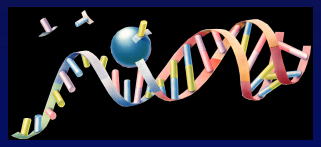
DNA Computing

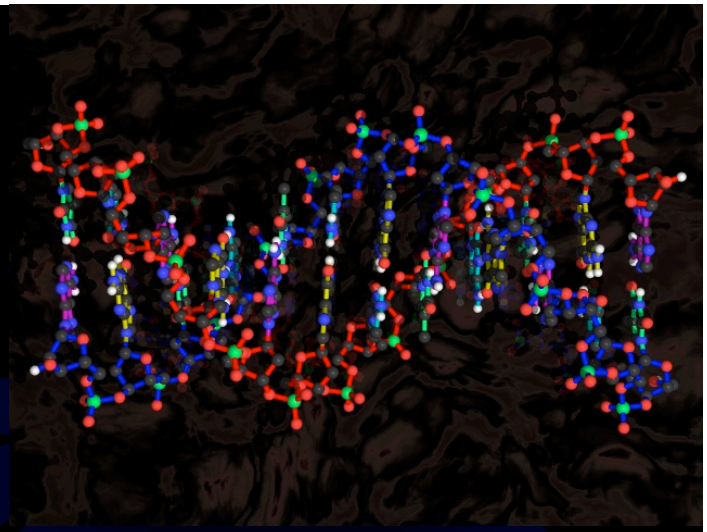


Information Processing with DNA Molecules

Christian Jacob

January 21, 2003.

Table of Contents

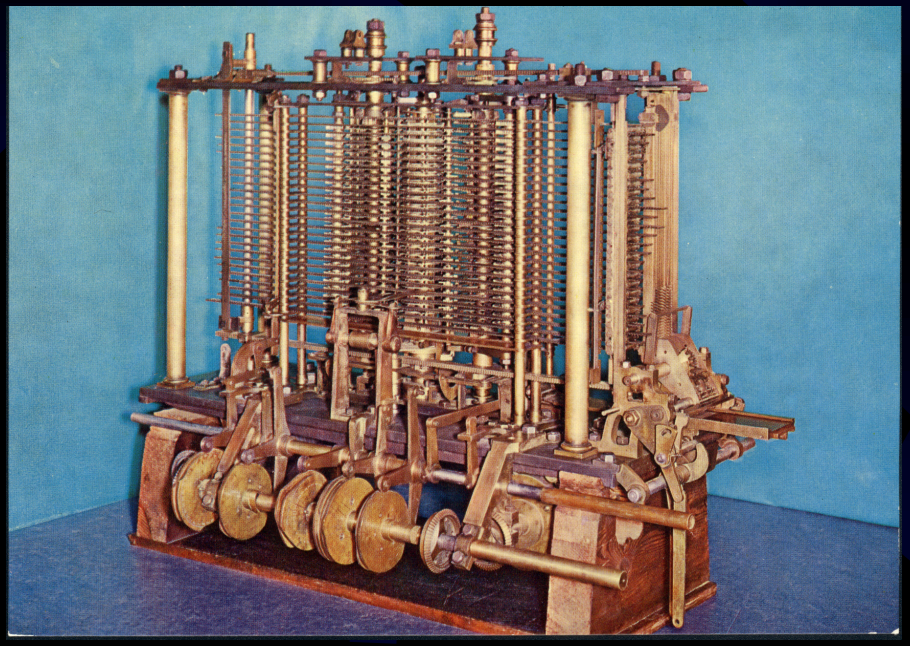


Why DNA Computing?

- From silico to carbon.
  From microchips to DNA molecules.
- Limits to miniaturization with current computer technologies.
- Information processing capabilities of organic molecules ...
  - replace digital switching primitives,
  - enable new computing paradigms.

Challenges of DNA Computing

- Biochemical techniques are not yet sufficiently sophisticated or accurate.
  - Compare Charles Babbage´s „Analytical Engine“ (1810-1820)



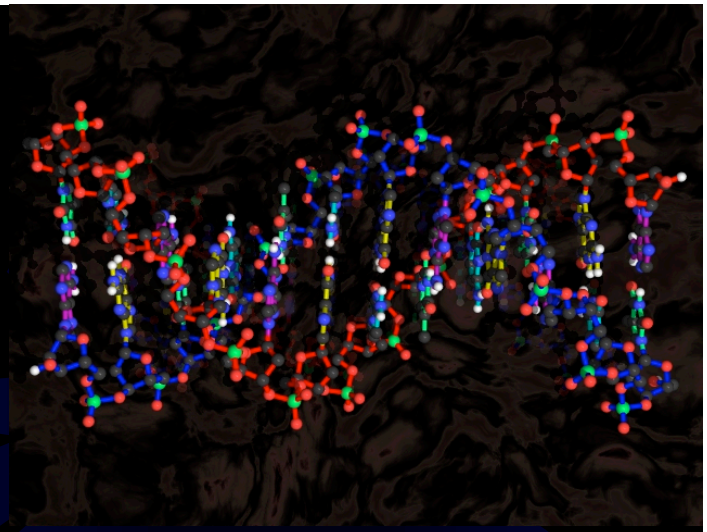
Key Features of DNA Computing

- Massive parallelism of DNA strands
  - high density of information storage
  - ease of constructing many copies
- Watson-Crick complementarity
  - feature provided „for free“
  - universal twin shuffle language

Still: Why DNA Computing?

- Further reasons to investigate DNA computing:
  - support for standard computation
  - better understanding of how nature computes
    - new data structures (molecules)
    - new operations
      - cut, paste, adjoin, insert, delete, ...
    - new computability models.

Table of Contents



The Structure of DNA

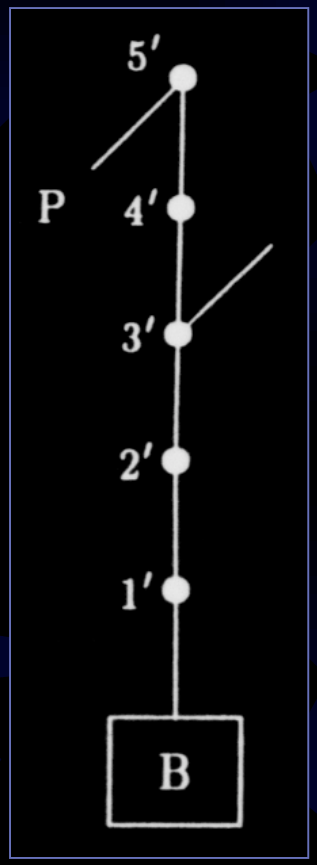
- DNA is a polymer („large“ molecule).
- DNA is strung together from monomers („small“ mols.): deoxyribonucleotides.

