kidneys and liver. With spleen damage, young children become more prone to bacterial infections. A particularly serious complication of sickle cell disease is high blood pressure in the blood vessels that supply the lungs (pulmonary hypertension) and can lead to heart failure.

**Structure of Hemoglobin**

Hemoglobin (Hb) is a red-pigmented protein found in erythrocytes (or RBCs) that is responsible for the transportation of oxygen and carbon dioxide within the body. Each Hb molecule consists of two chains: one alpha and one non-alpha (epsilon, gamma, delta, beta), each containing a non-protein (or heme) group that is in the shape of a ring, with an iron ion (Fe2+) in its center. Two hemoglobin molecules combine to produce functional HbA tetramer. Alpha globin genes are present on chromosome 16 (with two gene loci, HBA1 and HBA2, hence 4 copies per cell), while non alpha genes (one gene locus for beta globin, HBB, hence only 2 copies per cell) are present on chromosome 11. On chromosome 11, 2 gamma genes and an epsilon gene are present during embryogenesis. The two gamma genes form hemoglobin F or fetal hemoglobin. Thereafter, adult hemoglobin is present, which is composed of two alpha genes and two beta genes, designated as HbA (Hemoglobin A). And also on chromosome 11, delta is present which contributes a very small amount of hemoglobin to both adults and children (Fig. 22).

![Alpha Globin gene cluster](image1.png)  ![Chromosome 11](image2.png)

**Figure 22:** (a) Structure of alpha and beta globin gene cluster. Source: Author
(b) **Location of HBB gene on chromosome 11.**  

While the binding of oxygen actually occurs at the iron sites, all four globin chains must work together in order for the Hb to function well. When oxygen concentrations are high, Hb has a very high affinity for oxygen, but at low oxygen concentrations, its affinity for oxygen decreases (Fig. 23).
Figure 23: (a) Transport of oxygen by hemoglobin (Hb) in RBCs., (b) Structure of Hb molecule
(b) https://sites.jmu.edu/bio103shook/bbq4
Mutation which causes Sickle cell anemia

The mutations that cause sickle cell anemia have been extensively studied and demonstrate how the effects of mutations can be traced from the DNA level up to the level of the whole organism. Consider someone carrying only one copy of the gene. She does not have the disease, but the gene that she carries still affects her, her cells, and her proteins:

There are effects at the protein level

**Figure 24:** Normal hemoglobin (left) and hemoglobin in sickled red blood cells (right) look different; the mutation in the DNA slightly changes the shape of the hemoglobin molecule, allowing it to clump together.

Source: [https://evolution.berkeley.edu/evolibrary/article/0_0_0/mutations_06](https://evolution.berkeley.edu/evolibrary/article/0_0_0/mutations_06)

Sickle cell anaemia is caused by a point mutation in the beta globin gene (HBB), producing its abnormal version, known as hemoglobin S (HbS). Beta hemoglobin (beta globin) is a single chain of 147 amino acids.

**There are effects at the cellular level**: When red blood cells carrying mutant hemoglobin are deprived of oxygen, they become "sickle-shaped" instead of the usual round shape (see picture). This shape can sometimes interrupt blood flow.
There are negative effects at the whole organism level: Under conditions such as high elevation and intense exercise, a carrier of the sickle cell allele may occasionally show symptoms such as pain and fatigue.

There are positive effects at the whole organism level: Carriers of the sickle cell allele are resistant to malaria, because the parasites that cause this disease are killed inside sickle-shaped blood cells.

As a result of this mutation, even though the protein is still 147 residues in length, but valine (a non-polar amino acid) is inserted into the beta globin chain instead of glutamic acid (an electrically charged amino acid) (Fig. 25). Though, it seems only a single nucleotide change, but the results are destructive.

![Image of amino acid substitution](http://cnx.org/content/m44402/latest/?collection=col11448/latest CC)

Figure 25: Genetics of Sickle cell anemia: (a) and (b) show that a single amino acid substitution leads to sickle cell anemia. In normal hemoglobin, the amino acid at position seven is glutamate. In sickle cell hemoglobin, this glutamate is replaced by a valine.

