Biocomputing

Elements of Biotechnology

Content

- Introduce the methods used in the lab and in the cell itself to manipulate DNA
  - The availability of such tools made the rapid development of the field possible
  - It is not a lab manual, only present a number of available tools and the ideas supporting them
  - Essential to know about elementary lab tools also for mathematicians and computer scientists
- Topics: measuring DNA, testing for/isolating DNA molecules, separating and fusing DNA strands, lengthening, shortening, cutting, linking DNA, modifying nucleotides, multiplying, sequencing

Measuring the length of DNA molecules

- **Problem**: measure the length of a molecule without sequencing it
- Length – definition:
  - Single stranded DNA: the number of nucleotides (ex: 12 mer)
  - Double stranded DNA: the number of base pairs (ex: 12 bp)

Measuring the length of DNA

- **Central fact**: DNA molecules are negatively charged: in electric fields, they migrate towards + electrods
- The force needed to move DNA molecules – proportional to its length (due to friction)
- Idea: gel electrophoresis technique

The gel electrophoresis technique

- Pour some gel in a rectangular container with electrodes on the sides
- In the cooling process: a comb is inserted on the negative side
- Gel cooled down, comb removed: row of small wells

Gel electrophoresis

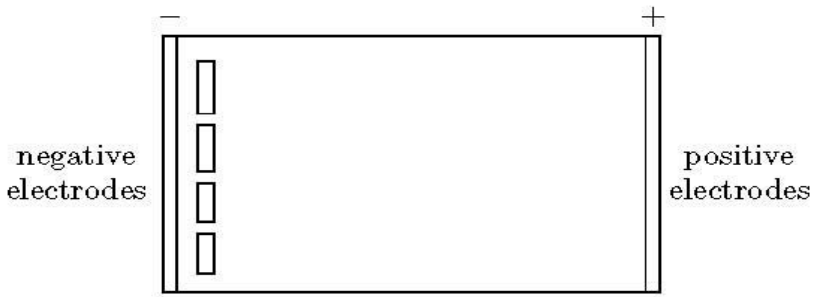

Figure 1.8: Gel prepared for electrophoresis

The gel electrophoresis technique

- The solution with the DNA to be measured brought in the wells
- Activate the electric field: DNA molecules travel towards +
- Gel acts as a molecular sieve – *big friction*
  – *small molecules move faster than big molecules*

The gel electrophoresis technique

- Deactivate the field when the first molecules have reached +
- Small molecules – longer distance than bigger molecules
- Molecules of same length – same distance

Gel electrophoresis

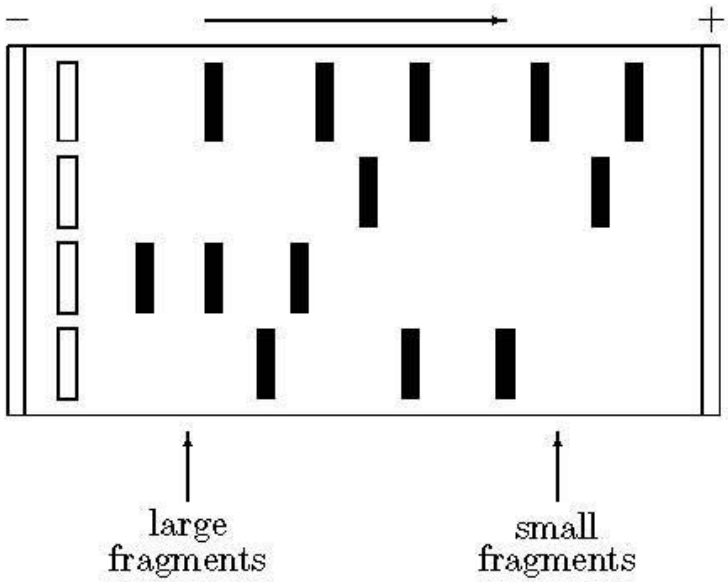

Figure 1.9: Gel electrophoresis

The gel electrophoresis technique

- Reading the resulting gel
- DNA molecules are colorless: mark them
- Two techniques for marking:
  - Staining with ethridium bromide → fluorescent mark under ultraviolet light
  - Attaching radioactive markers to the ends of DNA molecules → expose a film

Computing the length

- Based on the distance traveled

- Using a calibrating solution in one of the wells → compare the location of bands on the other paths with the calibrating path

Testing for known molecules

- **Problem**: test if some known single stranded molecules **s** exist in a given DNA solution; variant: isolate some of them if so
- **Solution**: Attach the complementary strands (*probes*) to a filter and pour the solution through the filter
- **s** molecules bind to their complements to form double strands (*annealing*) and stay on the filter

Testing for/isolating known molecules

- Other methods:
  - Attach probes to tiny glass beads, place them in a glass column and pour the solution over them
  - Attach probes to tiny magnetic beads, throw them in the solution and stir; attract them to one side using a magnet

Operations on DNA molecules

- Separating and fusing DNA strands
- Lengthening DNA
- Shortening DNA
- Cutting DNA
- Linking DNA
- Modifying nucleotides
- Multiplying DNA (cloning, PCR)
- Sequencing

Separating and fusing DNA strands

- Hydrogen bond between complementary bases is weaker than the covalent bond between consecutive nucleotides within one strand
- One can separate the two strands of a DNA molecule without breaking the single strands
- Separating the two strands: *denaturation*
- Fusing two complementary strands: *annealing/renaturation/hybridization*

Denaturation and annealing

- *Denaturation*: heat the DNA solution until DNA melts
  - Melting temperature: 85 – 95 C
  - Use of chemicals (e.g., *formamide*) lowers the melting temperature
  - Melting temperature of DNA = the temperature at which half of the molecules separate
- *Annealing*: solution cooled down (slowly), the separate strands fuse again by hydrogen bonds (also called *renaturation* or *hybridization*)