- DNA = Deoxyribo Nucleic Acid

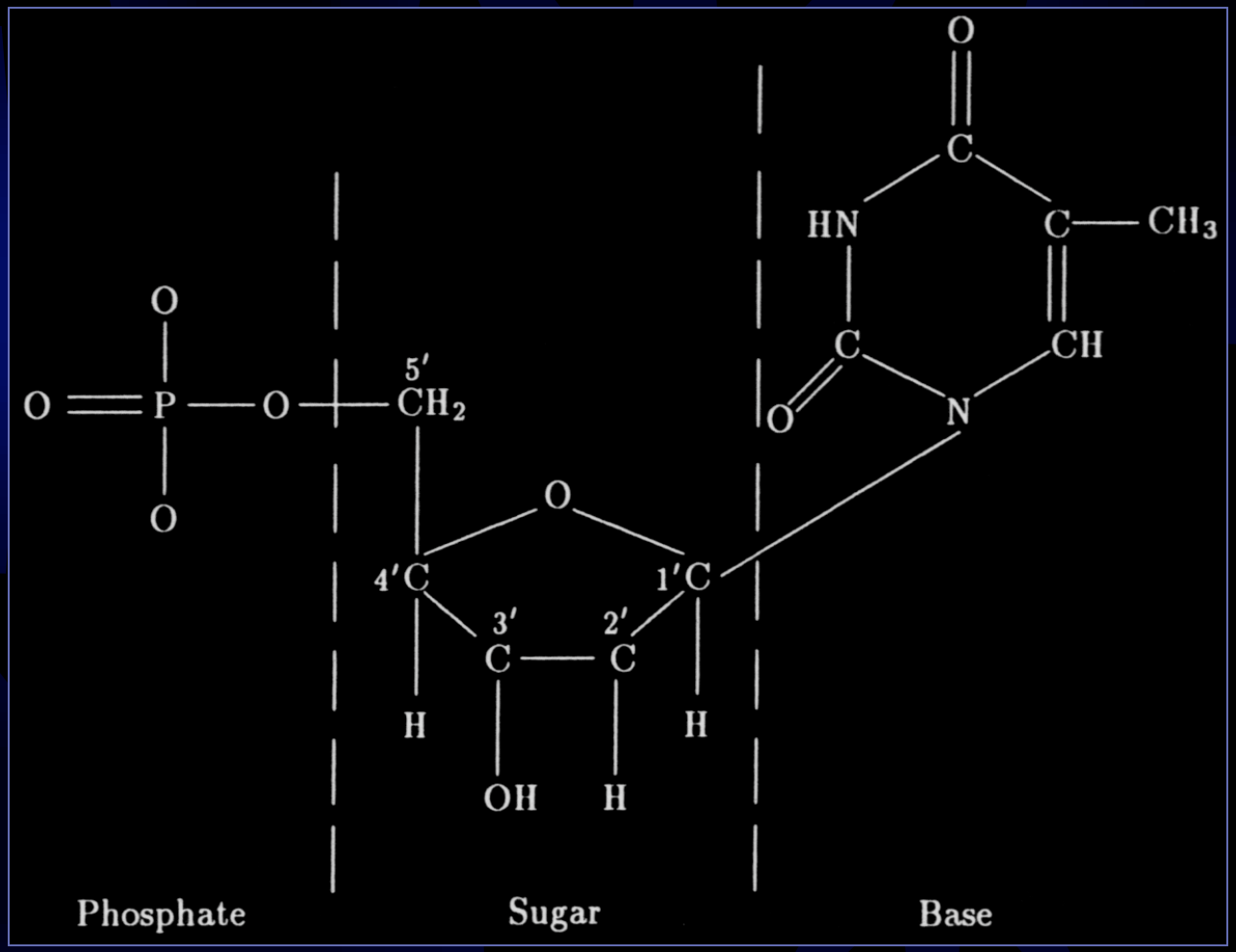
- DNA supports two key functions for life:
  - coding for the production of proteins,
  - self-replication.

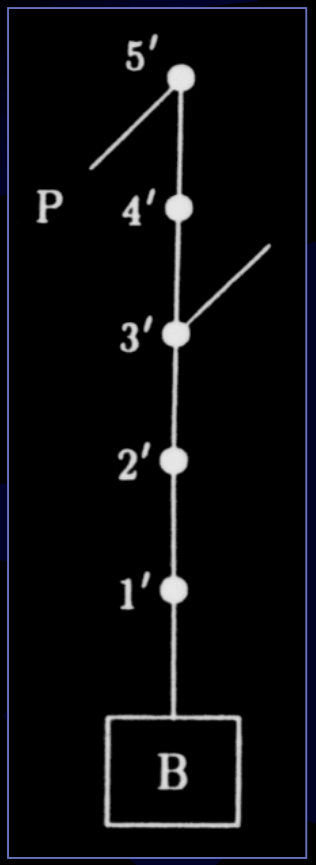
Structure of a DNA Monomer

- Each deoxyribonucleotide consists of three components:
  - a sugar — deoxyribose
    - five carbon atoms: 1´ to 5´
    - hydroxyl group (OH) attached to 3´ carbon
  - a phosphate group
  - a nitrogenous base.



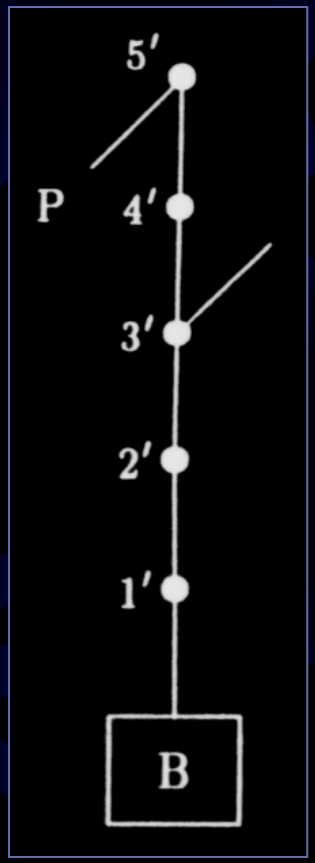
Chemical Structure of a Nucleotide





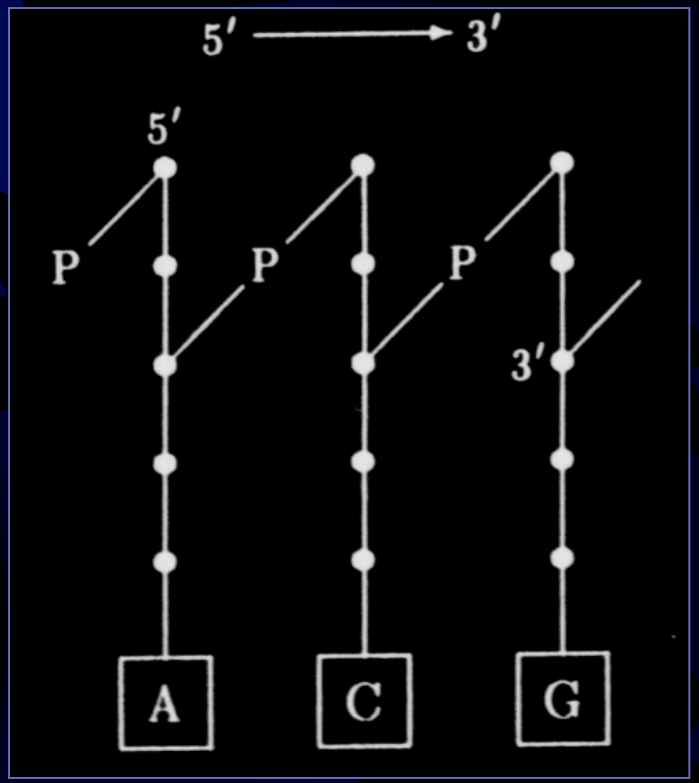
Structure of a DNA Monomer (2)

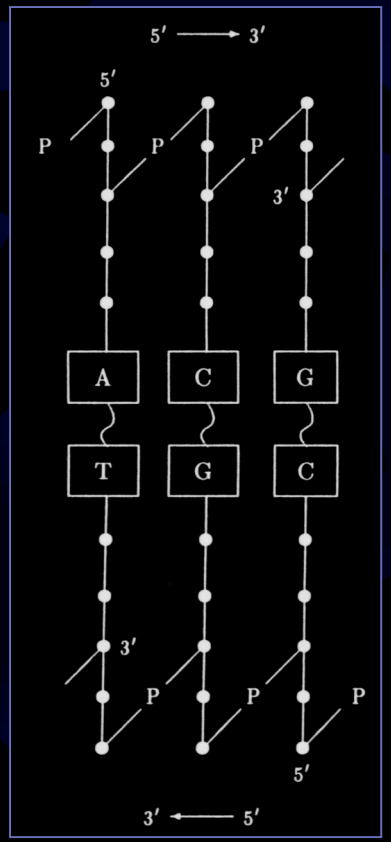
- DNA nucleotides differ only by their bases (B):
  - purines
    - Adenine A
    - Guanine G
  - pyrimidines
    - Cytosine C
    - Thymine T



Linking of Nucleotides

- The DNA monomers can link in two ways:
  - Phosphodiester bond
  - Hydrogen bond