Source: (a) [http://cnx.org/content/m44402/latest/?collection=col11448/latest CC](http://cnx.org/content/m44402/latest/?collection=col11448/latest CC)
(b) [http://bio-brainstorm.wikispaces.com/Chromosomes,alleles+%26+mutations CC](http://bio-brainstorm.wikispaces.com/Chromosomes,alleles+%26+mutations CC)

This is a chain of causation. What happens at the DNA level propagates up to the level of the complete organism. This example illustrates how a single mutation can have a large effect, in this case, both a positive and a negative one. But in many cases, evolutionary change is based
on the accumulation of many mutations, each having a small effect. Whether the mutations are large or small, however, the same chain of causation applies: changes at the DNA level propagate up to the phenotype.

**Additional Information: Video**

**Sickle Cell Anemia**

Please watch this interesting video and learn about the sickle cell anaemia disease

Source: [https://www.youtube.com/watch?v=9UpwV1tdxcs](https://www.youtube.com/watch?v=9UpwV1tdxcs)

There is another beneficial side of this sad plight, caused by the point mutation, i.e., the heterozygotes (HbA HbS) are less likely to succumb to malaria disease. Malaria epidemics in regions where sickle cell anemia was common caused a large number of deaths. However, carriers of the disease always survived malaria even if they became infected. It was found by Haldane in 1994 that the mutation is not beneficial to those who have two copies of the mutated gene in their cells (the HbS homozygotes); they suffer greatly and often die before reaching reproductive age. But heterozygotes (HbA HbS) do receive a benefit. Malaria is caused by the protist *Plasmodium falciparum*, which enters human blood when the person is bitten by a mosquito (genus *Anopheles*). Though, the exact connection between sickle cell anemia and malaria is not very clear, but there are different theories (beyond the scope of this chapter) which have been proposed.

**Figure 1.** Cartoon illustration of how AS heterozygotes are relatively protected from severe P. falciparum malaria. The upper part of the cartoon is a schematic diagram of what happens in red cells in a normal (Hb AA) person with malaria: after invasion of a red cell by a merozoite, this becomes a ring form, and this starts multiplying (schizogony); when a schizont is mature the infected red cell essentially bursts and releases new merozoites, each one of which can invade a new red cell. The lower part of the cartoon is a schematic diagram of what happens in red cells in an AS heterozygote with malaria: the red cell, which appears normal at the time of invasion, once infected undergoes sickling (probably as a result of deoxygenation and lowering pH caused by the parasite), and thus it falls easy prey to macrophages in the spleen, in other organs and even in the peripheral blood[7]. Phagocytosis of a parasitized red cell clearly interrupts the schizogonic cycle and thus the parasitemia can be kept under control.

**Figure 26:** In the upper part of figure, the typical life cycle stages of *P. falciparum* in normal individual (AA or HbA HbA) are illustrated, wherein, a merozoite after invasion of a red cell becomes a ring form, which starts multiplying (schizogony); then a schizont is mature and the infected red cell essentially bursts and releases new merozoites, each one of which can invade a new red cell. The lower part of the figure shows the changes in red cells in a heterozygote individual (AS or HbA HbS) with malaria: once RBCs are infected, they undergo sickling and thus it falls easy prey to macrophages in the spleen,
in other organs and even in the peripheral blood. Phagocytosis of a parasitized red cell clearly interrupts the schizogonic cycle and thus the parasitaemia can be kept under control.


In a study published by Luzzatto in 2012, it is reported that the parasite lowers the pH of human red blood cells after spending part of its life cycle in the cells. This lowered pH as well as deoxygenation probably causes the cells to sickle. These deformed red blood cells are separated and phagocytized by immune system cells; thus the protist is destroyed along with the sickled cells. Although this does not provide complete protection from malaria, it does lessen the severity of the disease (Fig. 26).

Moreover, the sickled red blood cells live for only ten to 20 days, (much lower than the 120 day lifespan of normal, healthy red blood cells). As a result, the life cycle of the parasite is affected in the body, thus stopping it from replicating, which may give the immune system some time to attack and reduce malaria completely.