Enzymes

- Enzymes: proteins that catalyze biochemical reactions
  - Very specific: most catalyze just a single chemical reaction
  - Highly efficient: speed up the chemical reaction by as much as a *trillion* times
- If no enzymes existed, bio-reactions would most likely be too slow to support life

Lengthening DNA

- The enzymes: *polymerases* - add nucleotides to an existing DNA molecule
- Requirements:
  - Existing single stranded template describing the chain of nucleotides to be added
  - Existing sequence (primer) bonded to a part of the template

Lengthening DNA

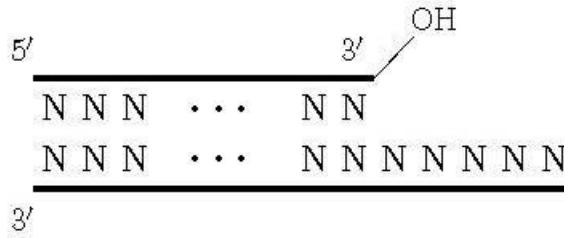

Figure 1.10: A DNA molecule w
an incomplete upper strand

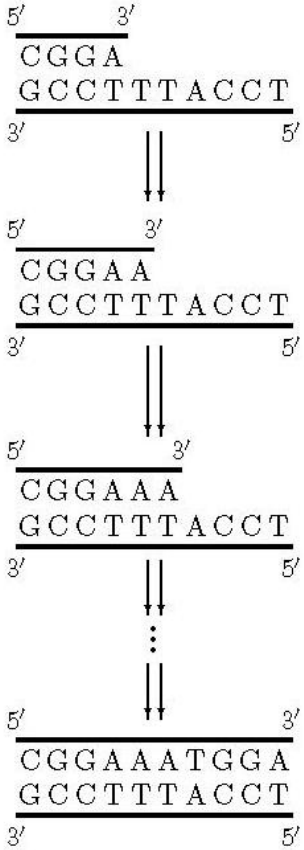

Figure 1.11: A polymerase in action

Lengthening DNA - polymerases

- Polymerases only extend the DNA molecules in the 5'-3' direction
- They extend repeatedly the 3' end, provided that the required nucleotides are available in the vicinity
- Given one strand, to produce the corresponding double stranded molecule, one needs the primer and the polymerases

Lengthening DNA

- Exceptions: *terminal transferase* does not need a template: extends double strands at both ends by some single stranded tails

Lengthening DNA

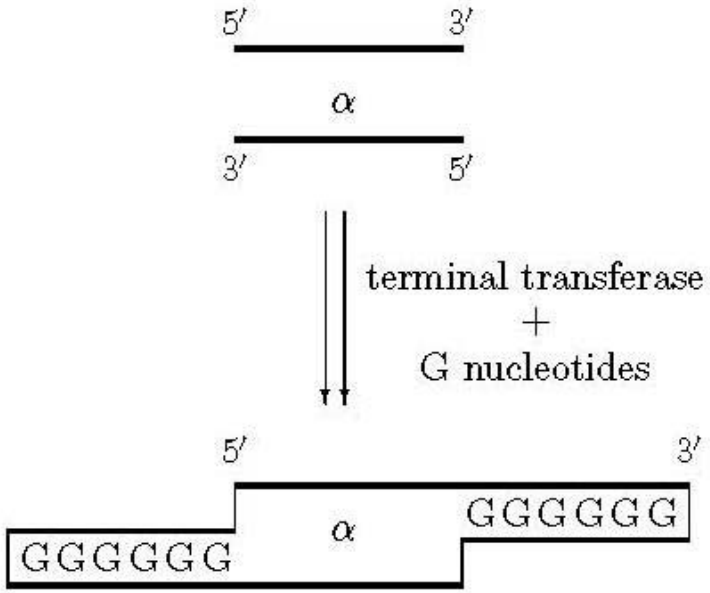

Figure 1.12: Transferase activity

Lengthening DNA

- Exceptions: *terminal transferase* does not need a template: extends double strands at both ends by some single stranded tails
- Engineering single strands is possible: a procedure leading to automation exists → "synthesizing robots"
  - Short such single strands are called *oligonucleotides* (*oligos* in short); they are essential in PCR

Shortening DNA

- The enzymes: *DNA nucleases* – degrade DNA
- Two types of enzymes:
  - *DNA exonucleases*: cleave (remove) one nucleotide at a time from the ends of the DNA
  - *DNA endonucleases*: destroy internal bonds in the DNA molecule (cutting DNA)

Shortening DNA

- *Exonucleases*:
  - some remove nucleotides from 5', others from 3'
  - some degrade single strands, others on double strands, other operate on both
- Examples:
  - Polymerases can act as exonucleases: error correction in DNA replication
  - Exonuclease III: 3'-nuclease
  - Bal 31