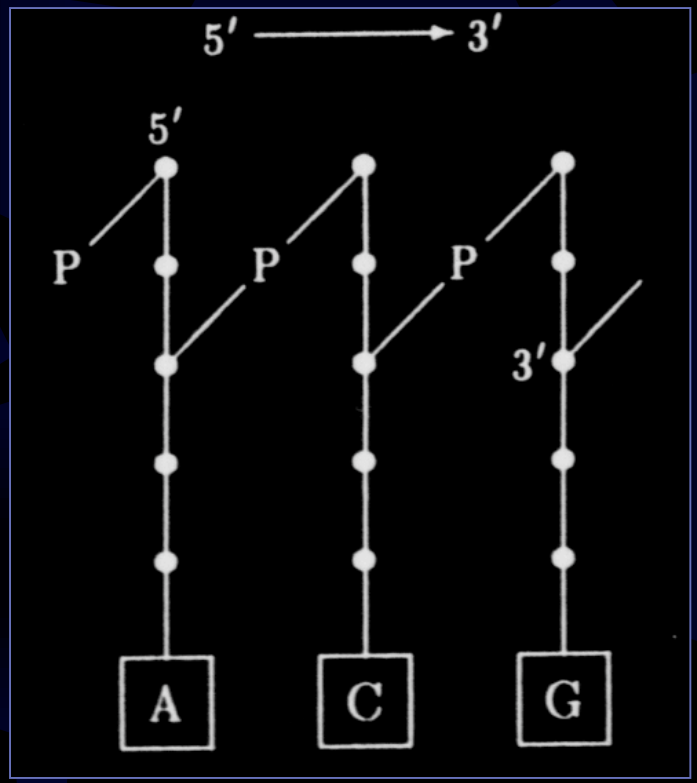


Linking of Nucleotides

Phosphodiester Bond

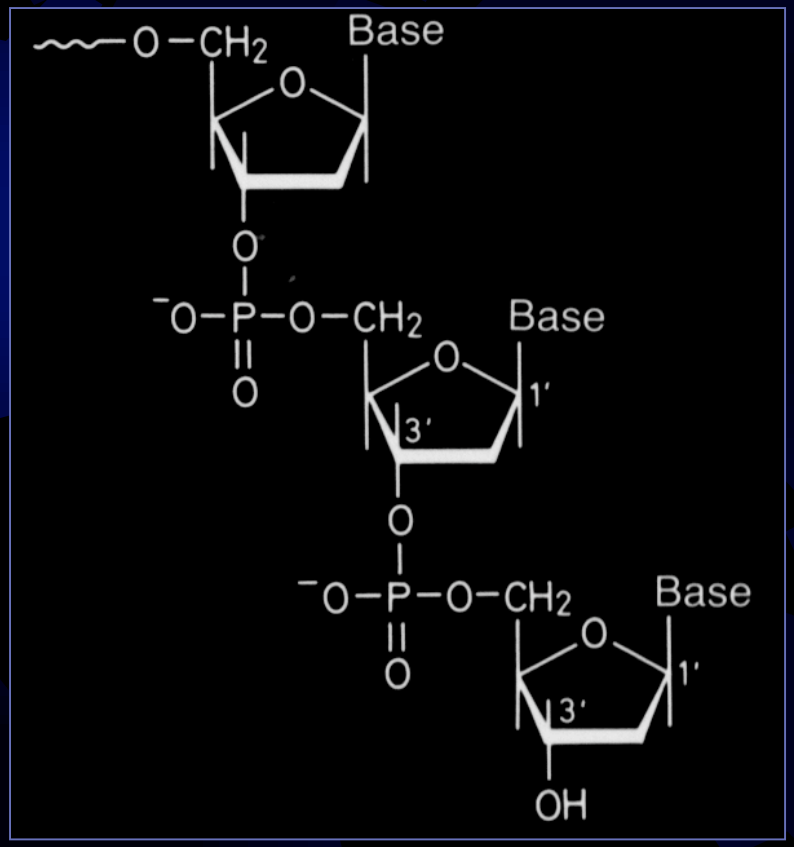
- The 5´-phosphate group of one nucleotide is joined with the 3´-hydroxyl group of the other
  - strong (covalent) bond
  - directionality:

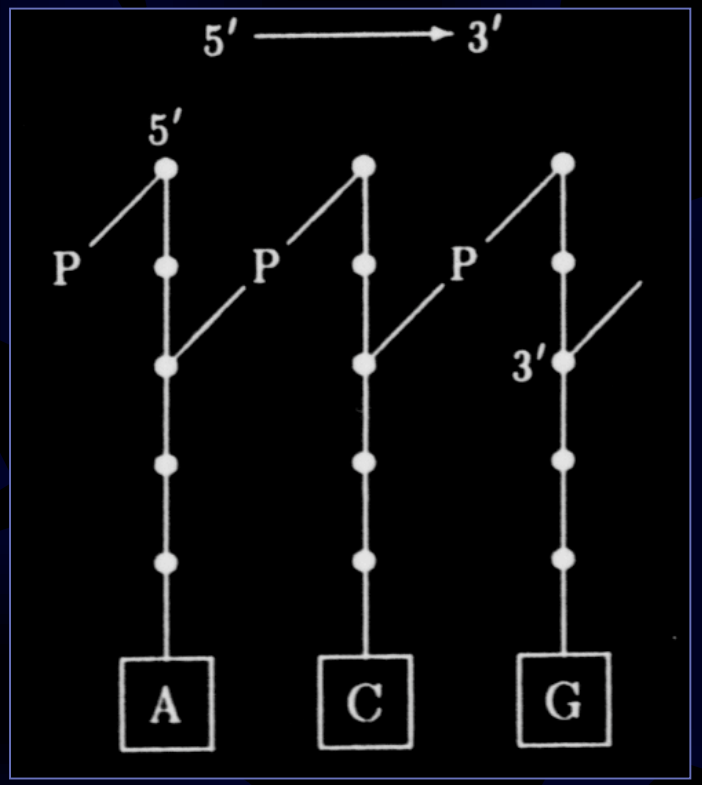
    5´—3´ or 3´—5´



Linking of Nucleotides

Phosphodiester Bond

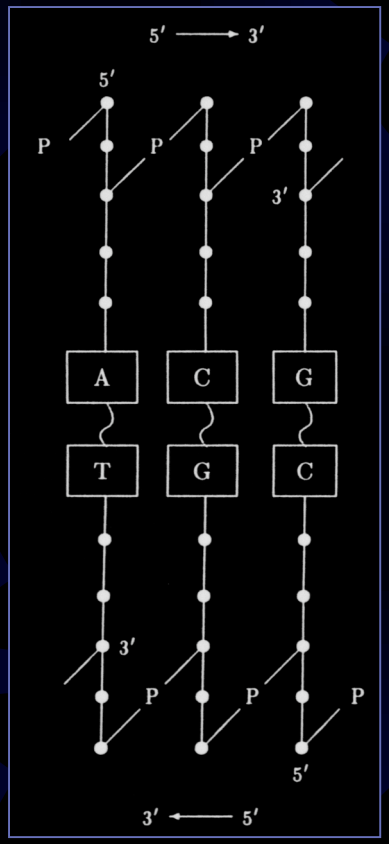




Linking of Nucleotides

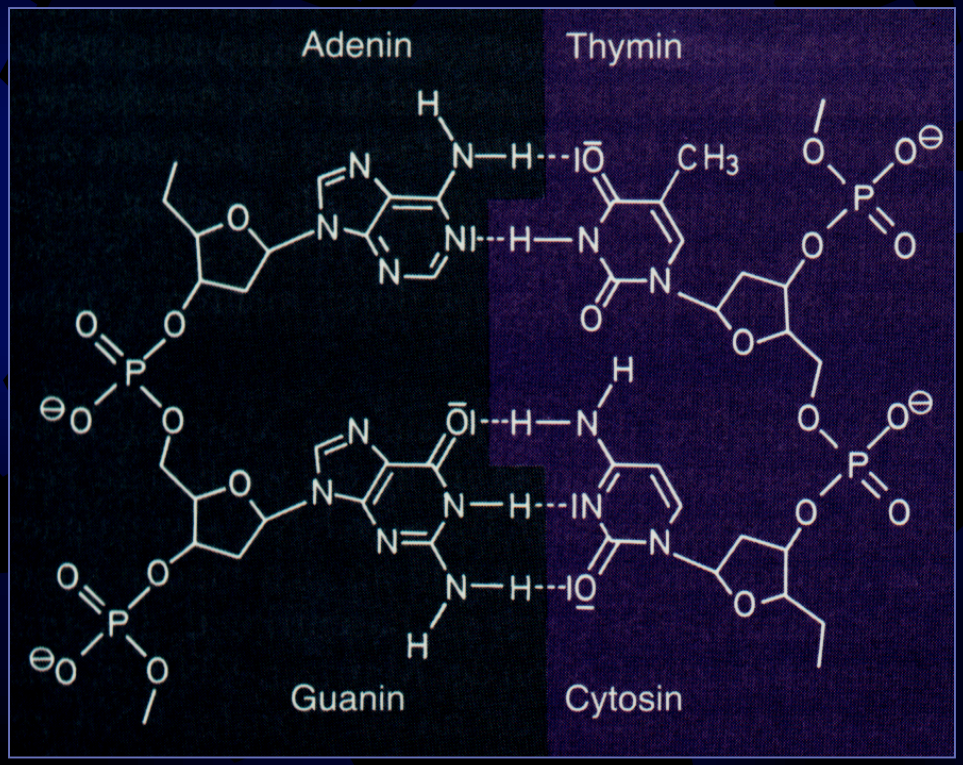
Hydrogen Bond

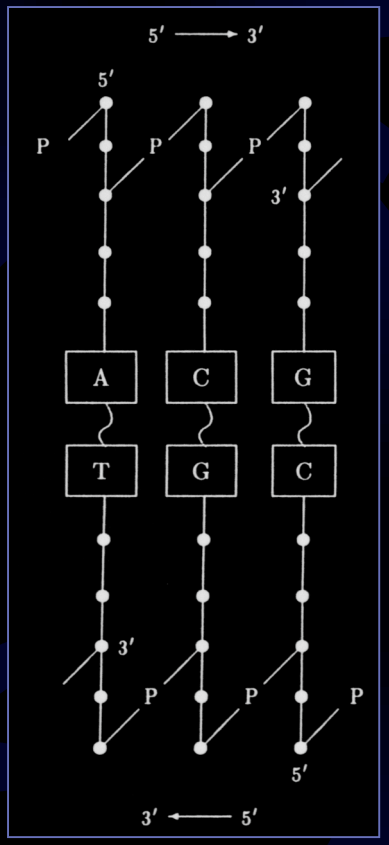
- The base of one nucleotide interacts with the base of another
  - base pairing (weak bond)
    - A — T (2 hydrogen bonds)
    - C — G (3 hydrogen bonds)
  - Watson-Crick complementarity
    - James D. Watson
    - Francis H. C. Crick
      - deduced double-helix structure of DNA in 1953
      - Nobel Prize (1962)



Linking of Nucleotides

Hydrogen Bond





DNA

Double Helix

- Longer streches keep the double strands together through the cumulative effect (the sum) of hydrogen bonds.
- Dense packing:
  - Bacteria: DNA molecule is 10,000 times longer than the host cell
  - Eucaryotes: „hierarchical“ packing

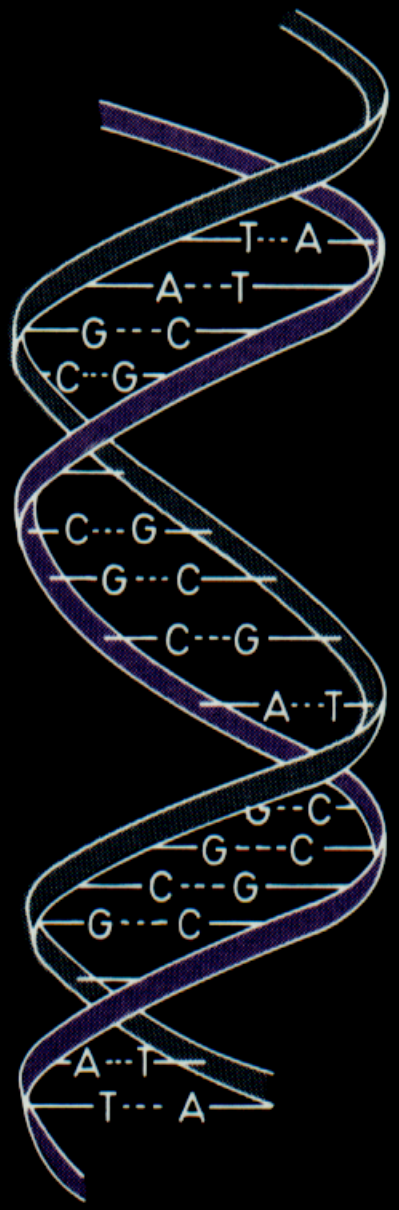
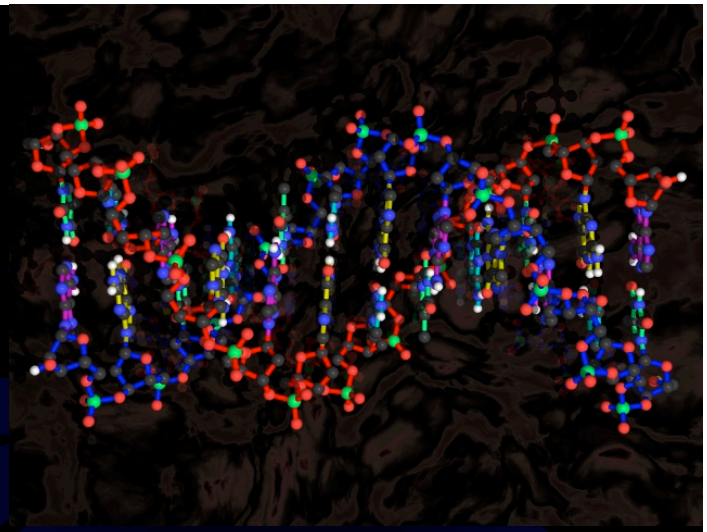


Table of Contents



Operations on DNA Molecules

- Separating and fusing DNA strands
- Lengthening of DNA
- Shortening DNA
- Cutting DNA
- Multiplying DNA

Separating and Fusing DNA Strands

- Denaturation: separating the single strands without breaking them
  - weaker hydrogen than phosphodiester bonding
  - heat DNA (85° - 90° C)
- Renaturation:
  - slowly cooling down
  - annealing of matching, separated strands

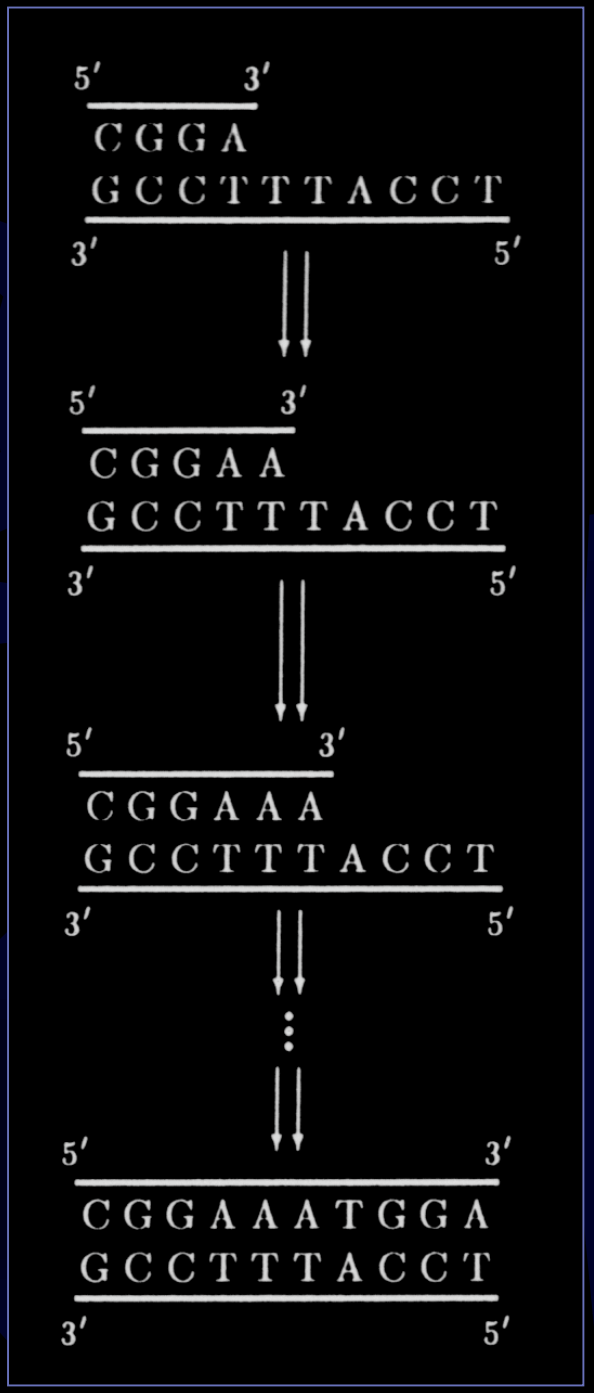
Enzymes

Machinery for Nucleotide Manipulation

- Enzymes are proteins that catalyze chemical reactions.
- Enzymes speed up chemical reactions extremely efficiently (speedup: 10^{12})
- Enzymes are very specific.
- Nature has created a multitude of enzymes that are useful in processing DNA.