**Additional Information: Video**

**Malaria and Sickle Cell Anemia**

Watch this interesting video and learn why it is difficult for malarial parasite to survive in sickle cell

Source: https://www.youtube.com/embed/Zsbhl2nVNE

**Additional Information:**

**Discovery of Sickle cell anaemia**

Sickle cell disease or anemia is a hereditary genetic disease, characterized by the presence of crescent or sickle shaped erythrocytes. This disease is associated with an eventful history. Dr. James Herrick (Fig. 27), who gave the first description of sickle shaped red blood cells in 1910. He was working as cardiologist and professor of medicine at Chicago College of Dental Surgery. One of his intern, Ernest Edward Irons, observed peculiar elongated and “sickle-shaped” cells in the blood of Walter Clement Noel, a 20-year-old first-year dental student from a wealthy black family in Grenada. Noel was admitted to the Chicago Presbyterian Hospital in December 1904 with complaints of pain episodes and symptoms of anemia. Later, the disease was named "sickle-cell anemia" by Vernon Mason in 1922.

Before the advent of this first reported US case, the African medical literature had previously reported this condition in the 1870s as "ogbanjes", meaning “children who come and go” because of very high infant mortality rate caused by this condition.

The hereditary nature of the disease was discovered in 1949 by Dr. James V. Neel. The association of sickle cell disease with hemoglobin was discovered by Linus Pauling and Harvey Itano in 1951 and the actual amino acid substitution by Vernon Ingram in 1956.
Figure 27: Dr. James B. Herrick
Methods used in testing Sickle Cell Anemia
(A) RFLP analysis

The sequence of the normal HbA allele contains a 7 bp restriction recognition site of \textit{MstII} (CCTGAGG) and CvnI (CCTNAGG, where N is any nucleotide), which is altered in the HbS allele (CCTGTGG). Thus, there is single-nucleotide substitution (A to T) in the second position of the sixth codon of this gene (Fig. 27).

The sequence of normal HbA allele includes three \textit{MstII} sites, while the mutation that causes sickle cell anemia (HbS) eliminates a restriction site for the enzyme \textit{MstII} in the \(\beta\)-globin gene and thus only 2 \textit{MstII} sites are left (Fig. 28). Thus, if the total genomic DNA of the tested individual is digested with \textit{MstII}, then the expected RFLP pattern will be as follows:

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. of bands on gel</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hb(^A) Hb(^A)</td>
<td>2</td>
<td>Normal hemoglobin</td>
</tr>
<tr>
<td>Hb(^A) Hb(^S)</td>
<td>3</td>
<td>Sickle Cell Trait/Carrier</td>
</tr>
<tr>
<td>Hb(^S) Hb(^S)</td>
<td>1</td>
<td>Sickle Cell Disease</td>
</tr>
</tbody>
</table>

Figure 27: Single-nucleotide substitution (A to T) in the second position of the sixth codon of HBB gene.
The difference between the standard $BA$ allele and the sickle-cell $BS$ allele is a single-nucleotide substitution (AT) in the second position of the sixth codon of this gene. The sequence of the standard $BA$ allele (CCTGAGG) happens to correspond to an $MstII$ restriction site (CCTNAGG), which is altered in the $BS$ allele (CCTGTTG). The beta-globin gene region includes two flanking $MstII$ sites (red lines).

In the genetic test for the $BS$ allele, total DNA from the individual tested is digested with $MstII$ and run in a Southern Blot. The blot is hybridized with a probe specific for the beta-globingene. If the standard $BA$ allele is present, the probe sticks to the two small $MstII$ fragments and produces two smaller bands. If the sickle-cell $BS$ allele is present, the probe sticks to the single large fragment and produces one larger band. Thus, a standard AA homozygote shows the two-band pattern, a SS homozygote (with sickle-cell anemia) shows the one-band pattern, and an AS heterozygote (with sickle-cell trait) shows all three bands. The pattern of molecular expression is therefore described as co-dominant.

Important: this particular test depends on the coincidence that the nucleotide substitution responsible for the sickle-cell allele happens to occur in such a way as to create an RFLP: the
absence of the \textit{MstII} site does not itself cause sickle-cell anemia, but is instead a genetic marker for the allele.

\textbf{(B) Southern Blot Hybridization}

A similar band pattern, as found by RFLP can be obtained if total DNA from the individual being tested is digested with \textit{MstII} and run in a Southern Blot. The blot is hybridized with a probe specific for the beta-globin gene. If the standard HbA allele is present, the probe sticks to the two small \textit{MstII} fragments and produces two smaller bands. If the sickle-cell HbS allele is present, the probe sticks to the single large fragment and produces one larger band. Thus, a standard AA homozygote shows the two-band pattern, a SS homozygote (with sickle-cell anemia) shows the one-band pattern, and an AS heterozygote (with sickle-cell \textit{trait}) shows all three bands.

