The Polymerase Chain Reaction (PCR) provides an extremely sensitive means of amplifying relatively large quantities of DNA.

First described in 1985, Nobel Prize for Kary Mullis in 1993.

The technique was made possible by the discovery of Taq polymerase, the DNA polymerase that is used by the bacterium Thermus aquaticus that was discovered in hot springs.

The primary materials, or reagents, used in PCR are:
- DNA nucleotides, the building blocks for the new DNA
- Template DNA, the DNA sequence that you want to amplify
- Primers, single-stranded DNAs between 20 and 50 nucleotides long (oligonucleotides) that are complementary to a short region on either side of the template DNA
- DNA polymerase, a heat stable enzyme that drives, or catalyzes, the synthesis of new DNA.

PCR

The cycling reactions:
There are three major steps in a PCR, which are repeated for 20 to 40 cycles. This is done on an automated Thermo Cycler, which can heat and cool the reaction tubes in a very short time.

Denaturation at around 94°C:
During the denaturation, the double strand melts open to single stranded DNA, all enzymatic reactions stop (for example the extension from a previous cycle).

Annealing at around 54°C:
Hydrogen bonds are constantly formed and broken between the single stranded primer and the single stranded template. If the primers exactly fit the template, the hydrogen bonds are so strong that the primer stays attached

Extension at around 72°C:
The bases (complementary to the template) are coupled to the primer on the 3’ side (the polymerase adds dNTPs from 5’ to 3’, reading the template from 3’ to 5’ side, bases are added complementary to the template).
**PCR**

The different steps of PCR:

- **Step 1**: denaturation
  - 1 minute 94 °C
- **Step 2**: annealing
  - 45 seconds 54 °C
- **Step 3**: extension
  - 2 minutes 72 °C
  - Only dNTPs

**PCR**: Polymerase Chain Reaction

30-40 cycles of 3 steps:

1. **Denaturation**: 1 minute at 94 °C
2. **Annealing**: 45 seconds at 54 °C
3. **Extension**: 2 minutes at 72 °C

**Exponential increase of the number of copies during PCR**

- **1st cycle**: 2 copies
- **2nd cycle**: 4 copies
- **3rd cycle**: 8 copies
- **4th cycle**: 16 copies
- **35th cycle**: $2^{35} = 34$ billion copies

Exponential amplification

(template DNA)
Every cycle results in a doubling of the number of strands DNA present.

After the first few cycles, most of the product DNA strands made are the same length as the distance between the primers.

The result is a dramatic amplification of the DNA that exists between the primers. The amount of amplification is 2 raised to the n power; n represents the number of cycles that are performed. After 20 cycles, this would give approximately 1 million fold amplification. After 40 cycles the amplification would be $1 \times 10^{12}$. 

**Verification of PCR product**

![Verification of PCR product on agarose or Sepharose gel](image)
**PCR and Contamination**

The most important consideration in PCR is contamination

Even the smallest contamination with DNA could affect amplification

For example, if a technician in a crime lab set up a test reaction (with blood from the crime scene) after setting up a positive control reaction (with blood from the suspect) cross contamination between the samples could result in an erroneous incrimination, even if the technician changed pipette tips between samples. A few blood cells could volitilize in the pipette, stick to the plastic of the pipette, and then get ejected into the test sample

Modern labs take account of this fact and devote tremendous effort to avoiding cross-contamination

---

**Optimizing PCR protocols**

**PCR can be very tricky**

While PCR is a very powerful technique, often enough it is not possible to achieve optimum results without optimizing the protocol

Critical PCR parameters:

- Concentration of DNA template, nucleotides, divalent cations (especially Mg$^{2+}$) and polymerase
- Error rate of the polymerase (Taq, Vent exo, Pfu)
- Primer design
Primer design

General notes on primer design in PCR

Perhaps the most critical parameter for successful PCR is the design of primers.

Primer selection
Critical variables are:
- primer length
- melting temperature (Tm)
- specificity
- complementary primer sequences
- G/C content
- 3’-end sequence

Primer length
- specificity and the temperature of annealing are at least partly dependent on primer length
- oligonucleotides between 20 and 30 (50) bases are highly sequence specific
- primer length is proportional to annealing efficiency; in general, the longer the primer, the more inefficient the annealing
- the primers should not be too short as specificity decreases

Primer design

Specificity

Primer specificity is at least partly dependent on primer length: there are many more unique 24 base oligos than there are 15 base pair oligos.

Probability that a sequence of length n will occur randomly in a sequence of length m is:

Example: the mtDNA genome has about 20,000 bases, the probability of randomly finding sequences of length n is:

<table>
<thead>
<tr>
<th>n</th>
<th>P_n</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>19.52</td>
</tr>
<tr>
<td>10</td>
<td>$1.91 \times 10^2$</td>
</tr>
<tr>
<td>15</td>
<td>$1.86 \times 10^3$</td>
</tr>
</tbody>
</table>
**Primer design**

**Complementary primer sequences**
- primers need to be designed with absolutely no intra-primer homology beyond 3 base pairs. If a primer has such a region of self-homology, "snap back" can occur
- another related danger is inter-primer homology: partial homology in the middle regions of two primers can interfere with hybridization. If the homology should occur at the 3' end of either primer, primer dimer formation will occur.

