Is Nature a Supercomputer?

Victor Mitrana

University of Bucharest
Faculty of Mathematics and Computer Science

mitrana@fmi.unibuc.ro

What is Natural Computing?

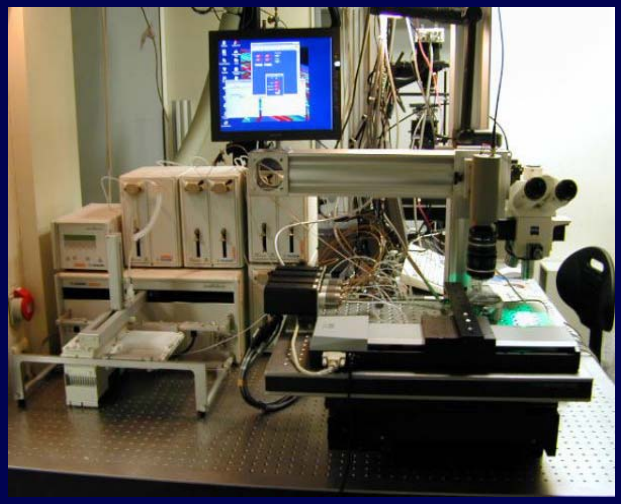
- Computing paradigms suggested by nature
- Bio-inspired computing
  - DNA computing (in vitro)
  - Molecular computing (in vivo)
  - Formal models based on bio-operations or bio-architecures
    - Splicing systems
    - Membrane computing
    - Networks of evolutionary processors
- Bio-computing
- Quantum computing
- Chemical computing
- Optical computing

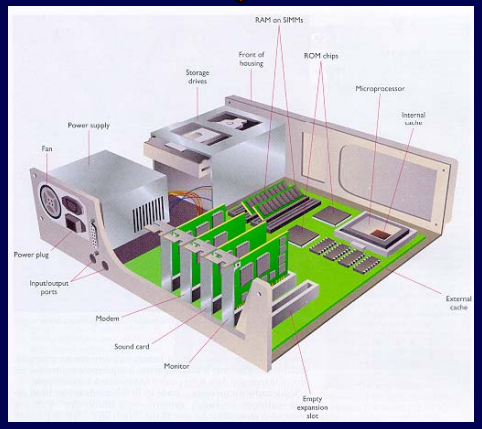
What is Bio-Inspired Computing?

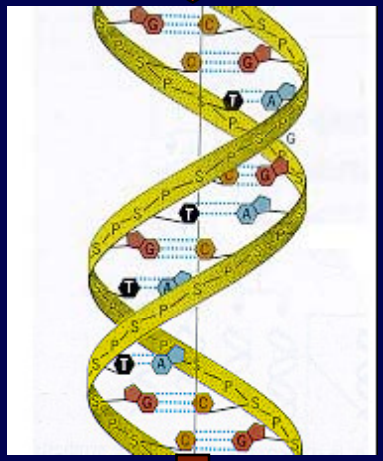
- The field of bio-inspired computing is concerned with understanding how complex biological molecules process information in an attempt to gain insight into new models of computation.

- Cells and nature ``compute´´ by reading and rewriting DNA or RNA by processes that modify sequence at the DNA or RNA level. DNA computing is interested in applying computer science methods and models to understand such biological phenomena and gain insight into early molecular evolution and the original of biological information processing.

What is DNA Computing?







011001101010001

ATGCTCGAAGCT

What is DNA Computing?

- In vitro
- Uses DNA molecules to store data
- Uses DNA molecules to compute by means of biochemical processes which manipulate DNA sequences

DNA computers vs. conventional computers

DNA-based computers	Microchip-based computers
slow at individual operations	fast at individual operations
can do billions of operations simultaneously	can do substantially fewer operations simultaneously
can provide huge memory in small space	smaller memory
setting up a problem may involve considerable preparations	setting up only requires keyboard input
DNA is sensitive to chemical deterioration	electronic data are vulnerable but can be backed up easily

Speed of DNA computing

Computer speed

- number of parallel processors
- number of steps each processor can perform per unit of time

DNA computer

- 3 grams of water contains 10^{22} molecules
- massively parallel

Electronic computer

- advantage in number of steps performed per unit of time

Density of DNA computing

information per space unit perform per unit of time

DNA computer

- 10^6 Gbits per cm^2 (1 bit per nm^3)

Electronic computer

- 1 Gbits per cm^2

Efficiency of DNA computing

DNA computer

- 10^{19} operations per Joule

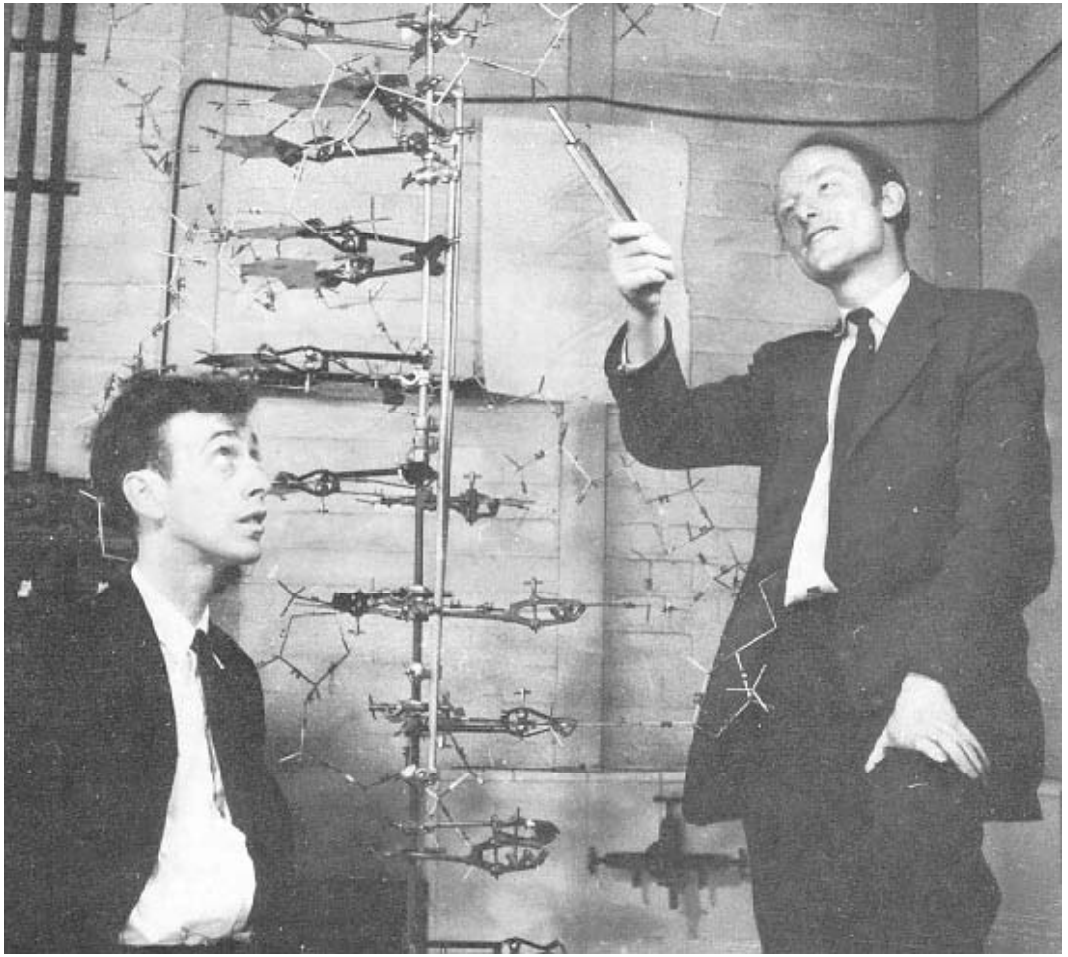
Electronic computer

- 10^{9} operations per Joule

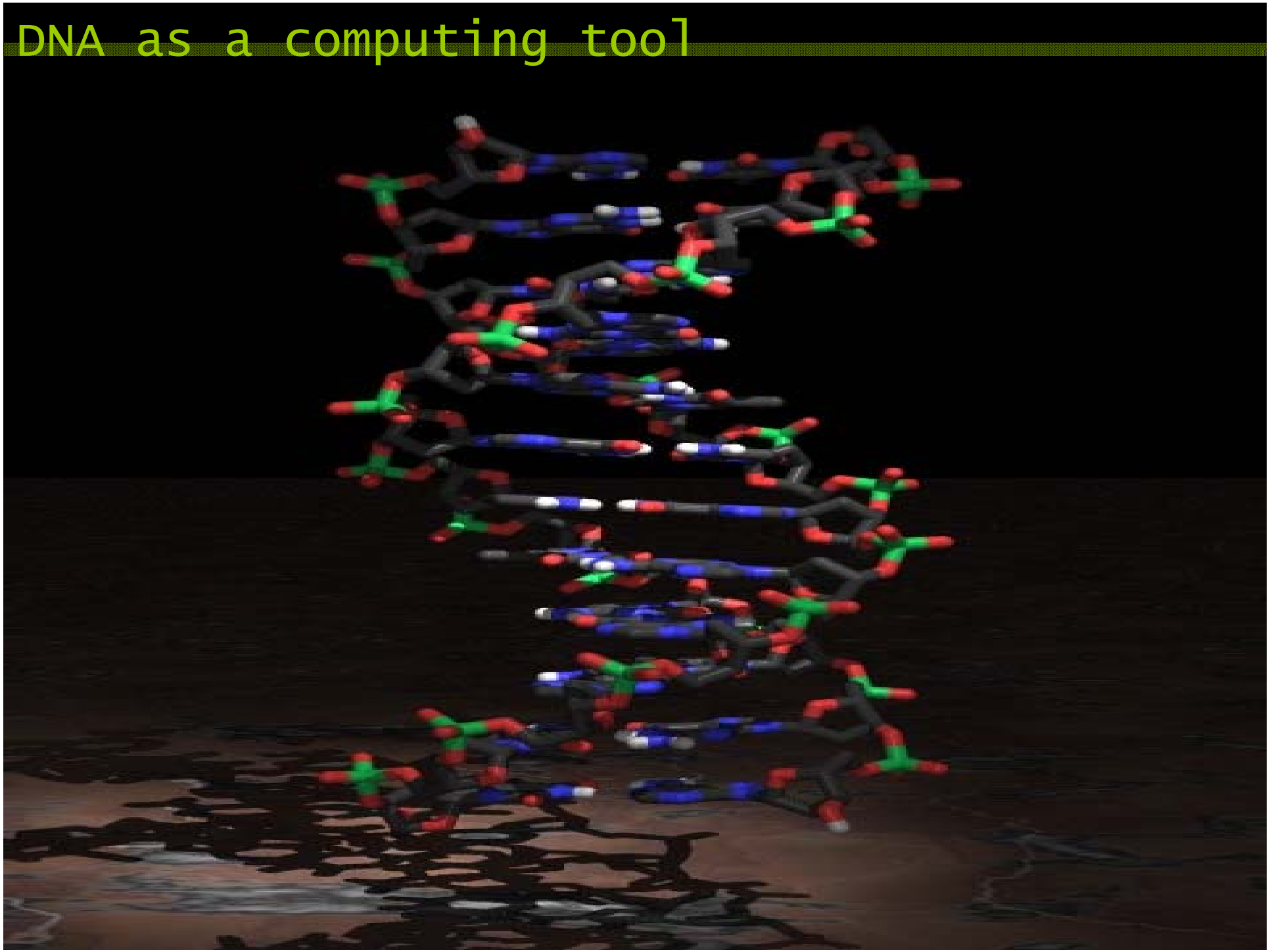
DNA computer

DNA (deoxyribonucleic acid)

Watson & Crick (1953): *Nature* 25: 737-738
Molecular Structure of Nucleic Acids: a structure for deoxyribose nucleic acid.
Nobel Prize, 1962.



DNA as a computing tool



DNA as computing tool

DNA sequences consist of

- A, C, G, T

Nucleotide:

- purine or pyrimidine base
- deoxyribose sugar
- phosphate group

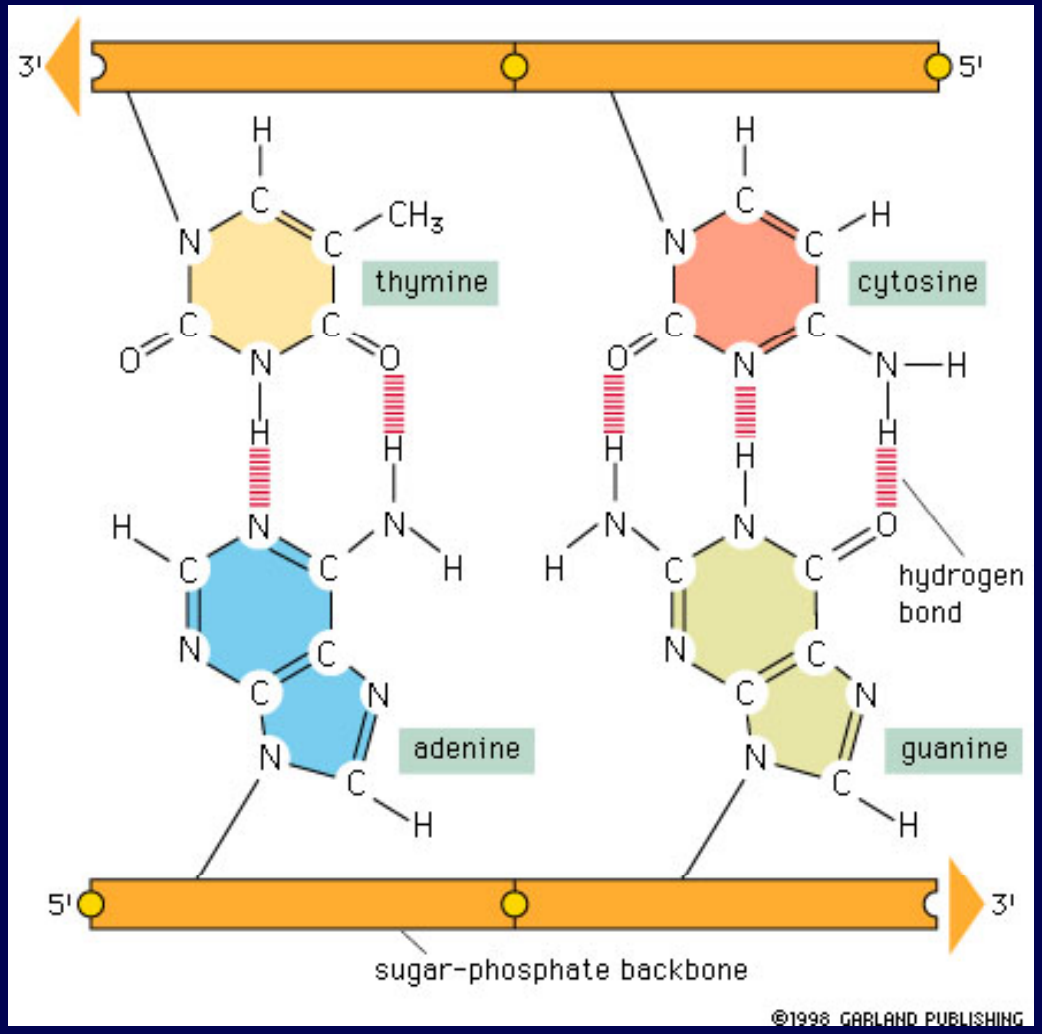
Purine bases

- A(denine), G(uanine)

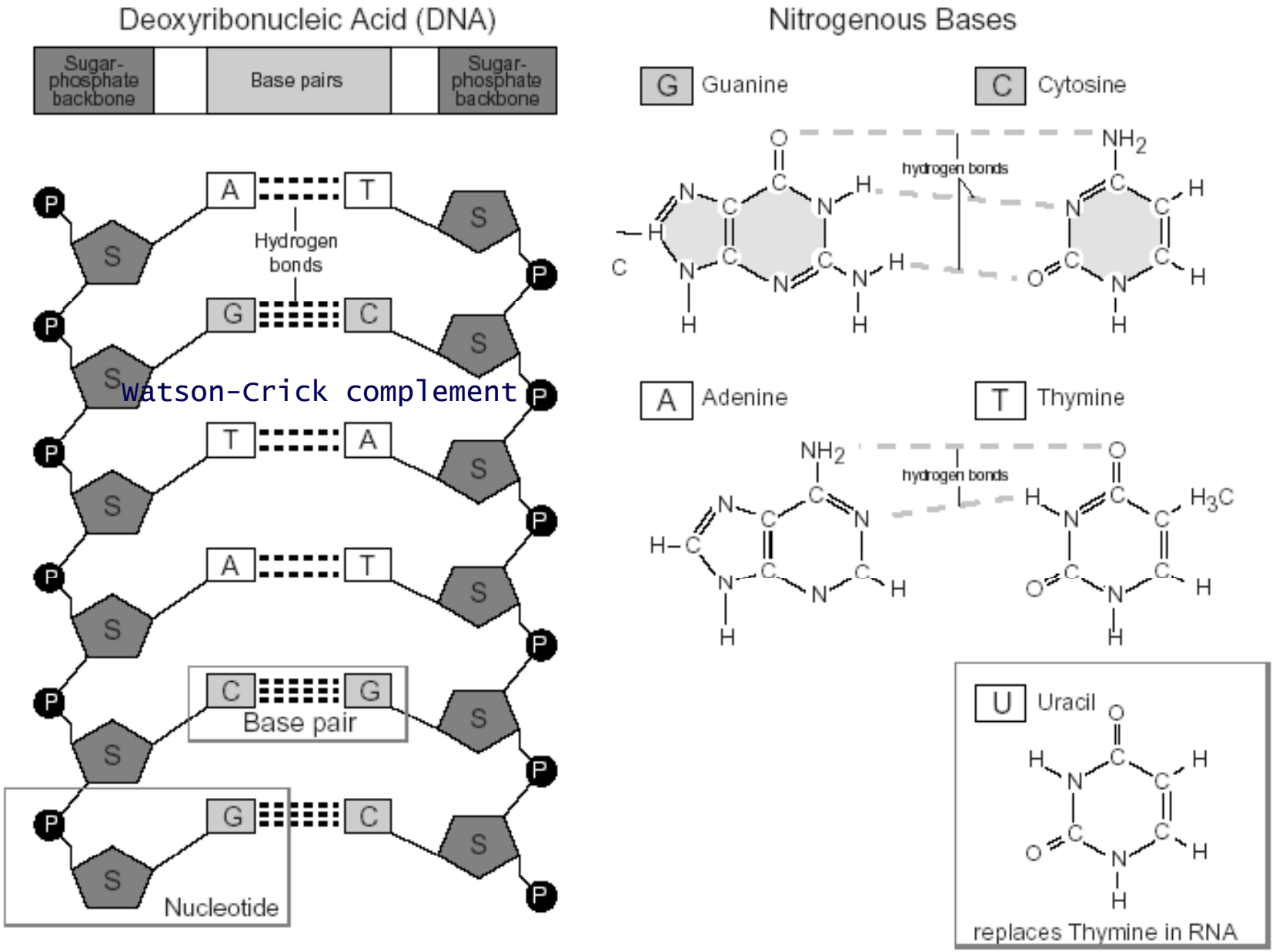
Pyrimidine bases

- C(ytosine), T(hymine)