Thus, it can be concluded that \textit{MstII} is a good genetic marker for detecting the sickle cell allele.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{gel_patterns.png}
\caption{Gel patterns. The left panel is an electrophoretic gel stained with ethidium bromide. Total DNA has been extracted from nine bacterial DNA clones (lanes #1-9), digested with a particular restriction endonuclease, and separated by electrophoresis. Starting about one-third of the way from the top, a bright smear is present, indicating the presence of a random collection of restriction fragments of various sizes. There are no discrete bands, as no one DNA sequence is present in more than one copy. The lane marked (L) is a molecular weight standard (a "Ladder"), with a series of DNA fragments at 100bp intervals. A cloned DNA fragment has been loaded in lane #10: its mobility corresponding to the four rung of the ladder indicates a size of about 400bp. The middle panel is a Southern Blot autoradiogram of the \textit{same gel}. The DNA in the gel is transferred ("blotted") to a filter. The filter is then hybridized with a radioactively-labelled DNA(a "probe") made from the DNA loaded in lane #10. The filter is then exposed to X-ray film. Where the probe DNA finds a complementary sequence in the blot, it base-pairs ("sticks") to that DNA. The radioactivity then exposes the film and produces a dark band on the X-ray film. The right panel is a schematic representation of the autoradiogram: the information content is the presence or absence of bands, and the size of the fragments. The bands in the autoradiogram show that a DNA sequence homologous...}
\end{figure}
to the probe DNA is present in clones ##3, 4, & 8, with the expected size, and absent in clones ##1, 2, 5, & 9. [The probe sticks to itself in #10, as expected]. This analysis indicates that the gene of interest has been successfully cloned in the first set of plasmids, which can now be analyzed further.

**Figure 30. Analysis of DNA by the Southern Blot technique**

DNA is applied to an agarose gel, and electrophoresis separates the fragments of DNA according to size. The gel is then placed atop a thin sponge wick resting in a dish of salt solution, and a special filter (typically nitrocellulose) is placed on top of the gel. A stack of absorbent material (typically paper towels) is placed on top of this stack. The absorbent material draws the salt solution from the dish into the wick and through the gel by capillary action, which transfers the DNA fragments into the filter. The procedure is called a "Southern transfer" after the scientist Eric M. Southern who invented the procedure. The filter now contains the DNA fragments in the same pattern as the gel, but is more easily manipulated.

The filter is placed in a standard "seal-a-meal" bag, containing a solution of radioactively-labelled DNA probe [~~~~~] for a particular gene sequence. The probe binds to the filter only where a complementary DNA sequence is located. After washing to remove unbound probe, a piece of X-ray film is placed over the hybridized filter and exposed for several hours to several
days. The radioactive label produces a black band on the film where it has stuck to the complementary DNA, producing an autoradiogram. If a labelled size marker has been used, the exact sizes of the fragments can be determined.

The Southern Blot technique is useful for identifying a DNA sequence that appears only once or twice in the genome, the typical situations with nuclear loci. Note that in this example, the electrophoresis gel shows a continuous "smear" of DNA of all fragment sizes, whereas the autoradiogram identifies exactly three fragments, of three different sizes.

![Sequence from a normal individual](image1)

![Sequence from the diseased individual](image2)

**Figure 2:** Sequencing of the beta globin gene shows that the proband (the affected individual) in this case carries a homozygous A>T mutation at the codon 6; thus, changing the amino acid from glutamic acid to valine. The upper panel shows the sequence from a healthy individual and the lower panel is from the affected individual.

**Figure 31: Sequence from a normal and diseased individual**

(C) Polymerase chain reaction-oligonucleotide ligation assay (PCR-OLA)

As the name suggests, oligonucleotide ligation (OLA) means the ligation or joining of two oligonucleotides (~ 20-mers) by DNA ligase. These oligonucleotides are labelled and used as probes and their joining is dependent on three important criteria:

- hybridization to the target DNA,
- probes to be positioned directly next to one another in a 5′ to 3′ orientation, and
- complete base-pair complementarity with the target DNA at the site of their join.

