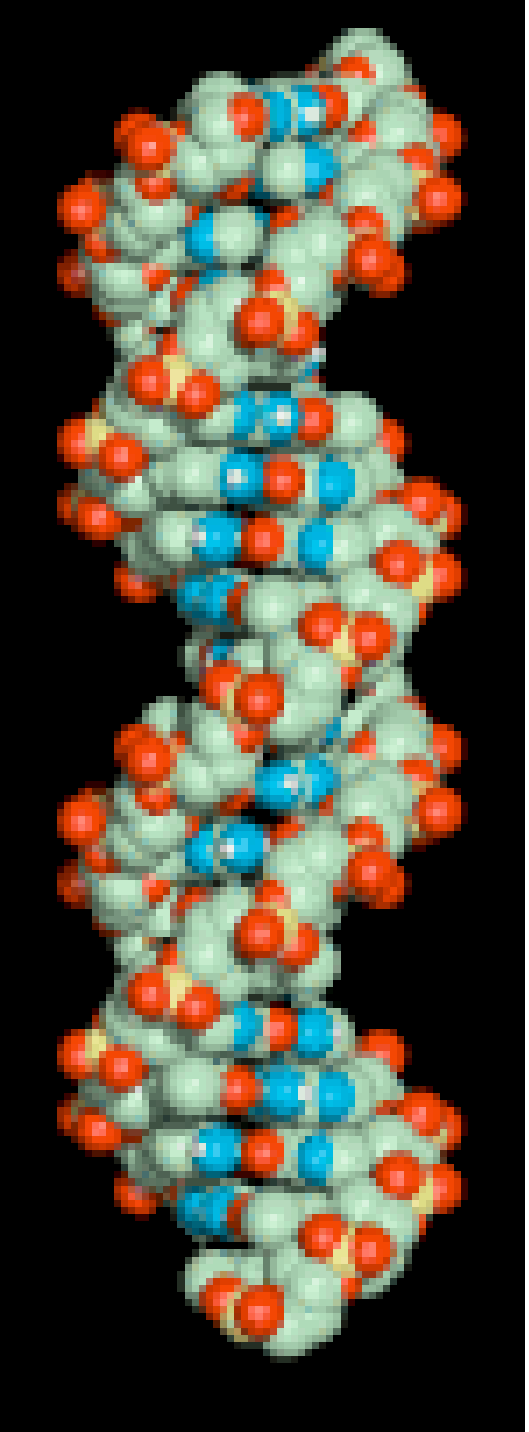


DNA Computing

Sources

Richard Spillman , Prateek Tandon , Lars Schäfer, Julian Miller, Thierry Metais, Westcott, Sanders Chong, Craig Jacobs, Daniel McFarlin, Russell Deaton, Charlie Moad, W. Joel Strait, Junghuei Chen, Byoung-Tak Zhang, Anton Lopatinsky, Nicholas Carter.

Biological Computing

Biochemical Computation

Computation Using DNA Molecules

(Bio)Molecular Computing

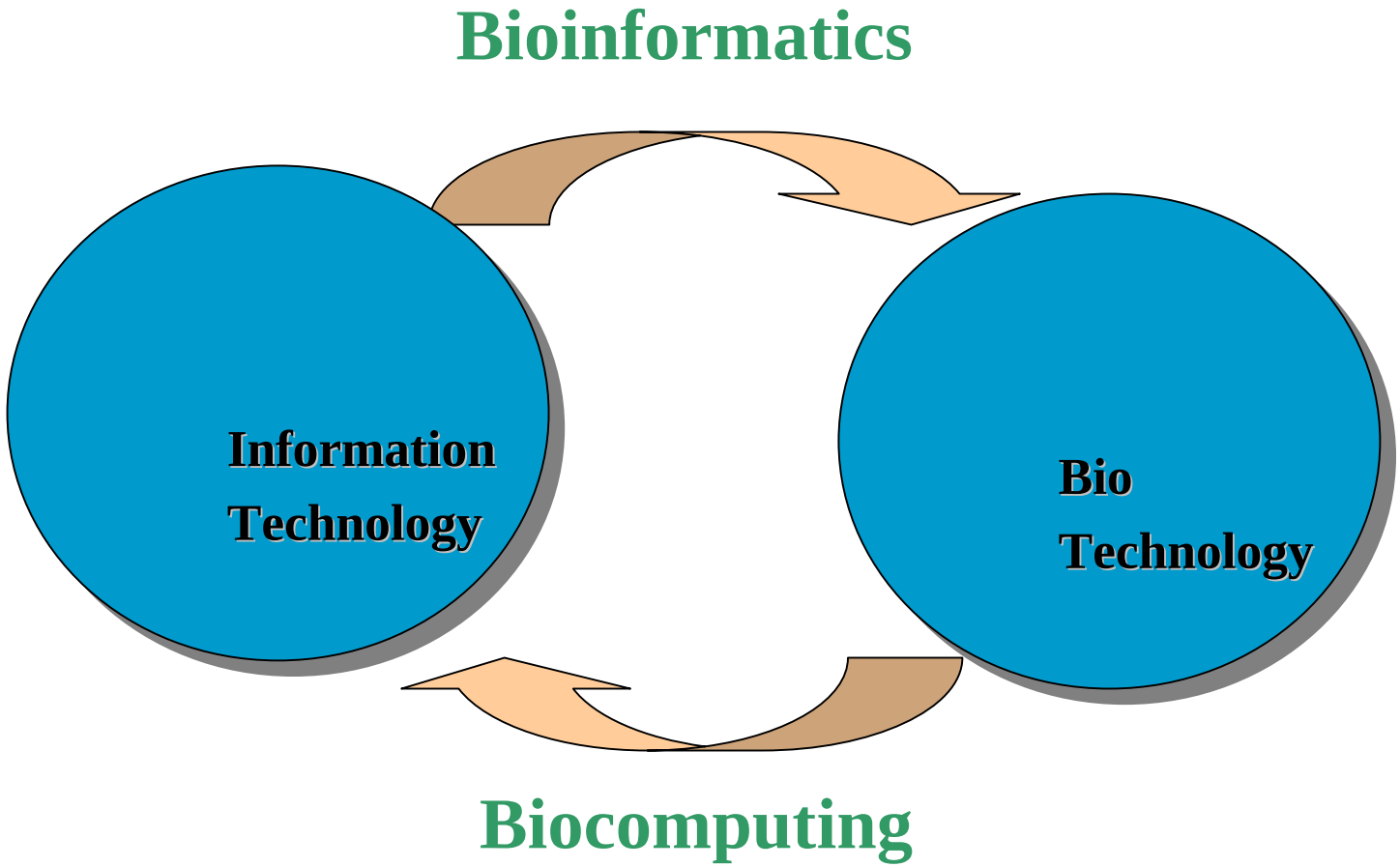
DNA Computing

Outline

- What Is Biological and what is Molecular Computing?
- What is DNA?
- DNA Computing Operators
- DNA Computing
- Examples
  - Hamiltonian Path of Adleman
  - Traveling Salesman
  - Molecular Theorem Proving
  - Molecular Optimization
- DNA versus Silicon
- The Difficulties of DNA Computers
- Applications of DNA Computing
- Future of DNA computing
- Conclusions
- Information Sources and Exam Problems

What is Molecular Computing?

Biocomputing vs. Bioinformatics



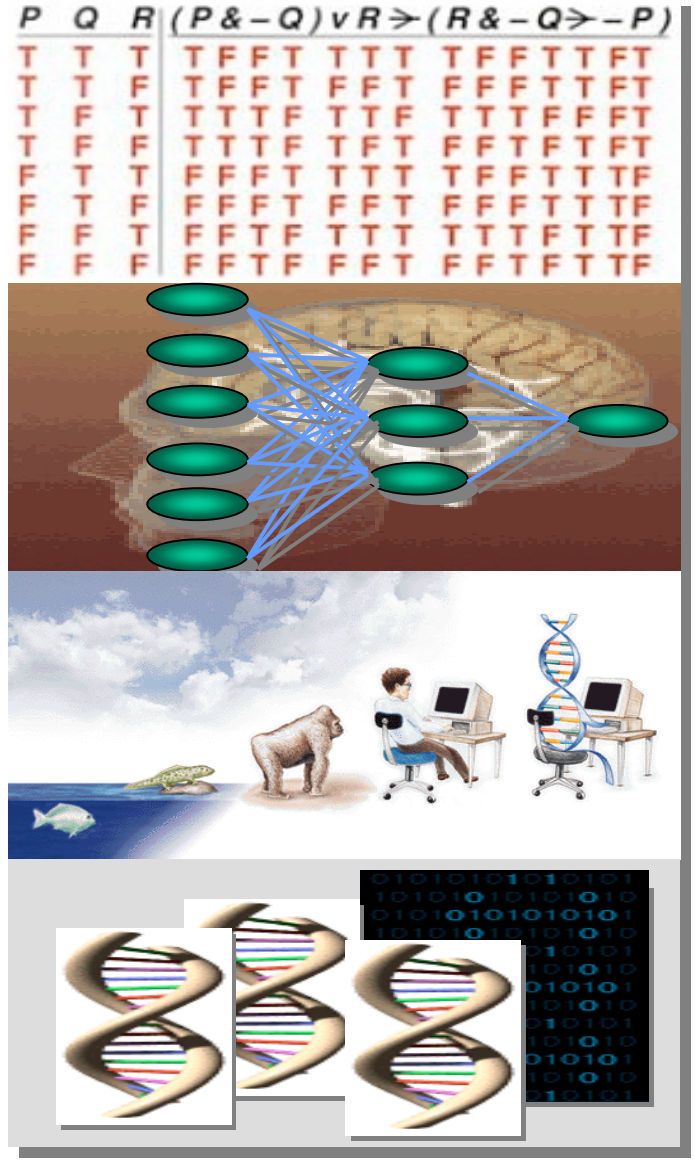
Artificial Intelligence (AI) and Biology

Symbolic AI
Rule-Based Systems

Connectionist AI
Neural Networks

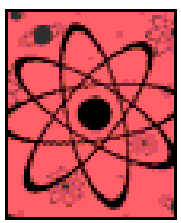
Evolutionary AI
Genetic Algorithms

Molecular AI:
DNA Computing

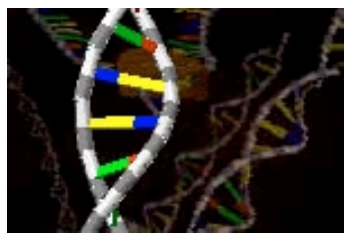


Nanotechnology

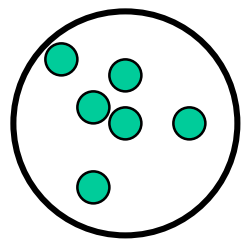
We are at the point of connecting machines to individual cells



Atoms
<1 nm



DNA
~2.5 nm



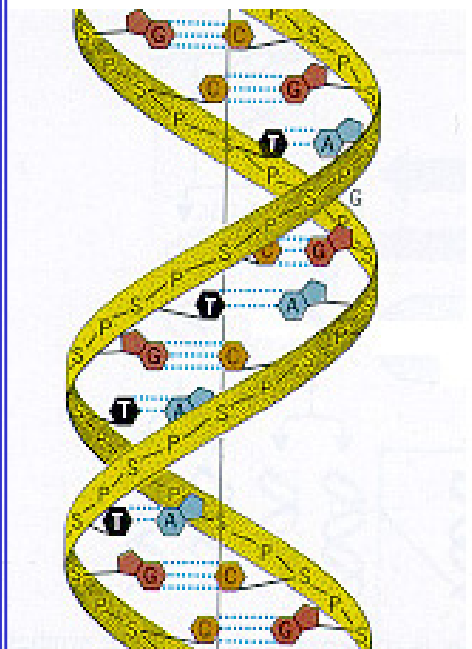
Cells
thousands of nm

Introduction to Molecular Biology

- In order to sustain life, biological systems have developed the ability to:
  - **to build three-dimensional molecules that assist in the critical functions of life (proteins, RNA).**
  - **to compress the information about how (and when) to build these molecules in a linear code (DNA).**

Broad Generalization: Proteins versus DNA

- **Protein molecules:**
  - do the work,
  - their 3D structure is critical.
- **DNA molecules:**
  - store the information,
  - don’t do much work,
  - their 3D structure is incidental.



Protein Molecules

- **Protein molecules are made up of linked subunits**
  - **These subunits are amino acids (also called protein residues for historical reasons).**
- **There are 20 different amino acids with different physical and chemical properties.**
- **The interaction of these properties allows a chain of the amino acids (upto 1000's long) to fold into a unique, reproducible 3D shape.**

DNA Action

- **The sequence of amino acids is determined by DNA**
- **DNA uses an alphabet of 4 letters (A,T,C,G), more commonly called bases.**
  - **Although the 4 letters have interesting chemical structure, for our purposes they are just information carriers.**
- **Long sequences of these 4 letters are linked together to create GENES and CONTROL INFORMATION.**

Enzymes

- Enzymes are biochemical catalysts
- Enzymes can:
  - cut,
  - copy,
  - paste,
  - error-check **DNA** (c.f. AND, OR, NOT, XOR, Rotate, Compare…?)
- Parallelism:
  - Enzyme molecules *do not act sequentially*
    - many enzyme molecules *act on many* DNA molecules at the same time

Mass Parallel operation of Enzymes

- Enzymes can work on many DNA nucleotides at once.
- The problem of replicating speeds up exponentially, and is a quite reliable process.

Genetic Code

- **Each of the twenty protein amino acids can be specified by 3 consecutive DNA bases.**

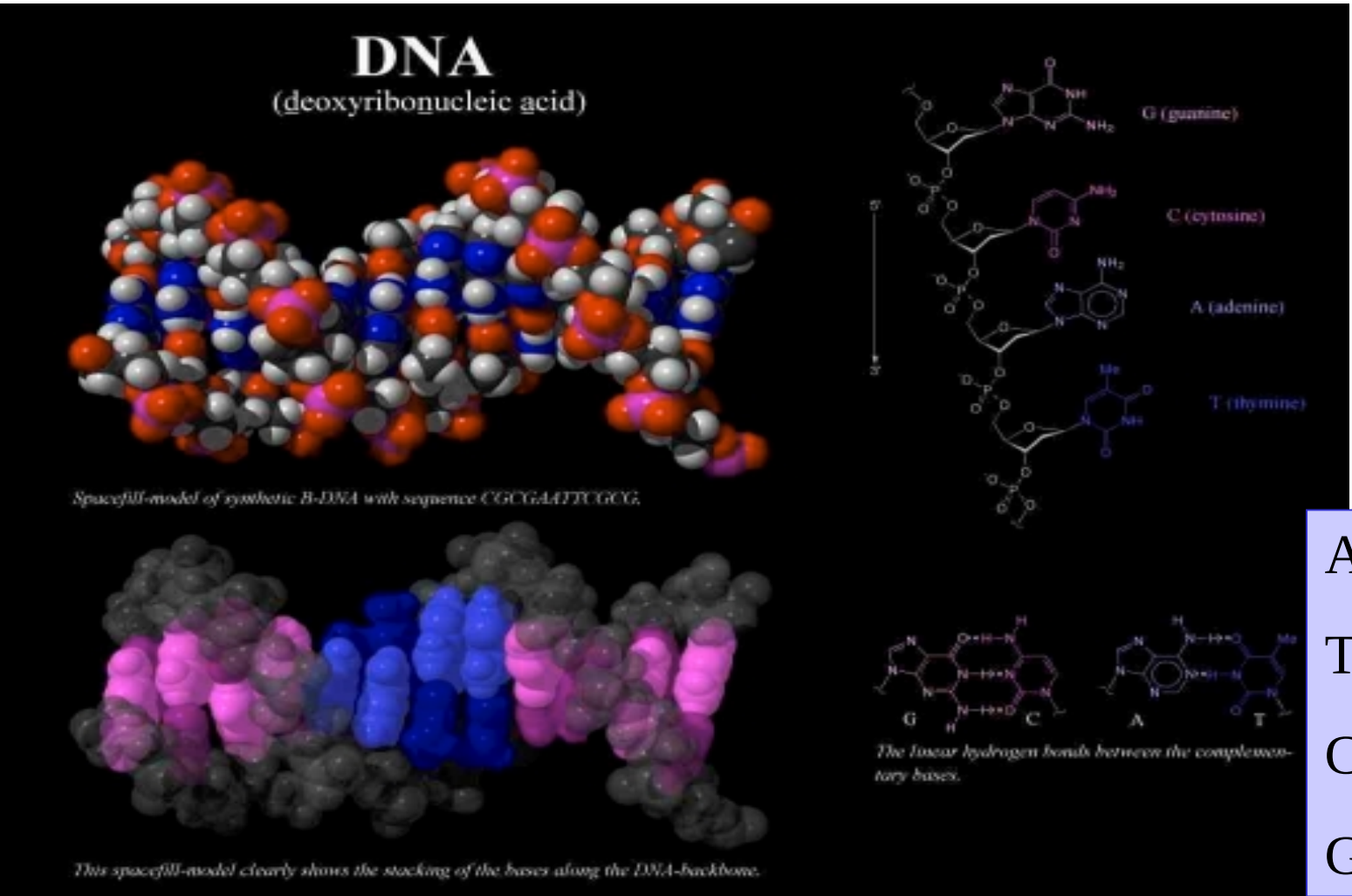
- **A cellular apparatus can:**
  - **"read" a sequence of DNA bases (three at a time) and**
  - **create the corresponding protein chain—which folds itself based on the amino acid properties.**