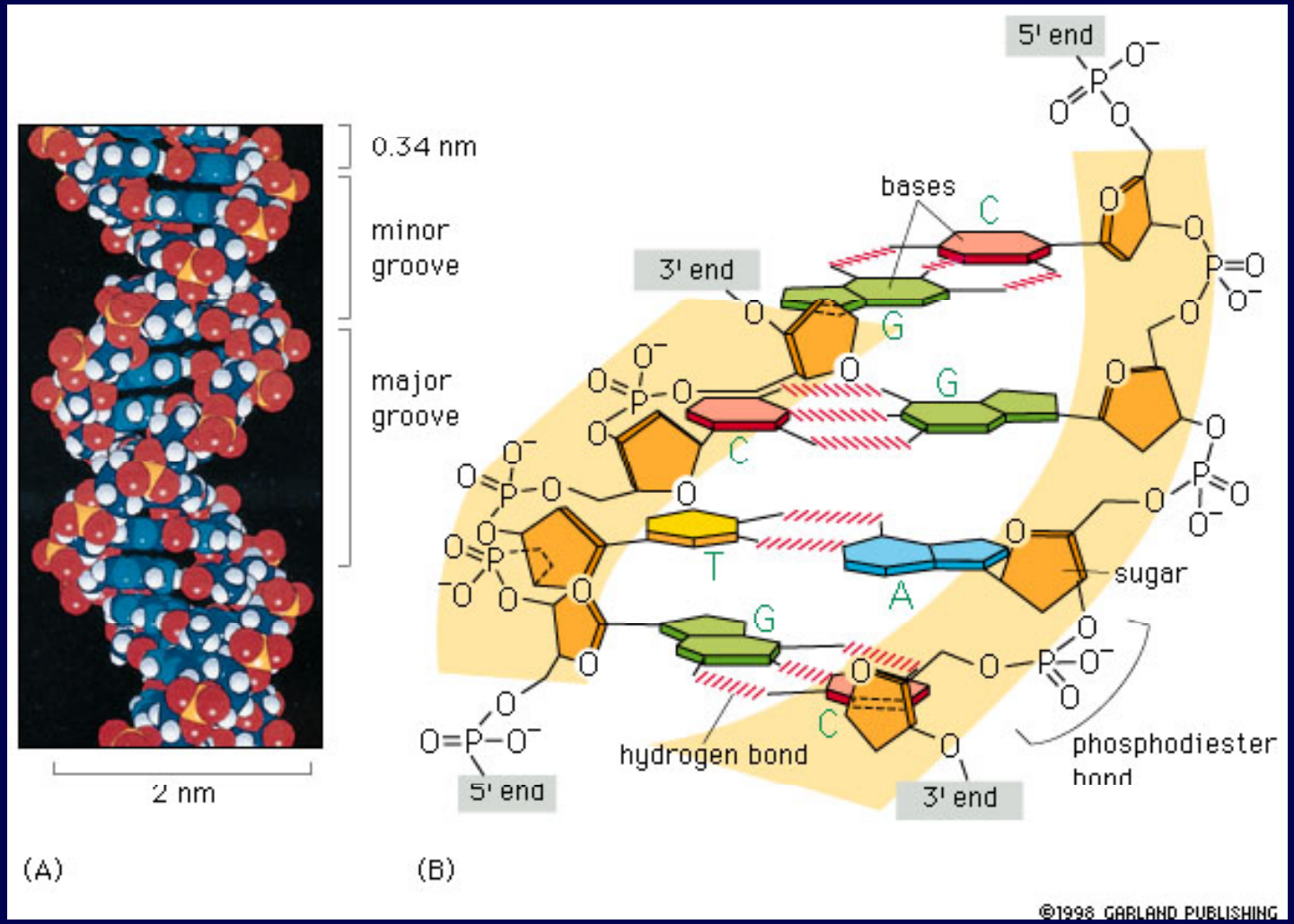
DNA as computing tool



Inter-strand hydrogen bonding



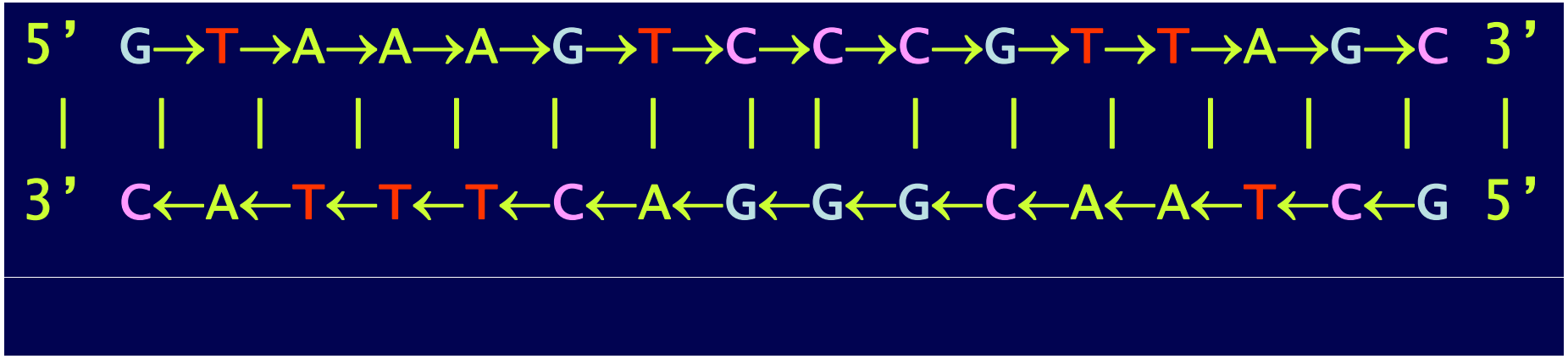
DNA as computing tool



Single stranded polynucleotide

5’ G→T→A→A→A→G→T→C→C→C→G→T→T→A→G→C 3’

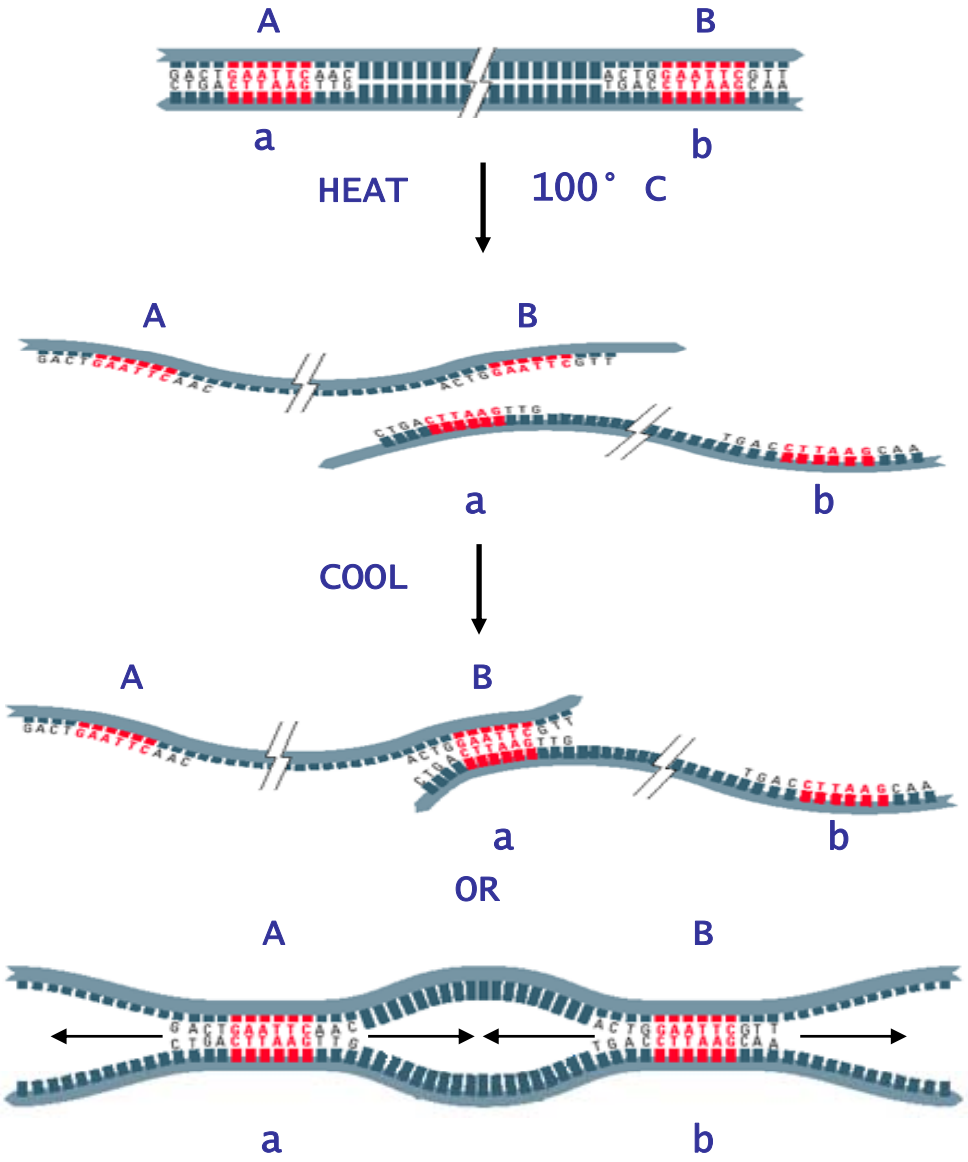
Double stranded polynucleotide



These are the techniques that are common in the microbiologist's lab and can be used to program a molecular computer. DNA can be:

- *synthezise* desired strands can be created
- *separate* strands can be sorted and separated by length
- *merge* by pouring two test tubes of DNA into one to perform union
- *extract* extract those strands containing a given pattern
- *melt/anneal* breaking/bonding two ssDNA molecules with complementary sequences
- *amplify* use of PCR to make copies of DNA strands
- *cut* cut DNA with restriction enzymes
- *rejoin* rejoin DNA strands with 'sticky ends'
- *detect* confirm presence or absence of DNA

Separating and fusing DNA strands

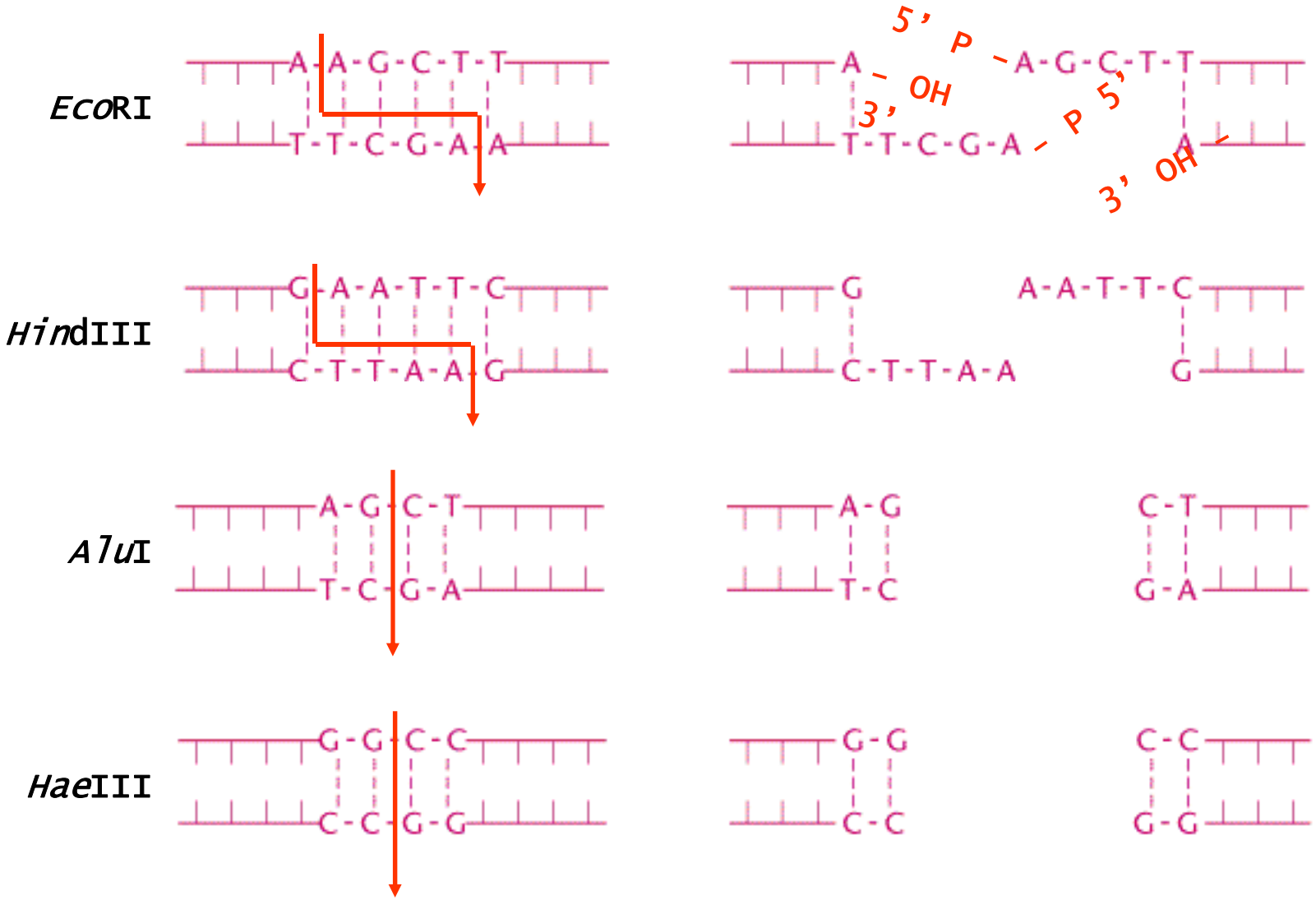


Enzymes

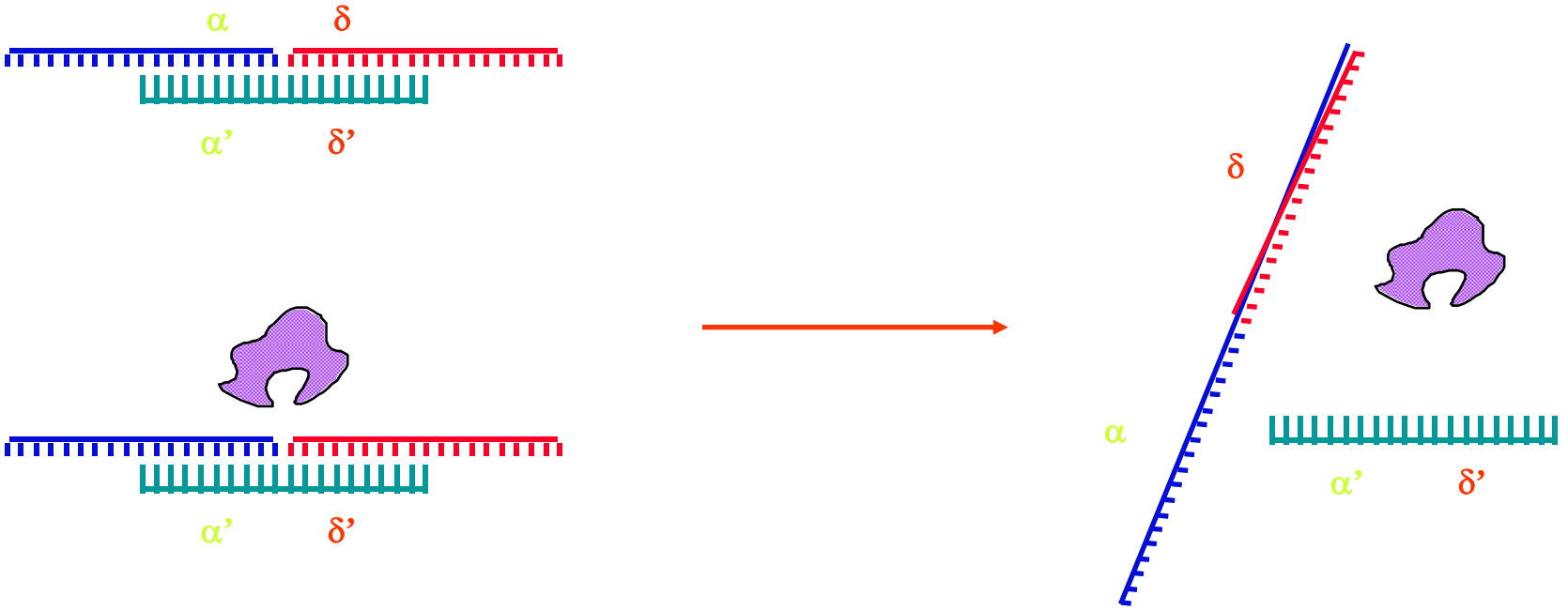
Machinery for Nucleotide Manipulation

- Enzymes are proteins that catalyze chemical reactions.
- Enzymes are very specific.
- Enzymes speed up chemical reactions extremely efficiently
- Nature has created a multitude of enzymes that are useful in processing DNA.