In this method, the gene of interest (or disease carrying gene) is first amplified using PCR. This amplified product is then used as a starting template for OLA, where it is denatured to form a single stranded target DNA. Then, two adjacent oligonucleotide probes and DNA ligase
enzyme are added for each ligation reaction. The oligonucleotides for OLA are designed so that the juxtaposition of the adjacent ends occurs at a previously identified mutation site on the template (Fig. 31). For example, if the point mutation of A to T in HBB gene is at position 106. Then, **Probe 1** should be designed in such a way that its last base at 3′ end is complementary to the nucleotide at position 106 of the normal sequence (i.e the probe sequence ending with A nucleotide and binding with complementary T nucleotide in a normal sequence). On the other hand, **Probe 2** should be designed in a way, that it starts at its 5′ end with a nucleotide that is complementary to the nucleotide immediately adjacent to position 106. The probes are subsequently hybridized to the target DNA in a head-to-tail (5′ to 3′) orientation. If the probes and the target DNA are fully complementary to each other, then the DNA ligase enzyme forms a phosphodiester bond between the probes. If there is even a single base pair mismatch on either side, or close to the nick between the probes, the efficiency of the enzyme is significantly decreased and the ligation is prevented; therefore it can be distinguished from a perfect match.

For example, **when Probe 1 and Probe 2 are hybridized** to mutant DNA in which the nucleotide at position 106 is altered, the nucleotide at the 3′ end of Probe 1 is mismatched and is not able to pair with nucleotide 106 in the target DNA sequence; while Probe 2, is perfectly aligned. In this case, DNA ligase cannot join Probe 1 and Probe 2 because of the single-nucleotide misalignment. With such two kinds of probes, it is possible to ascertain the genetic makeup of any tested individual at a particular site. For example, heterozygous individuals. To ascertain whether ligation has occurred between two labelled probes, an antigen-antibody reaction coupled with a colorimetric reaction is used.

For example, **Probe 1 is labelled at its 5′ end** with biotin and **Probe 2 is labelled at its 3′ end** with digoxigenin. DNA is denatured to release the hybridized probes after the hybridization and ligation steps are carried out, and the entire mixture is transferred to a small plastic well that has been coated with streptavidin. The well is then washed to remove unbound material, so only the biotin-labelled probe DNA remains bound.

If **anti-digoxigenin antibodies**, which have been previously coupled to alkaline phosphatase, are added to the well. After the unbound conjugated anti-digoxigenin antibodies are removed with an additional washing step, a colourless chromogenic substrate is added. The appearance of color in the well indicates that anti-digoxigenin antibodies have bound to digoxigenin and that the digoxigenin-labelled probe was ligated to the biotinylated probe. If no colour appears, then no ligation occurred. Thus, the PCR/OLA system is rapid, sensitive, and highly specific.
It has even been automated with a robotic workstation to carry out the steps of the assay procedures.
yield positive results with both pairs of probes.

Figure 31: Procedure of Polymerase chain reaction-oligonucleotide ligation assay (PCROLA)
Summary

- The field of medical diagnostics has been revolutionized with the application of molecular tools of recombinant DNA technology.
- Many lives have been saved with timely and accurate molecular diagnosis of many life threatening genetic diseases.
- Earlier, prenatal diagnosis was limited for only those diseases for which the biochemical effects were known.
- Additionally, such disorders could only be diagnosed if they were expressed in fetal tissues like amniotic fluid cells, chorionic villi, blood, skin, liver.
- Thus, due to such clinical limitations, genetic diseases like cystic fibrosis (CF) and Duchenne muscular dystrophy (DMD) could not be detected prenatally. But, with the use of molecular diagnostics, almost all the mendelian genetic disorders are responsive to prenatal diagnosis.
- The reason of such breakthrough is quite clear: every cell of human body contains the genetic makeup in the form of DNA, that is packaged in chromosome pairs, which further comprises of various genes. Thus, the biochemical basis of any disorder in general and gene in specific need not to be known for its diagnosis. For example, although genes encoding hemoglobin is expressed only by red blood cells, the DNA coding for hemoglobin is present in all cells, including amniotic fluid cells and chorionic villi.
- DNA-based analysis is based on the fact that there are practically indefinite numbers of nucleotide-sequence differences in the DNA of different individuals. These differences can be detected by restriction fragment length polymorphisms (RFLPs). For example, in cases where the disease-causing mutation is known, RFLP analysis may be used (e.g., sickle cell anemia).
- Additionally, techniques like Southern blot hybridization or dot blot and amplification of DNA using the polymerase chain reaction (PCR) can also be used as powerful tools for the diagnosis of various genetic disorders.
- Efficient recombinant DNA techniques have opened new avenues in the field of clinical diagnostics and medical genetics. These offer an obvious advantage, i.e., it may help in giving advice to families not to have children with harmful diseases or help identify people who are at a higher risk for genetic conditions that may be preventable.
But these genetic tests should be used carefully and interpretation of results should be done accurately.

