I. PCR

The **polymerase chain reaction** is a method of *in vitro* cloning, based on:

1) DNA polymerase requires a primer to initiate DNA synthesis
2) DNA denatures into single strands at high temperatures
3) An 18–24 nt sequence should be unique in the genome
4) H-bonding will stabilize 18–24 bp at 50–70 C

Components of a PCR reaction:

template: genomic DNA, plasmid, bloodstain, single cell, etc.
primers: two synthetic **oligonucleotides**, usually 18-24 nt long, complementary to opposite strands of the target sequence in the template:

```
5'-GGGCGATATGGTCCAGTCACCCAAGGTCAGTCCAACGGACTTAGTATTGCACCGAC-3'
3'-CCCGCTATACCAGGTCAGTGGGTTCCAGTCAGGTTGCCTGAATCATAACGTGGCTG-5'
```

primer 1 = 5'-GGGCGATATGGTCCAGTCA-3'
primer 2 = 5'-GTCGGTGCAATACTAAGTC-3'

dNTP mixture
DNA polymerase: DNA polymerase from *Thermophilus aquaticus* (*Taq* polymerase) is used because it is stable at 95 C
buffer: appropriate conditions for polymerase activity (Mg^{++}, correct pH, etc.)

Steps in a PCR cycle

1) **Denature:** template is heated to 95 C to denature it into singe strands
2) **Anneal:** cool template/primer mix to appropriate temperature (50-70 C) so that primers can anneal (hybridize) with template
3) **Extend:** shift to temp at which DNA polymerase functions best (72 C for *Taq*)

Each set of three steps is one cycle; each cycle doubles the amount of DNA product.
Generally a PCR run consists of 15-35 cycles.

Once the components are mixed together, you need only shift the temperature between denaturing, annealing, and extending temps. This is done in a thermal cycler (PCR machine).

Strengths of PCR

sensitivity: a sequence from a single DNA molecule can be amplified a million-fold
speed: a typical PCR run can be completed in less than two hours (cloning into a vector requires several days)
flexibility: many different applications

Limitations of PCR

sensitivity: it is easy to contaminate PCR reactions with extraneous DNA
size: upper limit of about 25 kb, but works best up to a few kb
information: usually, some sequence information is necessary to design primers

II. Applications of PCR

A. Diagnosis. The standard HIV test looks for antibodies to the virus, which can take many months to appear. PCR can detect minute quantities of the virus in the blood.

B. **Site-directed mutagenesis**. By altering the primer sequences, we can generate PCR products that are mutated in defined ways. Useful in analysis of gene and protein function.

C. **DNA fingerprinting** is used in forensics, determining parentage, etc.

VNTRs (variable number tandem repeats) or **SSLPs** (simple-sequence length polymorphisms) occur at microsatellites when the number of repeats varies within the population (*i.e.*, multiple alleles that differ in DNA length)

PCR with primers that flank the repeats will amplify a product whose size depends on the number of repeats. Each VNTR may have many alleles, and there are many different VNTRs in the genome.

D. **Degenerate PCR**. It is possible to clone a gene when the sequence of the polypeptide is known (or inferred), by designing primers based on the amino acid sequence.

```
Ser-Ala-Asp-Thr-Met-Trp-Pro-Leu
TCA GCA GAC ACA ATG TGG CCA CTA
TCC GCC GAT ACC         CCC CTC
TCG GCG     ACG         CCG CTG
TCT GCT     ACT         CCT CTT
AGC                     TTA
AGT                     TTG
(6) (4) (2) (4) (1) (1) (4) (6) = 4608 possibilities
    GCN GAY CAN ATG TGG CC       = 32-fold degenerate
```

E. Amplification of random segments. Using short or highly degenerate primers, one can amplify random segments of DNA without knowing the sequence. Useful in cloning Neanderthals and dinosaurs for Jurassic park.

III. DNA sequencing

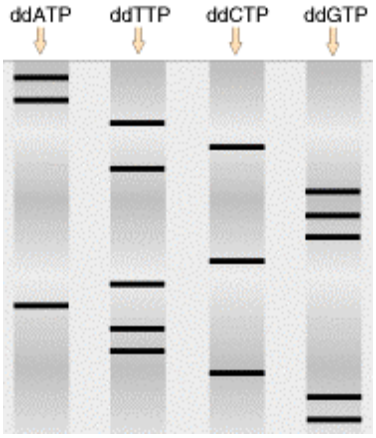
dideoxy chain-terminating method (Sanger):

```
5'-TGCGGGCTTATCGGGTCTAA-3'
3'-ACGCCCGAATAGCCCAGATT-5'
```

Design a primer for DNA polymerase (5'-TGCG-3'), denature DNA and anneal primer:

```
5'-TGCG-3'
3'-ACGCCCGAATAGCCCAGATT-5'
```

Add DNA polymerase, buffer, dNTPs, plus a small amount of **dideoxy-ATP** (A*); at every position of T in the template, either A or A* will be put into the growing chain – A* lacks the 3'-OH required to make the phosphodiester bond, so each time A* is added, the chain stops.



5'-TGCGGGCTTA*-3' 10 nt
5'-TGCGGGCTTATCGGGTCTA*-3' 19 nt
5'-TGCGGGCTTATCGGGTCTAA*-3' 20 nt

Separate these fragments by size on a **polyacrylamide** gel (better resolution of small fragments than agarose), under **denaturing** conditions (so the template and new strand run separately. The DNA fragments make a "ladder" of bands, each telling us the position of an A in the strand being synthesized (lane on left).

Repeat with C*, G*, and T*.

The sequence of the strand the primer is on (the strand that is complementary to the template) is read from bottom to top of gel.

The DNA polymerase cannot have 3' → 5' exonuclease (proofreading), because that can remove the dideoxy nucleotides.

From one set of sequencing reactions, one can typically read sequence beginning about 5-30 nt from the primer, and ending about 600 nt away.

IV. Sequencing technologies and strategies

dye terminators: instead of radioactive dNTPs, use ddNTPs with fluorescent tags, a different color in each dideoxy reaction. Then all four reactions can be run on a single lane, with the colors read by a laser as each band runs off the bottom of the gel.

automated sequencers use cycle sequencing (like PCR; no amplification because there is only one primer, but each template molecule is used multiple times) and dye terminators; the results are read and analyzed by a computer and analyzed

If we don't know the sequence, how can we design a primer??? We use a primer to a sequence in the vector (*i.e.* in the polylinker).

Usually we want to sequence a segment of DNA longer than 600 bp. We have three choices:

a. "subclones" of different fragments and sequence the ends of each, using vector primers
b. sequence once for 600 bp, then design and synthesize a new primer to the end of the that sequence; repeat until the end is reached
c. **shotgun sequencing**: break the DNA into random small fragments and clone (essentially this is making a library of the DNA to be sequence, which might be a BAC or cosmid clone, or perhaps and entire genome); determine the sequence of randomly selected fragments, until about 5X coverage (*e.g.*, for a 10 kb segment, make 500 bp clones, sequence 100 randomly selected clones); use a computer to assemble the fragments back into a single sequence.