Restriction endonucleases



DNA ligation



Ligase joins 5' phosphate to 3' hydroxyl

Tom Head

Department of Mathematical Sciences
Binghamton University

Areas of interest

Algebra

Computing with biomolecules

Formal representations of communication

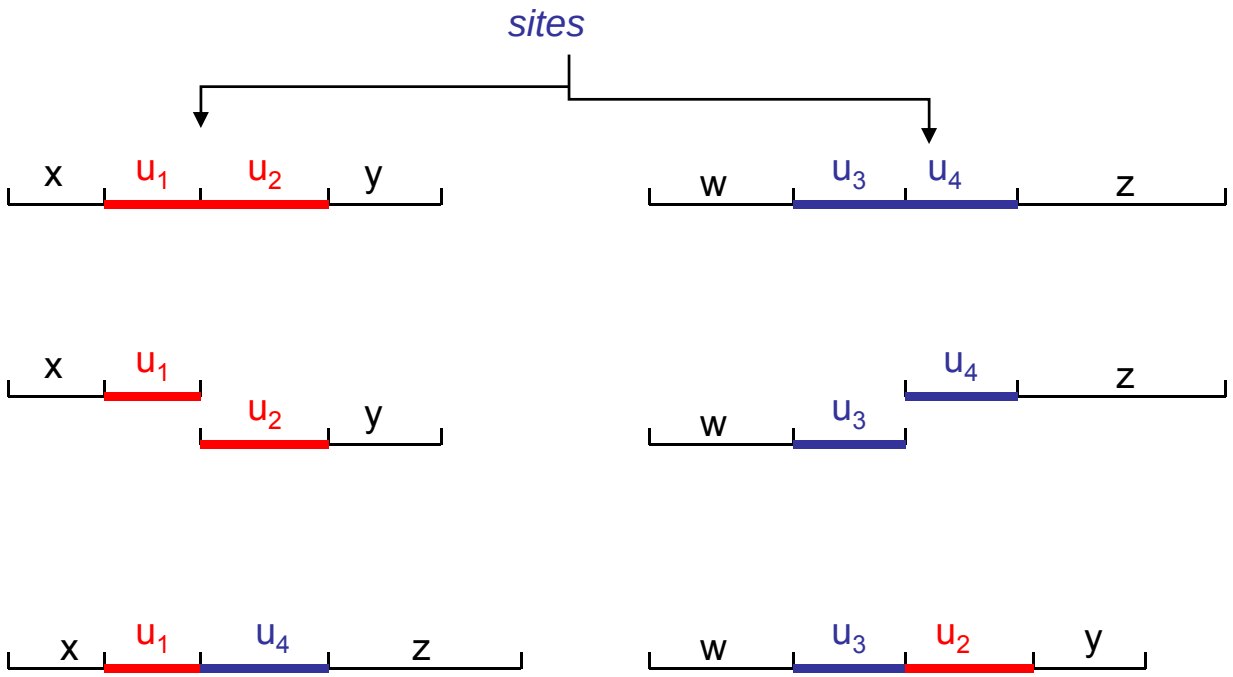
http://www.math.binghamton.edu/tom/

Formal splicing

σ: (x u_1u_2 y, wu_3u_4 z)

r = u_1#u_2 \$ u_3#u_4 rule

(x u_1 u_4 z , wu_3 u_2 y)



Pattern recognition

cut

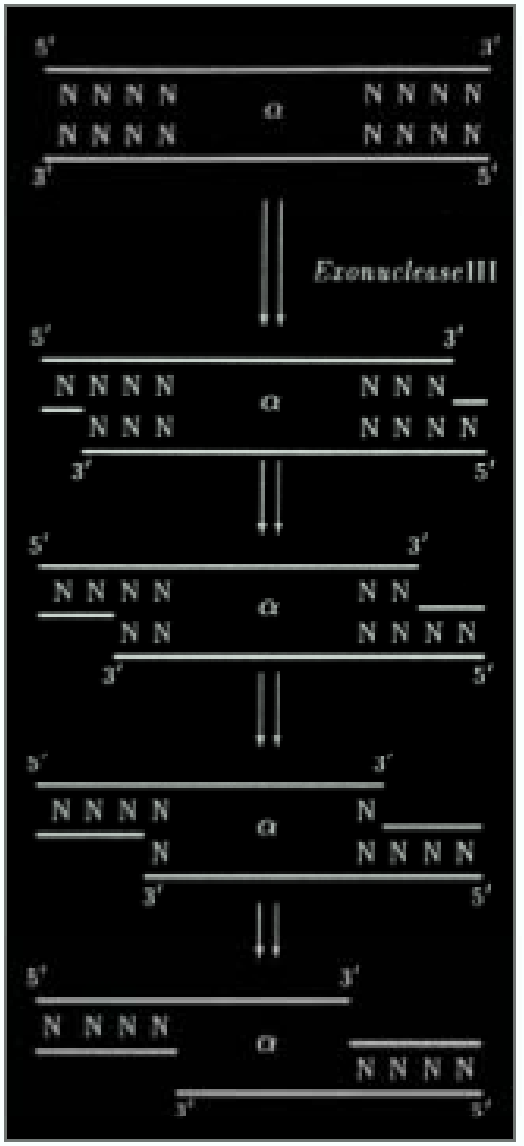
paste

Shortening DNA

DNA nucleases are enzymes that degrade DNA.

DNA exonucleases

- cleave (remove) nucleotides one at a time from the ends of the strands
- **Example:** Exonuclease III 3´-nuclease degrading in 3´-5´direction

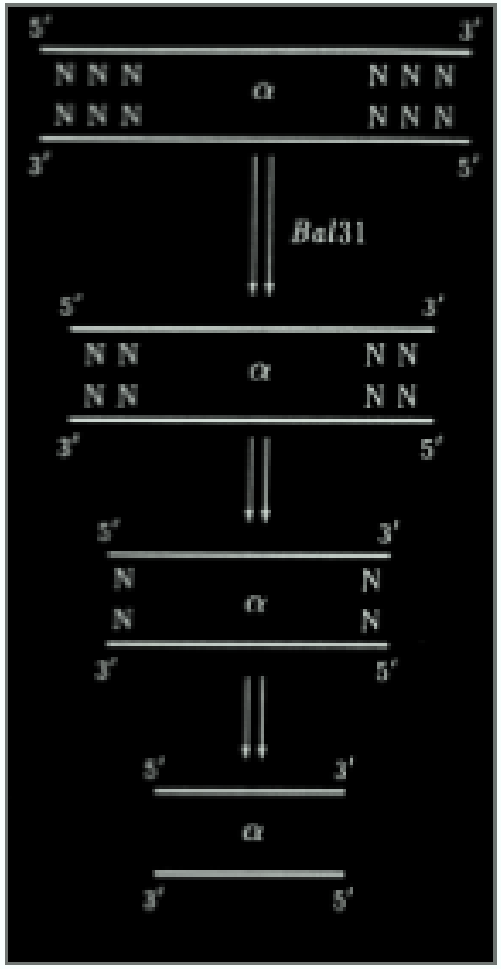


Shortening DNA

DNA nucleases are enzymes that degrade DNA.

DNA exonucleases

- cleave (remove) nucleotides one at a time from the ends of the strands
- **Example:** Bal31 removes nucleotides from both strands



Cutting DNA

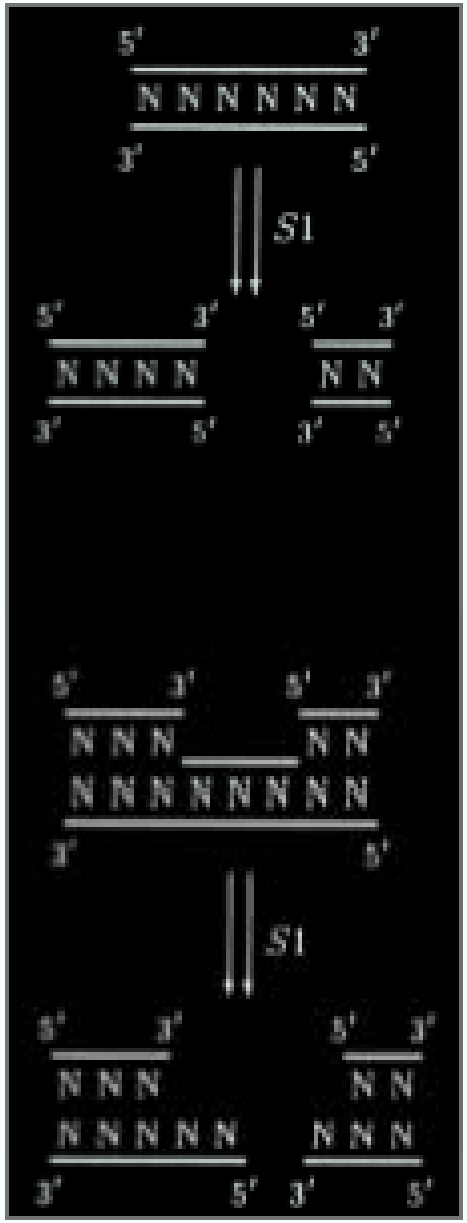
DNA nucleases are enzymes that degrade DNA.

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)

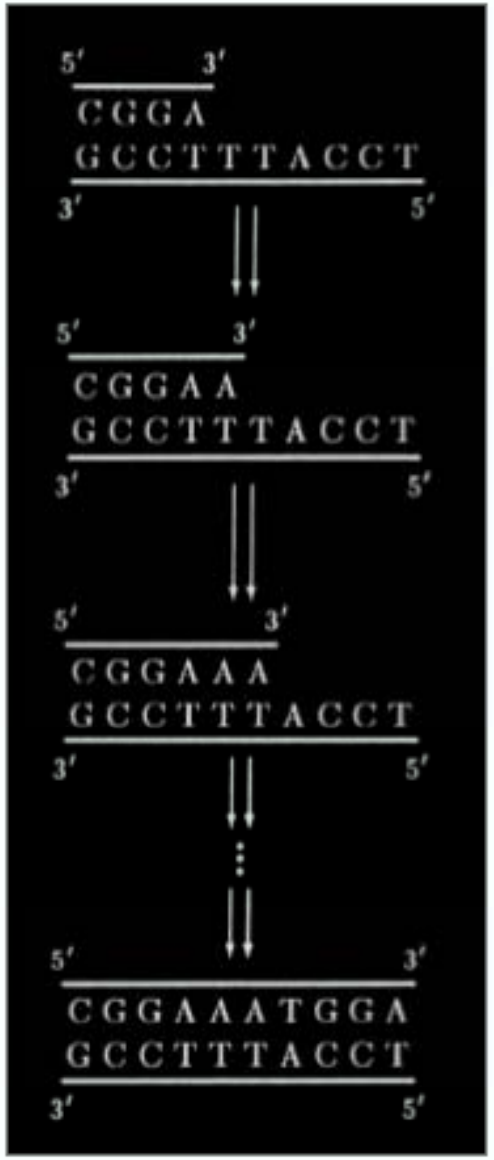


Lengthening DNA

- 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.



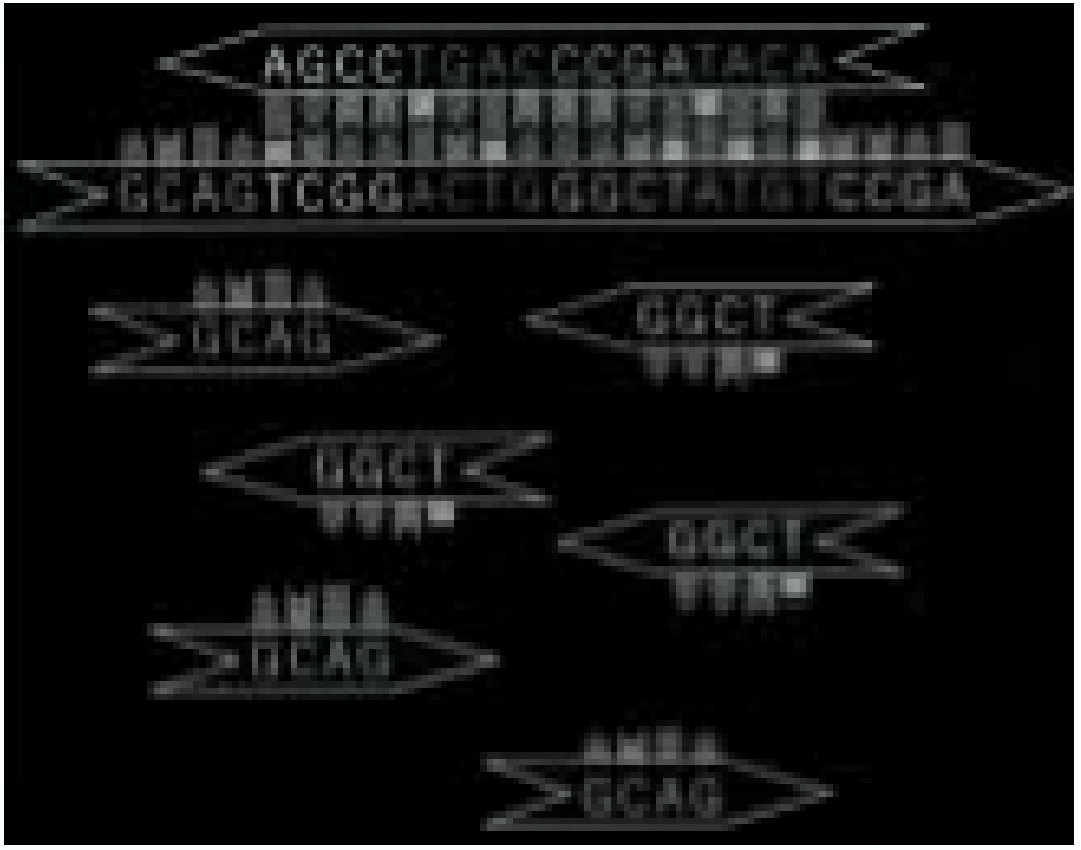
Multiplying DNA

- 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 copy machine

PCR - initialisation

Start with a solution containing the following ingredients:

- the target DNA molecule
- primers (synthetic oligo-nucleotides), complementary to the terminal sections
- polymerase, heat resistant nucleotides