Moreover, a single laboratory test should not be used for the presence or absence of a mutant or disease carrying gene, instead, additional tests should be done to supplement the results. But, there are also some limitations to these DNA based tests, like, genetic testing can provide only limited information about an inherited condition.

The test often can’t determine if a person will show symptoms of a disorder, how severe the symptoms will be, or whether the disorder will progress over time. Moreover, there is a possibility for laboratory errors, which might be due to sample misidentification, contamination of the chemicals used for testing, or other factors.

Then, there is also the possibility of some sort of negative stigma attached to having a carrier status for a particular disease.

Sometimes, if the genetic cause of a disease is identified in a patient, it does not necessarily guarantee that there is a cure for the disease. A treatment or therapy may not yet have been developed for that particular disease.

Thus, though the technology of molecular diagnostics is indeed a powerful technique, but it should be used with caution and care.

**Important definitions**

**Alleles:** A pair of contrasting characters, each of two or more alternative forms of a gene that arise by mutation and are found at the same place on a chromosome.

**Amplification:** It is a process in which the polymerase chain reaction (PCR) is repeated for a number of cycles with changing temperatures to exponentially increase the number of copies of a specific target region of a gene.

**Autosomal dominant inheritance:** The gene in question (associated with some phenotype or disease carrying mutation) is located on one of the autosomes, or non-sex chromosomes and a single copy of the gene is enough to cause the disease.

**Autosomal recessive inheritance:** The gene in question (associated with some phenotype or disease carrying mutation) is located on one of the autosomes (chromosome pairs 1 to 22) and two copies of the gene are necessary to have the trait, one inherited from the mother, and one from the father.

**Carrier:** It is an individual who is capable of passing on a genetic mutation associated with a disease but may or may not display disease symptoms. Carriers are associated with diseases inherited as recessive traits.

**Chorea:** It comes from the Greek word 'choreia, meaning dancing. It is the most visible feature of Huntington’s disease, characterized by involuntary, rapid and jerky body movements.
Cystic fibrosis transmembrane conductance regulator (CFTR): It forms a Cl channel in many organs, including the intestines, pancreas, lungs, sweat glands, and kidneys and controls the regulation of other transport systems. Mutations in the CFTR gene cause cystic fibrosis.

Epigenetics: Epigenetics is the study of heritable phenotype changes that do not involve alterations in the DNA sequence. The Greek prefix epi- in epigenetics implies features that are "on top of" or "in addition to" the traditional genetic basis for inheritance.

Fluorescence in situ hybridization (FISH): It is a cytogenetic technique used to visualize specific segments on chromosomes with the help of fluorescent probes. It is used to diagnose specific chromosomal and genetic abnormalities.

Genetic disorder: A genetic disorder is a disease that is caused by an abnormality in an individual's genetic material like DNA. Abnormalities can be mutation in one or more genes, or they can involve the addition or subtraction of entire chromosomes.

Huntingtin: It is a disease causing gene, abbreviated as HTT, which is linked to Huntington's disease, a neurodegenerative disorder. The HTT gene provides instructions for making a protein called huntingtin (originally known as IT-15).

Marker: A marker is a piece of DNA that lies near a gene and is usually inherited with it. Helpful in identifying disease.

Molecular Diagnostics: diagnostic tests that identify a disease, susceptibility for a disease, or progress in treating a disease by detecting specific molecules such as DNA, antibodies, and proteins.

Oligonucleotide ligation assay (OLA): It refers to the ligation or joining of two oligonucleotides (~ 20-mers) by DNA ligase and its detection using a colorimetric reaction.

Pathogenesis: It is a term which describes the mechanism by which a disease manifests in an individual, describing its origin and development.

Penetrance: It is the frequency of expression of an allele in a population of individuals. For example, if 8/10 of individuals carrying an allele express the trait, the trait is said to be 80% penetrant.