- **The 64 mappings of 3 bases to 1 amino acid is called the GENETIC CODE**
- **This code is universal (on earth...).**
- Present in nearly all living organisms
  - Some bacteria and viruses use RNA

Cells (and DNA) as Computers and Memories

- **Cells process and store information**
  - **DNA stores the information for the construction of proteins**
  - **This information is preserved and past on to other generations**
  - **DNA serves as memory:**
    - **RNA performs a memory access,**
    - **ribosomes serve as processors**

- The life's molecule:



A = adenine

T = thymine

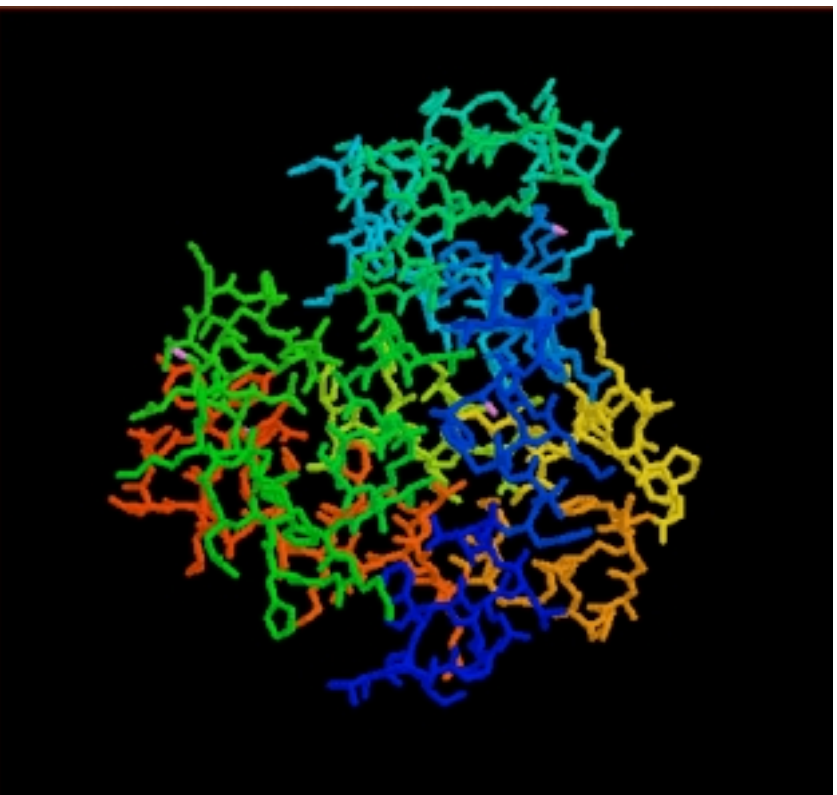
C = cytosine

G = guanine

Myoglobin Gene

- **Gene for Myoglobin**
  - ctgcagataa ctaactaaag gagaacaaca acaatggttc tgtctgaaggtgaatggcag ctggttctgc atgtttgggc taaagttgaa gctgacgtcg ctggtcatgg tcaggacatc ttgattcgac tgttcaaatc tcatccggaa actctggaaa aattcgatcg tttcaaacat ctgaaaactg aagctgaaat gaaagcttct gaagatctga aaaaacatgg tgttaccgtg ttaactgccc taggtgctat ccttaagaaa aaagggcatc atgaagctga gctcaaaccg cttgcgcaat cgcatgctac taaacataag atcccgatca aatacctgga attcatctct gaagcgatca tccatgttct gcattctaga catccaggta acttcggtgc tgacgctcag ggtgctatga acaaagctct cgagctgttc cgtaaagata tcgctgctaa ctgggttacc agggttaatg aggtacc
  - **BASE COUNT 155 a 108 c 115 g 129 t**

- **Myoglobin Protein sequence**
  - **MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEM KASEDLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHV LHSRHPGNFGADAQGAMNKALELFRKDIAAKYKELGYQG**



Myoglobin Protein

A = adenine

T = thymine

C = cytosine

G = guanine

What is Biological Computing?

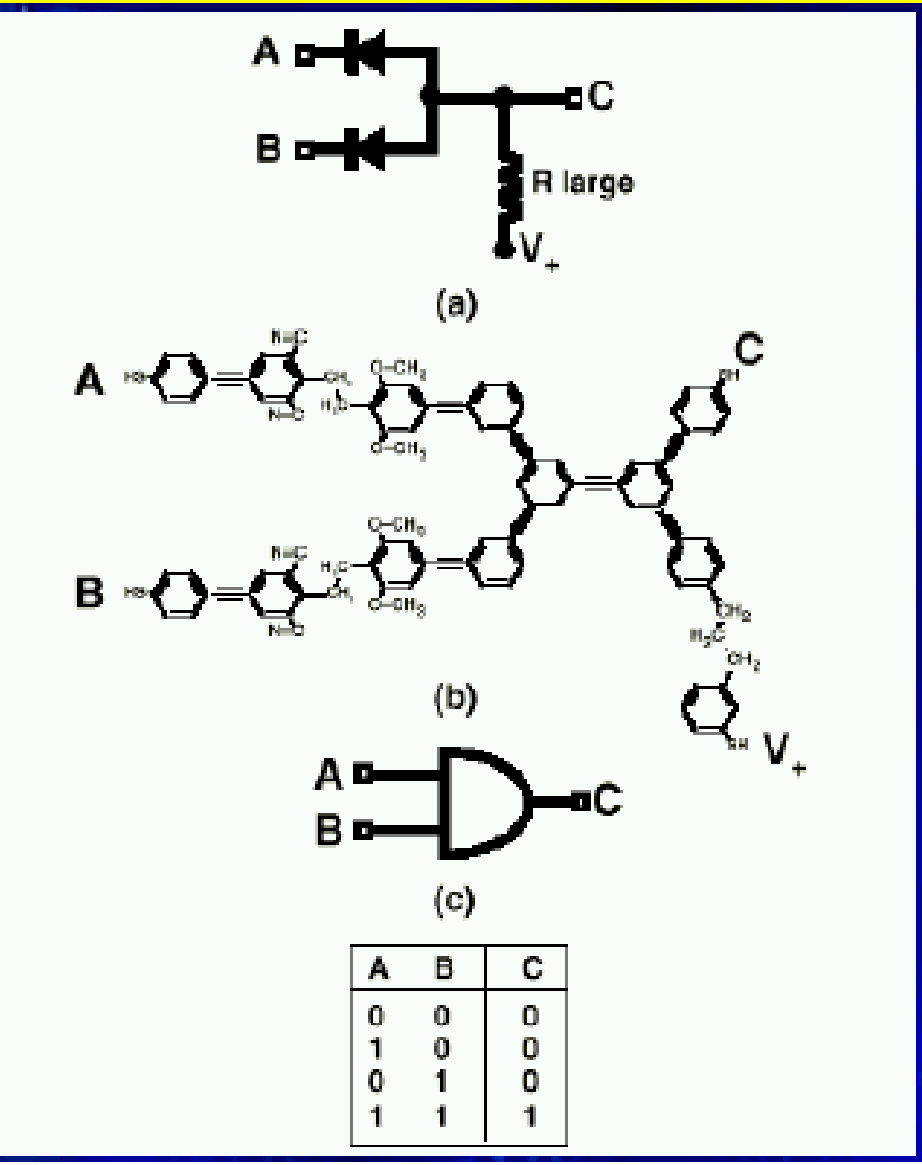
- Using biological processes to perform computation
  - Most efforts focus on DNA-based reactions
- Tremendous parallelism
- Slow cycle times
  - Biological gates may operate at mili-Hertz frequencies
- Leverage ability of biological processes to manipulate matter at the molecular level
- Biological Computing is also an Interface with an organism that is studied and/or manipulated

Approach 1: Construct Logic Gates Within Cells

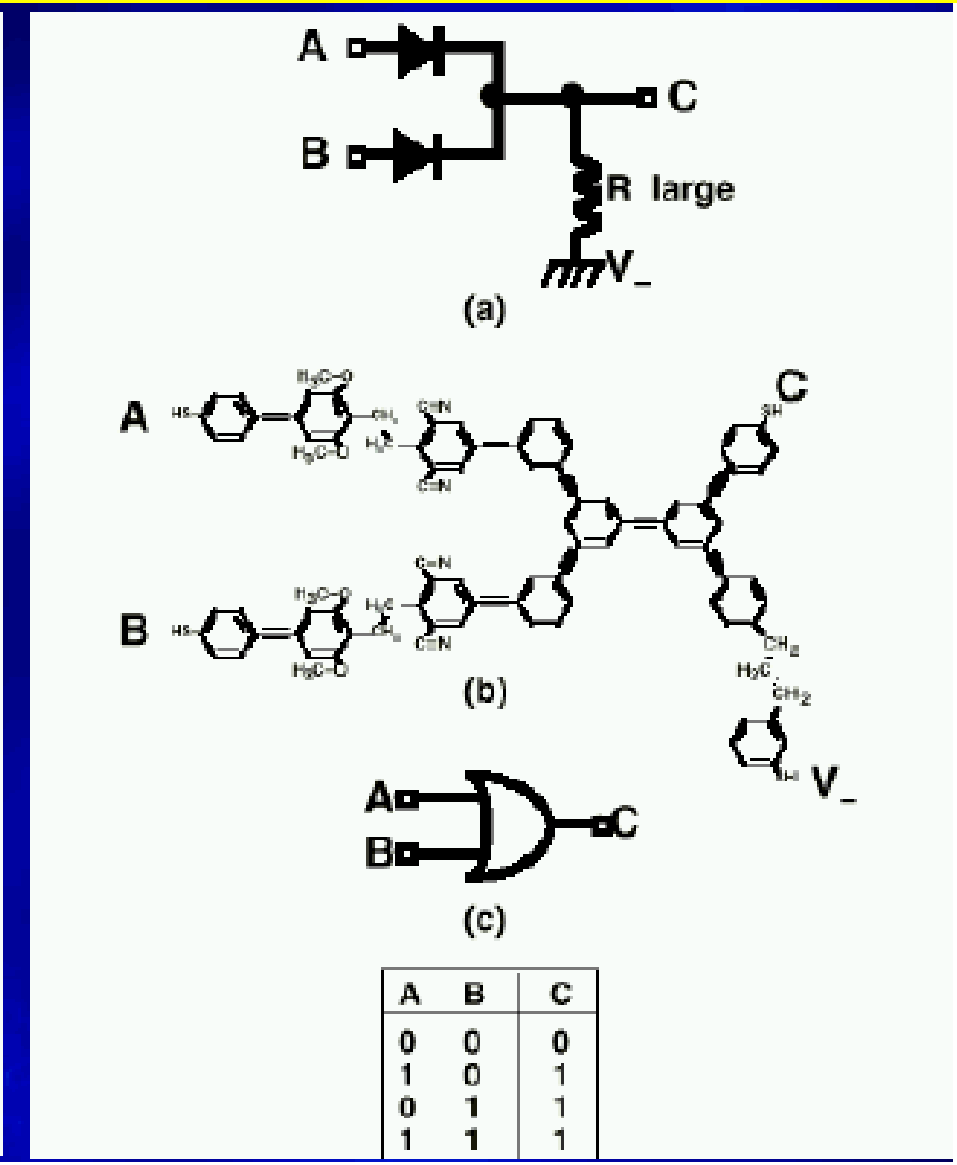
- **Basic question:** Can we use biological processes within cells to implement logic gate-like functions?

- **Things to consider:**
  - What to use to carry signals?
    - Need to be able to generate
    - Need to be able to use as input
    - Need underlying process with high gain for 0-1 mapping
  - How to input signals to the system?
  - How to read outputs?

Molecular Electronics: AND Gate, OR Gate



A	B	C
0	0	0
1	0	0
0	1	0
1	1	1



A	B	C
0	0	0
1	0	1
0	1	1
1	1	1

- Picture form MITRE Corp

First Approach: Results

- Use levels of specific proteins as signals
  - Have minimum and maximum concentrations
  - Can “program” cells to produce them
  - Relatively independent – multiple signals within a cell
- Chemical reactions give good **inverter-like** response
  - High gain
  - Saturation at 0 and 1
  - Relative insensitivity to operating conditions
  - Use different **dimers** to control ***“gain”***

- Issues
  - Slow (mili-Hertz)
  - No demonstratiof actual implementation
- Application
  - Use of cells to fabricate nanotech-scale devices

Approach 2: Use DNA to Perform Computations

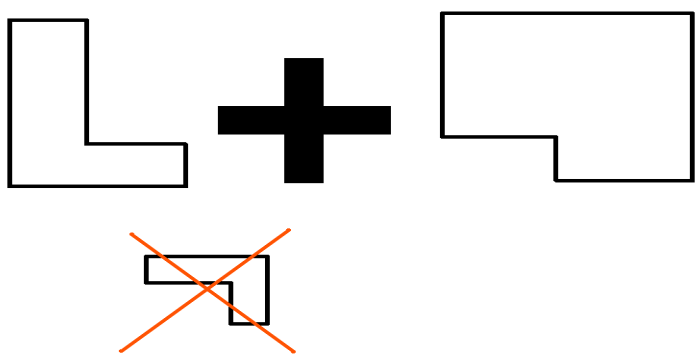
- This is what people generally talk about when they talk about biological computing
- Basic Idea:
  - Encode all possible inputs to an operation as strands of DNA,
  - then sort out input strands that ***match a particular criterion***
    - Useful for "search" problems, which tend to be NP-complete
    - Vast parallelism – many strands of DNA/tube
    - Relatively slow -- ~15 minutes/step

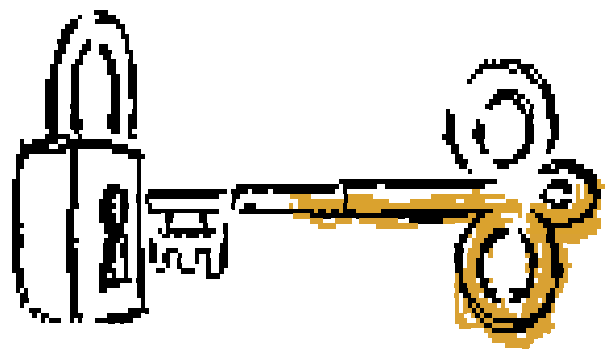
What is Molecular Computing?

The use of biological molecules, primarily DNA, DNA analogs, and RNA, for computation purpose.

What is Molecular Computing?

- **Answer:** Systems in which **macromolecules** mediate information-processing: ***e.g. biological organisms***
- Macromolecules are mostly proteins, DNA
- Underlying principle: Macromolecules can **recognize** *specific molecules* using physiochemical content (***lock-key*** model)





What is DNA computing ?

- History of DNA computing.
  - Around 1950 first idea (precursor Feynman)
  - First important experiment 1994: Leonard Adleman

- **Massive Parallelism** through simultaneous biochemical reactions
  - The concept of base - used for computing
  - In a liter of water, with only 5 grams of DNA we get around **10^{21} bases** !
  - Each DNA strand represents a processor !
- ***In Vitro*** Selection and Evolution
- **Satisfiability** and **Hamiltonian Path** **as examples**

What is DNA computing ?