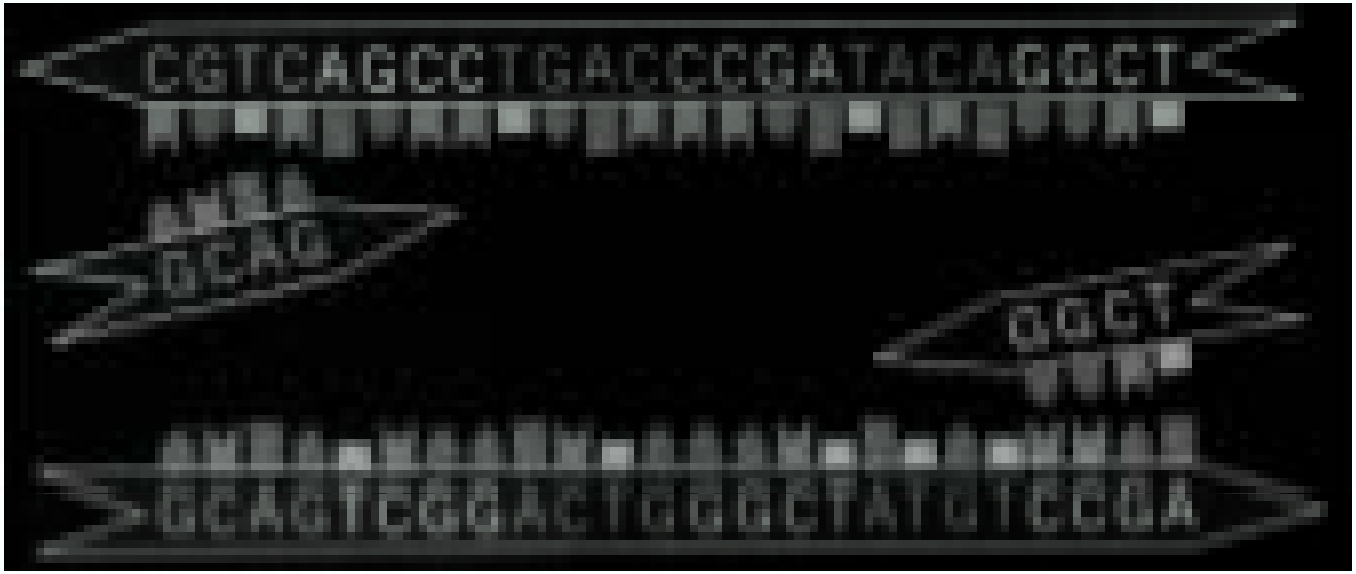


PCR – denaturation

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

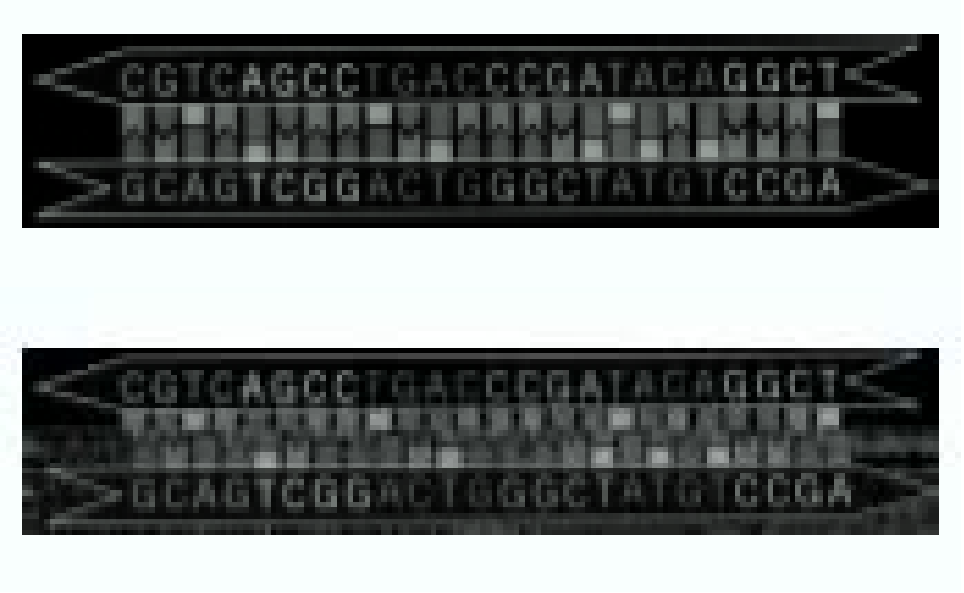
PCR - priming

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



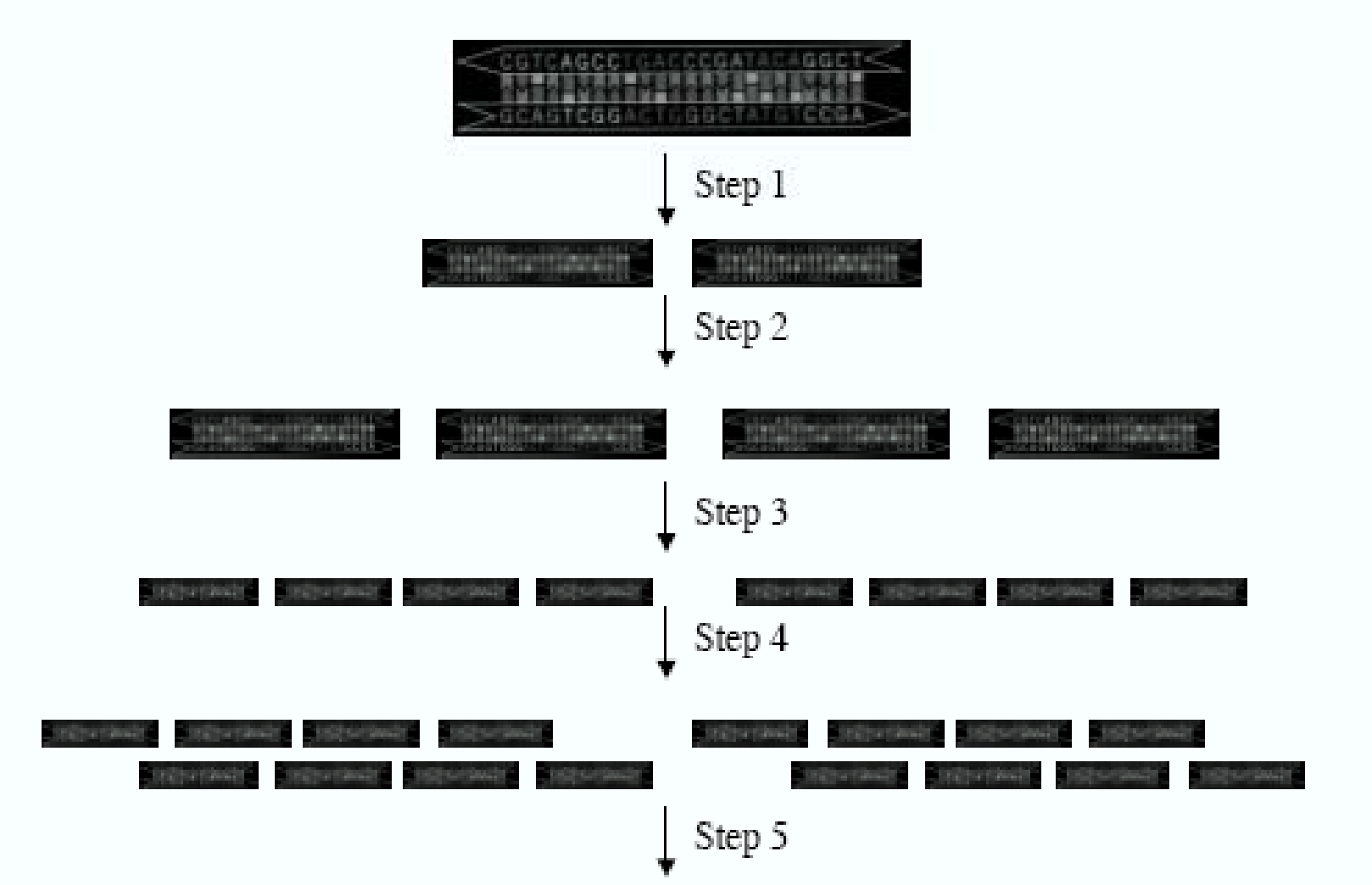
PCR - extension

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



PCR – copying

2n copies after n steps



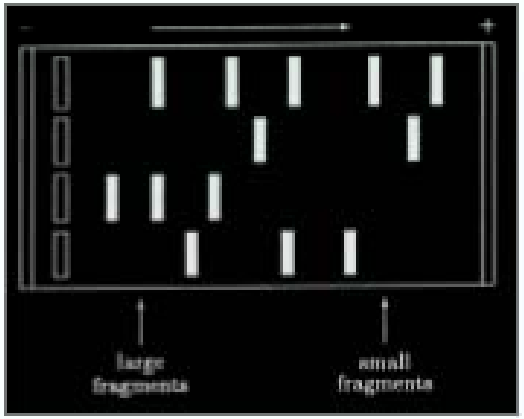
Gel electrophoresis

Measuring the Length of DNA Molecules

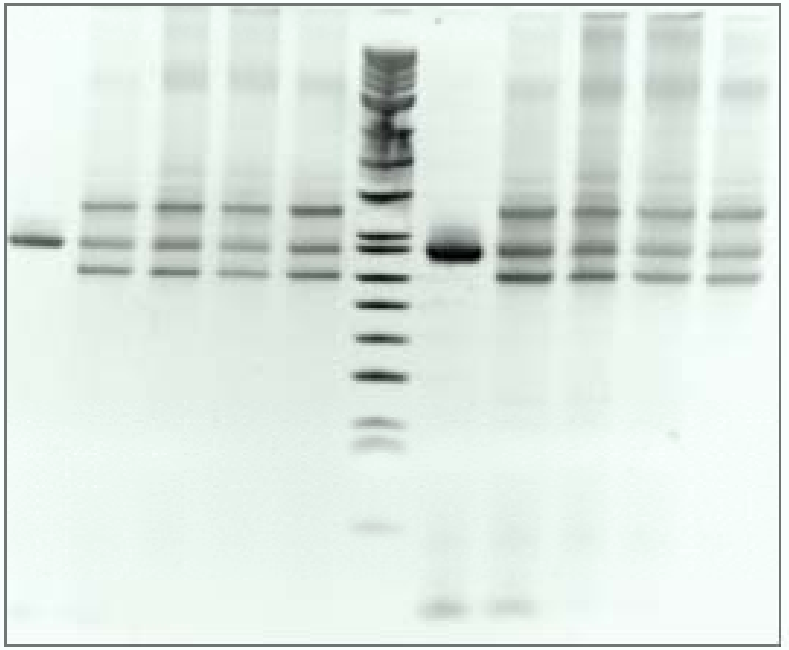
- DNA molecules are negatively charged.
- Placed in an electric field, they will move towards the positive electrode.
- The negative charge is proportional to the length of the DNA molecule.
- The force needed to move the molecule is proportional to its length.
- A gel makes the molecules move at different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

Gel electrophoresis

Schematic representation of gel electrophoresis

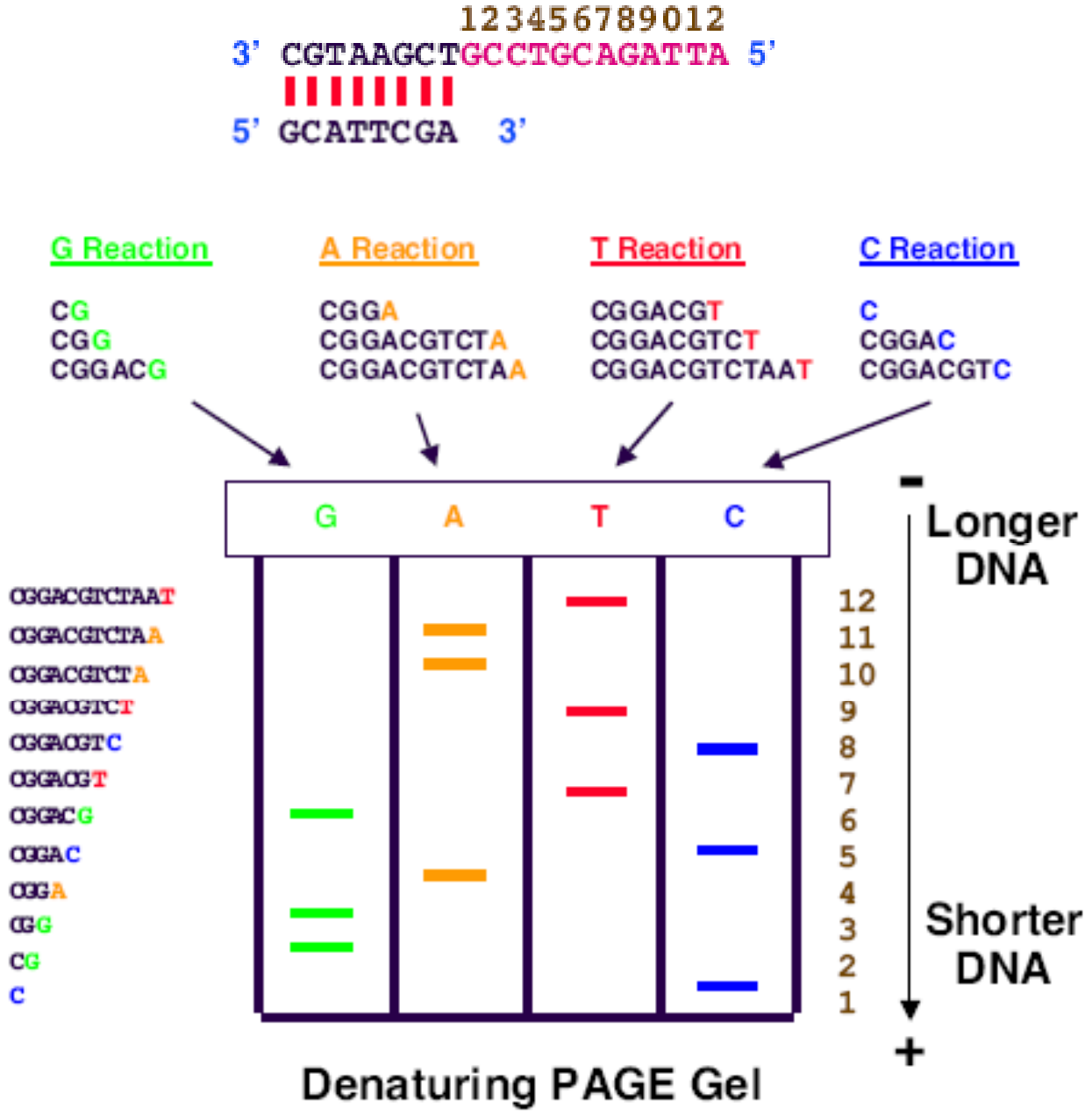


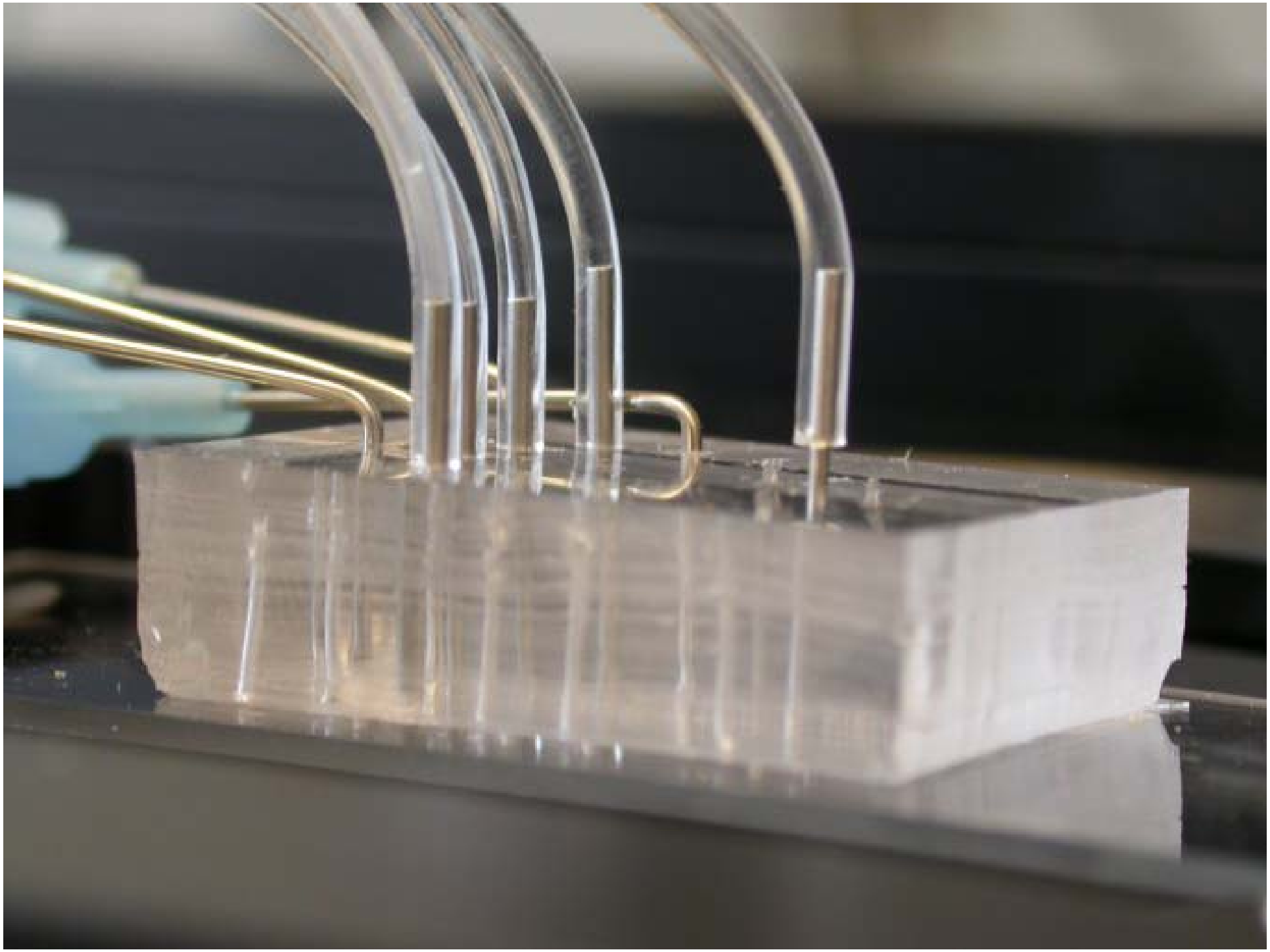
Radioactive marker

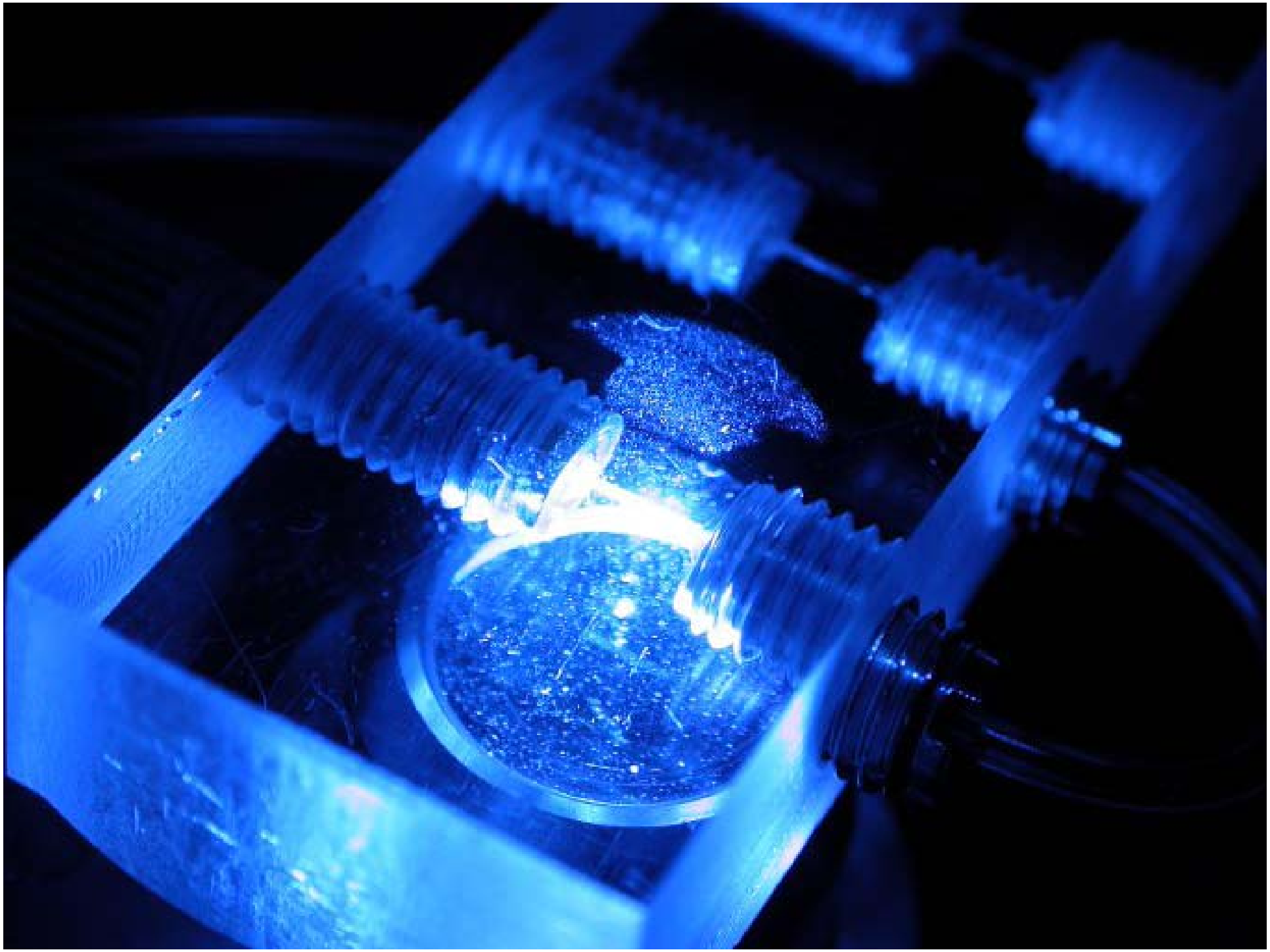


Ethidium bromide marker

Sequencing – reading the results



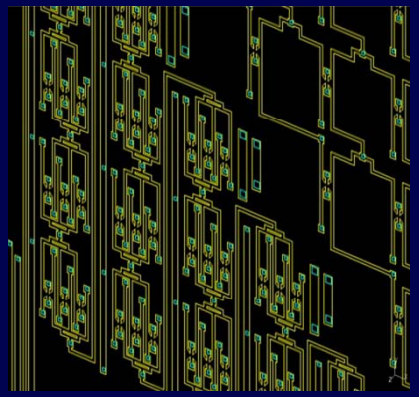




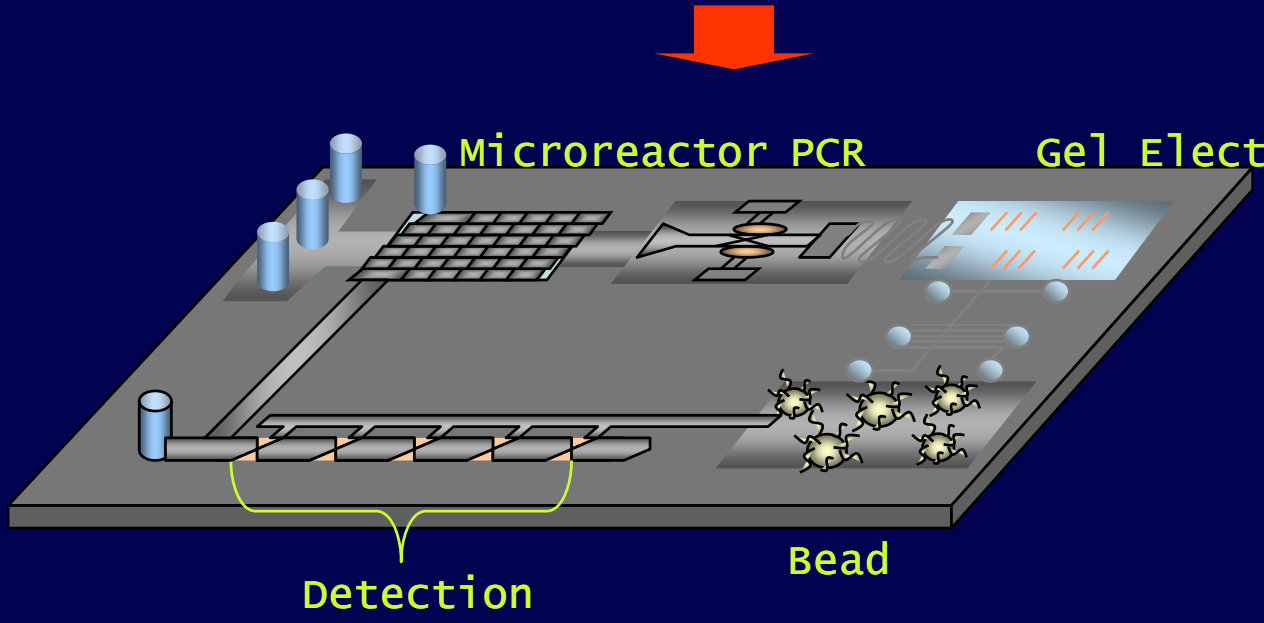
Molecular computer on a chip

DNA computing algorithm

+



MEMS (Microfluidics)