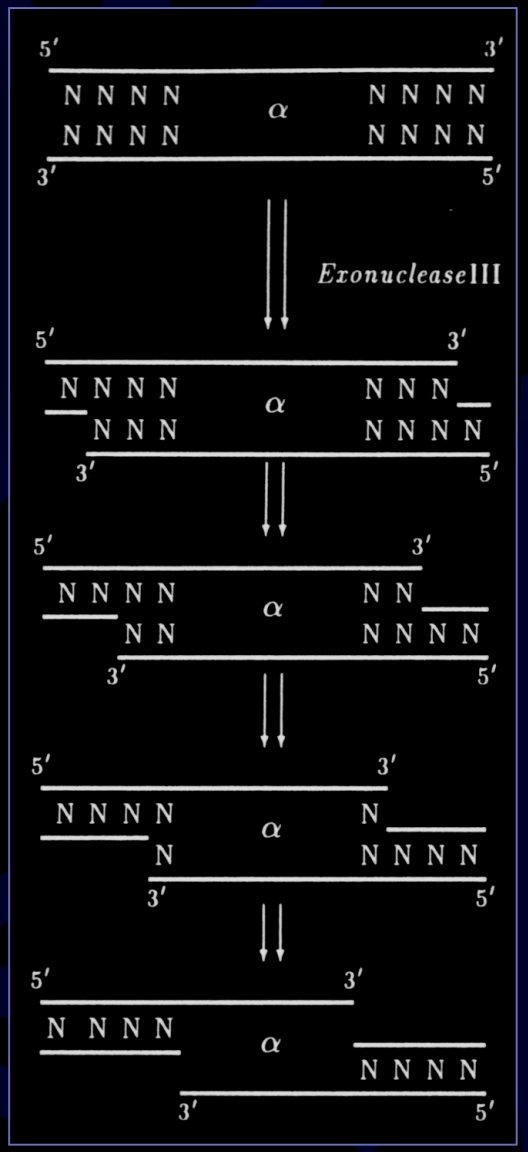
Lengthening DNA

- DNA polymerase enzymes add nucleotides to a DNA molecule
- Requirements:
  - single-stranded template
  - primer,
    - bonded to the template
    - 3´-hydroxyl end available for extension
      - Note: Terminal transferase needs no primer.



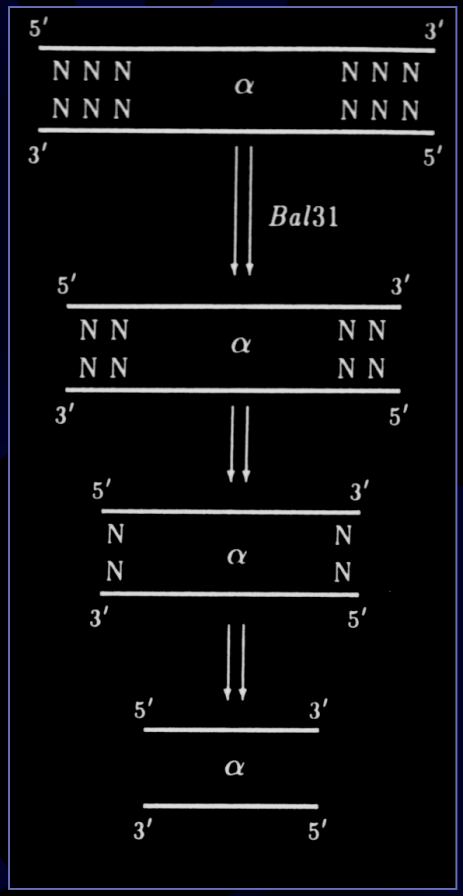
Shortening DNA

- DNA nucleases are enzymes that degrade DNA.
  - DNA <u>exonucleases</u>
    - cleave (remove) nucleotides one at a time from the ends of the strands
    - Example: Exonuclease III
      3´-nuclease
      degrading in 3´-5´direction



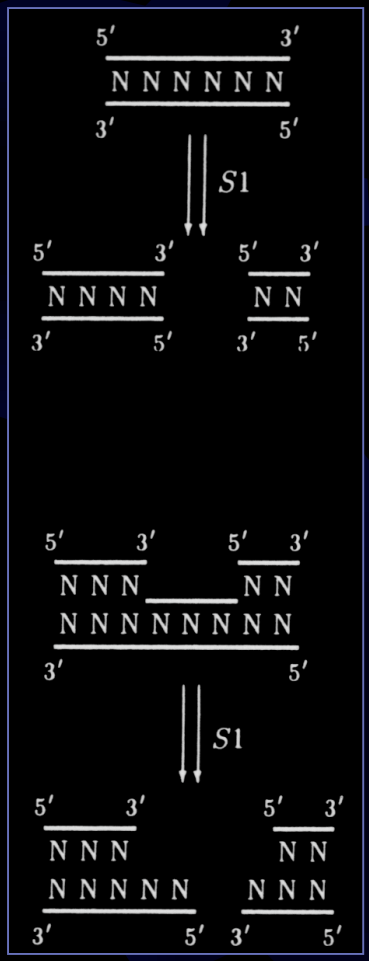
Shortening DNA

- DNA nucleases are enzymes that degrade DNA.
  - DNA <u>exonucleases</u>
    - cleave (remove) nucleotides one at a time from the ends of the strands
    - Example: Bal31 removes nucleotides from both strands



Cutting DNA

- DNA nucleases are enzymes that degrade DNA.
  - DNA endonucleases
    - destroy internal phosphodiester bonds
    - Example: S1 cuts only single strands or within single strand sections
  - Restriction endonucleases
    - much more specific
    - cut only double strands
    - at a specific set of sites (EcoRI)



Multiplying DNA

- Amplification of a „small“ amount of a specific DNA fragment, lost in a huge amount of other pieces.
- „Needle in a haystack“

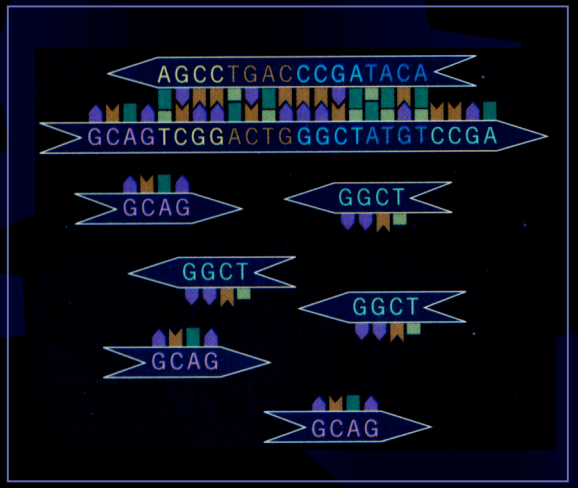
- Solution: PCR = Polymerase Chain Reaction
  - devised by Karl Mullis in 1985
  - Nobel Prize

  - a very efficient molecular Xerox machine

PCR

Step 0: Initialization

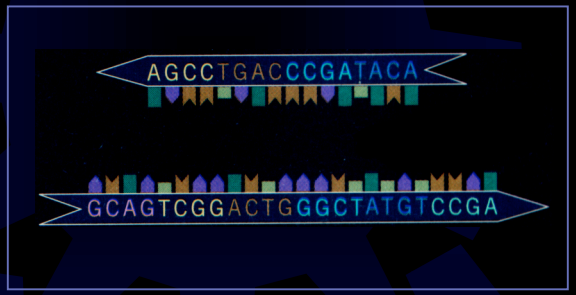
- Start with a solution containing the following ingredients:
  - the target DNA molecule
  - primers (synthetic oligonucleotides), complementary to the terminal sections
  - polymerase, heat resistant
  - nucleotides



PCR

Step 1: Denaturation

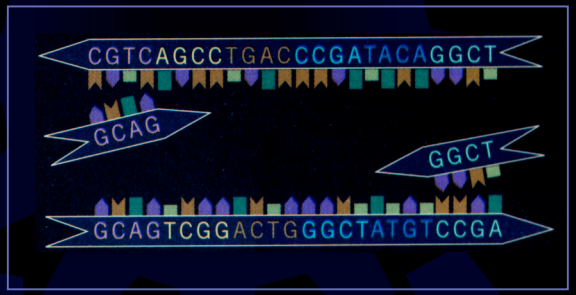
- Solution heated close to boiling temperature.
- Hydrogen bonds between the double strands are separated into single strand molecules.