Exonucleases

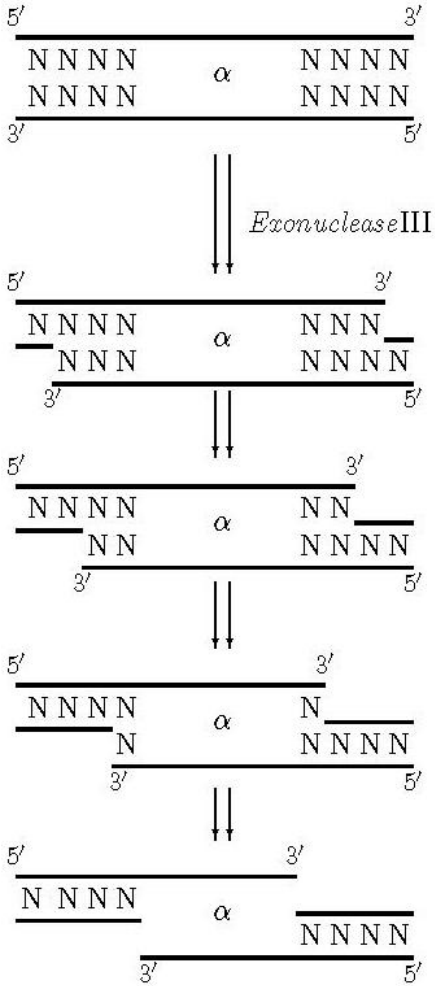

Figure 1.13: Exonuclease III in action

Exonucleases

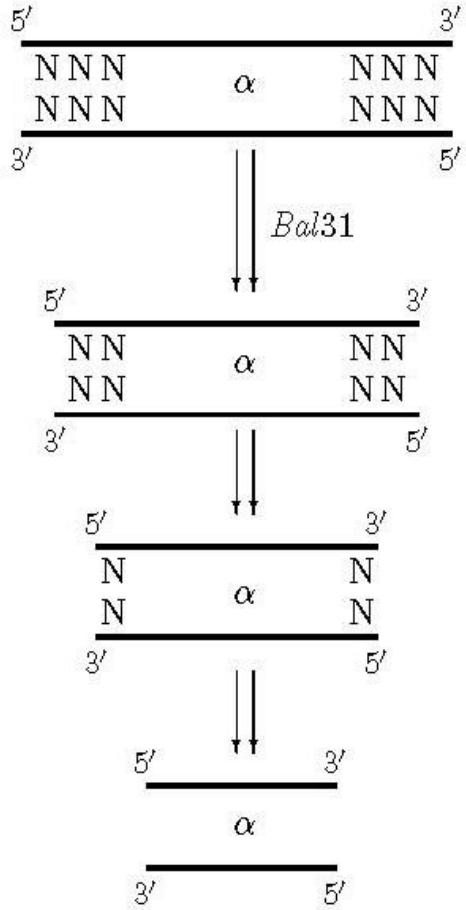

Figure 1.14: Exonuclease *Bal*31 in action

Cutting DNA

- *Endonucleases*: destroy internal covalent bonds in DNA molecules
- Can be quite specialized: what/where/how they cut
- Examples:
  - S1 – not site specific
  - DnaseI – not site specific

Endonucleases

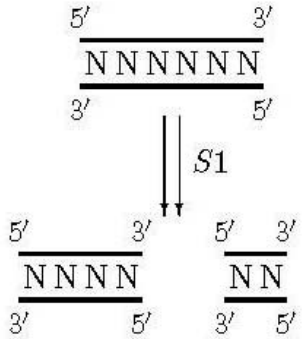

Figure 1.15: $S1$ endonuclease in action (i)

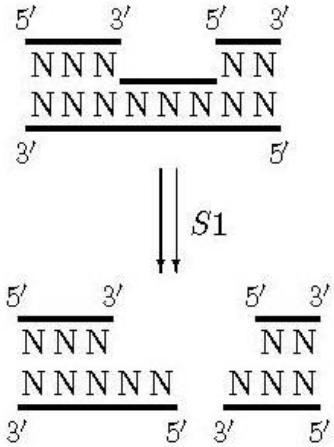

Figure 1.16: $S1$ endonuclease in action (ii)

Restriction enzymes

- Restriction endonucleases: much more specific
- Cut only double stranded molecules, only at very specific set of sites
- Bind to the *recognition site* and then cut (sometimes even outside the binding site)
- The cut: blunt or staggered
- Examples:
  - EcoRI: 5'-GAATTC
  - XmaI: 5'-CCCGGG
  - SmaI: 5'-CCCGGG