- Molecular level (just greater than 10^{-9} meter)
- DNA can perform computations
  - Huge information **storage density**
    - Stores information (up to 10^{8} TByte in 1 liter)
  - Allows manipulation
- DNA is tiny and "fast"
  - ***New computing paradigm***:
    - **Try all** possible solutions (parallel)
    - **Remove** wrong solutions

DNA Computing is A Different Approach

- **What DNA computing is:**
  - a completely new method among a few others

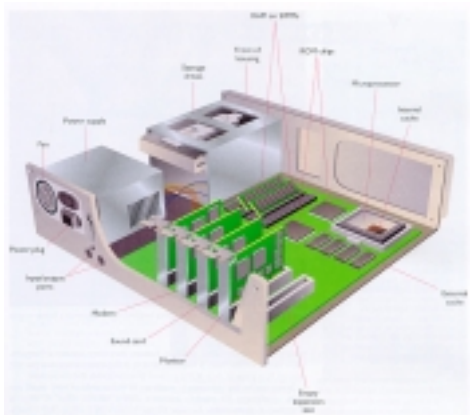
    (e.g., quantum computing) of general computation alternative to electronic/semiconductor technology
  - uses biochemical processes based on **DNA**

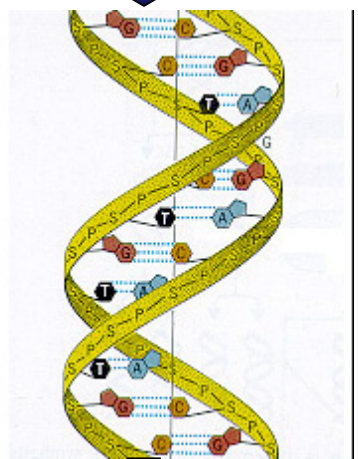
- **What DNA computing isn't:**
  - not to confuse with ***bio-computing***

    which applies biological laws (evolution, selection)

    to computer algorithm design.

DNA Computing





011001101010001

ATGCTCGAAGCT

DNA Computing: What are the Advantages?

- 1. Computes in parallel
- 2. Compact information storage
- 3. Low energy consumption

….it is possible thanks to…….

- Great advances in molecular biology and laboratory techniques:
  - PCR (Polymerase Chain Reaction)
  - DNA Microarrays
  - New enzymes and proteins
  - Better understanding of biological molecules
- Ability to produce massive numbers of DNA molecules with specified sequence and size
- DNA molecules interact through template matching reactions

What is DNA?

What is DNA?

- DNA contains the basic information used to replicate cells and proteins
- **Protein manufacture** is a two-step process:
  - **1.** Information in DNA is copied into RNA
  - **2.** RNA copy then used to make proteins
- The most common control point is the ***DNA->RNA copying***
  - **Protein binding** *to particular sites* within DNA interferes with copying of adjacent information into RNA
- Proteins are continually degraded within cells
  - To **maintain** a given protein concentration, cell has to **keep producing it**

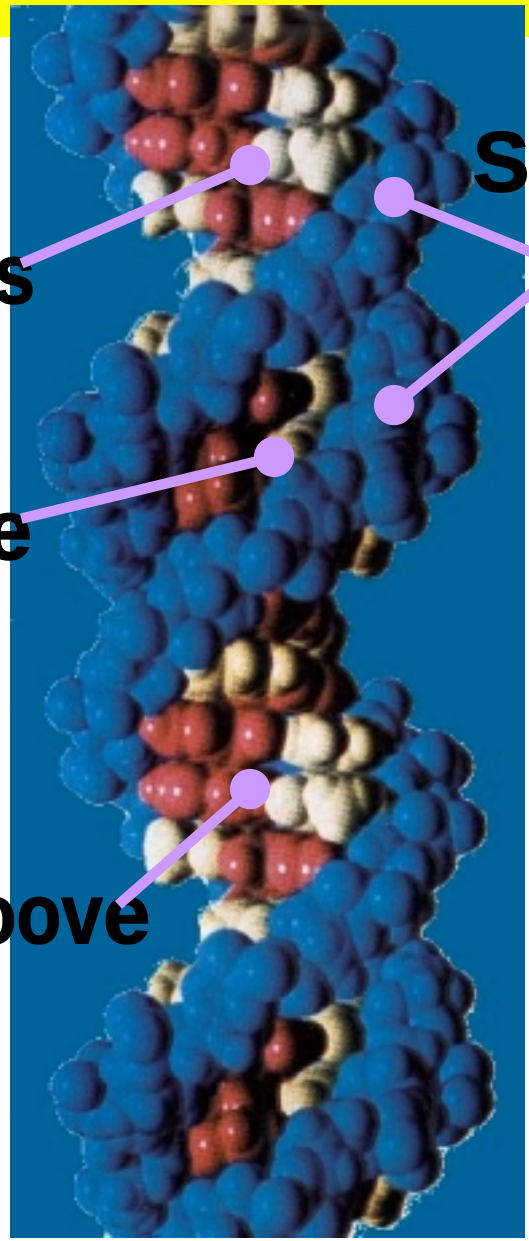
Various People that are of interest to DNA study

- Rosalind Franklin
- Maurice Wilkins
- Linus and Peter Pauling
- Probably others

- James D. Watson
- Francis H. C. Crick

There are **complementarities** between bases (Watson-Crick).
(A)←→(T)
(C)←→(G)

The Double Helix



Sugar/Phosphate Backbone

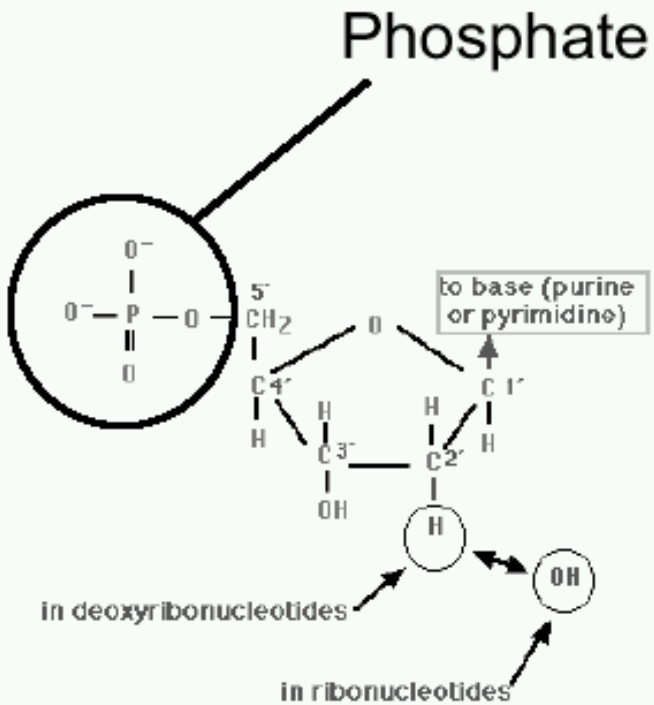
Base Pairs

Minor Groove

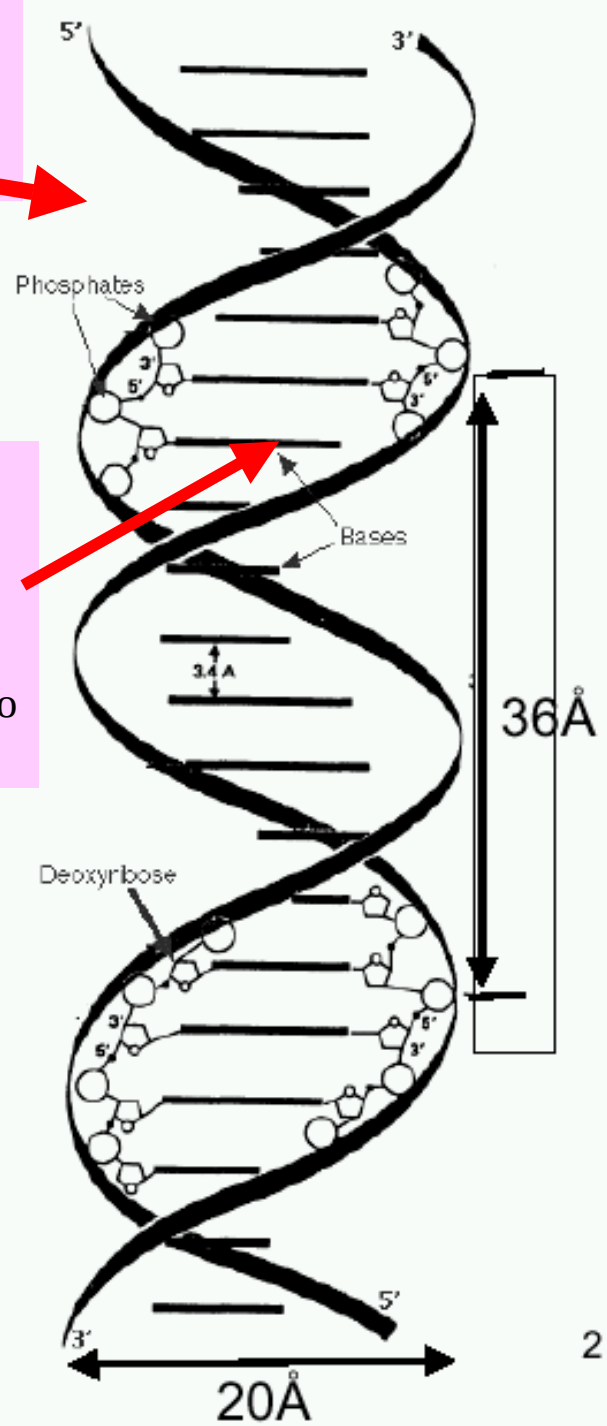
Major Groove

DNA is the hereditary molecule in every biological cell. Its shape is like a twisted rubber ladder (i.e. a double helix).

The rungs of the ladder consist of two bonded molecules called **bases,** of 4 possible types, labeled **G, C, A, T.**

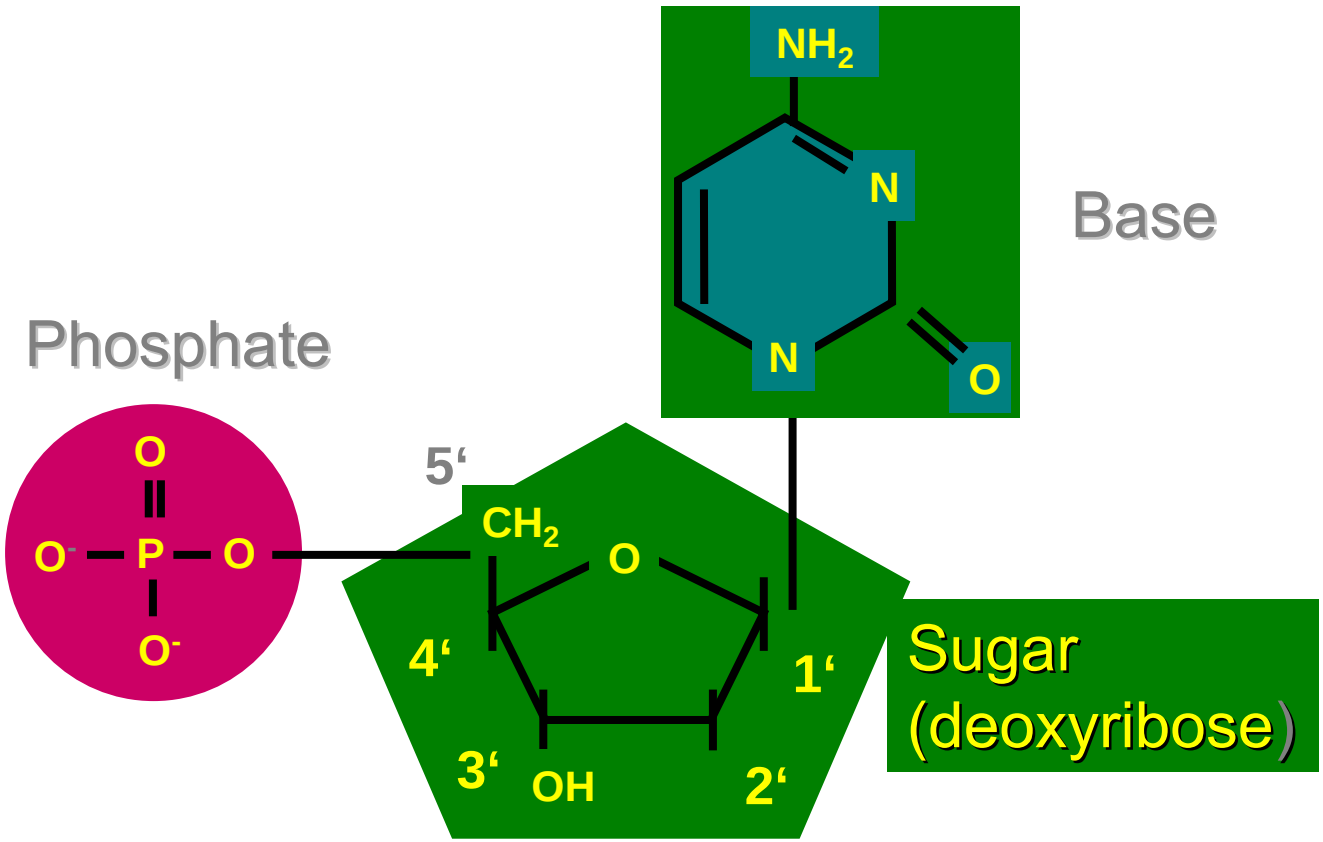


Base Pairs are on the inside.
Base Pairs are perpendicular to the helical axis



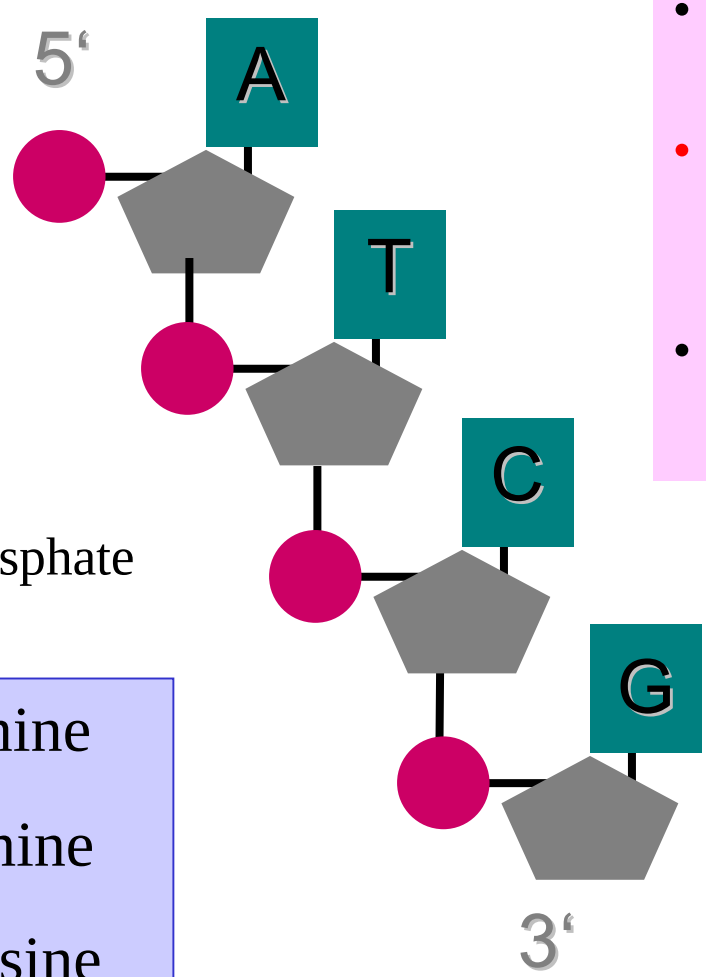
Double Helix

DNA Structure



Nucleotide (DeoxyriboNucleic Acid)