PCR

Step 2: Priming

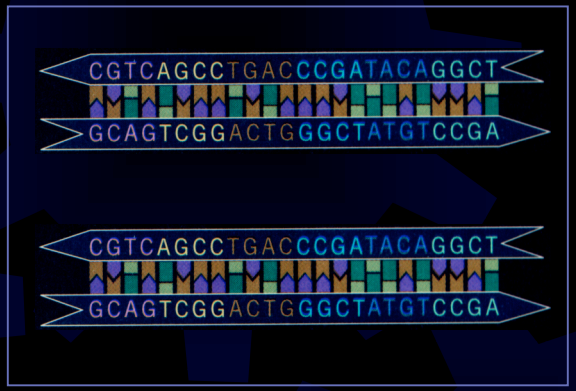
- The solution is cooled down (to about 55° C).
- Primers anneal to their complementary borders.



PCR

Step 3: Extension

- The solution is heated again (to about 72° C).
- A polymerase will extend the primers, using nucleotides available in the solution.
- Two complete strands of the target DNA molecule are produced.



PCR

Efficient Xeroxing: 2^n copies after n steps

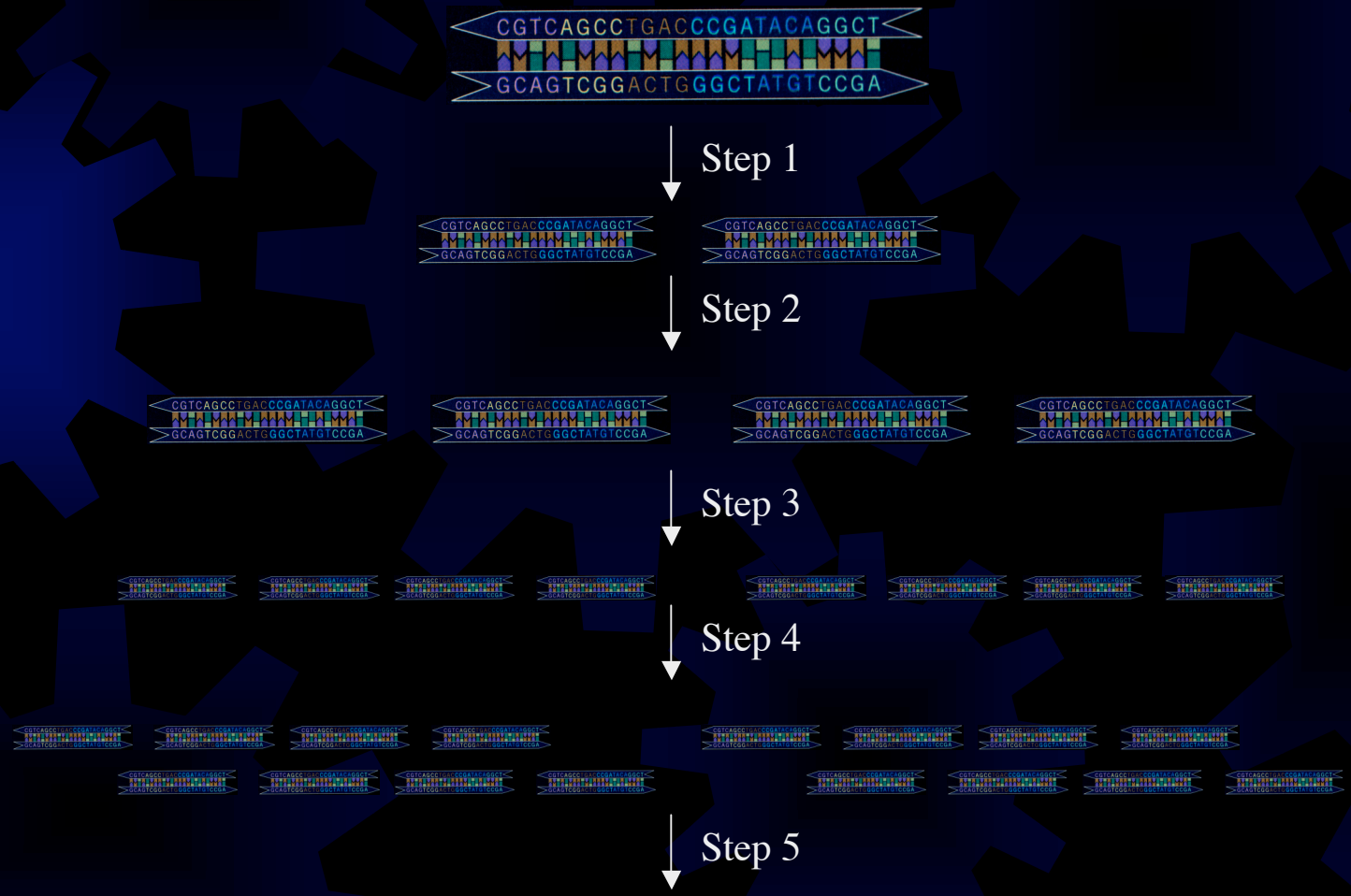
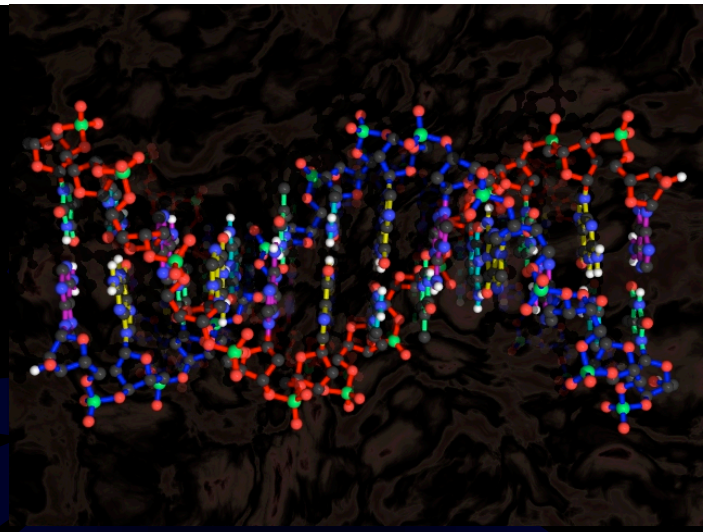
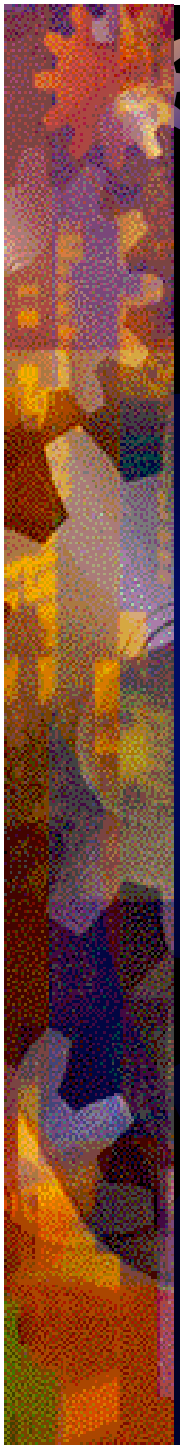