Restriction enzymes

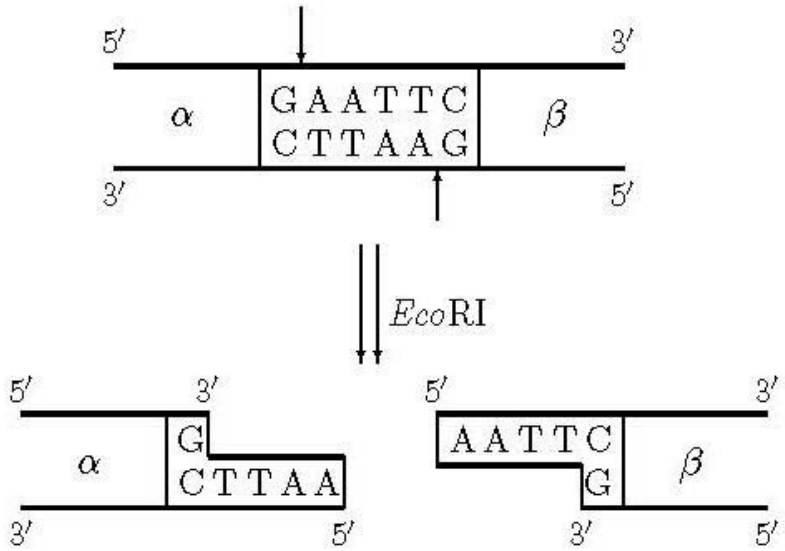

Figure 1.17: *Eco*RI in action

Restriction enzymes

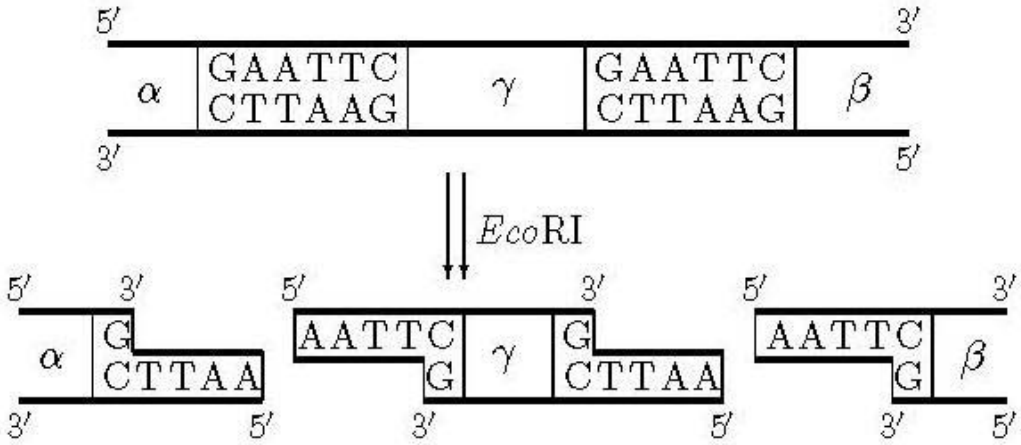

Figure 1.18: Multiple cut by *Eco*RI

Restriction enzymes

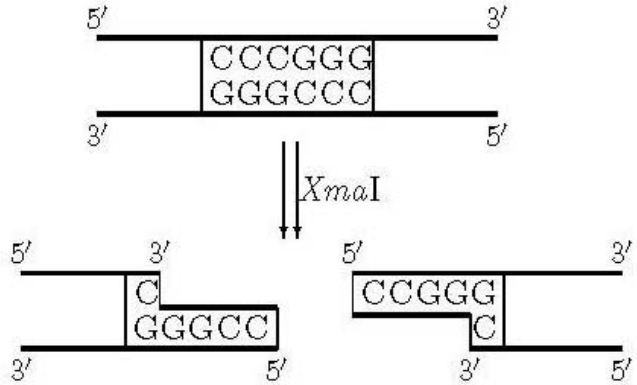

Figure 1.19: *Xma*I in action

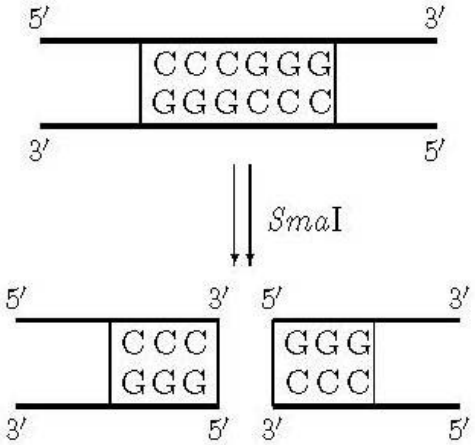

Figure 1.20: *Sma*I in action

Restriction enzymes

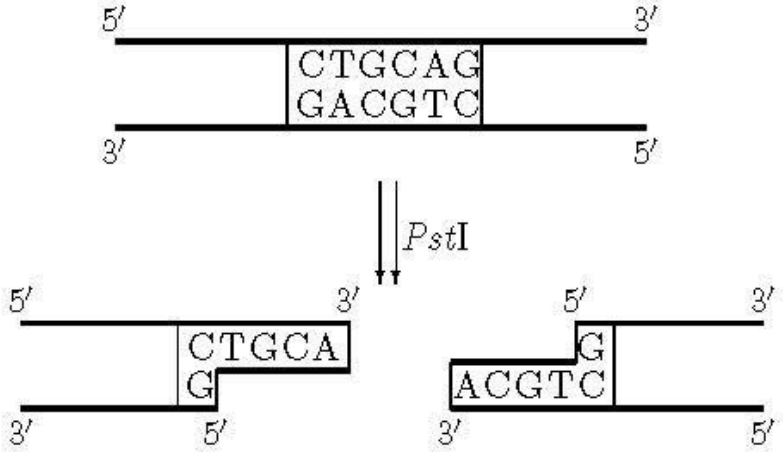

Figure 1.21: *Pst*I in action

Restriction enzymes

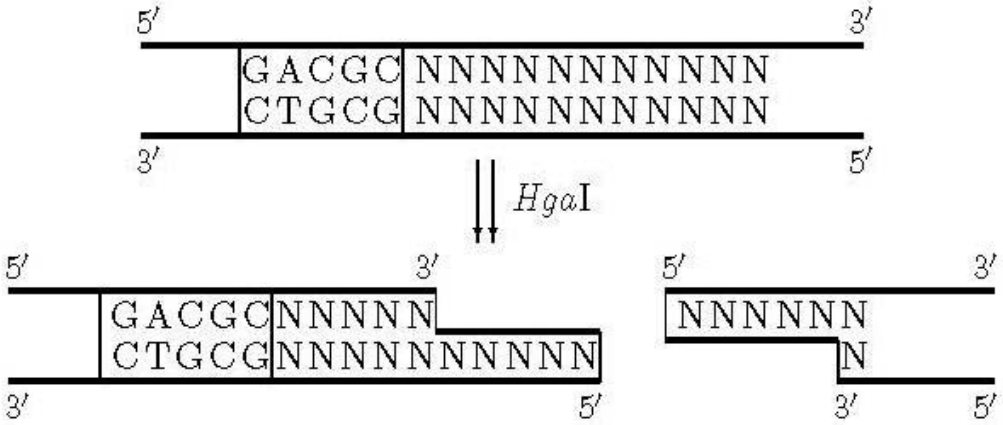

Figure 1.22: *Hga*I in action

Linking DNA

- Fuse together two pieces of DNA
- Add the necessary chemical links
- *Examples*
  - Hydrogen bonding
  - Ligation (enzyme: *ligase)*
  - Hybridization: link two complementary single strands
  - Blunt end ligation