Point mutation: It is a mutation involving the substitution, addition, or deletion of a single nucleotide base in a gene.

Polar Zipper: It is a motif present in proteins that consists of repeats of polar amino acid residues like glutamine, glutamate, lysine or histidine or even Asp-Arg or Glu-Arg repeats, arranged like a zipper. These residues promote binding between protein subunits or between proteins by hydrogen bonds.

Restriction fragment length polymorphism (RFLP): It is a molecular technique which is used to trace out the difference in homologous DNA sequences by the presence of fragments of different lengths after digestion of the DNA samples in question with specific restriction endonucleases. RFLPs can be used to trace inheritance patterns and diagnose diseases.

Sequence-specific oligonucleotide probe (SSO): It is a short piece of synthetic DNA which is complementary to the sequence of a target DNA. It is commonly used in Dot blot assay.

Characteristics of a Detection System A good detection system should have 3 qualities:

- Sensitivity
- Specificity
- Simplicity

- Sensitivity means that the test must be able to detect very small amounts of target even in the presence of other molecules.
B.Sc. (Hons) Zoology
Biotechnology
Sem VI

• Specificity: the test yields a positive result for the target molecule only.
• Simplicity: the test must be able to run efficiently and inexpensively on a routine basis.
B.Sc. (Hons) Zoology
Biotechnology
Sem VI

Exercises
Q1. Explain molecular diagnosis.
Q2. What is a diagnostic biomarker?
Q3. What is CFTR channel? How is it related to the genetic disease of cystic fibrosis?
Q4. What is ΔF508 and G551D mutation? How does it lead to cystic fibrosis?
Q5. Explain the diagnostics of mutated CFTR gene using dot blot hybridization.
Q6. Explain the pathogenesis of a triplet genetic disorder.
Q7. What are the causes of chorea?
Q8. Explain the genetic basis of Huntington’s disease.
Q9. Explain some of the molecular tools available to detect ‘polyQ’ disease.
Q10. Differentiate between sickle cell trait and sickle cell disease.
Q11. How does a point mutation in HBB gene leads to sickle cell anaemia?
Q12. Explain the detection of sickle cell disease using RFLP analysis.
Q13. Explain the principle of PCR-Oligonucleotide Ligation Assay and discuss its use in diagnosis of sickle cell disease.
Q14. What is the difference between HbA and HbS?
Q15. Explain the structure of CFTR and HTT gene.
Q16. How PCR can be used to detect CAG repeats in HD? Discuss its limitations.
Q17. What are SSOs? Explain their use in diagnosing cystic fibrosis disease.
Q18. What are the advantages of DNA based genetic tests over the traditional methods of diagnosing diseases?
Q19. Enumerate the limitations of molecular diagnostics.
Q20. Which of the following is not part of the triad of symptoms associated with Huntington’s disease?
   a) Chorea
   b) Weight gain
   c) Psychiatric and behavioural changes
   d) Cognitive decline
Q21. What is one of the most noticeable symptoms of CF?
   a) Increased energy
   b) Weight gain
   c) Thick sticky mucus
   d) Anaemia
Q22. A person carrying just one abnormal Sickle Cell gene is said to have, which of the following?
   a) Sickle Cell Trait
   b) Anaemia
   c) Cancer
   d) Asthma
Q23. A married couple plans their family, but both are carriers of mutant CFTR gene. They are referred for genetic counselling. It would be correct to tell them that:
   a) The probability that their child will be unaffected is 1 in 4.
   b) The probability that their children will be the carriers of disease is 2 in 4
c) The probability that their child will be affected is 1 in 4.
d) None of the above

Q24. Expand the following abbreviations:
   i. HTT
   ii. HBB
   iii. CFTR
   iv. PolyQ
   v. PCR-OLA
   vi. (CAG)n
   vii. SSO

Q25. Define the following terms:
   1. Chorea
   2. Huntingtin
   3. Autosomal Dominant Inheritance
   4. Point mutation
   5. Biomarkers

References


Links.
3. https://www.stanford.edu/group/hopes/cgi-bin/wordpress/?p=3581
5. https://evolution.berkeley.edu/evolibrary/article/0_0_0/mutations_06
Acknowledgement: IKS is thankful to Dr Anita K Verma for the sharing the required information and study materials.