Table of Contents



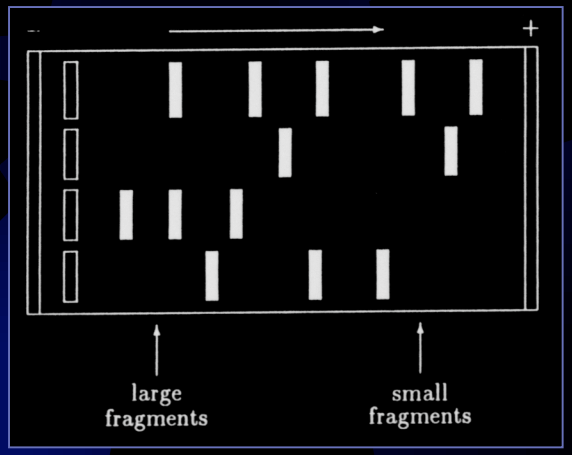


Measuring the Length of DNA Molecules

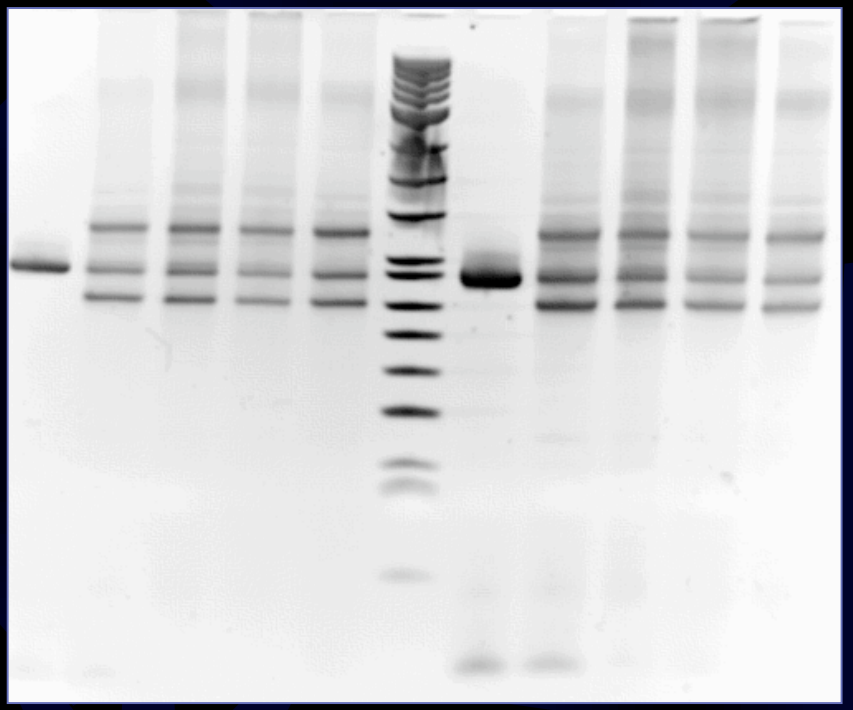
Gel Electrophoresis

- DNA molecules are negatively charged.
- Placed in an electric field, they will move towards the positive electrode.
  - The negative charge is proportional to the length of the DNA molecule.
  - The force needed to move the molecule is proportional to its length.
- A gel makes the molecules move at different speeds.
- DNA molecules are invisible, and must be marked (ethidium bromide, radioactive)

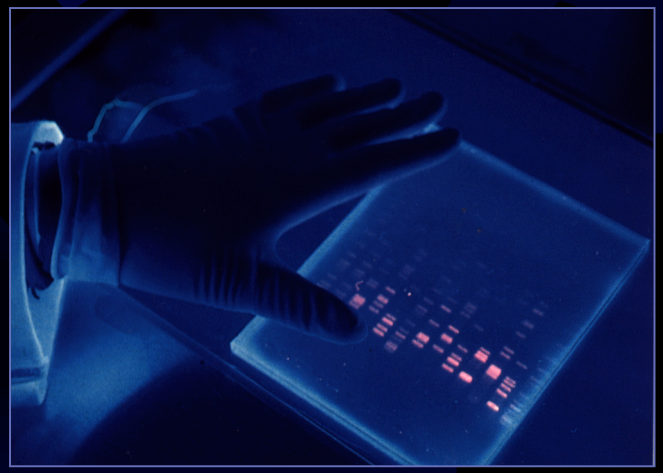
Schematic representation of gel electrophoresis



Radioactive marker



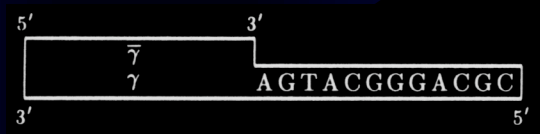
Ethidium bromide marker



Sequencing a DNA Molecule

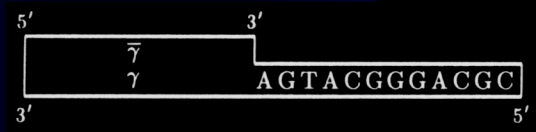
- Sequencing:
  - reading the exact sequence of nucleotides comprising a given DNA molecule
  - based on
    - the polymerase action of extending a primed single stranded template
    - nucleotide analogues
      - chemically modified
      - e.g., replace 3´-hydroxyl group (3´-OH) by 3´-hydrogen atom (3´-H)
      - dideoxynucleotides:
        - ddA, ddT, ddC, ddG
      - Sanger method, dideoxy enzymatic method

Sequencing — Part 1



- Objective
  - We want to sequence a single stranded molecule □.
- Preparation
  - We extend □ at the 3´ end by a short (20 bp) sequence □, which will act as the W-C complement for the primer *compl*(□).
    - Usually, the primer is labelled (radioactively, or marked fluorescently)
  - This results in a molecule □´= 3´- □□.

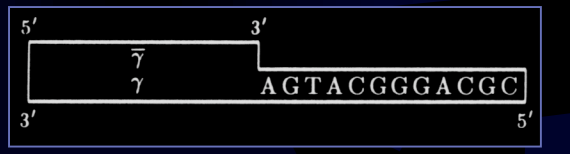
Sequencing — Part 2

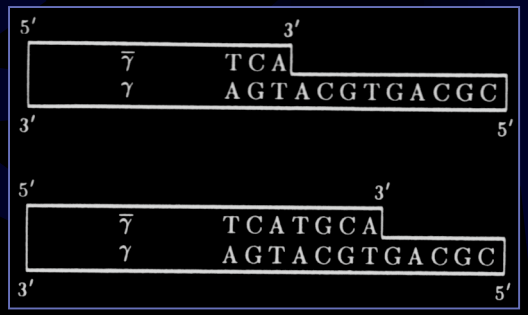


- 4 tubes are prepared:
  - Tube A, Tube T, Tube C, Tube G
  - Each of them contains
    - □ molecules
    - primers, *compl*(□)
    - polymerase
    - nucleotides A, T, C, and G.
  - Tube A contains a limited amount of ddA.
  - Tube T contains a limited amount of ddT.
  - Tube C contains a limited amount of ddC.
  - Tube G contains a limited amount of ddG.

Reaction in Tube A

- The polymerase enzyme extends the primer of □′, using the nucleotides present in Tube A: ddA, A, T, C, G.
  - using only A, T, C, G:
    - □′ is extended to the full duplex.
  - using ddA rather than A:
    - complementing will end at the position of the ddA nucleotide.



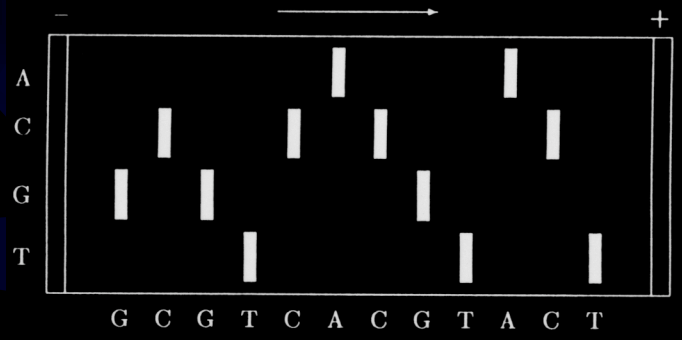


Resulting Sequences in Tubes

- Tube A:
  - TCATGCACTGCG
  - TCA
  - TCATGCA
- Tube T:
  - TCATGCACTGCG
  - T
  - TCAT
  - TCATGCACT
- Tube C:
  - TCATGCACTGCG
  - TC
  - TCATGC
  - TCATGCAC
  - TCATGCACTGC
- Tube G:
  - TCATGCACTGCG
  - TCATG
  - TCATGCACTG