Hydrogen bonding

- Complementary base pairing

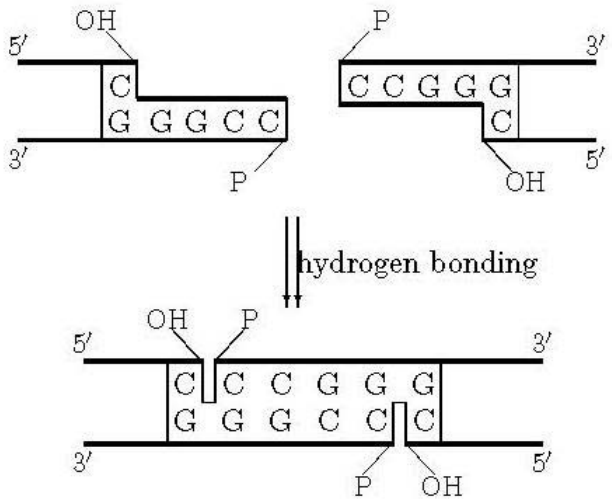

Figure 1.23: Complementary base pairing

Ligation

- Fix the phosphodiester bonds in single strands (*nick*)

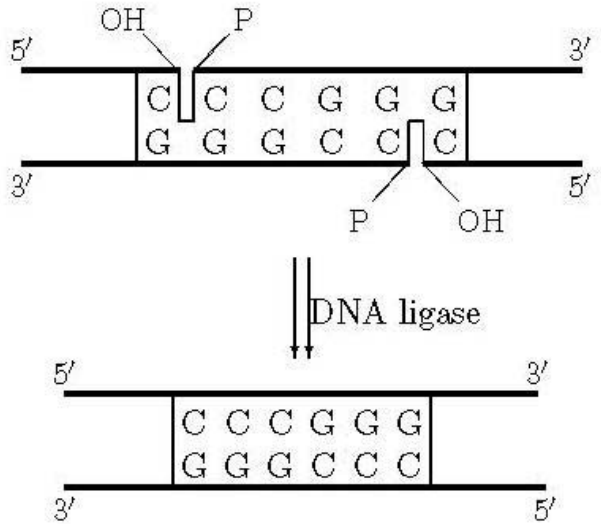

Figure 1.24: Ligation

Hybridization

- Double strands produced out of two single strands (*sticky ends*)