DNA Structure - Single Strand



Sugar-phosphate backbone

A = adenine

T = thymine

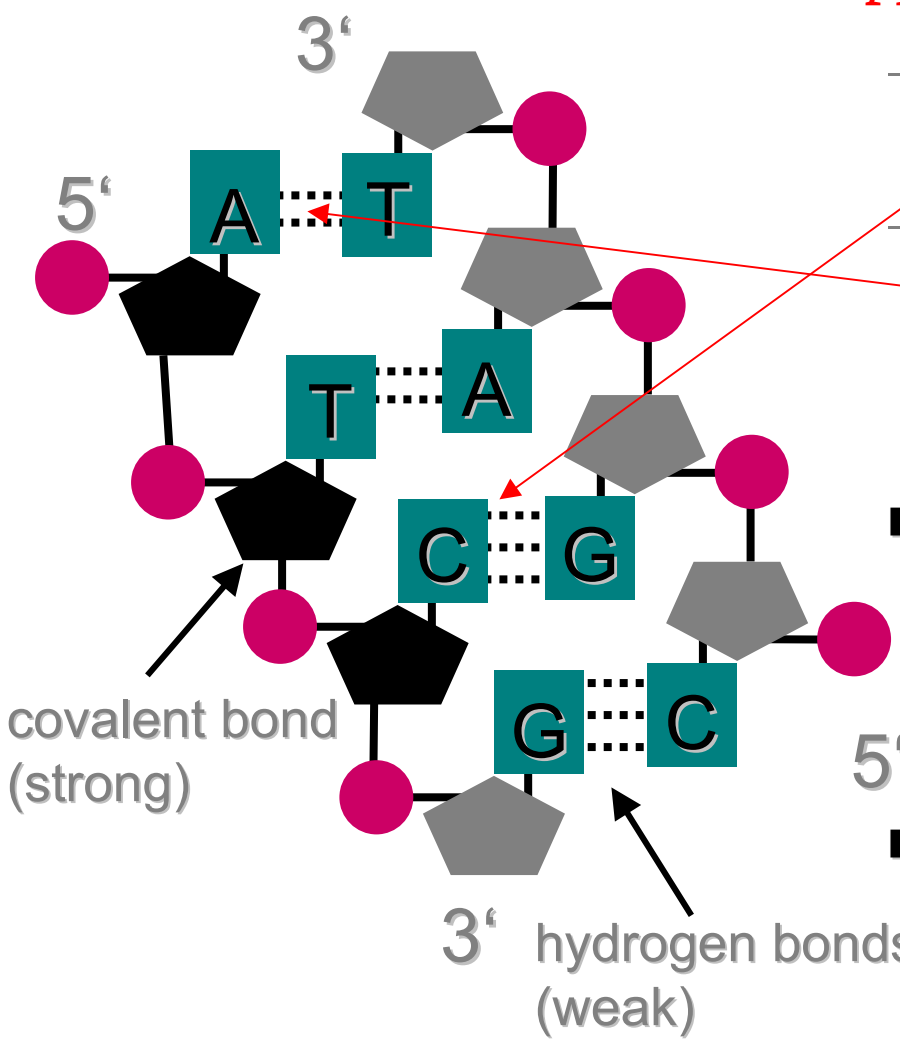
C = cytosine

G = guanine

- 4 bases (A, T, C, G) 4 states
- Bases sequenced *in different orders* to represent different information
- Double strands are ***complementary***:
  - **A** and **T**,
  - **G** and **C** (*Base pairs*)
- Complementarity serves purpose of error checking

- Distinctive ends
  - 5‘ (5-prime) end
  - 3‘ end
- Sequence
  - 5‘-ATCG-3‘

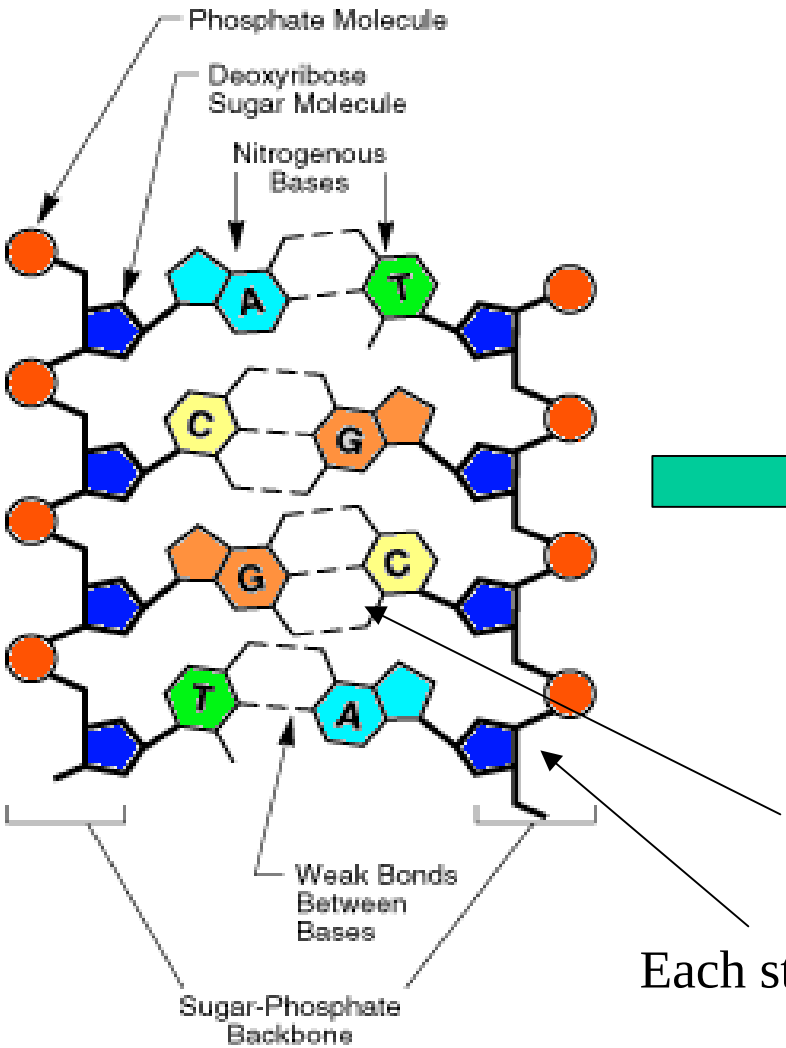
DNA Structure – Double Strand



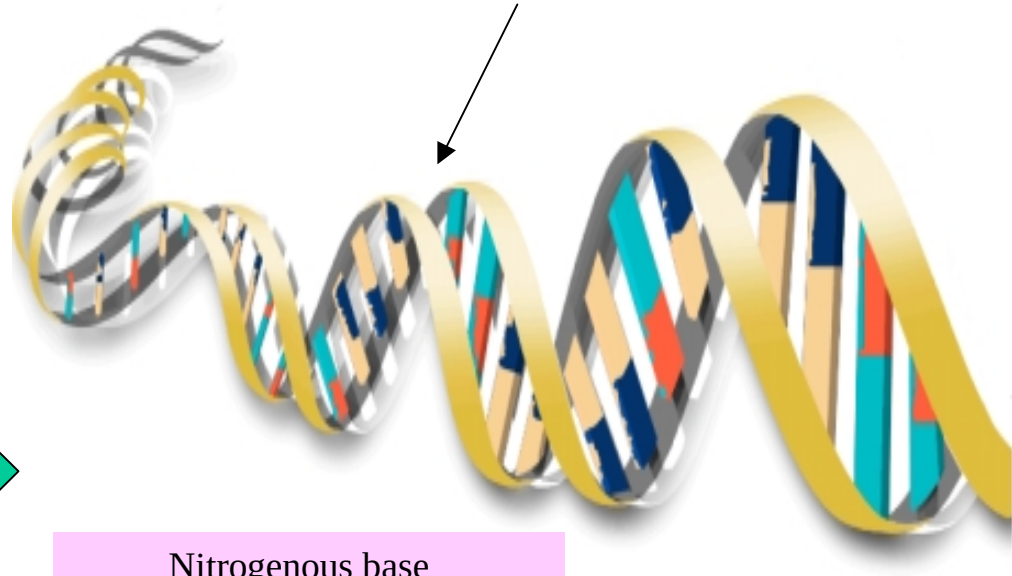
- **Annealing**
  - G lines up with C
    - 3 hydrogen bonds
  - A lines up with T
    - 2 hydrogen bonds

- Watson-Crick complement
  - 5‘-ATCG-3‘
  - 3‘-TAGC-5‘
- "Double Helix"
  - 2 intertwined strands

What is DNA? Conclusion



The DNA is a double stranded molecule



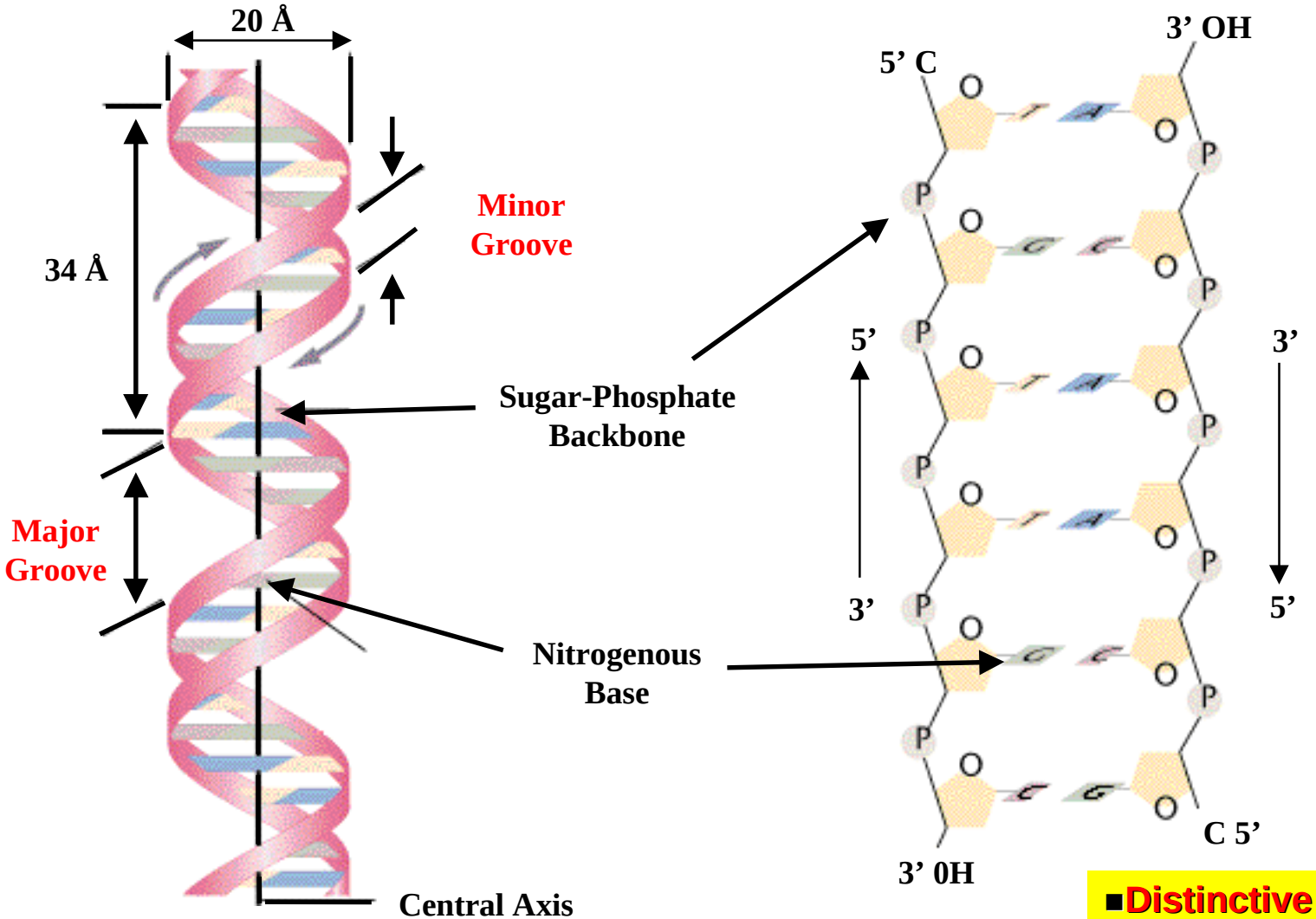
Nitrogenous base
- Adenine (A)
- Thymine (T)
- Guanine (G)
- Cytosine (C)

The linkage between bases has a **direction**

Each strand is based on 4 bases (A,T,C,G):

bases are linked through a sugar (desoxyribose)

More detailed look at PHYSICAL STRUCTURE OF DNA and distinctive ends 5', 3'



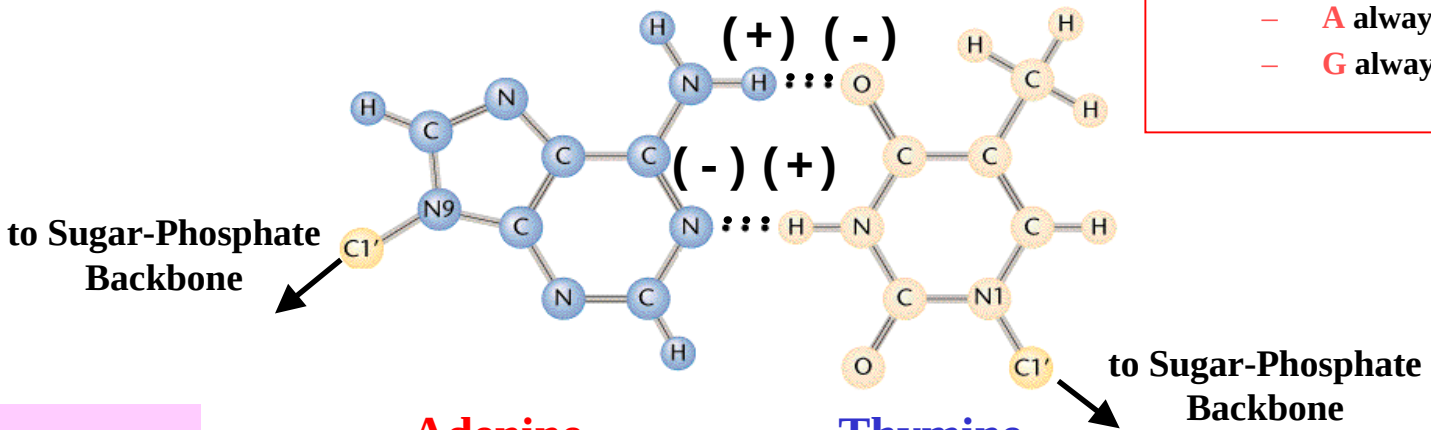
- **Distinctive ends**
  - 5' (5-prime) end
  - 3' end

Nitrogenous Bases

- Two Types
  - Pyrimidines
    - six-membered ring
    - Cytosine and Thymine
  - Purines
    - five and six-membered rings sharing a side
    - Guanine and Adenine

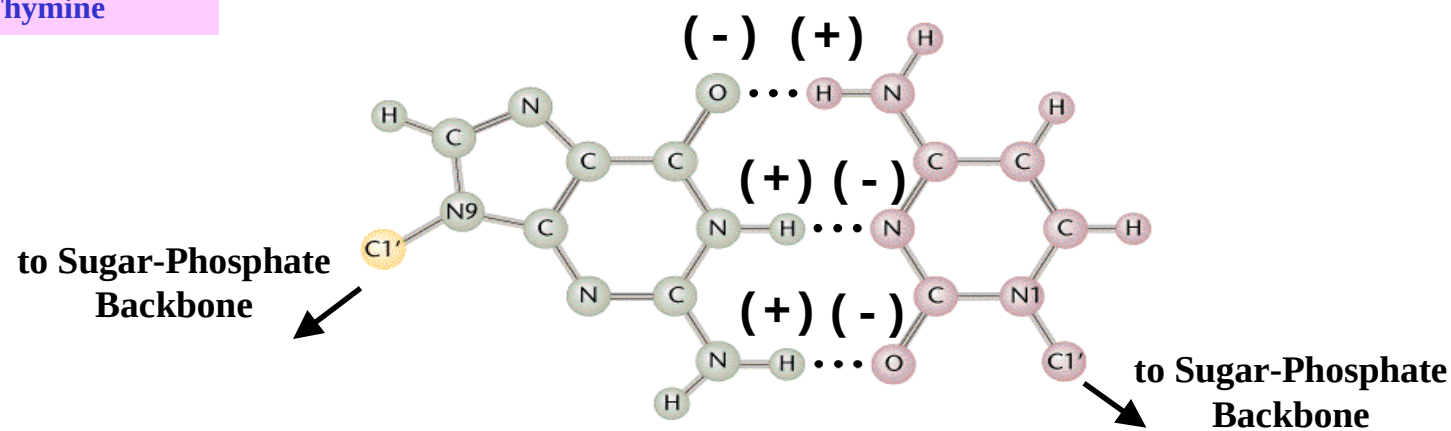
Inter-strand Hydrogen Bonding

- **Chargaff's Rule**
  - **Amount of A=T and amount of C=G**
  - **Purine:Pyrimidine = 1**
- **Purines always pair with pyrimidines**
  - **A always pairs with T**
  - **G always pairs with C**



Pyrimidines
Cytosine and Thymine

Adenine **Thymine**



Guanine **Cytosine**

••• **Hydrogen Bond**

DNA as a Data Structure

- DNA is base 4 (G, T, C, A)
- Nucleotides are spaced every 0.35nm.
- 18 Mbits/inch and over 1 million Gbits/in^2.
- A factor of 100,000 times smaller than a typical hard drive.