Final Reading of the Strands

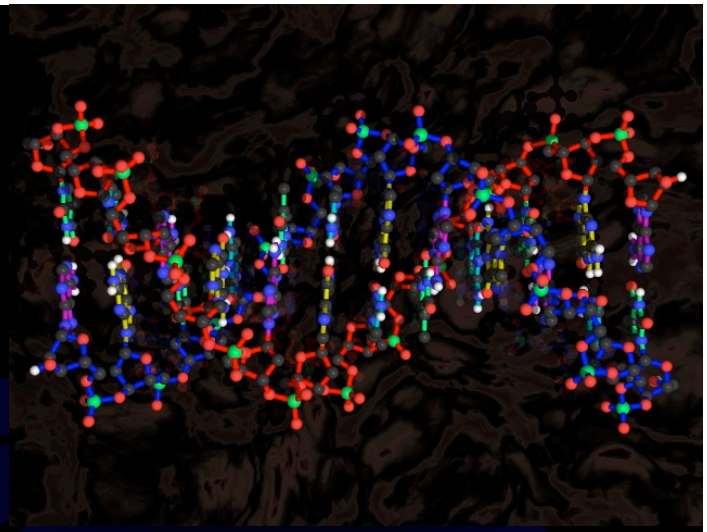
Gel electrophoresis:



- Tube A:
  - TCATGCACTGCG
  - TCA
  - TCATGCA
- Tube T:
  - TCATGCACTGCG
  - T
  - TCAT
  - TCATGCACT
- Tube C:
  - TCATGCACTGCG
  - TC
  - TCATGC
  - TCATGCAC
  - TCATGCACTGC
- Tube G:
  - TCATGCACTGCG
  - TCATG
  - TCATGCACTG

- We read:
  - T
  - TC
  - TCA
  - TCAT
  - TCATG
  - TCATGC
  - TCATGCA
  - TCATGCAC
  - TCATGCACT
  - TCATGCACTG
  - TCATGCACTGC
  - TCATGCACTGCG

Table of Contents

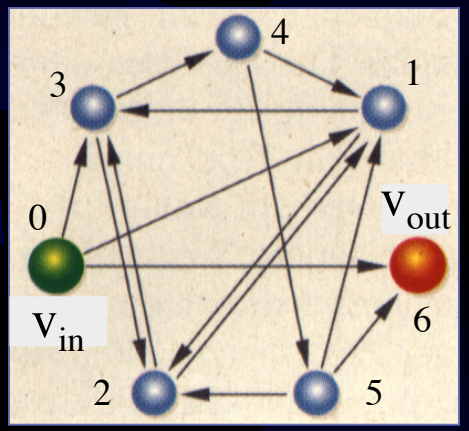


Adleman´s Experiment

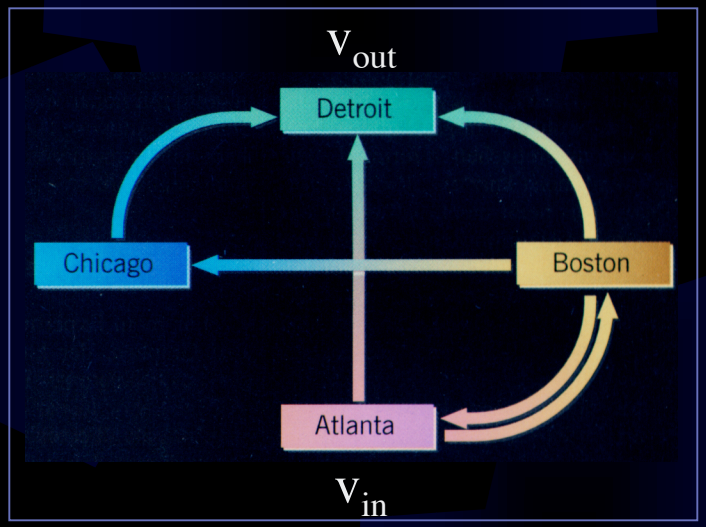
- In 1994 Leonard M. Adleman showed how to solve the Hamilton Path Problem, using DNA computation.

- Hamiltonian Path Problem:
  - Let G be a directed graph with designated input and output vertices, v_{in} and v_{out}.
  - Find a (Hamiltonian) path from v_{in} to v_{out} that involves every vertex in G exactly once.

Hamiltonian Path Example



- Adleman´s graph
- The only Hamiltonian path for this graph is:
  - 0—1—2—3—4—5—6



- Simplified graph
- Hamiltonian path:
  - Atlanta
  - Boston
  - Chicago
  - Detroit

Adleman´s Algorithm

- Input: A directed graph G with *n* vertices, and designated vertices v_{in} and v_{out}.
- Step 1: Generate paths in G randomly in large quantities.
- Step 2: Reject all paths that
  - do not begin with v_{in} and
  - do not end in v_{out}.
- Step 3: Reject all paths that do not involve exactly *n* vertices.
- Step 4: For each of the *n* vertices *v*:
  - reject all paths that do not involve *v*.
- Output: YES, if any path remains; NO, otherwise.

Vertex and Edge Encodings

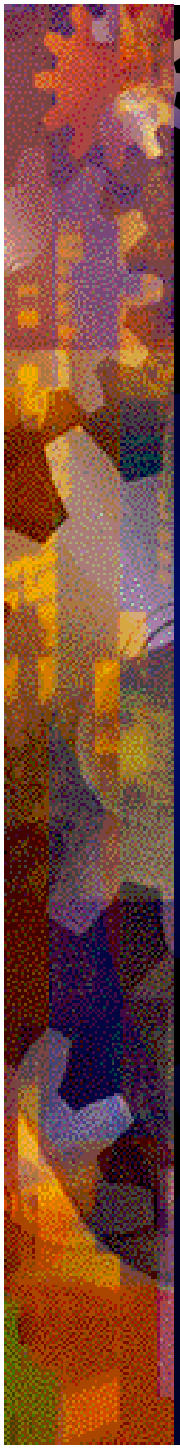
- Each city v_i is encoded by two sub-sequences: $v_i = v_i' \; v_i''$
  Each flight e_{ik} from v_i to v_k is encoded by: $e_{ik} = v_i'' \; v_k'$

Town	DNA Name	Complement
Atlanta	ACTTGCAG	TGAACGTC
Boston	TCGGACTG	AGCCTGAC
Chicago	GGCTATGT	CCGATACA
Detroit	CCGAGCAA	GGCTCGTT

Flight	DNA Flight Number
Atlanta – Boston	GCAGTCGG
Atlanta – Detroit	GCAGCCGA
Boston – Chicago	ACTGGGCT
Boston – Detroit	ACTGCCGA
Boston – Atlanta	ACTGACTT
Chicago – Detroit	ATGTCCGA

DNA Computation

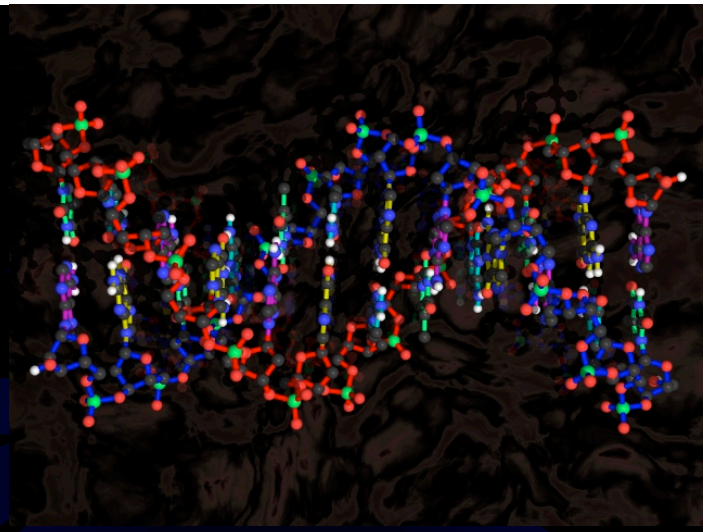
- The town complements and DNA flight numbers are used for computation.
- DNA molecules are put in a hydrous solution.
- Addition of ligase ensures catalysis of phosphodiester bonds.
- Shaking the test tube makes many DNA strands collide and interact.
- ~10^{14} computations are carried out in a single second.
- The solution strand has to be filtered from the test tube:
  - GCAG TCGG ACTG GGCT ATGT CCGA
  - Atlanta Boston Chicago Detroit



„DNA Computer“
Performance Evaluation

- Information density:
  - 10^{15} CDs per cm^3
- Massively parallel information processing:
  - 10^6 ops / sec for PCs
  - 10^{12} ops / sec for supercomputers
  - 10^{20} ops / sec possible for DNA
  - DNA computers would be > 1,000,000 times faster than any computer today.
- Energy efficiency:
  - $2 * 10^{19}$ operations per joule for DNA
  - 10^9 operations for silicon-based computers

Table of Contents



References

- Paun, G., Rozenberg, G., and Salomaa, A., *DNA Computing*, Springer,1998.