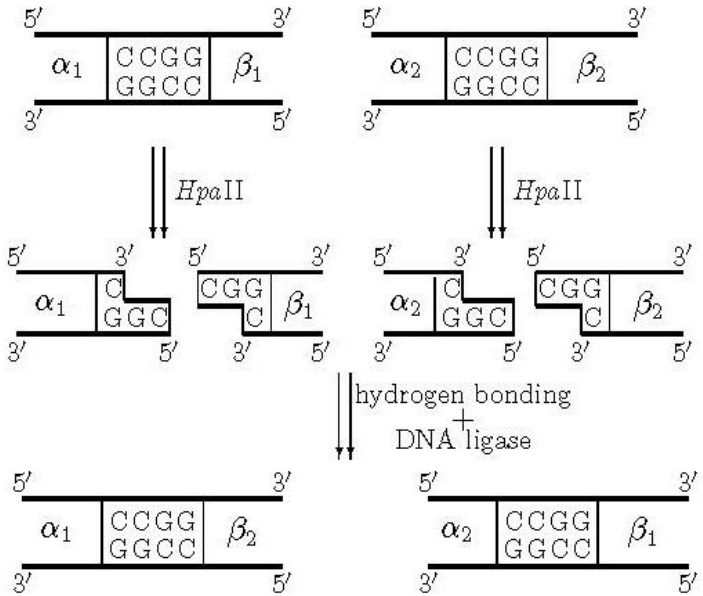

Figure 1.25: Hybridization

Blunt ligation

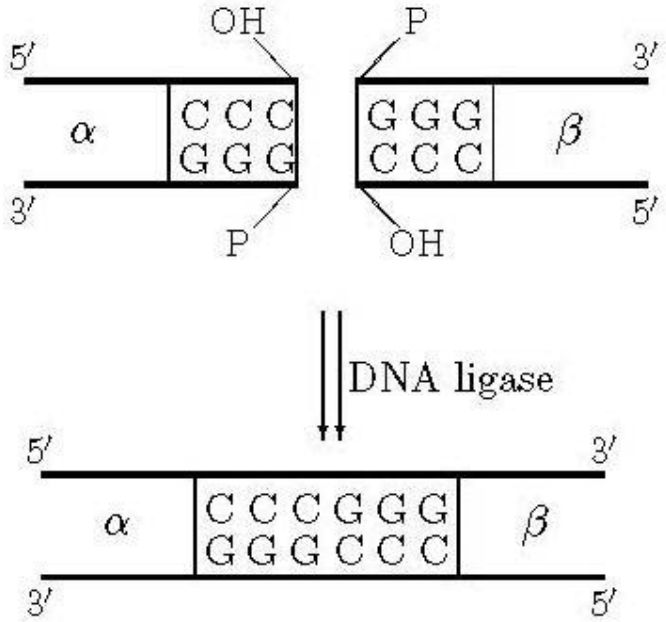

Figure 1.26: Blunt ligation

Blunt ligation

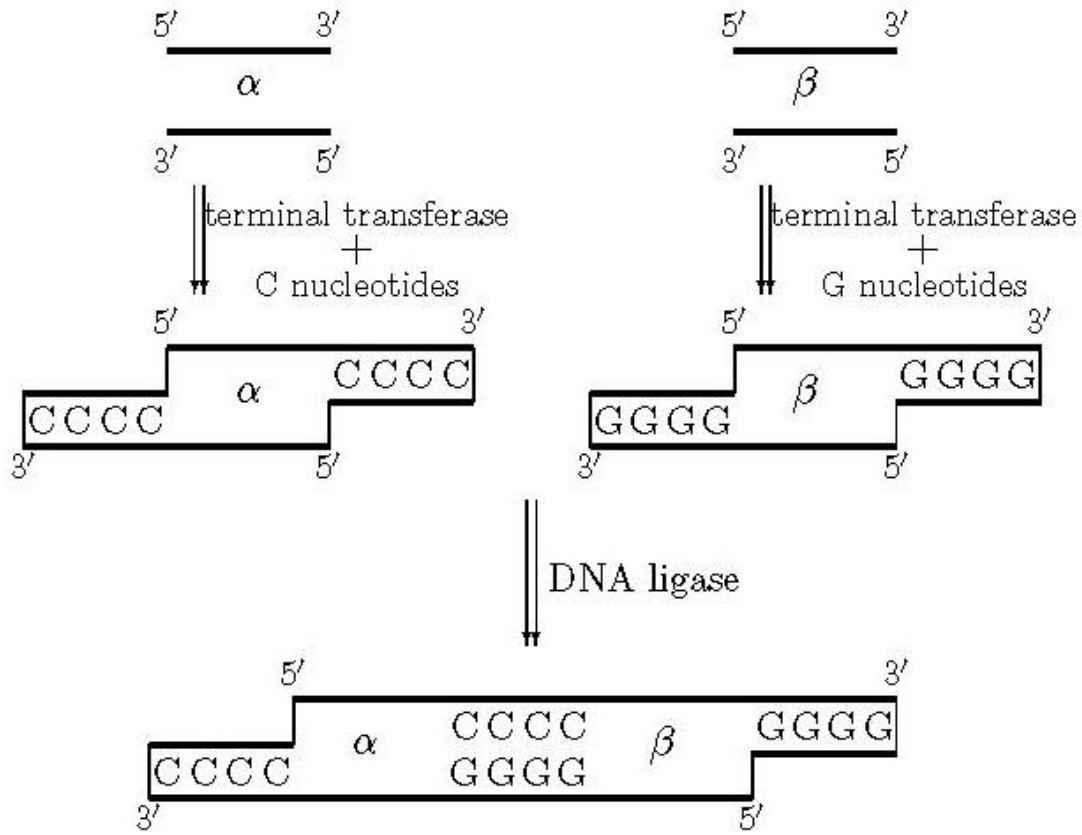

Figure 1.27: Joining blunt ended molecules using homopolymer tailing

Modifying nucleotides of DNA

- The enzymes: *modifying enzymes*
- Examples:
  - *Methylases*: used *in vivo* as partners of restriction enzymes – the same recognition site
    - It adds a methyl group to one of the nucleotides
  - *Alkaline phosphatase*: modifies the 5' end so that the molecule cannot ligate with itself
  - *Polynucleotide kinase*: reverses the effect of alkaline phosphate

Example

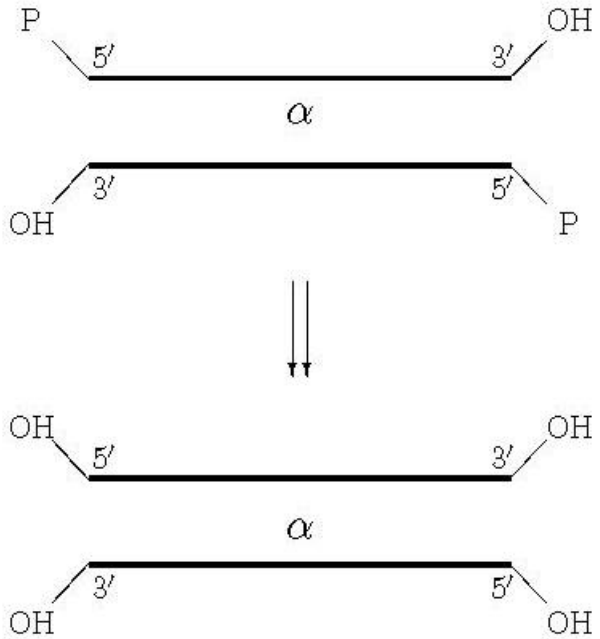

Figure 1.28: Alkaline phosphatase in actior

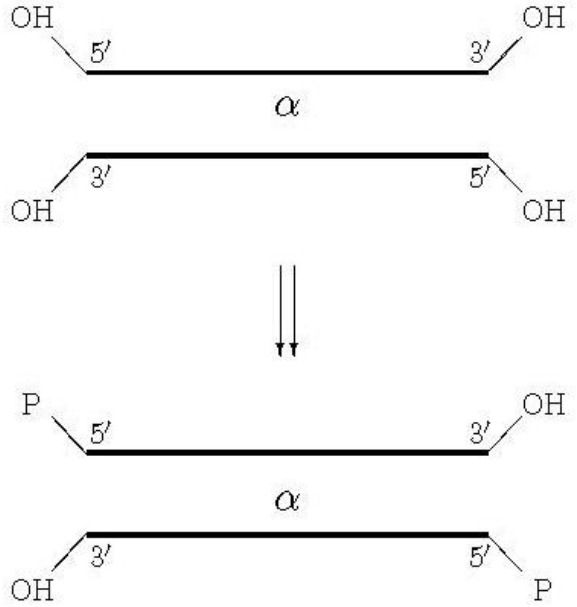

Figure 1.29: Polynucleotide kinase in action

Multiplying DNA

- Major problem in biochemical research: obtain sufficient quantities of some substance
- Solution: cloning techniques, PCR

Cloning

- Clone: a collection of identical organisms, all replicas of a single ancestor
- Cloning: insert the DNA fragment in the DNA of some fast reproducing cell and the fragment will be replicated with the cell
  - Use viruses or hybridization

Cloning

- Separate the cell with the DNA of interest from the others using an antibiotic
- This technique provides high quantities of DNA fragments
- It also gives a mean to preserve the fragments for long time (keeping alive the “host”)

Polymerase chain reaction - PCR

- 1985, Kary Mullis: Nobel prize
- *In vitro* multiplication method
- Made possible by the technique of designing oligos
- Very efficient: produce in a short time billions of copies of even a single strand