- Quarternary data structure
- Billions of bases (≈bits)
- 100,000 times higher data density that traditional storage media

- Lots of "agents". (strands)
- Tools provided by nature. (enzymes)
- How can we use all this?

- If we want to use DNA as an information bulk, we must be able to manipulate it .
- However we are talking of handling molecules…

Using Enzymes for DNA manipulations:

- **ENZYMES** = Natural **CATALYSERS**.
- So instead of using physical processes, we would have to *use natural ones*, more effective:
  - for **lengthening: polymerases**…
  - for **cutting**: **nucleases** (exo/endo-nucleases)…
  - for **linking**: **ligases**…
- ***Serialization:*** 1985: Kary Mullis → **PCR**

  Thanks to this reaction we get millions of identical strands, and we are allowed to think of massive parallel computing.

Conclusion: DNA Computing Takes Advantage of these:

- **1.** Our ability to produce massive numbers of DNA molecules with specific properties (size, sequence)
- **2.** The natural proclivity of specific DNA molecules to chemically interact **according to defined rules** to produce new molecules
- **3.** Laboratory techniques that allow the

  isolation/identification of product molecules with specific properties
  - PCR - Polymerase Chain Reaction,
  - Ligation,
  - Gel Electrophoresis,
  - etc.

DNA operations

Computation with DNA: operations

- **Construct a DNA molecule** for each potential solution:
  - **1.** Generate candidate solutions in parallel
  - **2.** Use molecular operations to eliminate invalid solutions

- **Five basic "mathematical" operations**:
  - ***Extract***: separates DNA containing a particular substring
  - ***Merge***: mixes two tubes of DNA to label solutions
  - ***Detect***: checks if there are any labeled DNA strands
  - ***Copy***: amplifies the labeled strands
  - ***Append***: Attaches a string to the end of every molecule in a tube

DNA Computer: Physical Operations

- Storage: DNA/RNA strands – more information/molecule (efficient packing)

- Amplification (PCR),
- Cutting (Enzymes),
- Adjoining (Hybridization/ligation),
- Filtering (biotin-avidin beads)
- Reading: Sequencing, Gel-Electrophoresis

- Input: DNA strand creation, solution mixing

The bio-lab technology: DNA processes

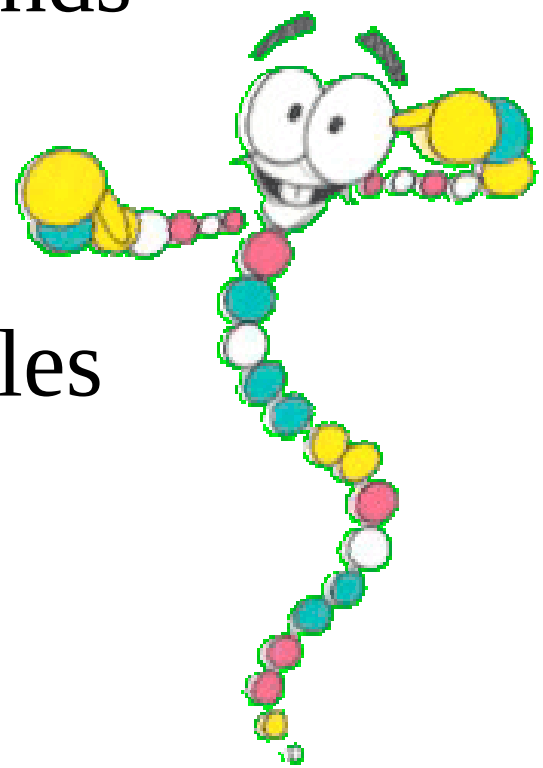
- **1.** Hybridization
- **2. Ligation**
- **3.** Polymerase Chain Reaction (**PCR**)
- **4.** Gel **electrophoresis**
- **5.** Affinity separation (**Bead**)
- **6. Enzymes**: restriction enzyme…

DNA Operations:

P1. Separating DNA strands (denaturation)

DNA operations:

- **P2. Binding** together DNA strands
  - (**renaturation** or **annealing**)
- **P3.** Completing *sticky ends*
- **P4. Synthesizing** DNA molecules

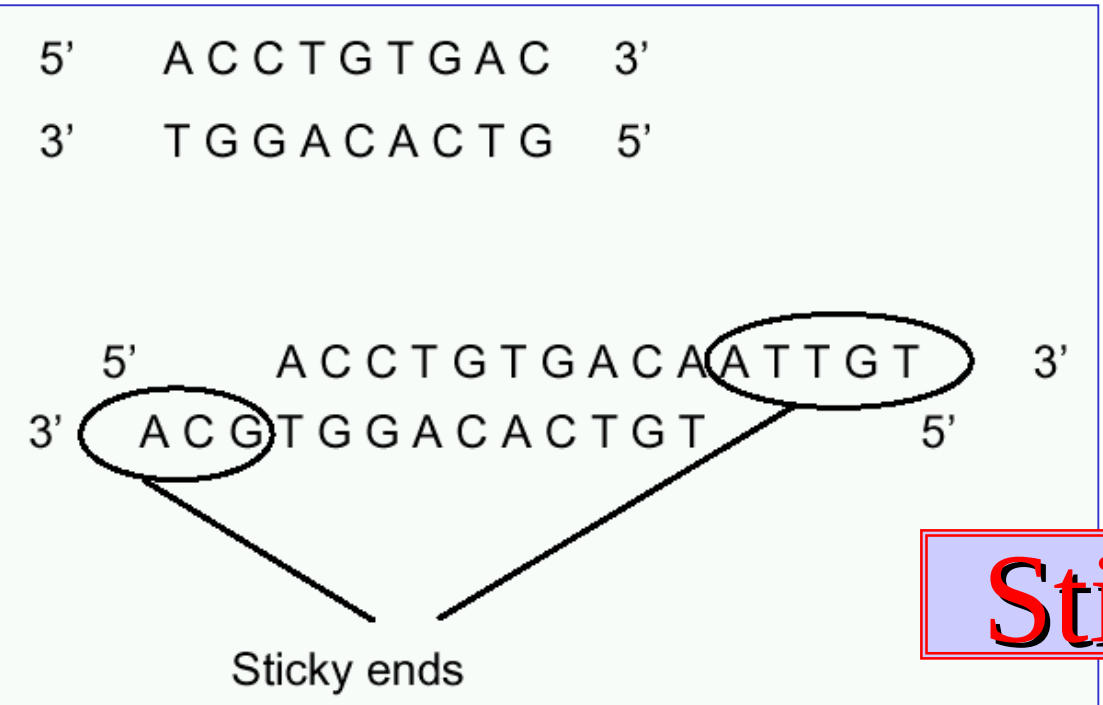


….what are sticky ends ?…...

P3. Completing Sticky Ends

If a single **strand (string)** of DNA is placed in a solution with isolated bases of A, G, C, T, then those bases will **pair off** with the bases in the string, and form a complementary string, e.g.

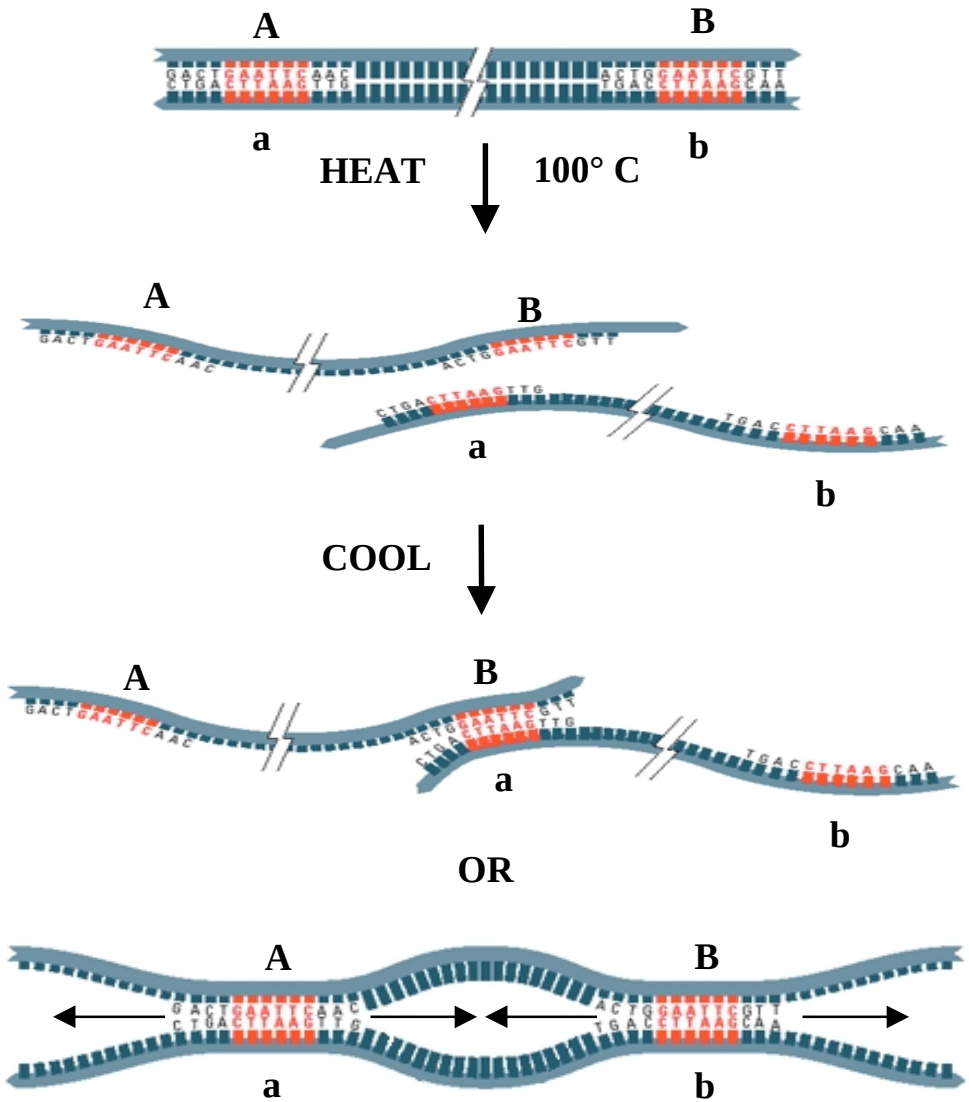
G A T T C A G A G A T T A T
C T A A G T C T C T A



DNA Code

Sticky Ends

1. STRAND HYBRIDIZATION

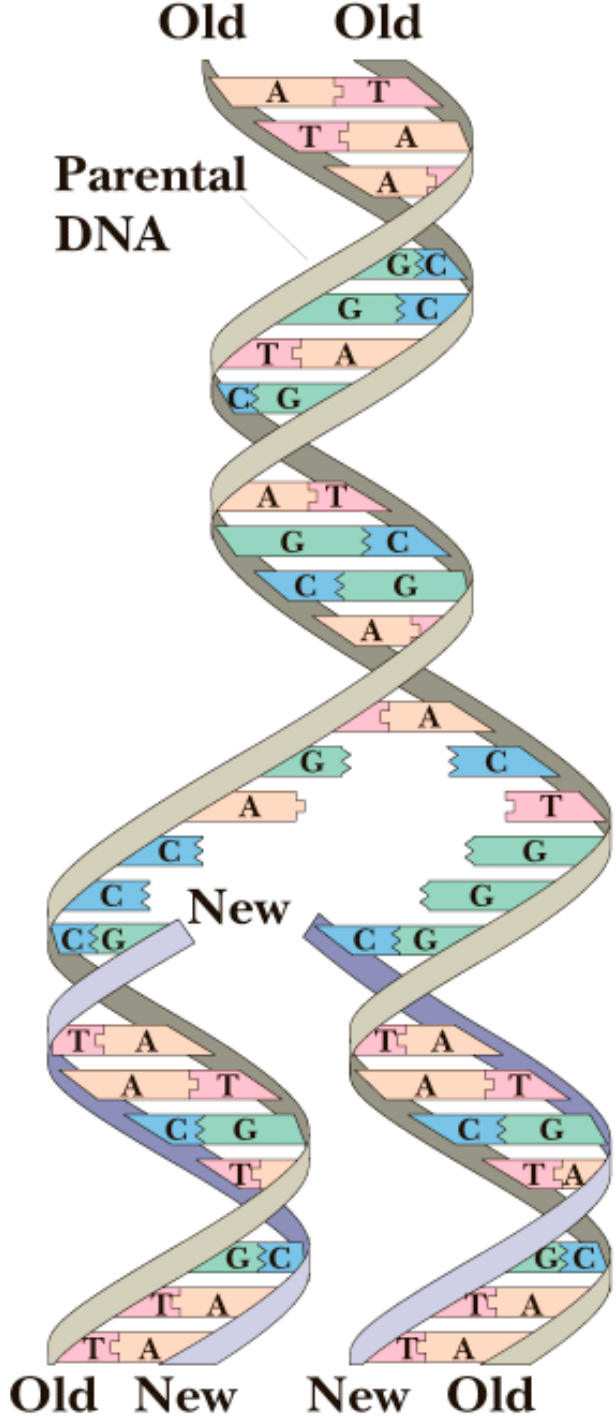


DNA operations

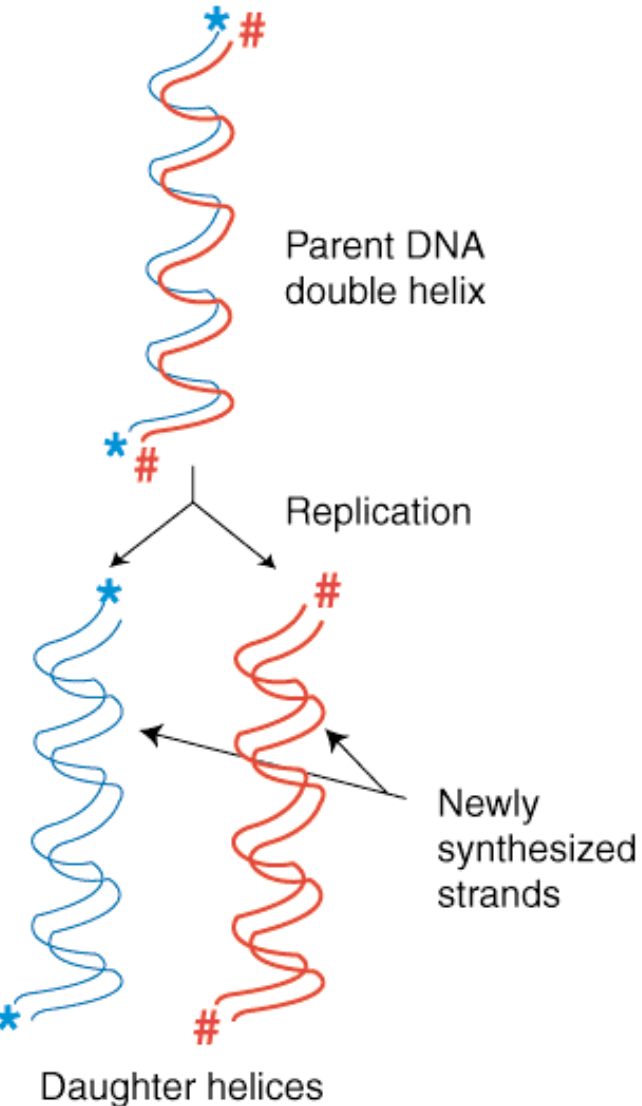
- **O1. Linking** DNA molecules
- **O2. Inserting** or **deleting** short subsequences
- **O3. Multiplying** (***replicating***) DNA molecules