**G/C content**
- ideally a primer should have a near random mix of nucleotides, a 50% GC content
- there should be no PolyG or PolyC stretches that can promote non-specific annealing

**3' end sequence**
- the 3' terminal position in PCR primers is essential for the control of mis-priming
- inclusion of a G or C residue at the 3' end of primers helps to ensure correct binding (stronger hydrogen bonding of G/C residues)

**Melting temperature (T_m)**
- the goal should be to design a primer with an annealing temperature of at least 50° C
- the relationship between annealing temperature and melting temperature is one of the "Black Boxes" of PCR.
- a general rule-of-thumb is to use an annealing temperature that is 5°C lower than the melting temperature.
- the melting temperatures of oligos are most accurately calculated using nearest neighbor thermodynamic calculations with the formula:
  \[ T_m = H \left[ S + R \ln \left( \frac{c}{4} \right) \right] - 273.15°C + 16.6 \log_{10} [K^+] \]
  (\(H\) is the enthalpy, \(S\) is the entropy for helix formation, \(R\) is the molar gas constant and \(c\) is the concentration of primer)
- a good working approximation of this value can be calculated using the Wallace formula:
  \[ T_m = 4x(\#C+\#G) + 2x(\#A+\#T)° C \]
- both of the primers should be designed such that they have similar melting temperatures. If primers are mismatched in terms of Tm, amplification will be less efficient or may not work; the primer with the higher Tm will mis-prime at lower temperatures; the primer with the lower Tm may not work at higher temperatures.
DNA Sequencing

Sequencing methods

- The process of determining the order of the nucleotide bases along a DNA strand is called **DNA sequencing**

- In 1977 two separate methods for sequencing DNA were developed: the **chain termination method** or **cycle sequencing** (Sanger et al.) and the **chemical degradation** method or **Maxam-Gilbert sequencing** (Maxam and Gilbert)

- Both methods were equally popular to begin with, but, for many reasons, the cycle sequencing method is the method more commonly used today

- This method is based on the principle that single-stranded DNA molecules that differ in length by just a single nucleotide can be separated from one another using polyacrylamide gel electrophoresis

---

**Cycle Sequencing**

*Concept: If we know the distance of each type of base from a known origin, then it is possible to deduce the sequence of the DNA.*

*For example, if we knew that there was an:*

- A at positions 2, 3, 11, 13 ... and
- G at positions 1, 12, ... and
- C at positions 6, 7, 8, 10, 15... and
- T at positions 4, 5, 9, 14...*

*then we can reconstruct the sequence*

*Obtaining this information is conceptually quite simple. The idea is to cause a termination of a growing DNA chain at a known base (A, G, C or T) and at a known location in the DNA.*

*In practice, chain termination is caused by the inclusion of a small amount of a single dideoxynucleotide base in the mixture of all four normal bases (e.g. dATP, dTTP, dCTP, dGTP and ddATP). The small amount of ddATP would cause chain termination whenever it would be incorporated into the DNA. The incorporation of ddATP would be random and thus all possible chains that end in 'A' will exist.*
DNA Sequencing

The different steps in cycle sequencing

**Sequencing**

30 cycles of 3 steps:

1. **Denaturation**
   - 1 minute 94°C

2. **Annealing**
   - 15 seconds 50°C
   - 1 primer !!!

3. **Extension**
   - 4 minutes 60°C
   - mixture of 4dNTP's and dNTP's

DNA Sequencing

Chain termination

For example, if the primer/hexanucleotide is: AGAA

The products will be: "AGAA...AAAGAAA...AAGAG...AGA...GAGAGA..."
The separation of the sequencing fragments can be measured by gel electrophoresis. Each reaction would be loaded into a separate well on a gel, and the fragments would be separated based on size.

**DNA Sequencing**

**Gel electrophoresis**

To measure the sizes of the fragments, each of the four reactions would be loaded into a separate well on a gel, and the fragments would be separated by gel electrophoresis.

**DNA Sequencing**

**Sequencing systems**

**Long Read DNA Sequencing**

**LICOR DNA 4300**
DNA Sequencing

Snapshots of the detection of the fragments on the sequencer
DNA Sequencing

The linear amplification of the gene in sequencing

1st cycle
6 template strands

2nd cycle
12 complementary strands

Linear amplification

PCR product
30 cycles - 189 complementary strands

- Primer fits only on one strand
- On incorporation of a fluorescently labelled dNTP (complementary with the base on the template) the chromosome stops

(After) Saiki et al 1985