The PCR technique

- We want to amplify a DNA molecule alpha with known borders beta and gamma

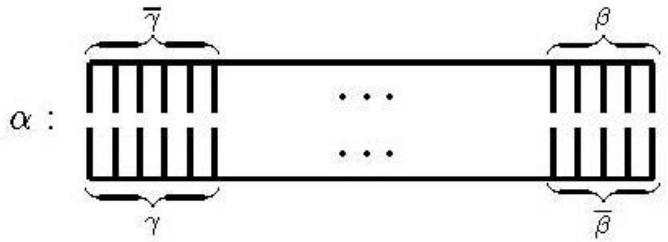

Figure 1.30: DNA with borders

The PCR technique

- We want to amplify a DNA molecule alpha with known borders beta and gamma
- **Process**: repeat a cycle of *denaturation, priming, extension*
- **Start**: prepare a solution containing alpha (the target), oligonucleotides beta' and gamma' (complements of beta and gamma), polymerase enzymes, and many nucleotides (A, C, G, T)

The PCR technique

- **Denaturation**: heat the solution to 85-95 C: alpha denatures into two single strands $alpha_1$ and $alpha_2$

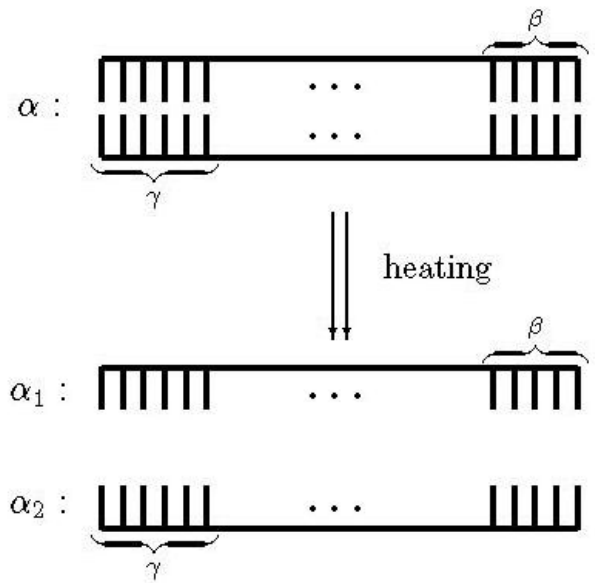

Figure 1.31: Denaturation

The PCR technique

- **Priming**: Cool down the solution to 55C: the primers beta' and gamma' anneal to their complementary borders

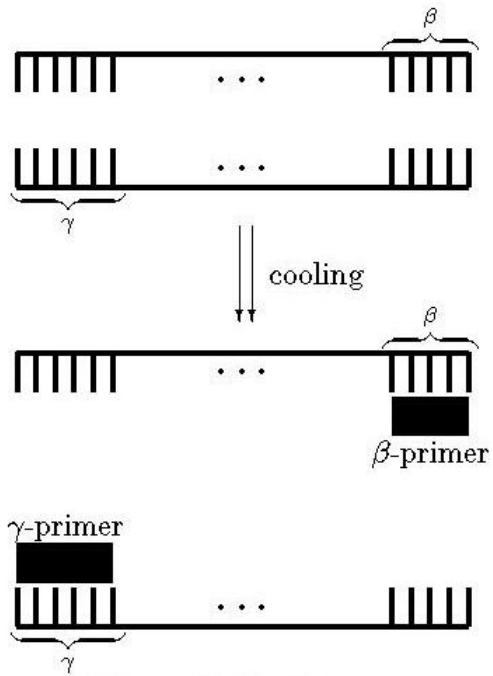

Figure 1.32: Priming

The PCR technique

- **Extension**: Heat the solution to 72C: polymerase extend the primers to produce two double stranded DNA molecules alpha

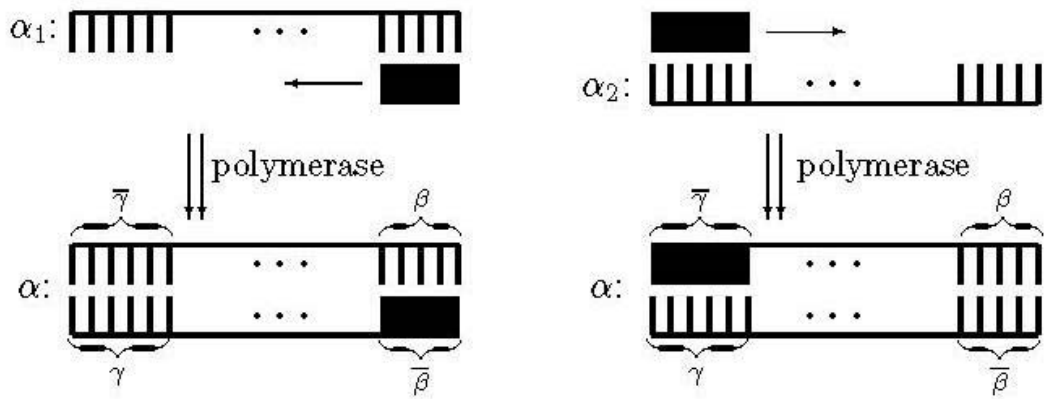

Figure 1.33: Extension

The PCR technique

- Repeat the cycle n times: 2^n copies (in principle): highly efficient bio-copymachine!
- A single cycle takes about 5 minutes: obtain billions of copies in several hours (days for cloning)
- The polymerase must be heat resistant – nature's solution: thermophilic bacteria
- Important observation:
  - one needs to know the borders of the DNA segment to be copied