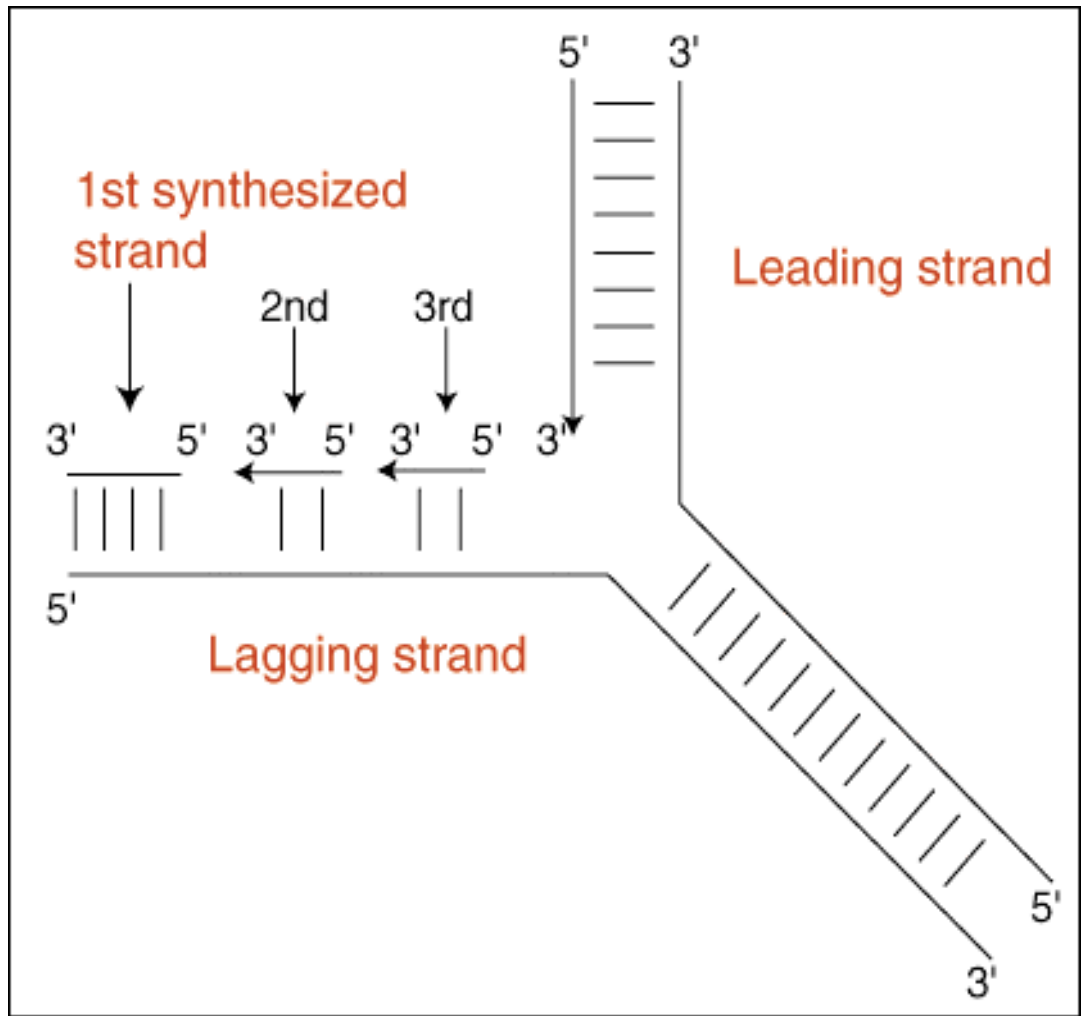
O3. DNA Replication

- DNA replication is "semi-conservative"
  - 2 daughter strands each contain half the parent strand
- Many enzymes are involved
  - **Polymerase(s), Ligase, Helicase**
- Enzymes operate in parallel

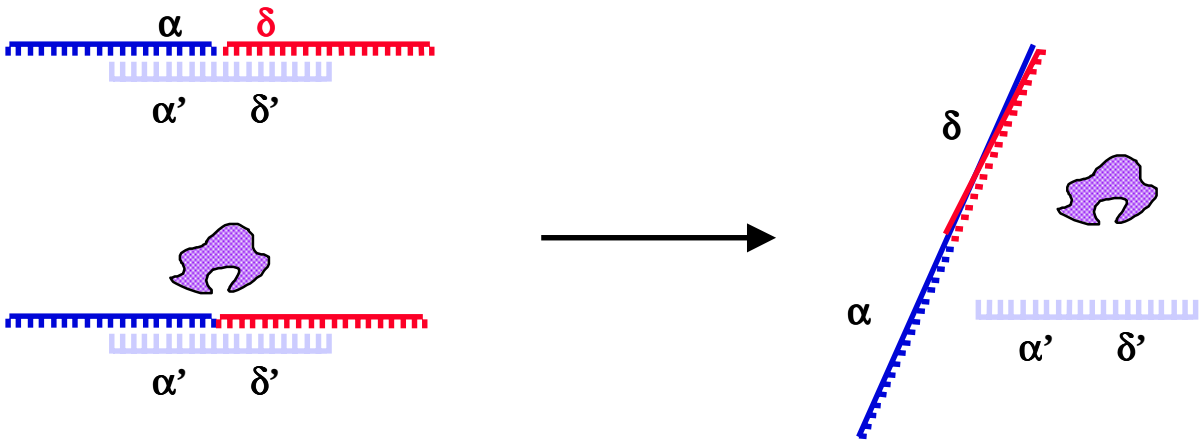


DNA Replication





2. DNA LIGATION



Ligase Joins 5' phosphate to 3' hydroxyl

anneal = bind together

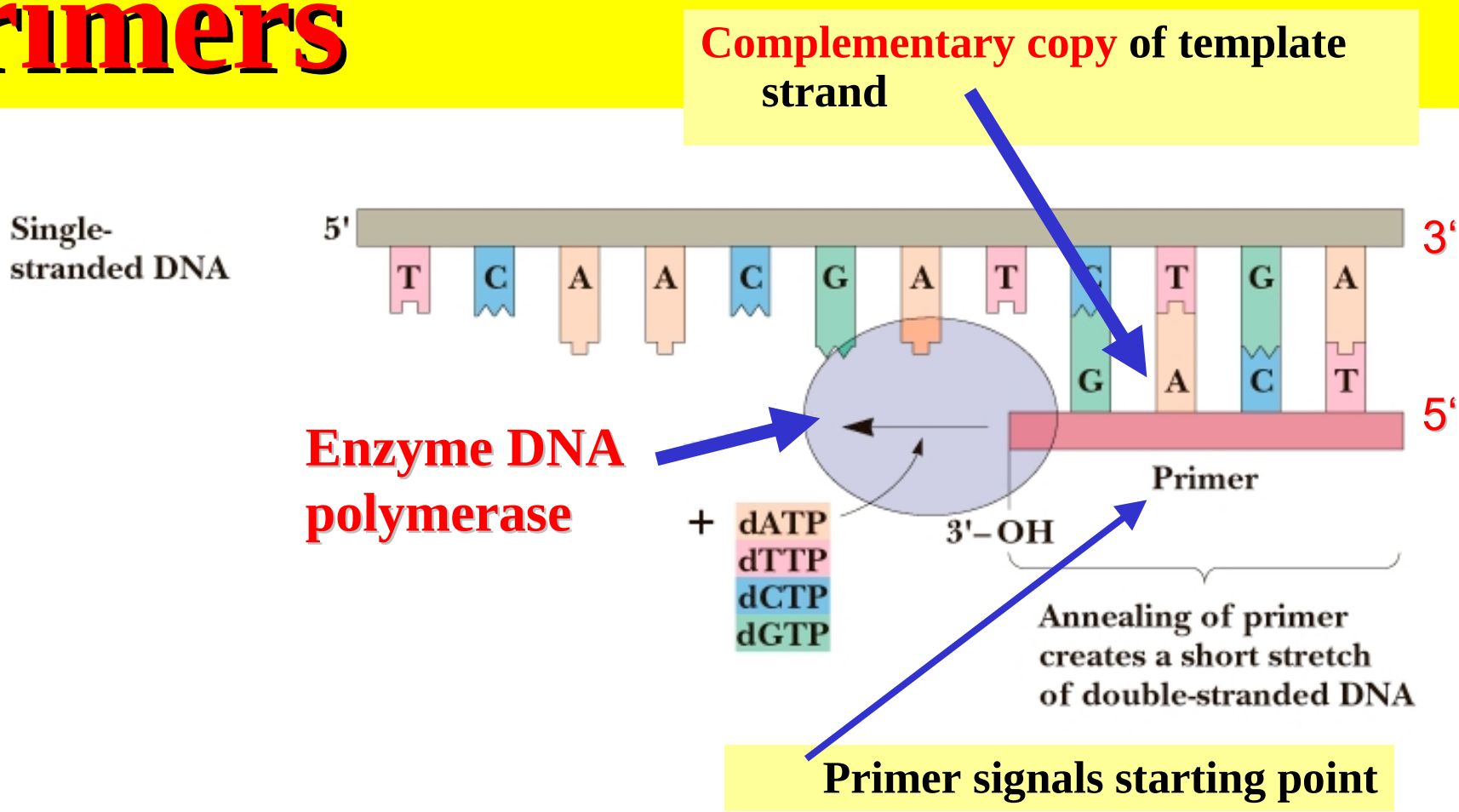
- **Enzyme T4 DNA *ligase***
  - **Anneal:**
    - **Overhanging complementary ends (sticky ends)**
    - **Hydrogen bonds**
  - **Ligate - a biochemical process of ligation:**
    - **Covalently bond 3‘ and 5‘ ends**

3. PCR: Polymerase Chain Reaction

- Process by which DNA is amplified
- Uses controlled biological processes similar to those which occur in the cell
- Can be done selectively
  - Using primers