Defining a rule set

- Given the nature of DNA, we can easily determine a set of rules to operate on DNA.
- Defining a Rule Set allows for *programming* the DNA much like programming a computer.
- The rule set assume the following:
  - DNA exists in a test tube
  - The DNA is in single stranded form

Leonard Adleman

Turing Award 2002

Department of Computer Science

Areas of interest

- Method for Obtaining Digital Signatures and Public-Key Cryptosystems
- Distinguishing Prime Numbers From Composite Numbers
- The First Case of Fermat's Last Theorem
- Primality Testing And Two Dimensional Abelian Varieties Over Finite Fields
- Molecular Computation of Solutions To Combinatorial Problem

http://www.usc.edu/dept/molecular-science/fm-adleman.htm

Richard Lipton

Theoretical Computer Science College of Computing, Georgia Tech

Areas of interest

- Algorithms and Complexity Theory
- Cryptography
- DNA Computing

http://www.cc.gatech.edu/computing/Theory/theory.html

Merge

- Merge simply merges two test tubes of DNA to form a single test tube.
- Given test tubes N_1 and N_2 we can merge the two to form a single test tube, N, such that N consists of N_1 U N_2.
- Formal Definition: *merge*(N_1, N_2) = N

Amplify

- Amplify takes a test tube of DNA and duplicates it.
- Given test tube N_1 we duplicate it to form test tube N, which is identical to N_1.
- Formal Definition: $N = \mathit{duplicate}(N_1)$

Detect

- Detect looks at a test tube of DNA and returns true if it has at least a single strand of DNA in it, false otherwise.
- Given test tube N, return TRUE if it contains at least a single strand of DNA, else return FALSE.
- Formal Definition: *detect*(N)

Separate / Extract

- Separate simply separates the contents of a test tube of DNA based on some subsequence of bases.

- Given a test tube N and a word *w* over the alphabet {A, C, G, T}, produce two tubes +(N, *w*) and –(N, *w*), where +(N, *w*) contains all strands in N that contains the word *w* and –(N, *w*) contains all strands in N that doesn't contain the word *w*.

- Formal Definition:
  - N ← +(N, *w*)
  - N ← -(N, *w*)

Length - separate

- Length-Separate takes a test tube and separates it based on the length of the sequences
- Given a test tube N and an integer n we produce a test tube that contains all DNA strands with length less than or equal to n.
- Formal Definition: $N \leftarrow (N, \leq n)$

Position - Separate

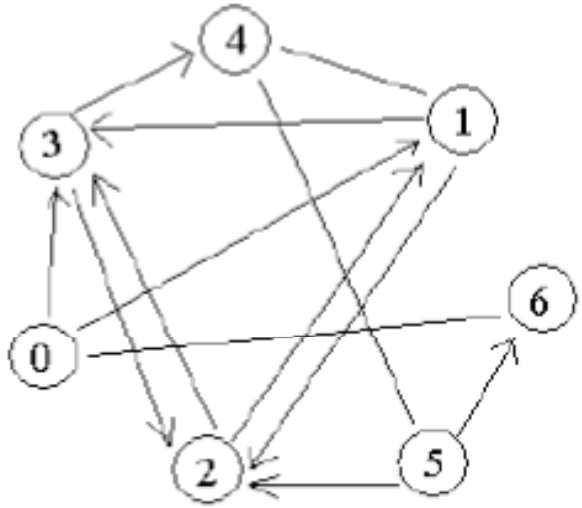
- Position-Separate takes a test tube and separates the contents of a test tube of DNA based on some beginning or ending sequence.

- Given a test tube N_1 and a word *w* produce the tube N consisting of all strands in N_1 that begins/ends with the word *w*.

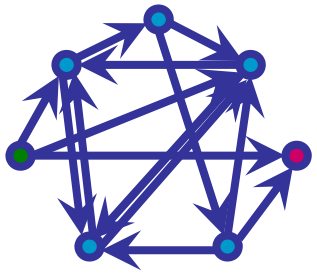
- Formal Definition:
  - $N \leftarrow B(N_1, w)$
  - $N \leftarrow E(N_1, w)$

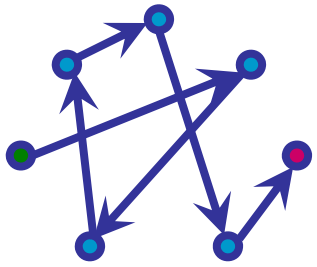
Back to Adleman's experiment

Now that we have some simple rules at our disposal we can easily create a simple program to solve the Hamiltonian Path problem for a simple 7-node graph as outlined by Adelman.



The 1994 experiment

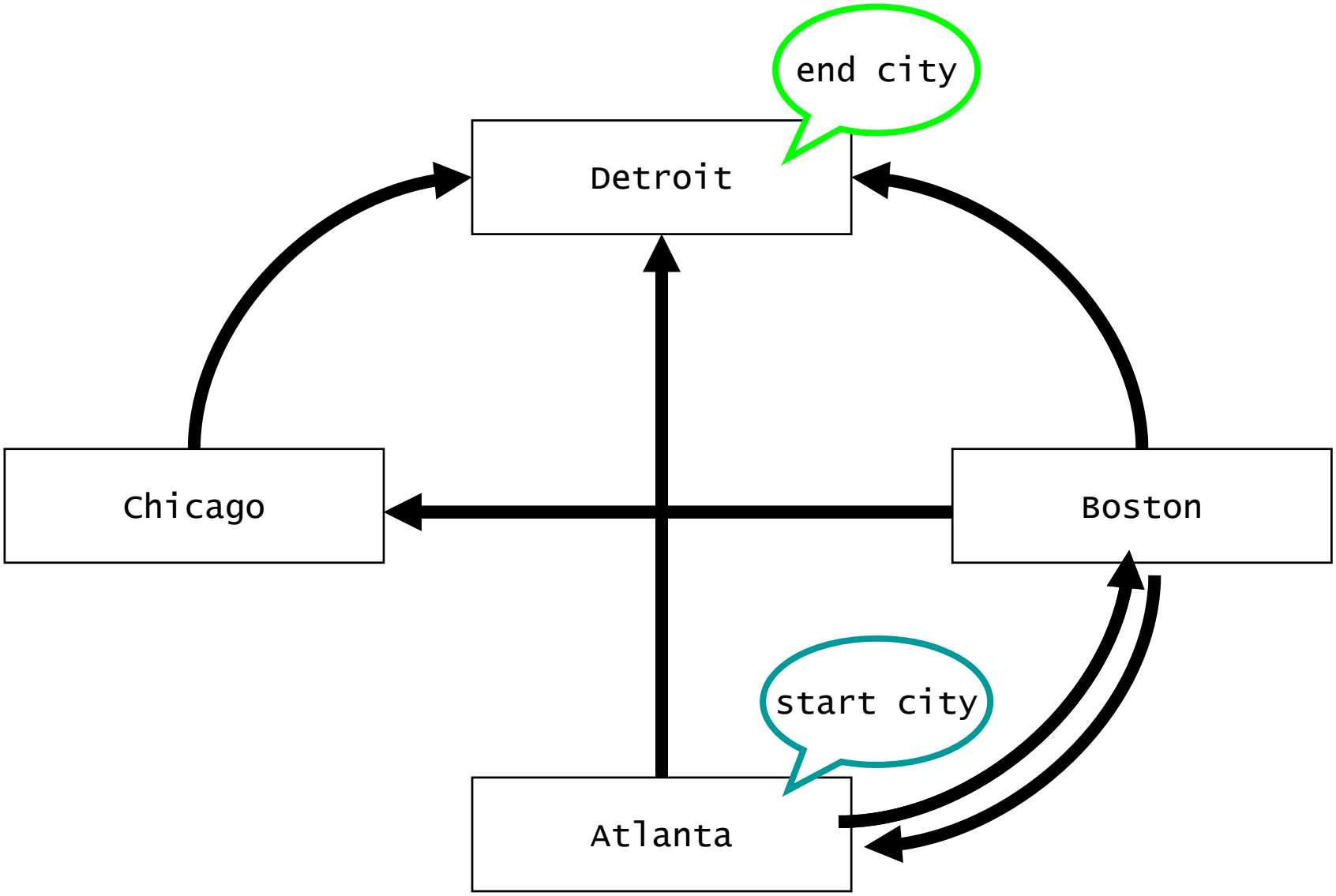




The program

```
1  input(N)
2  N ← B(N, s0)
3  N ← +(N, s6)
4  N ← +(N, ≤ 140)
5  for i = 1 to 5 do
   begin
       N ← +(N, si)
   end
6  detect(N)
```

Solving a Hamiltonian path problem

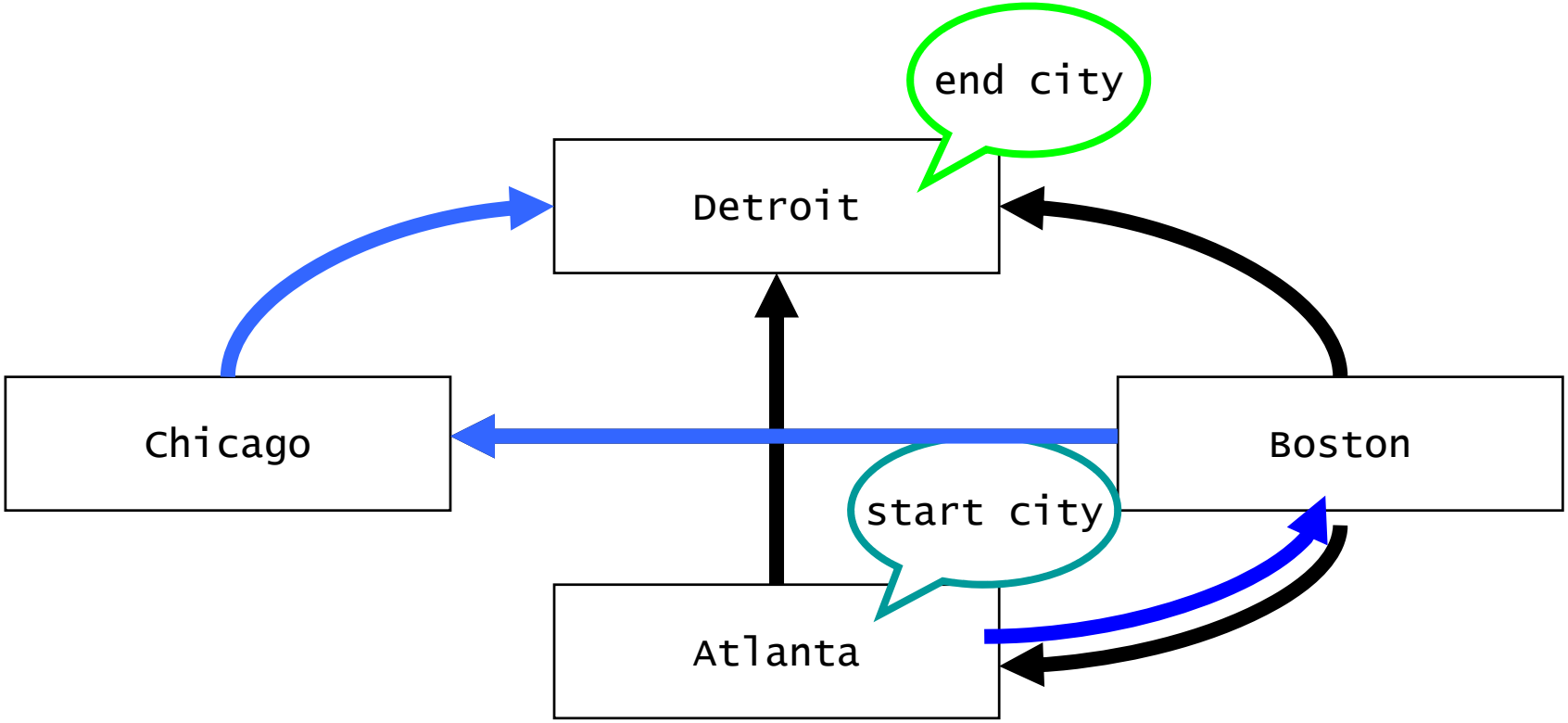


City coding

CITY	DNA NAME	COMPLEMENT
ATLANTA	ACTT*GCAG*	TGAA*CGTC*
BOSTON	TCGG*ACTG*	AGCC*TGAC*
CHICAGO	GGCT*ATGT*	CCGA*TACA*
DETROIT	CCGA*GCAA*	*GGCTCGTT*

CONNECTING PATH	DNA PATH
ATLANTA-BOSTON	*GCAG*TCGG
ATLANTA-DETROIT	*GCAG*CCGA
BOSTON-CHICAGO	*ACTG*GGCT
BOSTON-DETROIT	*ACTG*CCGA
BOSTON-ATLANTA	*ACTG*ACTT
CHICAGO-DETROIT	*ATGT*CCGA

Possible paths



Atlanta-Boston	Boston-Chicago	Chicago-Detroit

Atlanta*	Boston*	Chicago*	Detroit*

Maximal clique problem

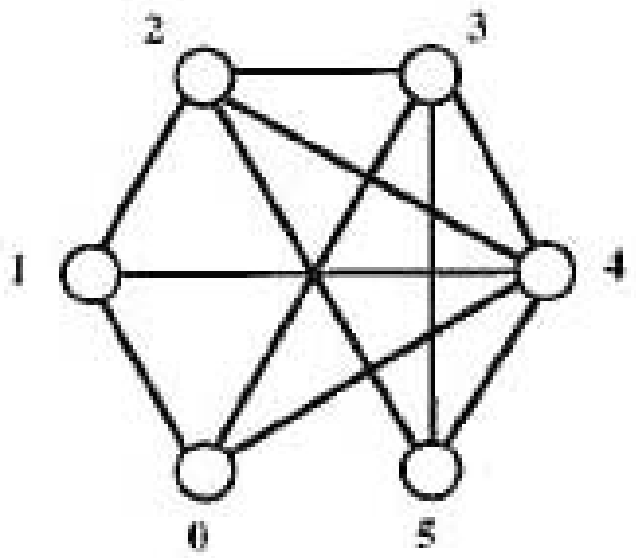
- **Clique**

  defined as a set of vertices in which every vertex is connected to every other vertex by an edge

- **Maximal clique problem**

  Given a network containing *N* vertices and *M* edges, ***how many vertices are in the largest clique?***

  Finding the size of the largest clique has been proven to be an NP-complete problem



(a)

Algorithm

Step 1 Make the **complete data pool**

For a graph with N vertices, each possible clique is represented by an N-digit binary number

1: a vertex in the clique

0: a vertex out of the clique

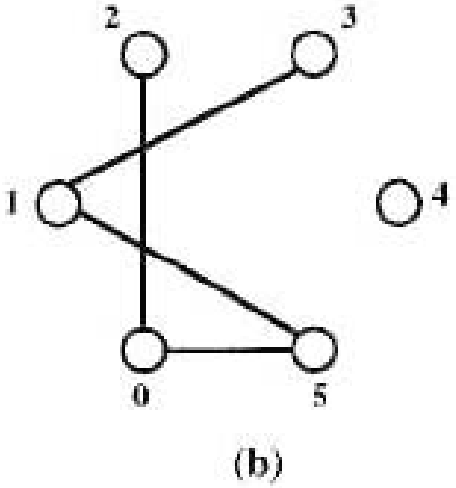
i.e. clique (4,1,0)⇒binary number 010011

Step 2

Find pairs of vertices in the graph that are not connected by an edge

(0,2) (0,5) (1,5) (1,3)

The complementary graph



(b)

Algorithm

Step 3

Eliminate from the complete data pool all numbers containing connections in the complementary graph

$\Rightarrow$ xxx1x1 or 1xxxx1 or 1xxx1x or xx1x1x

Step 4

Sort the remaining data pool to find the data containing the largest number of 1's

$\Rightarrow$ the clique with **the largest number of 1's** tells us the **size** of the maximal clique

Conclusions

- DNA Computing uses DNA molecules to computing or storage materials.
- DNA computing technology has many interesting properties, including
  - Massively parallel, solution-based, biochemical
  - Nano-scale, biocompatible
  - high energy efficiency
  - high memory storage density
- DNA computing is in very early stage of development.

What is Molecular Computing?