Sequencing

- Learning the exact sequence of nucleotides of a DNA molecule
- The most popular method of sequencing: the *Sanger method*
  - *1940s: Sanger, insulin sequencing (Nobel prize)*
  - *1977: Sanger, Gilbert, DNA sequencing (Nobel prize)*
  - *1985: Mullis, PCR (Nobel prize)*
- Main tool: nucleotide analogues – nucleotides chemically altered so that no other nucleotide can attach to their 3' end: ddA, ddC, ddG, ddT (dideoxynucleotides)

Sequencing – Sanger method

- Problem: sequence a single stranded molecule *alpha*
- Extend it to 3’ by *gamma* and add the primer *gamma*’ (labeled): let *beta* be the new molecule

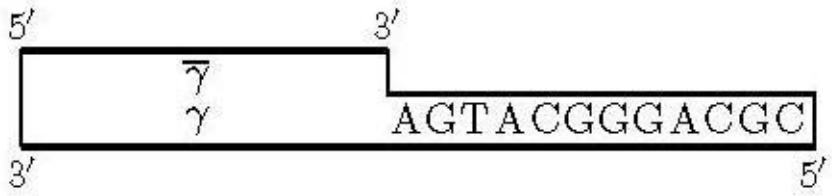

Figure 1.34: β' molecule

Sequencing – Sanger method

- Problem: sequence a single stranded molecule *alpha*
- Extend it to 3’ by *gamma* and add the primer *gamma*’ (labeled): let *beta* be the new molecule
- Prepare 4 tubes (A,C,G,T) containing:
  - *beta* molecules
  - Primers *gamma*’
  - Nucleotides

Sequencing

- If only nucleotides are used: produce full duplex

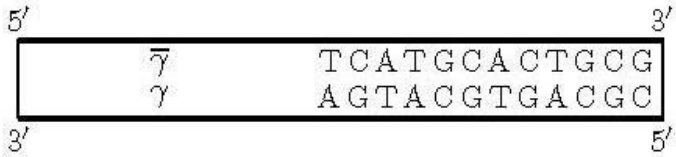

Figure 1.35: Full duplex

Sequencing

- If only nucleotides are used: produce full duplex
- Solution: tube X contains a *limited* amount of nucleotide analogues ddX, for all X in the set {A,C,T,G}

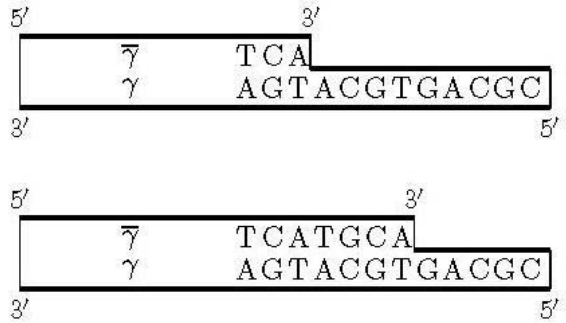

Figure 1.36: Incomplete molecules in Tube A

Example: alpha=3’-AGTACGTGACGC

- Tube A:
  - 5’- *gamma*’TCATGCACTGCG
  - 5’- *gamma*’TCA
  - 5’- *gamma*’TCATGCA
- Tube T:
  - 5’-*gamma*’TCATGCACTGCG
  - 5’- *gamma*’T
  - 5’- *gamma*’TCAT
  - 5’- *gamma*’ TCATGCACT
- Tube C:
  - 5’- *gamma*’TCATGCACTGCG
  - 5’- *gamma*’TC
  - 5’- *gamma*’TCATGC
  - 5’- *gamma*’TCATGCAC
  - 5’- *gamma*’TCATGCACTGC
- Tube G:
  - 5’- *gamma*’TCATGCACTGCG
  - 5’- *gamma*’TCATG
  - 5’- *gamma*’TCATGCACTG

Sequencing – Sanger method

- Gel electrophoresis using 4 wells - one for each tube
- Read the bands (the primers were marked)
- Obtain the sequence

Result: 5’-TCATGCACTGCG

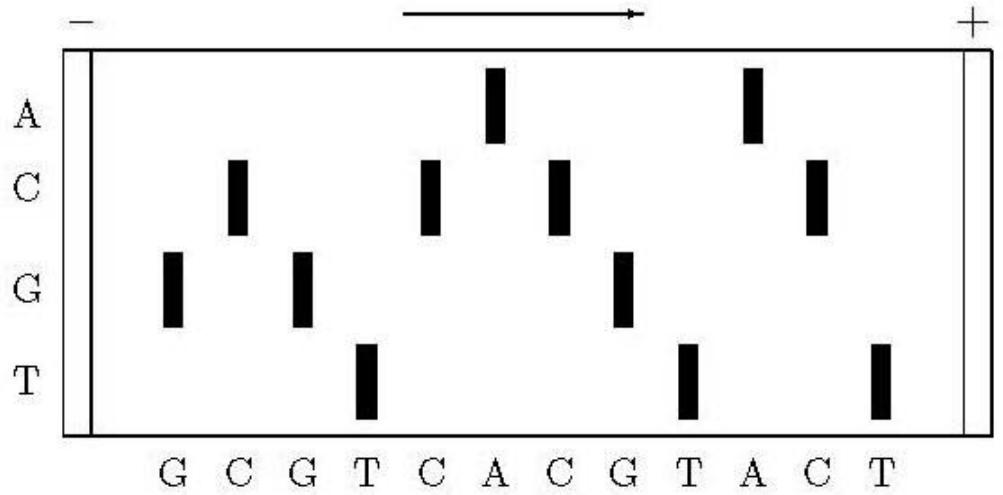

Figure 1.37: Sequencing ladder

Observations

- This was only a very simplified presentation
  - Use polymerase not having the associated exonuclease activity: do not cut the ddX
  - Limited amount of nucleotide analogues: too many make the polymerase before then end of the strand
  - Impossible to sequence long molecules with this method
- Sequencing a longer molecule: divide it in overlapping short segments, sequence them and then reassemble the original sequence: computational (math) problem !