Replication of DNA

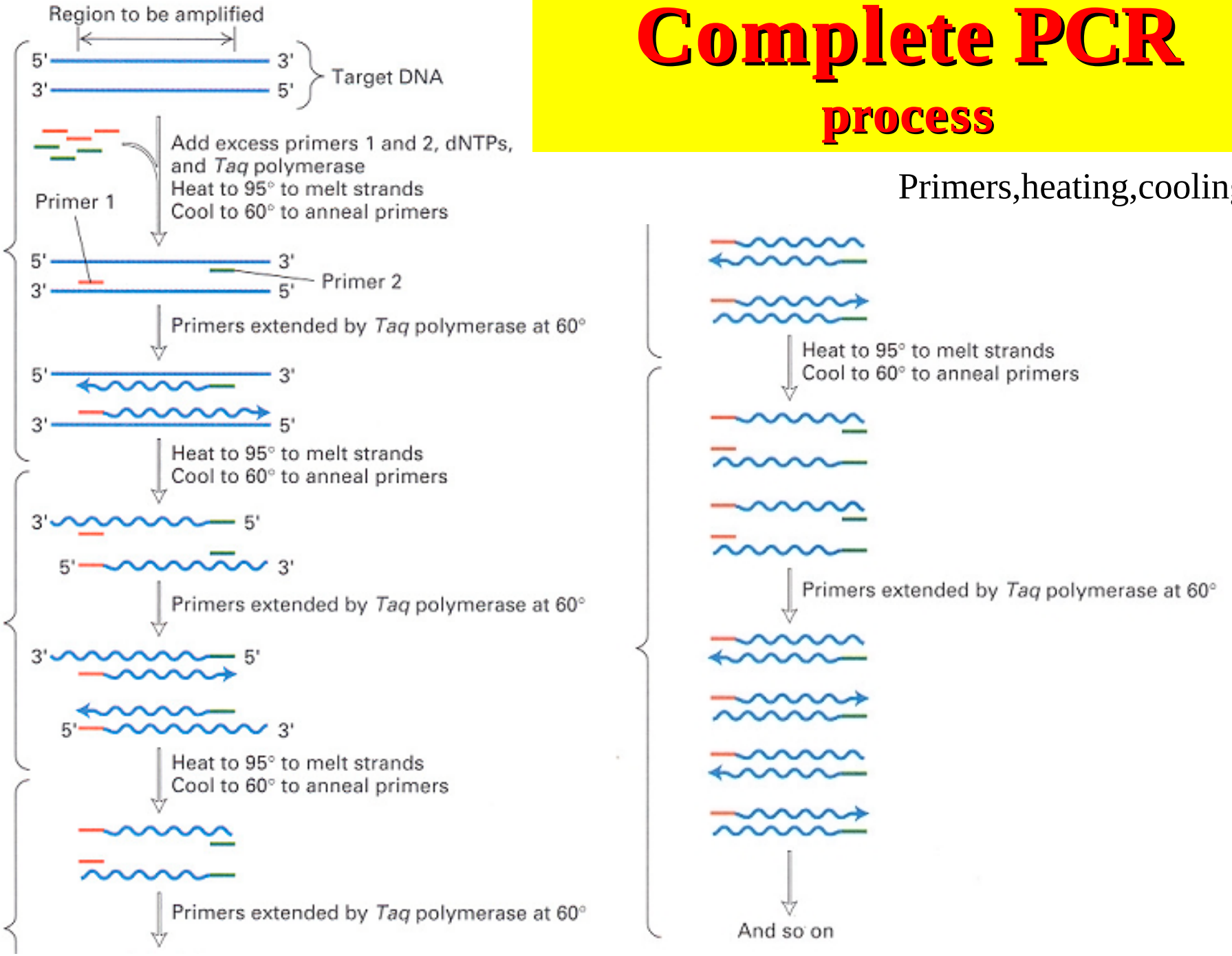
DNA Polymerase: using primers



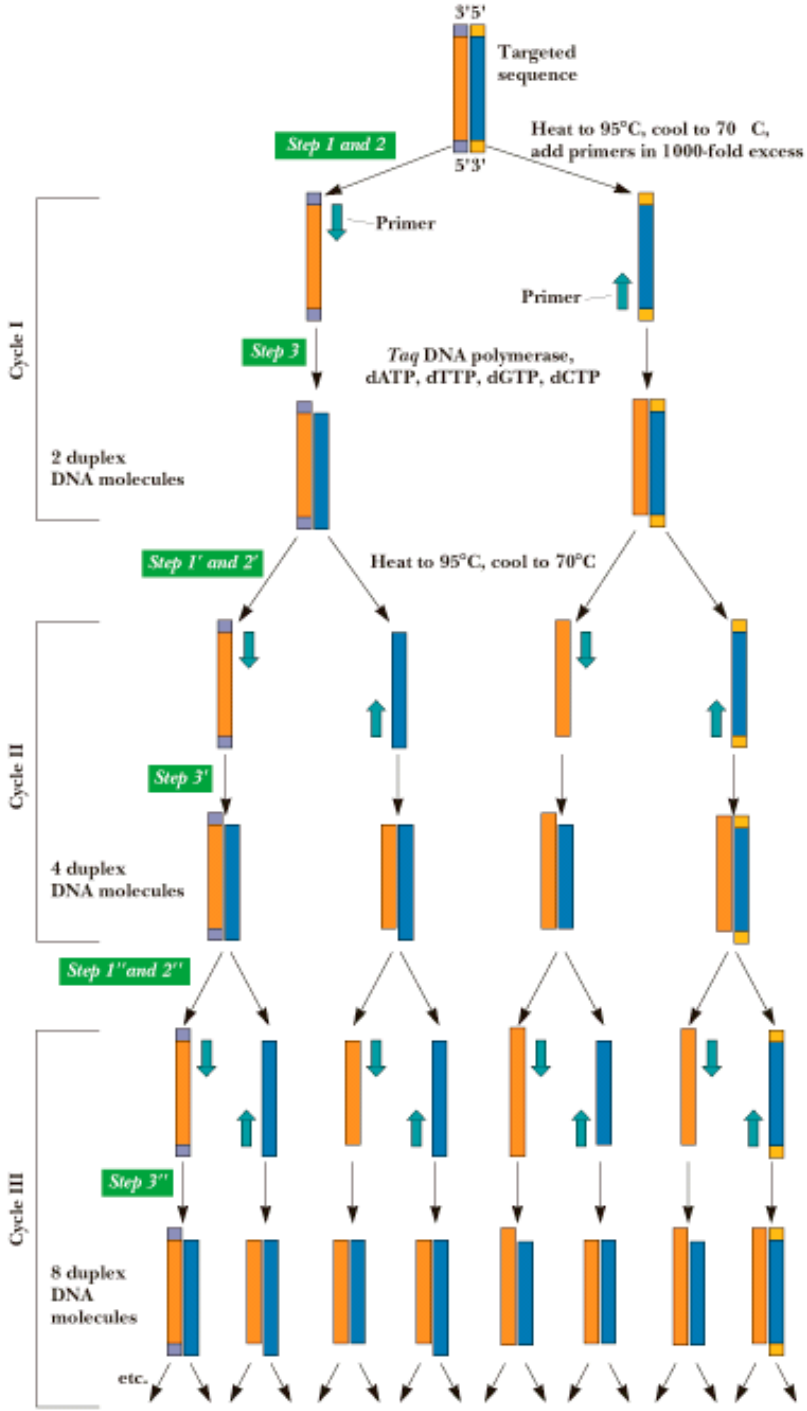
Saunders College Publishing

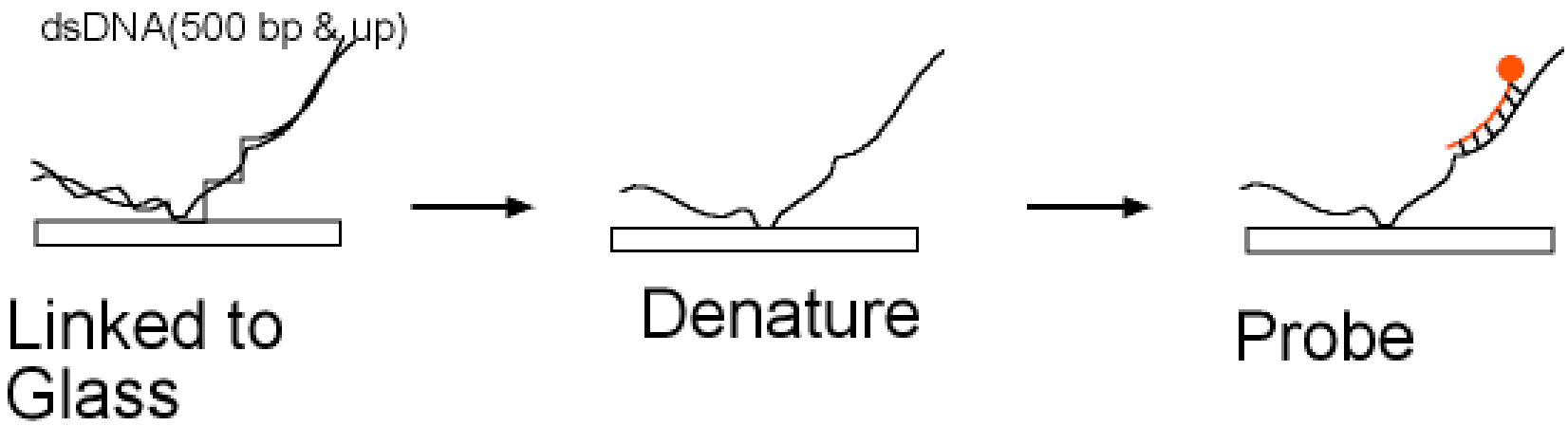
Complete PCR process

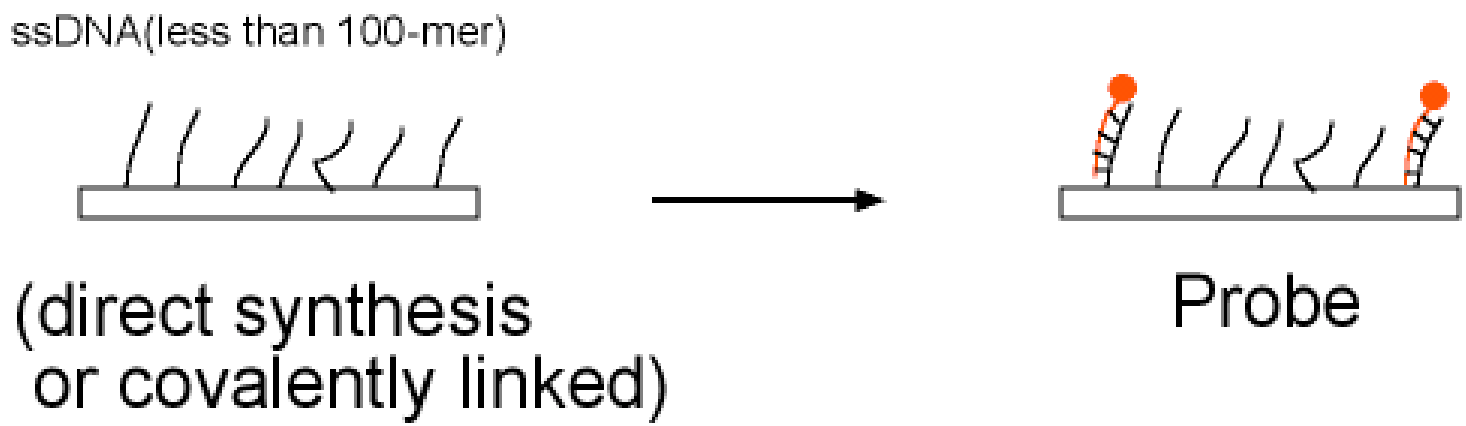
Primers,heating,cooling,



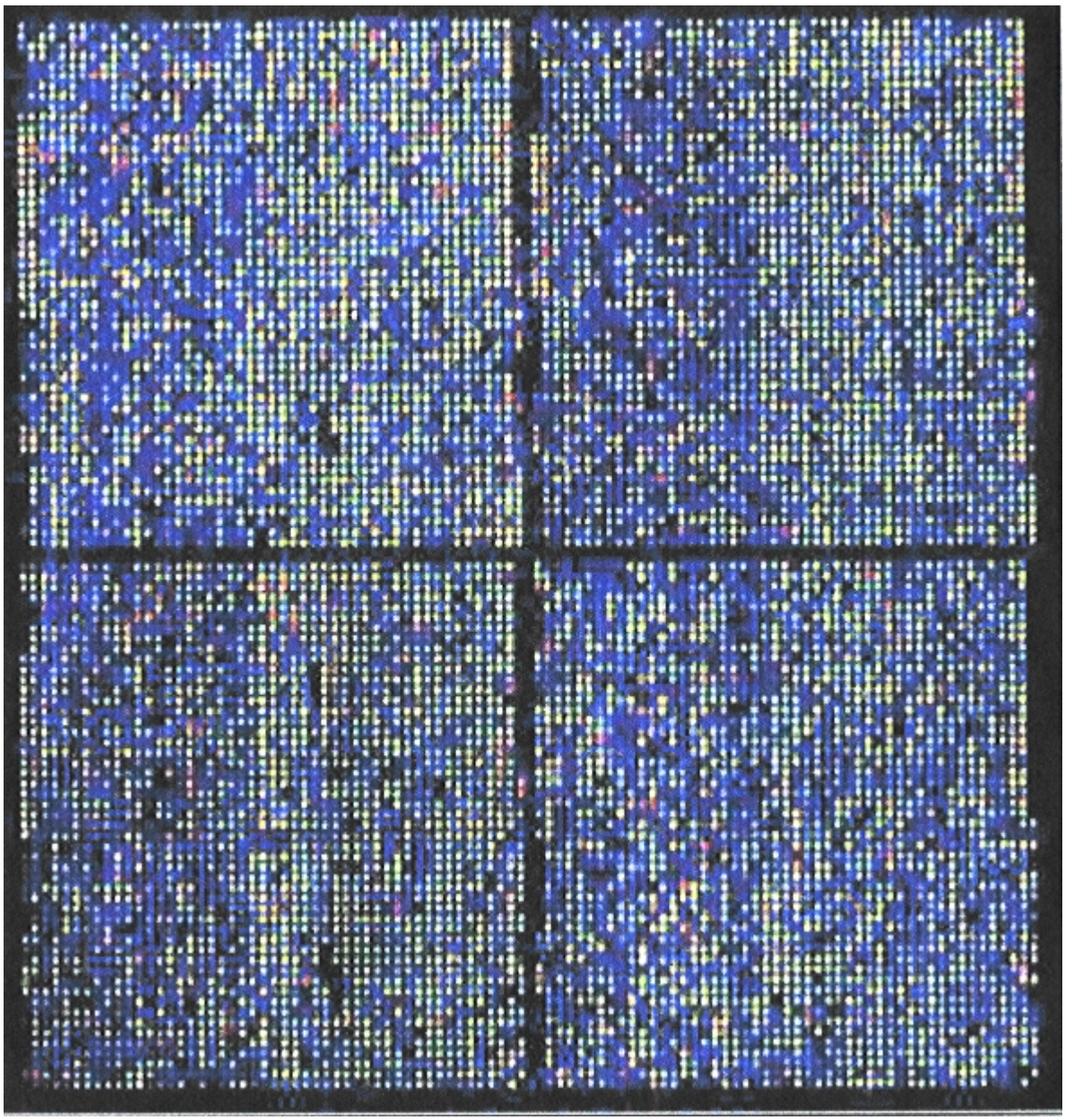
POLYMERASE CHAIN REACTION




dsDNA(500 bp & up)
Linked to Glass
Denature
Probe


ssDNA(less than 100-mer)
(direct synthesis or covalently linked)
Probe

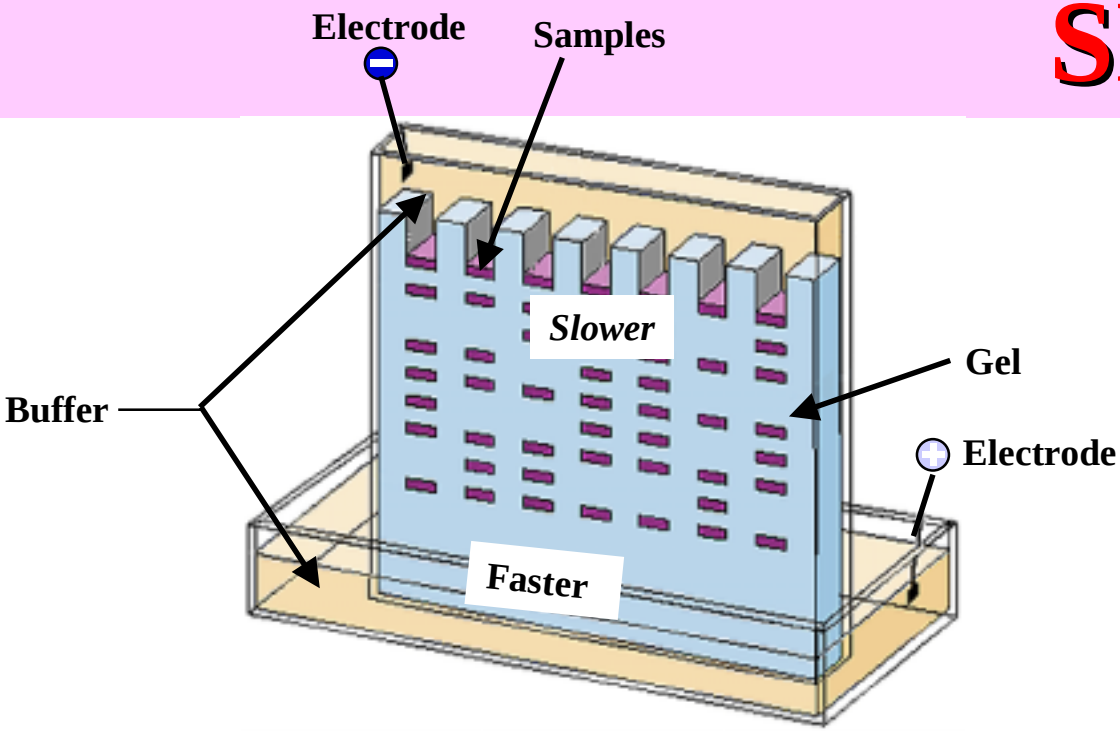
bp = base pair

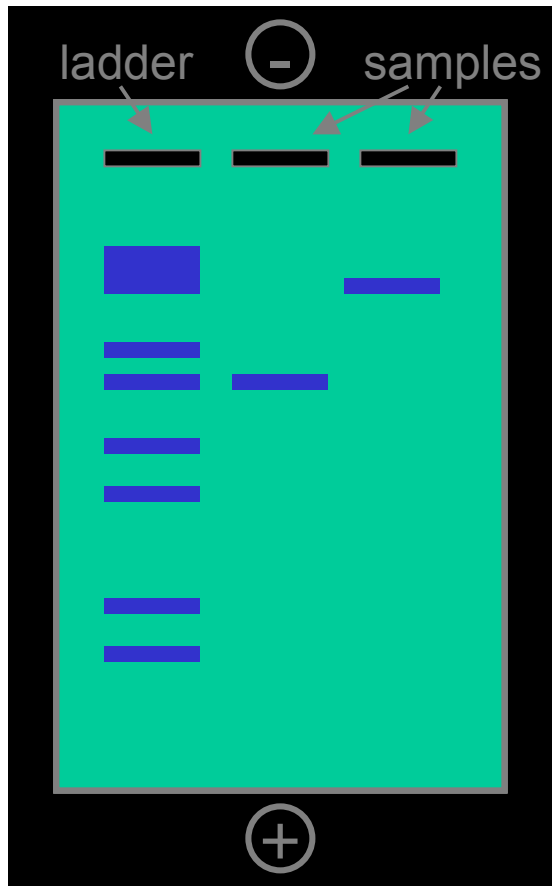


DNA operations: filtering, separation by length, reading

- Filtering
- Separation by length
- Reading

4. GEL ELECTROPHORESIS - SIZE SORTING

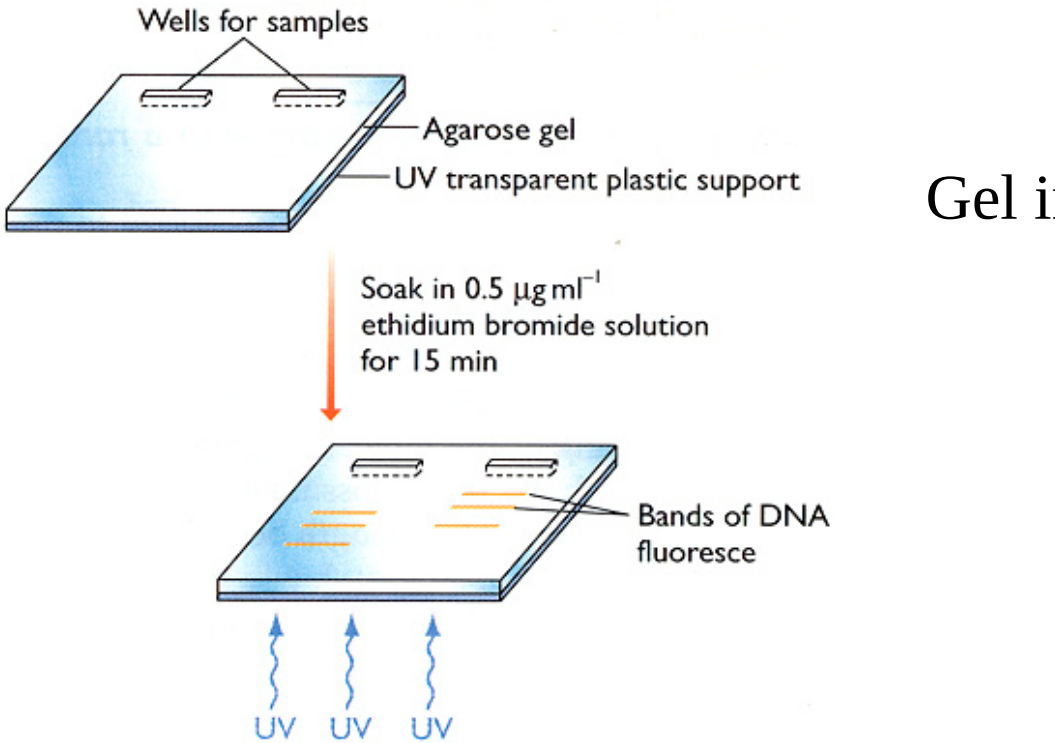




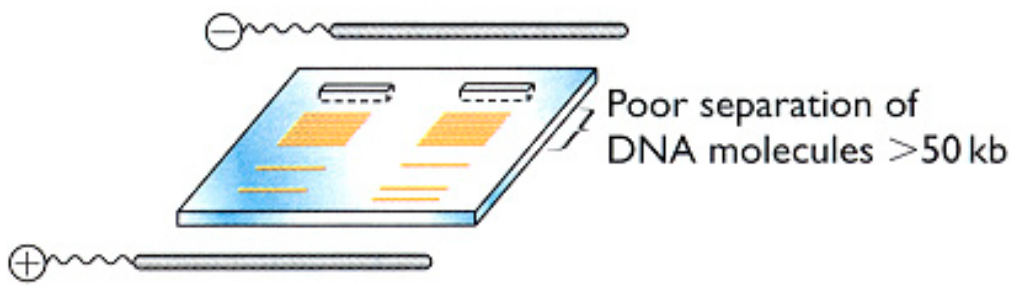
- Separation of DNA depending on
  - Size
  - Electrical charge