- In vivo
- Uses DNA molecules to store data
- Uses cells to compute by means of processes which take place in living organisms

Laura Landweber

Dept. of Ecology and Evolutionary Biology
Princeton University

Areas of interest

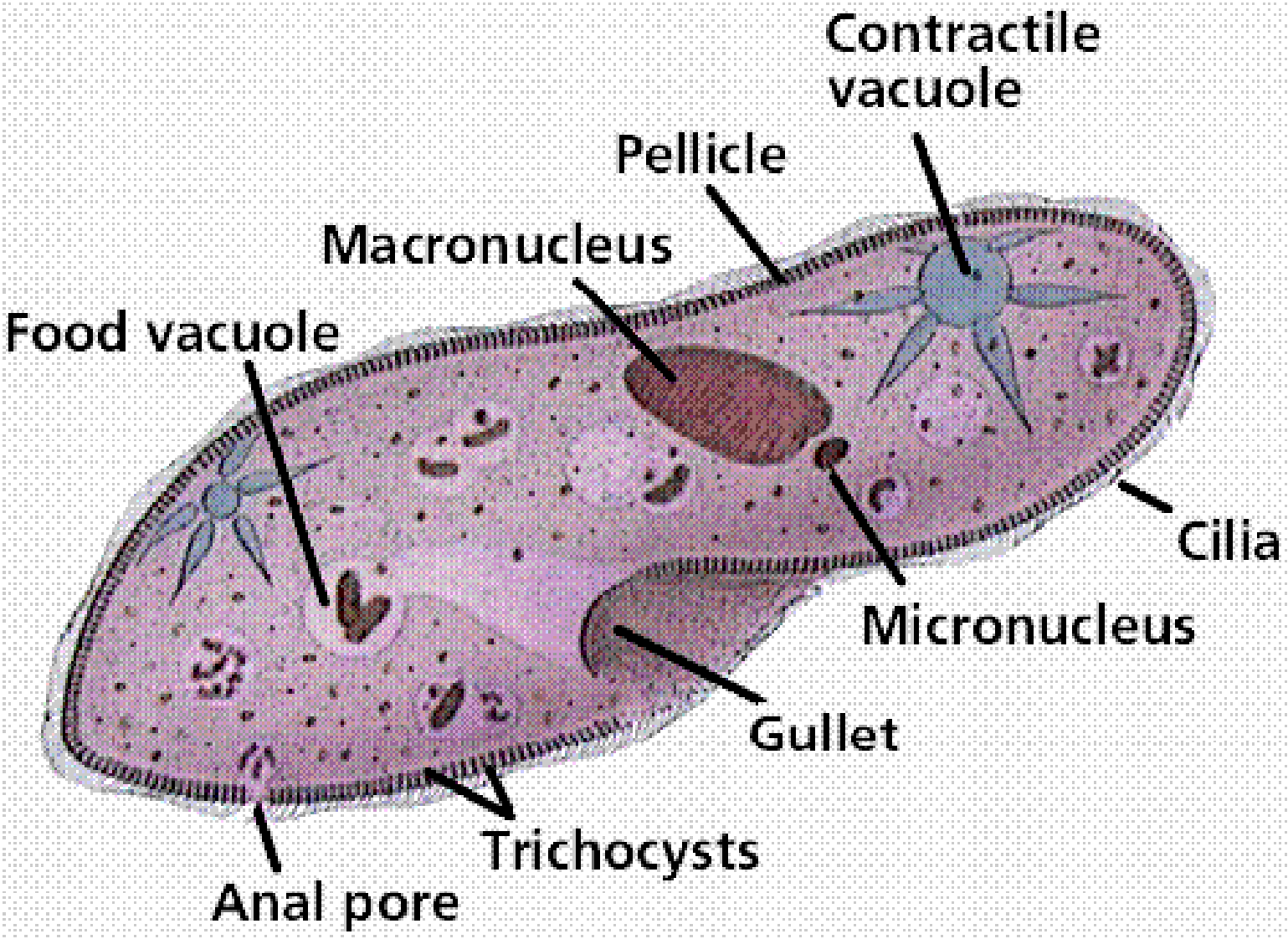
- Origins of Genes, Genomes
- the Genetic Code
- Early Pathways of RNA Evolution
- Scrambled Genes
- RNA Editing
- Gene Scrambling
- DNA Computing

http://www.princeton.edu/~lfl/

The ciliate

- Very ancient (~ 2 . 10^9 years ago)
- Very rich group (~ 10000 genetically different organisms)
- Very important from the evolutionary point of view

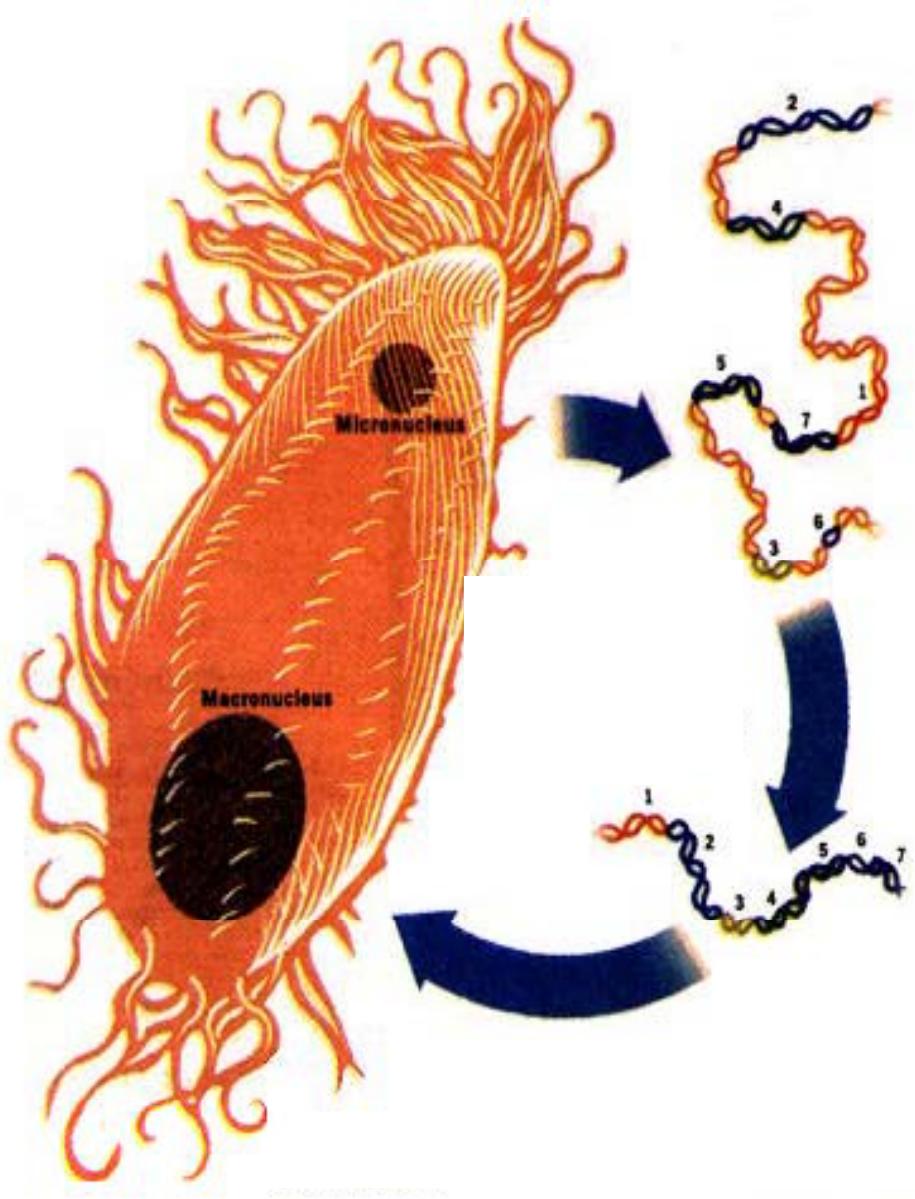
The ciliate


Contractile vacuole
Pellicle
Macronucleus
Food vacuole
Cilia
Micronucleus
Gullet
Trichocysts
Anal pore

Nuclei

- Micronucleus the small nucleus containing a single copy of the genome that is used for sexual reproduction
- Macronucleus the large nucleus that carries up to several hundred copies of the genome and controls metabolism and asexual reproduction (cell growth and proliferation)

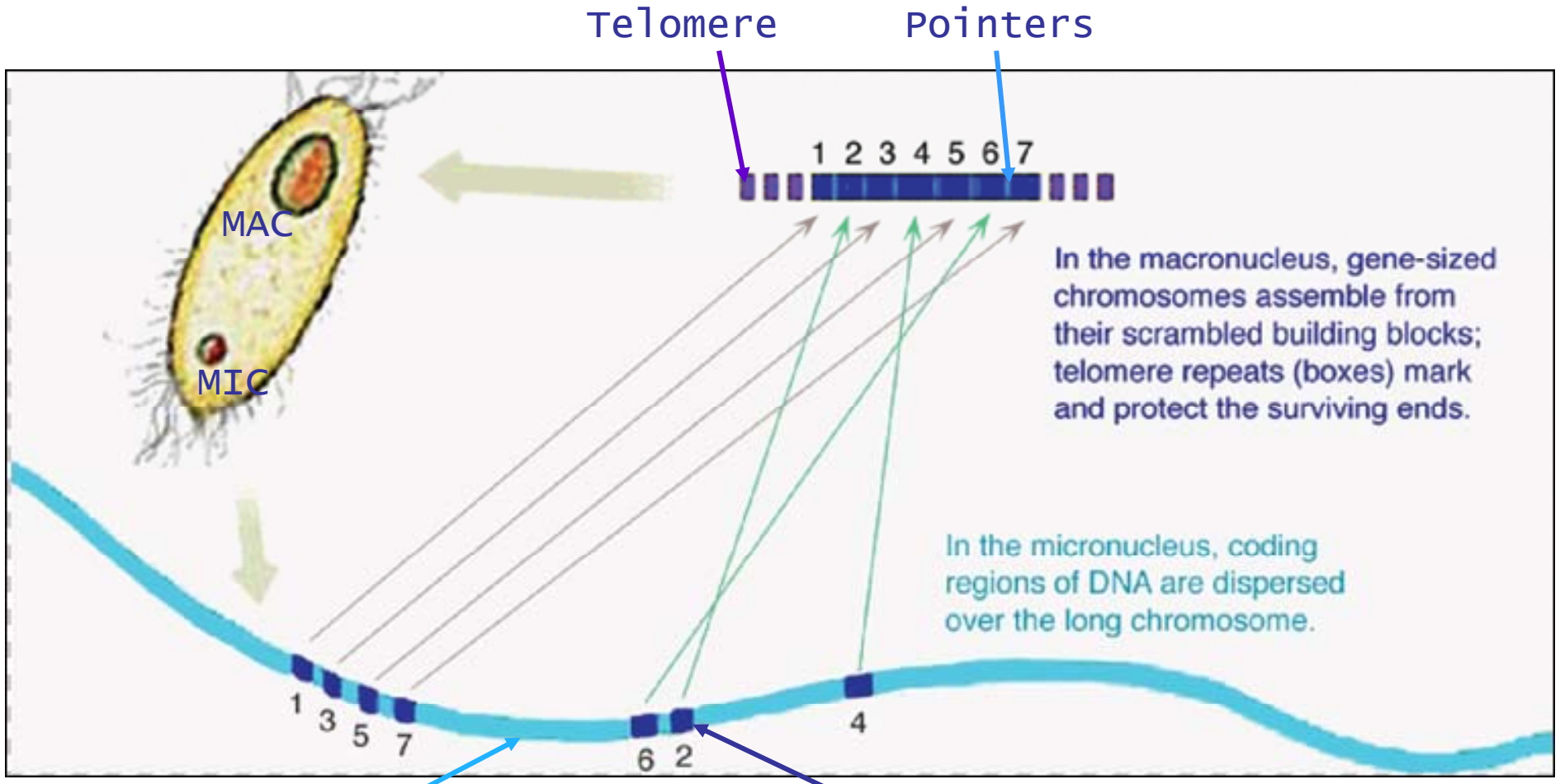
The ciliate



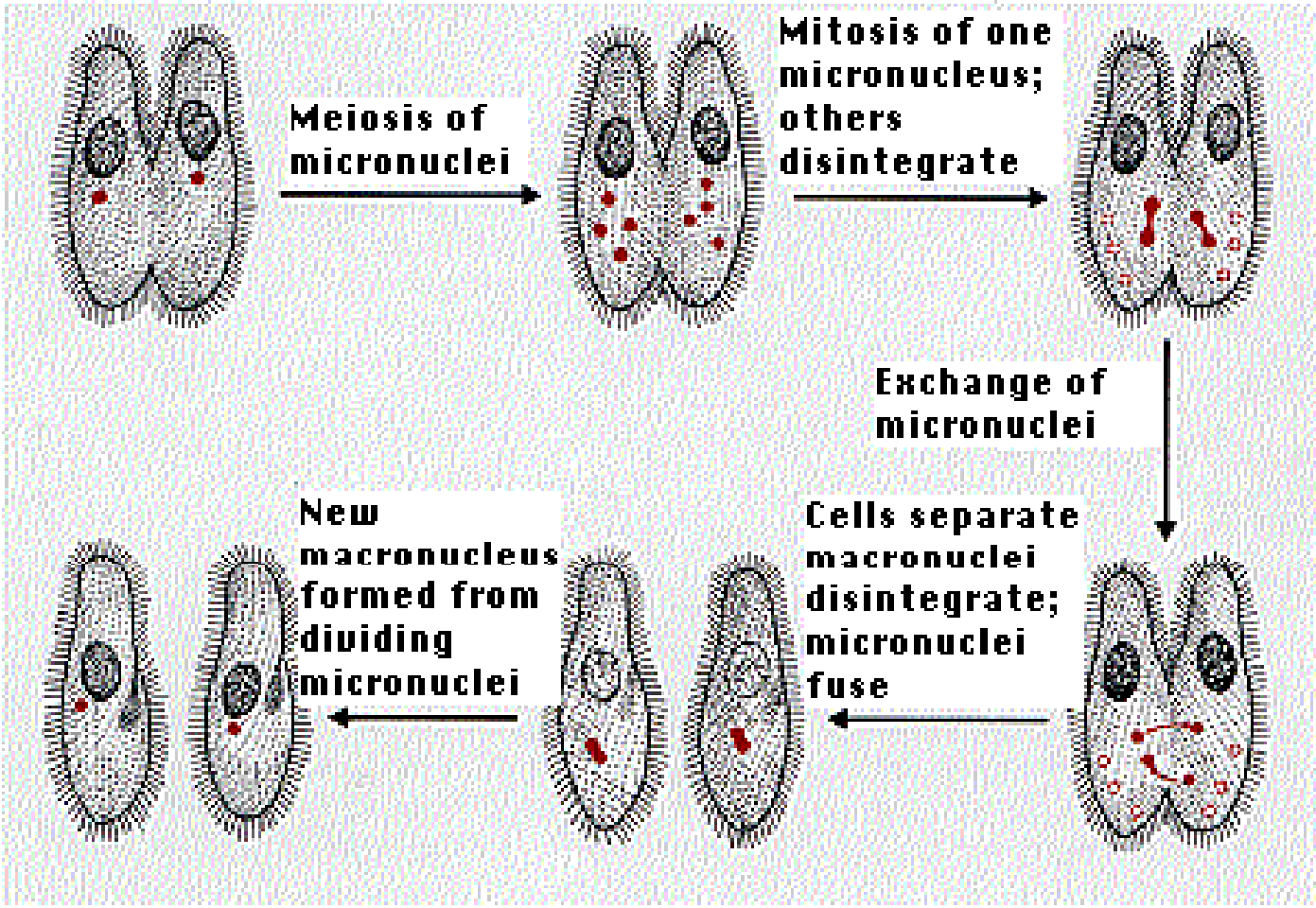
In initial micronucleus, DNA is‘junky’ and scrambled, but it contains the description of the working mode of the macronucleus in an encoded way.

It reassembles itself in proper sequence by means of computer-like acrobatics (unscrambling, throwing out genetic ‘junk’)–in macronucleus It reorganizes the material by removing non-coding sequences and placing the coding sequences in the correct order. Approx. 95% is discarded.