- Fill with agarose gel (0.5 – 5%)
- Load slots with samples and ladder
- Apply current
- Identify and extract

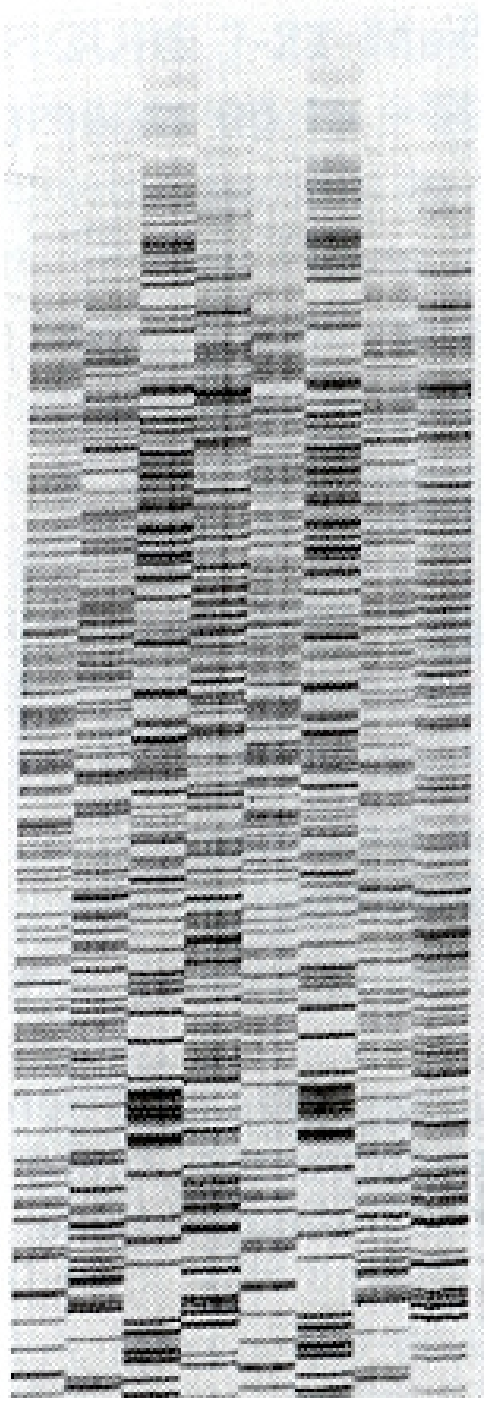
Gel Electrophoresis



(A) Standard agarose gel electrophoresis

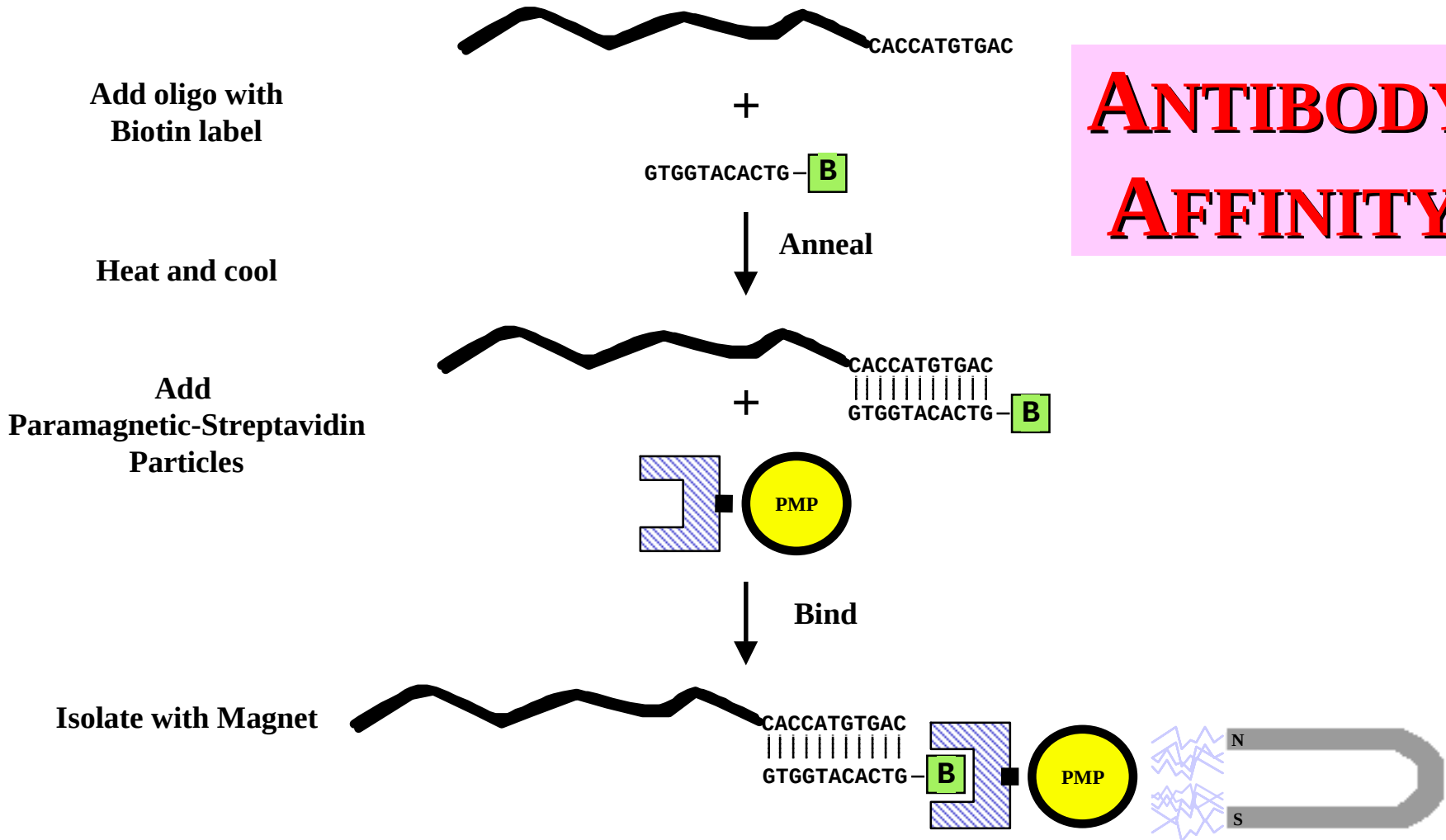


Gel image

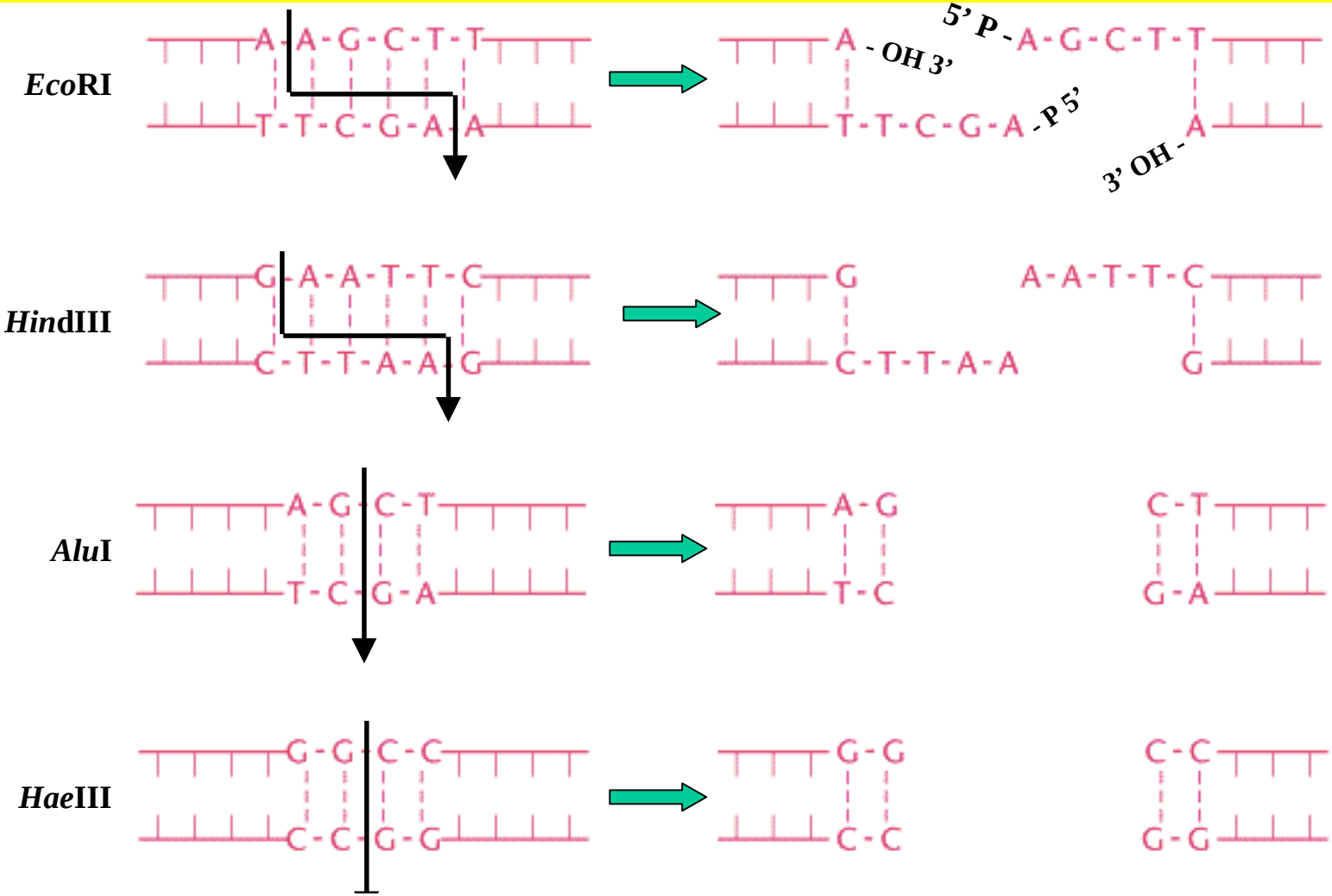


5. AFFINITY SEPARATION

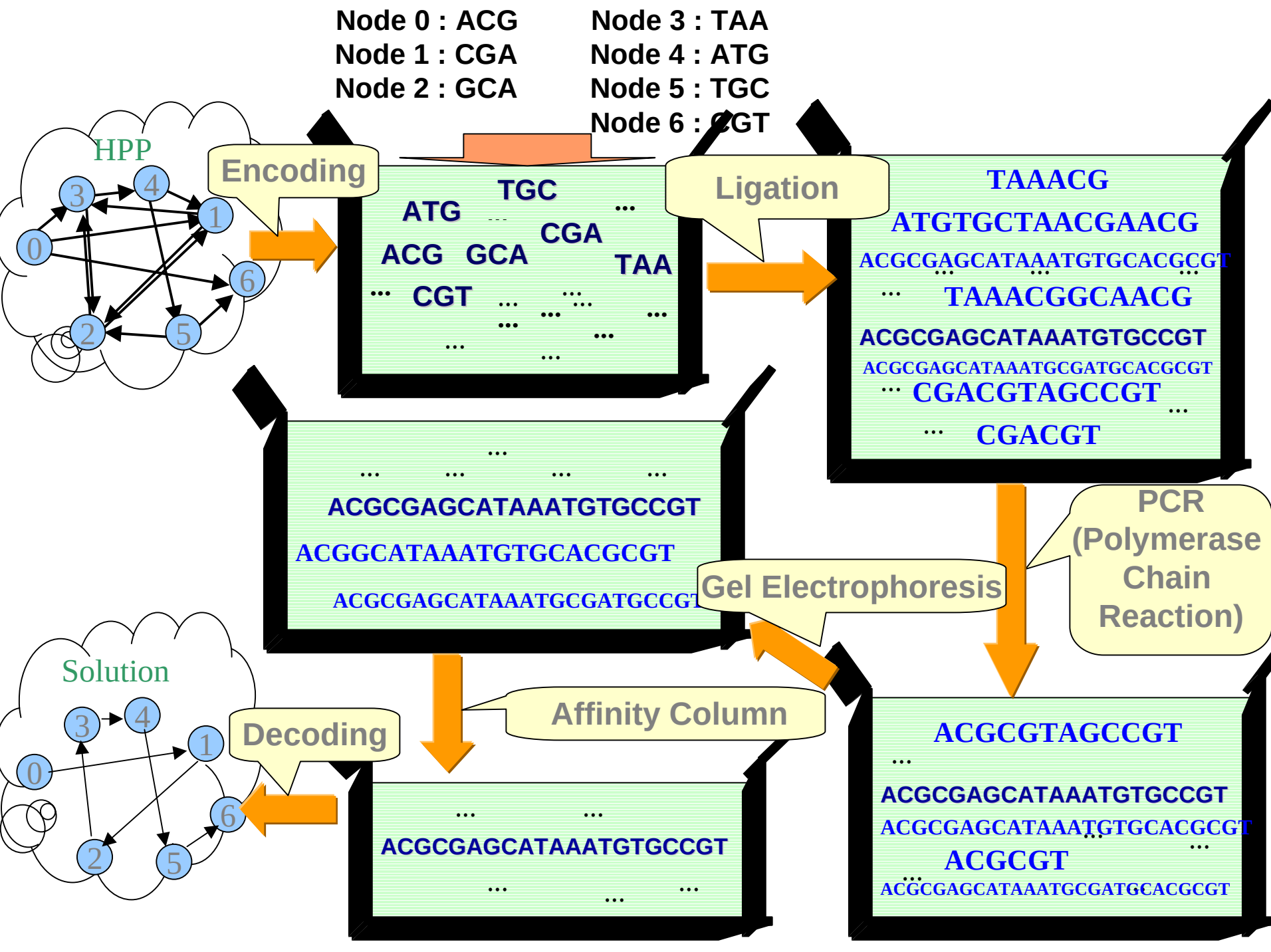
ANTIBODY AFFINITY



6. RESTRICTION ENDONUCLEASES



DNA operations: Shortening and Cutting DNA molecules


Node 0 : ACG
Node 1 : CGA
Node 2 : GCA
Node 3 : TAA
Node 4 : ATG
Node 5 : TGC
Node 6 : CGT
HPP
Encoding
TGC
ATG
CGA
ACG
GCA
TAA
CGT
Ligation
TAAACG
ATGTGCTAACGAACG
ACGCGAGCATAAATGTGCACGCGT
TAAACGGCAACG
ACGCGAGCATAAATGTGCCGT
ACGCGAGCATAAATGCGATGCACGCGT
CGACGTAGCCGT
CGACGT
ACGCGAGCATAAATGTGCCGT
ACGGCATAAATGTGCACGCGT
ACGCGAGCATAAATGCGATGCCGT
Gel Electrophoresis
PCR (Polymerase Chain Reaction)
Solution
Decoding
Affinity Column
ACGCGTAGCCGT
ACGCGAGCATAAATGTGCCGT
ACGCGAGCATAAATGTGCACGCGT
ACGCGT
ACGCGAGCATAAATGCGATGCACGCGT
ACGCGAGCATAAATGTGCCGT