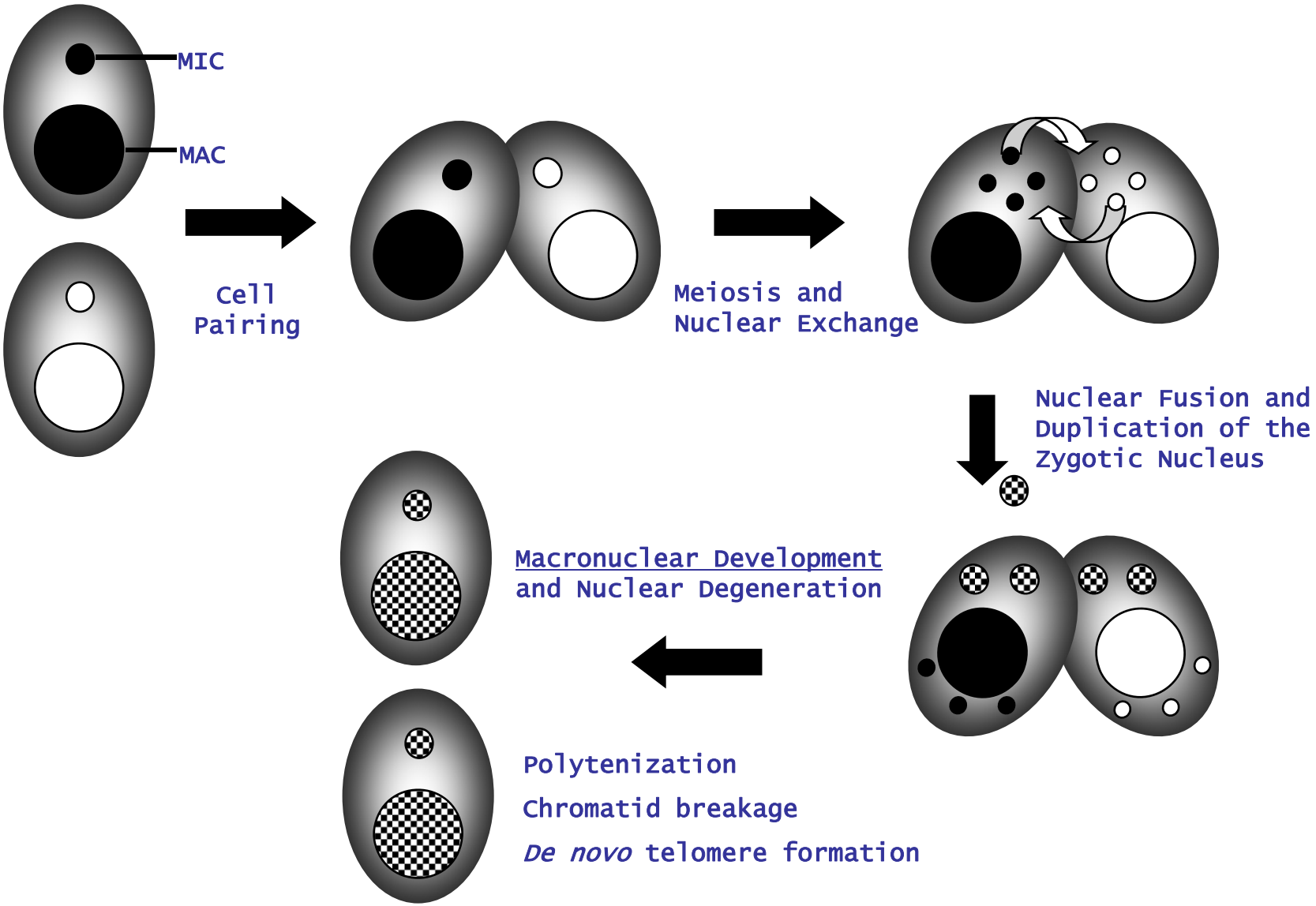
The complexity of spirotrich biology



The ciliate, meiosis

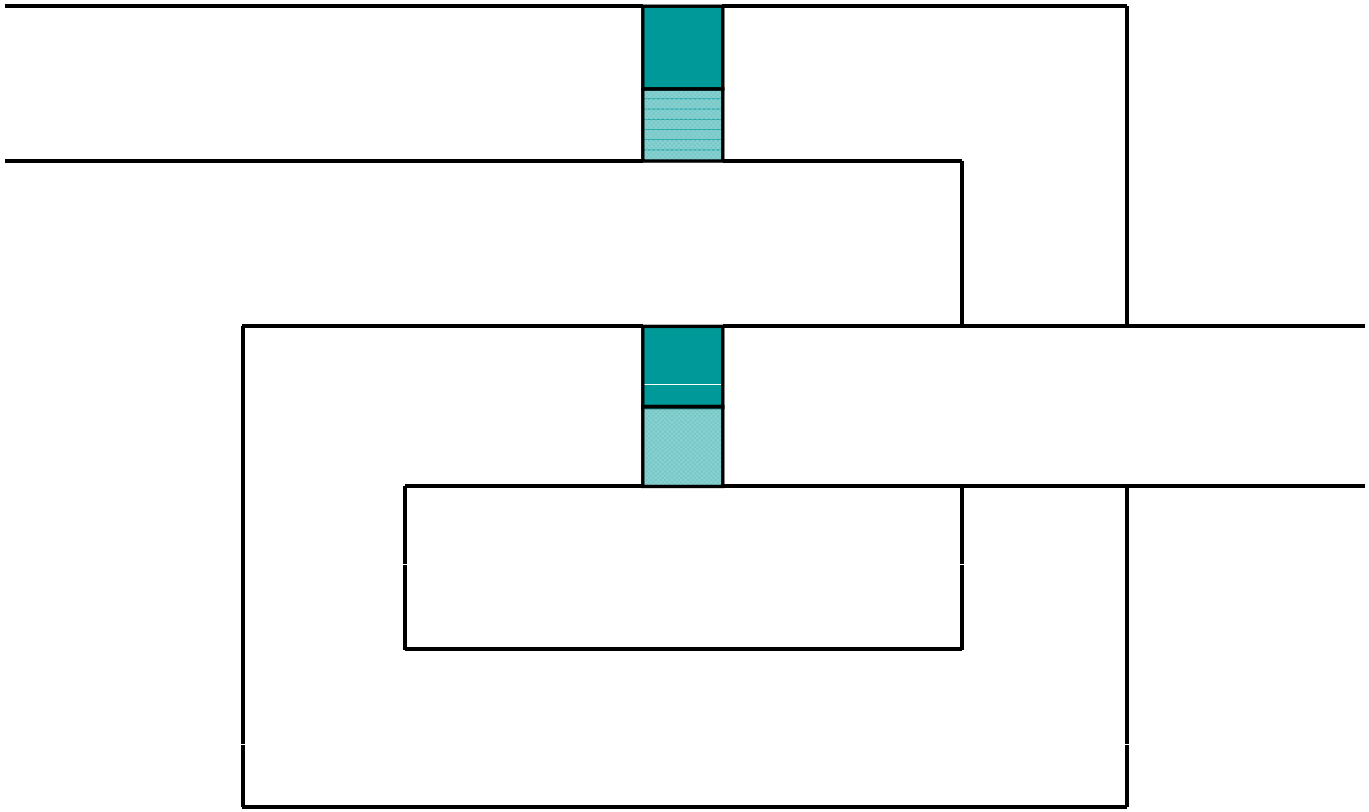


The ciliate, reproduction

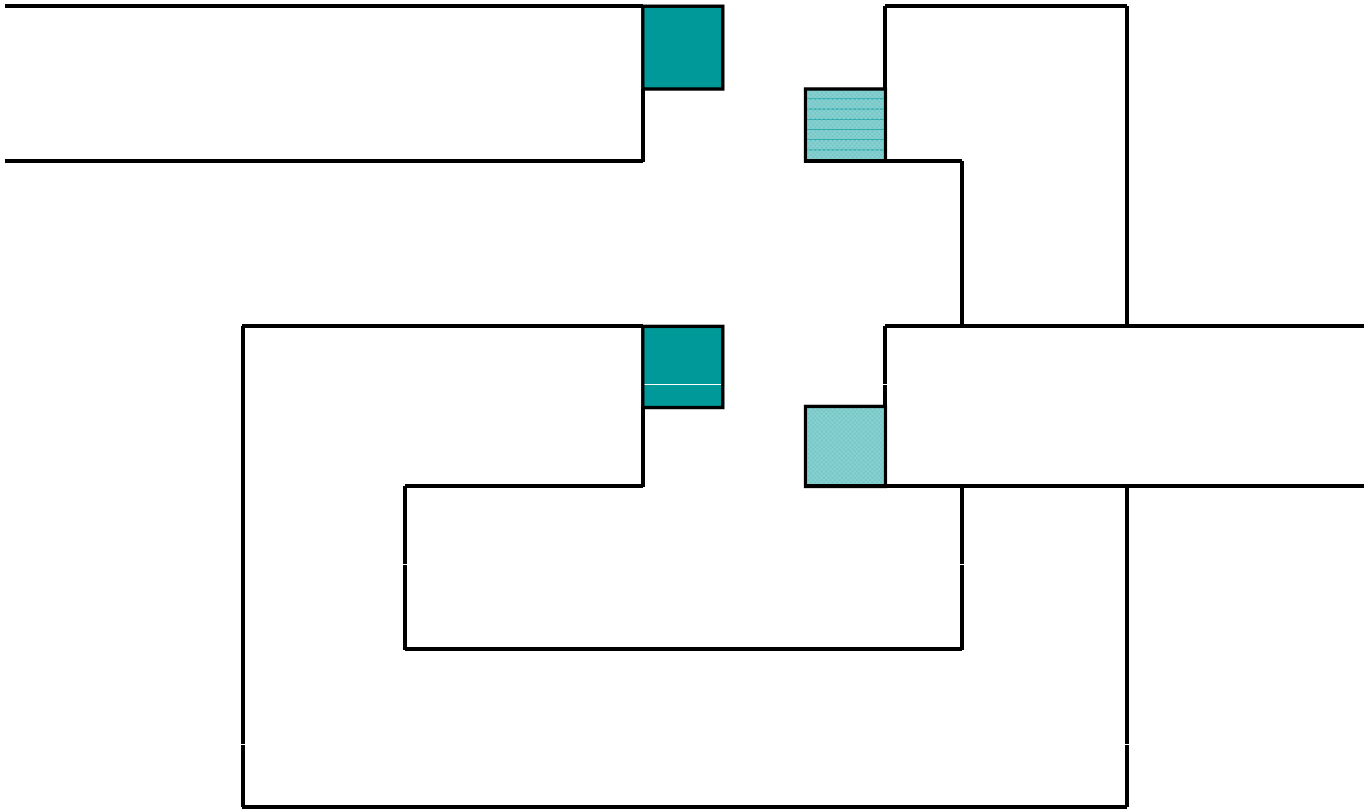


Modified from Larry Klobutcher & Carolyn Jahn Ann. Review Microbiology, 2002

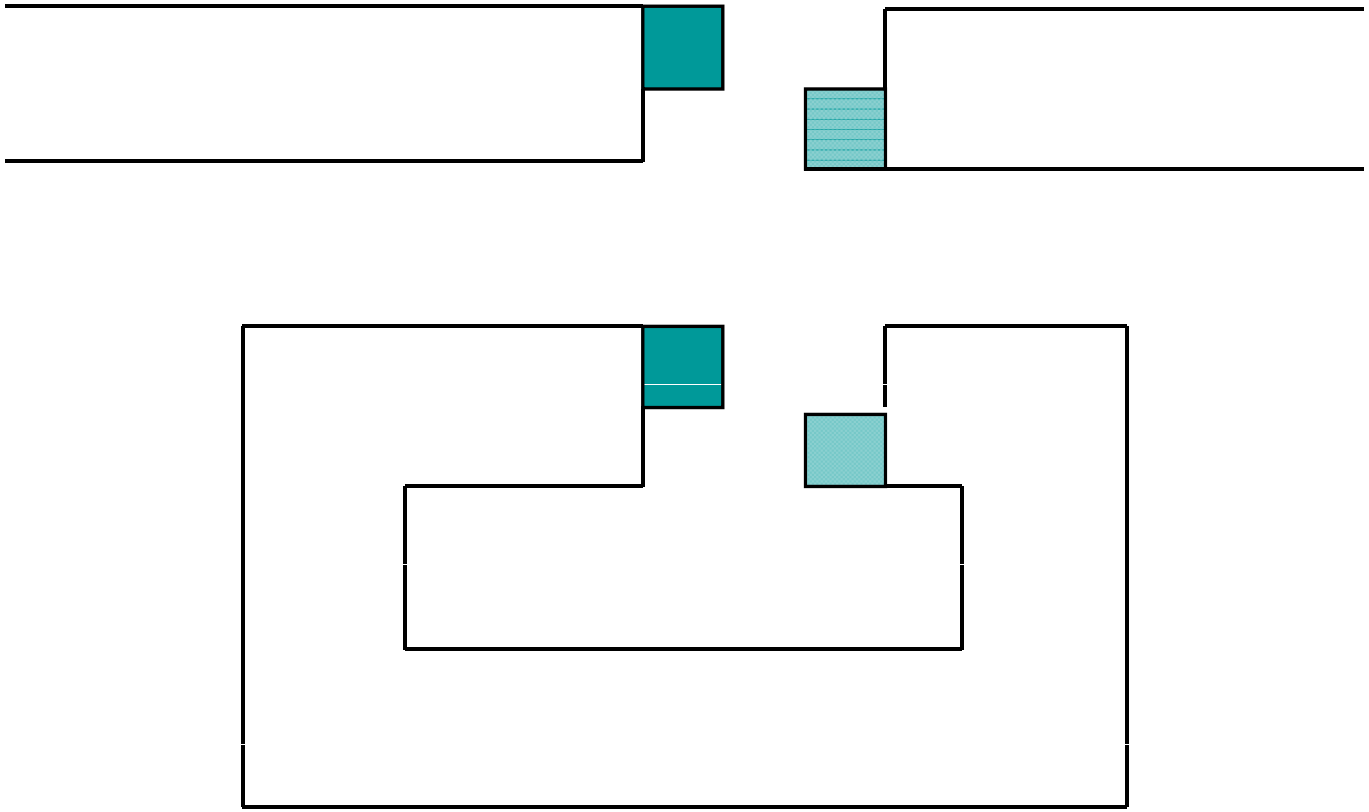
Molecular operators (ld)



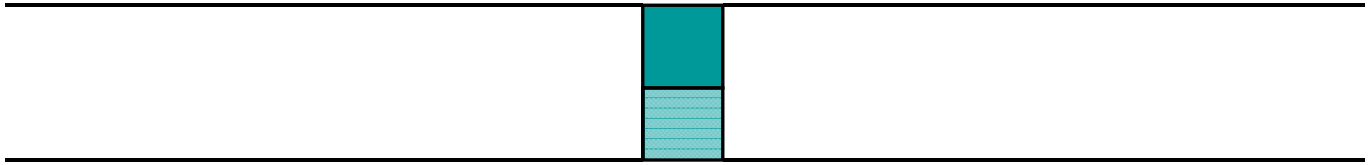
Molecular operators

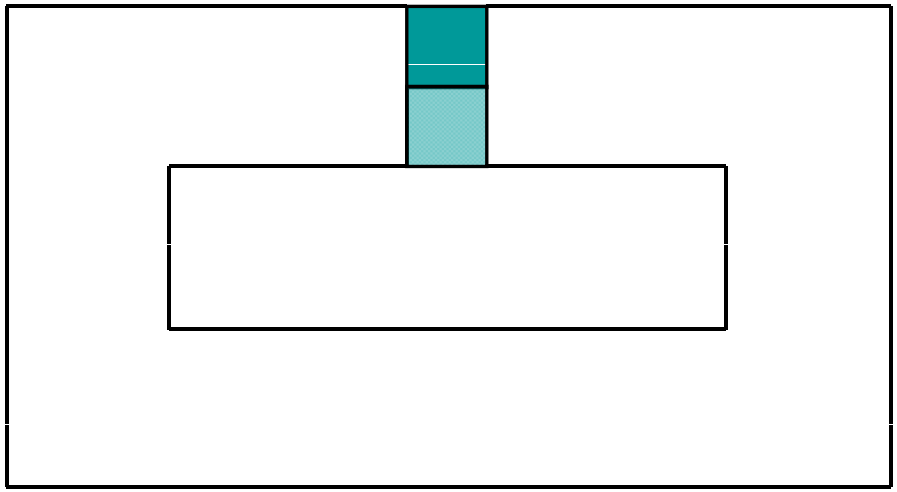


Molecular operators

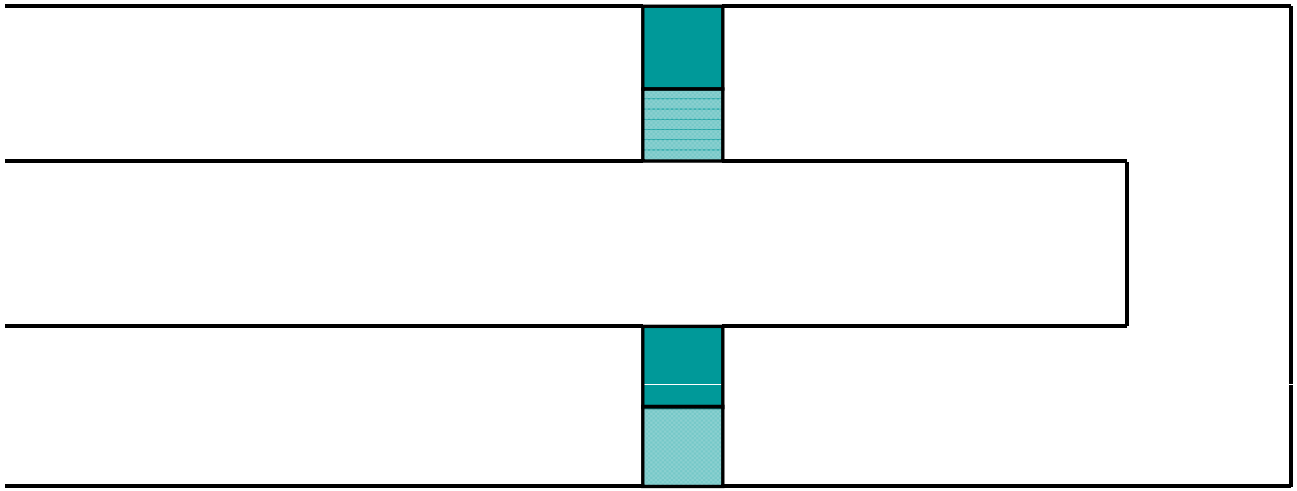


Molecular operators

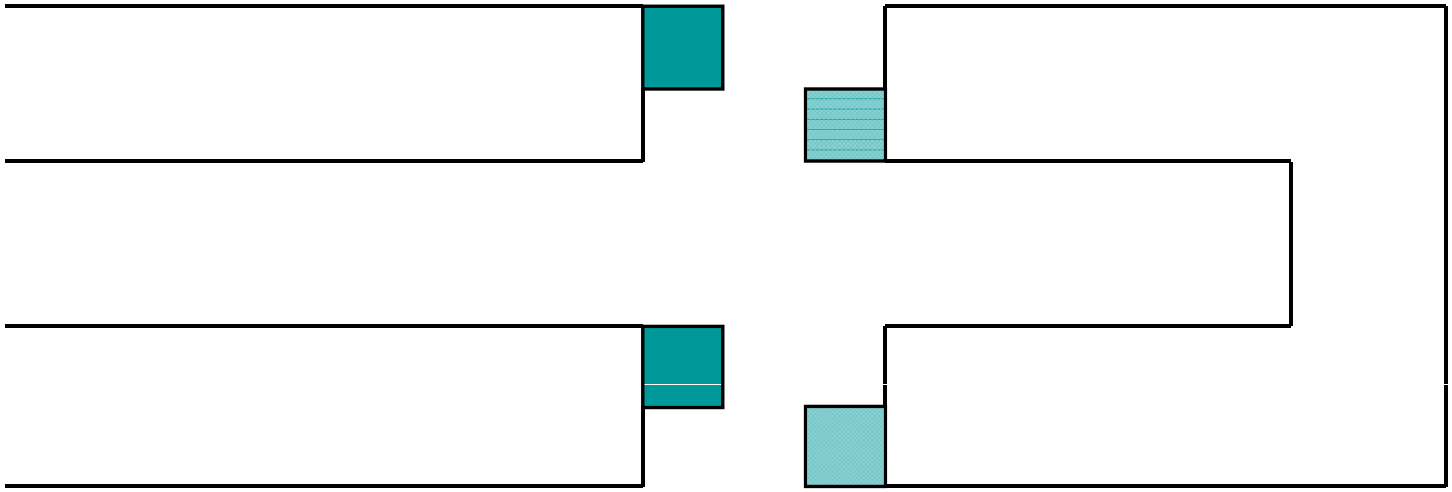




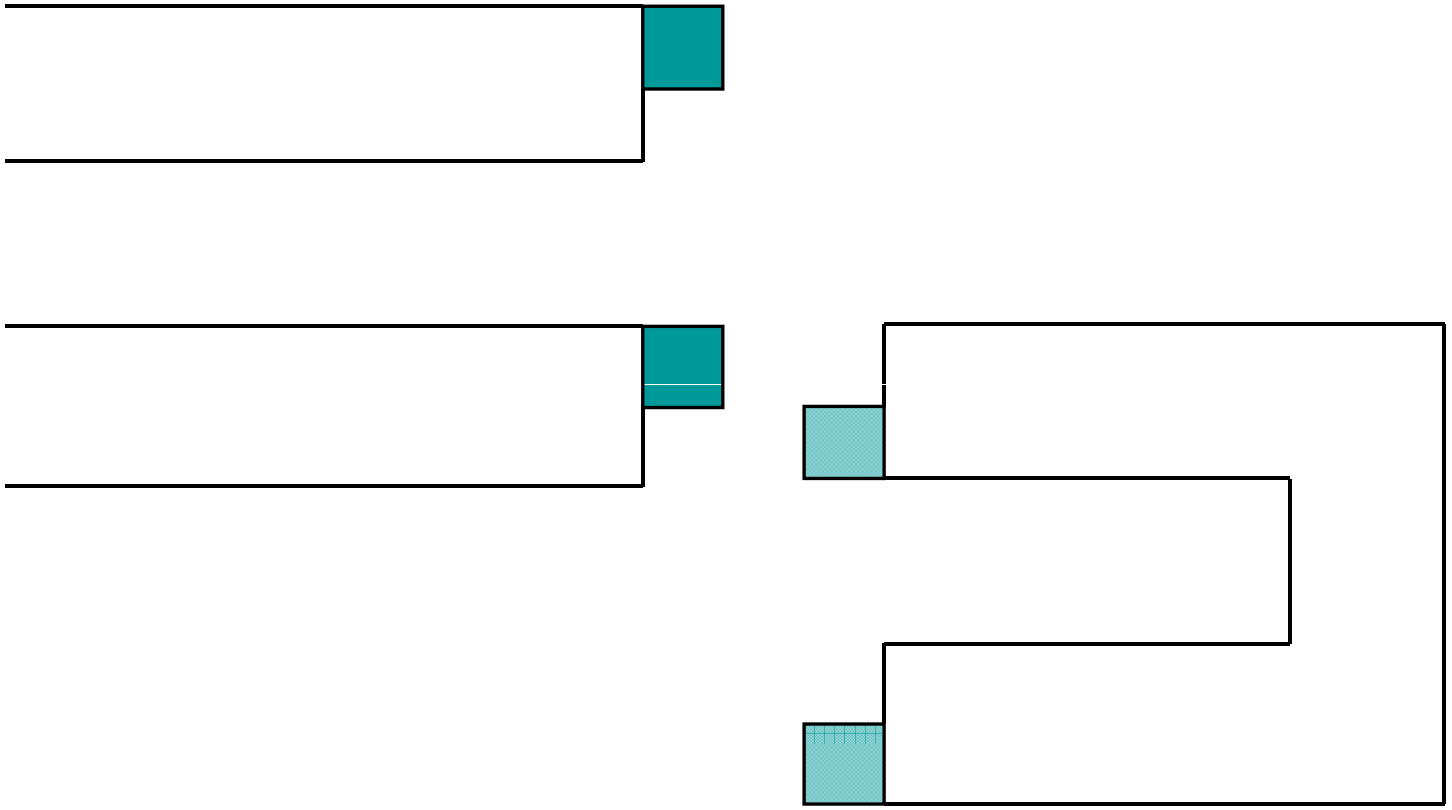
Molecular operators (hi)



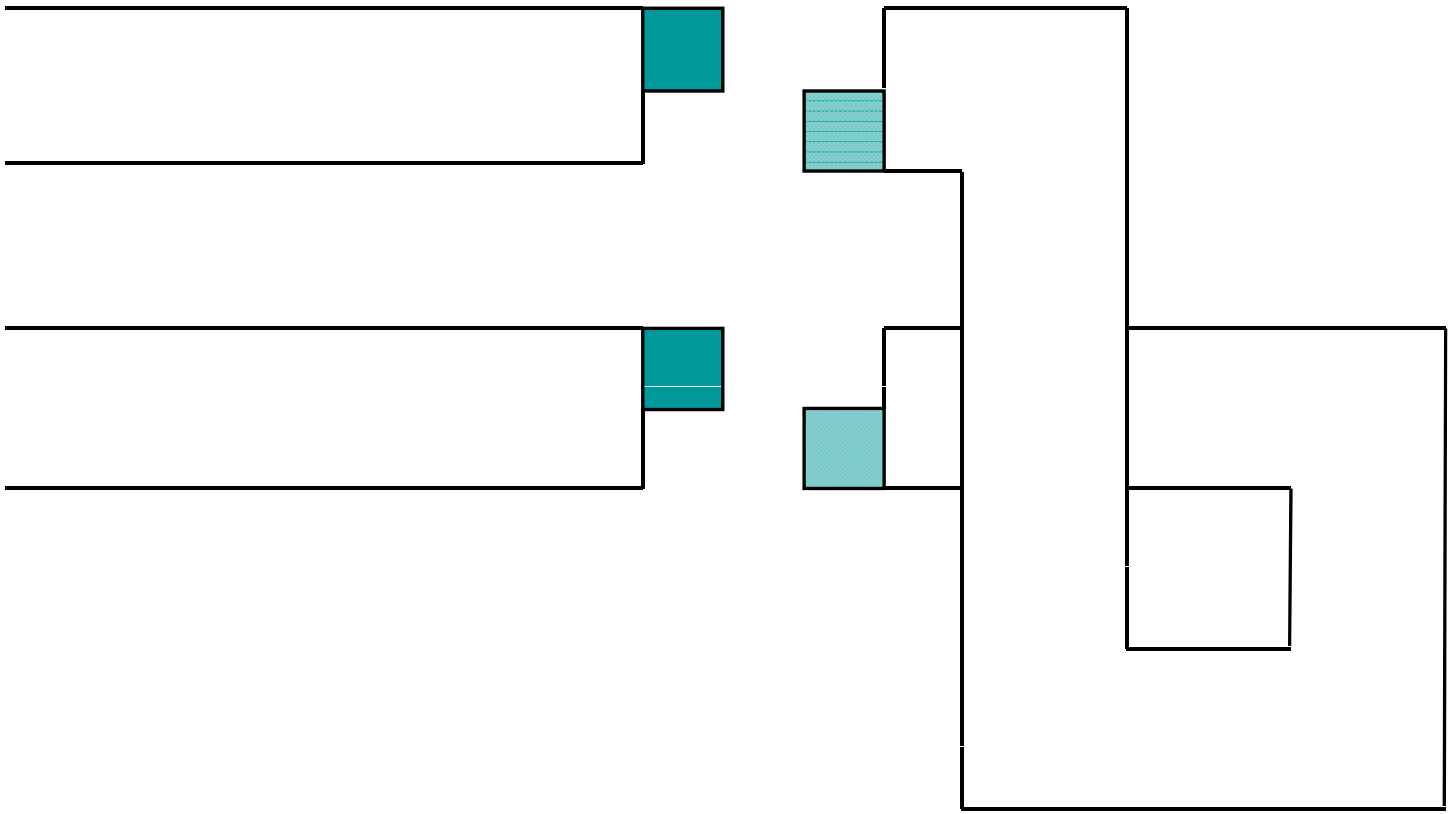
Molecular operators



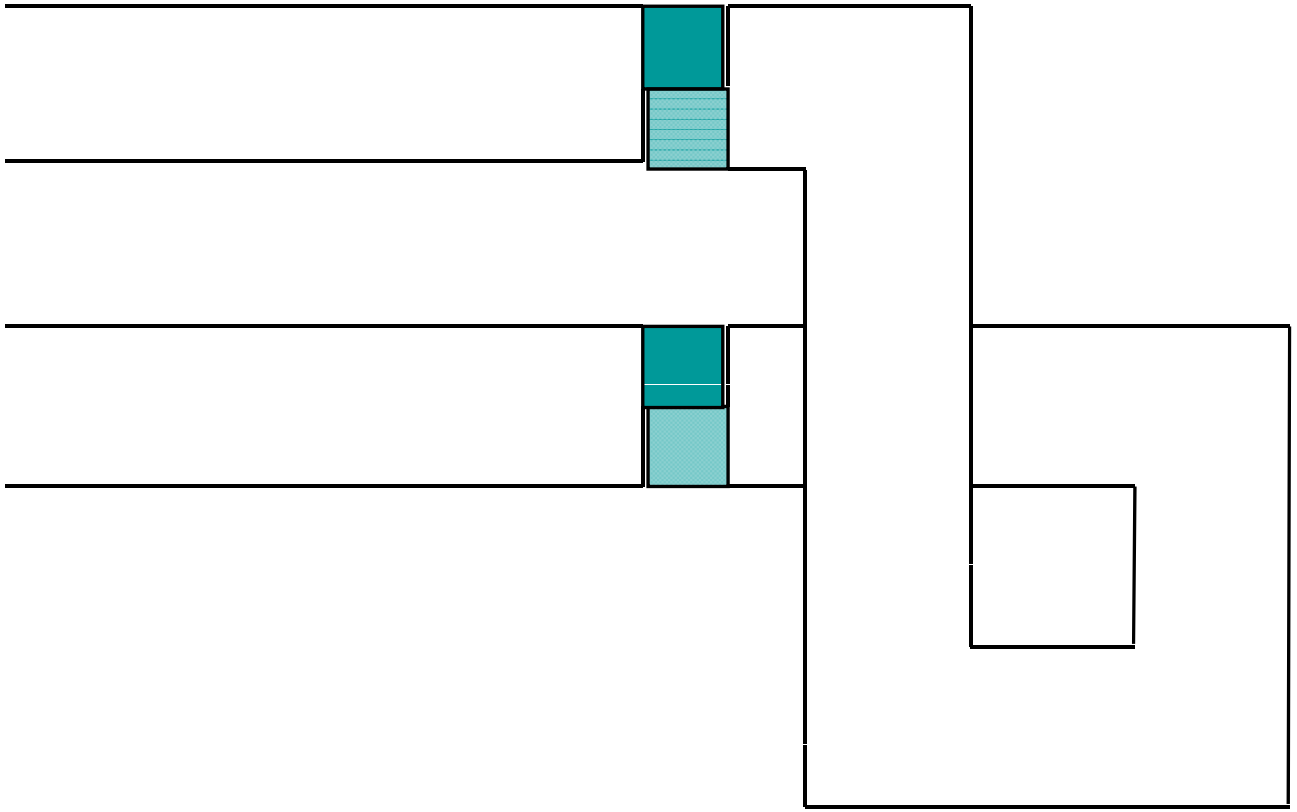
Molecular operators



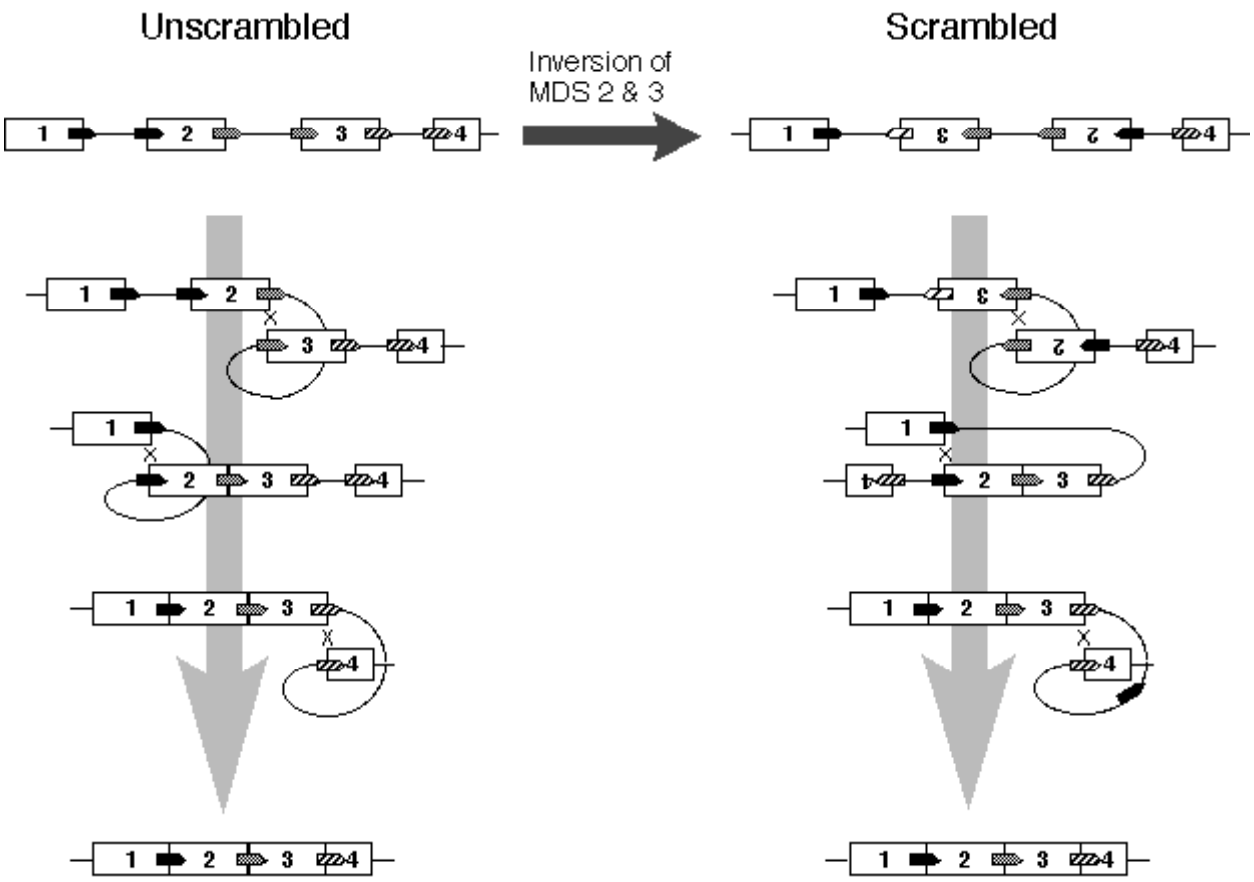
Molecular operators



Molecular operators



Ciliate computing



Formal model

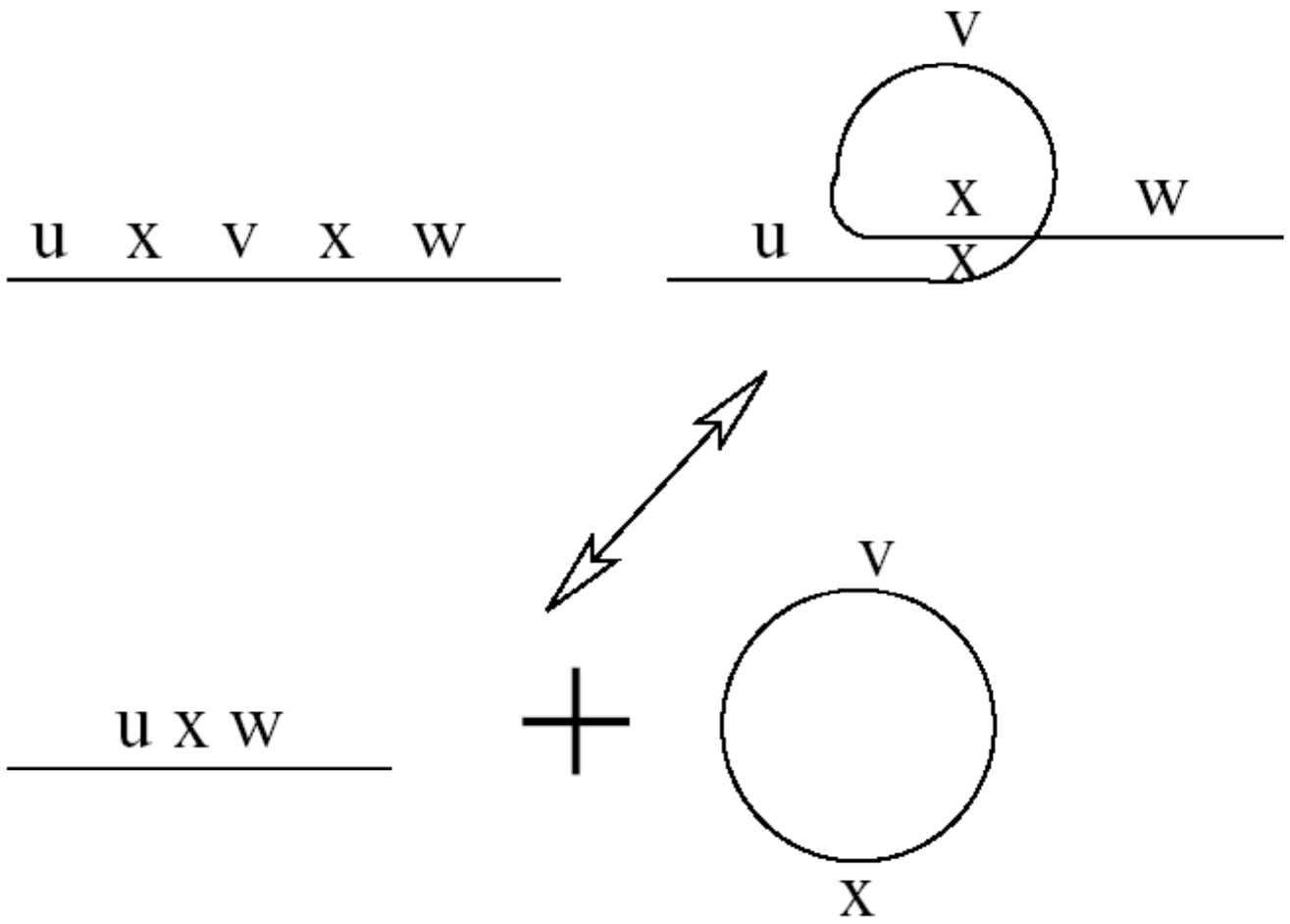
- Formal Language Model

$$uxwxv == uxv + [wx]$$

  Where u=u'p, w=qw'=w''p', v=q'v'

- Intramolecular recombination. The guide is x.
- Intermolecuar recombination. Strand Exchange.

Formal model



Ciliate computing

(J.Kari, L. Kari, L. Landweber 1999)

$ux_{12}\alpha_2x_{23}e_1x_{34}\alpha_4t_ee_2t_s\alpha_1x_{12}e_3x_{23}\alpha_3x_{34}v$

$ux_{12}e_3x_{23}\alpha_3x_{34}v,\ [\alpha_2x_{23}e_1x_{34}\alpha_4t_ee_2t_s\alpha_1x_{12}]$

$ux_{12}e_3x_{23}\alpha_3x_{34}v,\ [e_1x_{34}\alpha_4t_ee_2t_s\alpha_1x_{12}\alpha_2x_{23}]$

$ux_{12}e_3x_{23}e_1x_{34}\alpha_4t_ee_2t_s\alpha_1x_{12}\alpha_2x_{23}\alpha_3x_{34}v$

$ux_{12}e_3x_{23}e_1x_{34}v,\ [\alpha_4t_ee_2t_s\alpha_1x_{12}\alpha_2x_{23}\alpha_3x_{34}]$

$ux_{12}e_3x_{23}e_1x_{34}v,\ [t_s\alpha_1x_{12}\alpha_2x_{23}\alpha_3x_{34}\alpha_4t_ee_2]$

$t_s\alpha_1x_{12}\alpha_2x_{23}\alpha_3x_{34}\alpha_4t_e,\ e_2,\ ux_{12}e_3x_{23}e_1x_{34}v$

Ciliate computing

We use our knowledge of the first step to develop a model for the guided homologous recombinations and prove that such a model has the computational power of a Turing machine, the accepted formal model of computation. This indicates that, in principle, these unicellular organisms may have the capacity to perform at least any computation carried out by an electronic computer.

Thank You