Molecular Nanobiointelligence Computers

National Cancer Center, June 21, 2005

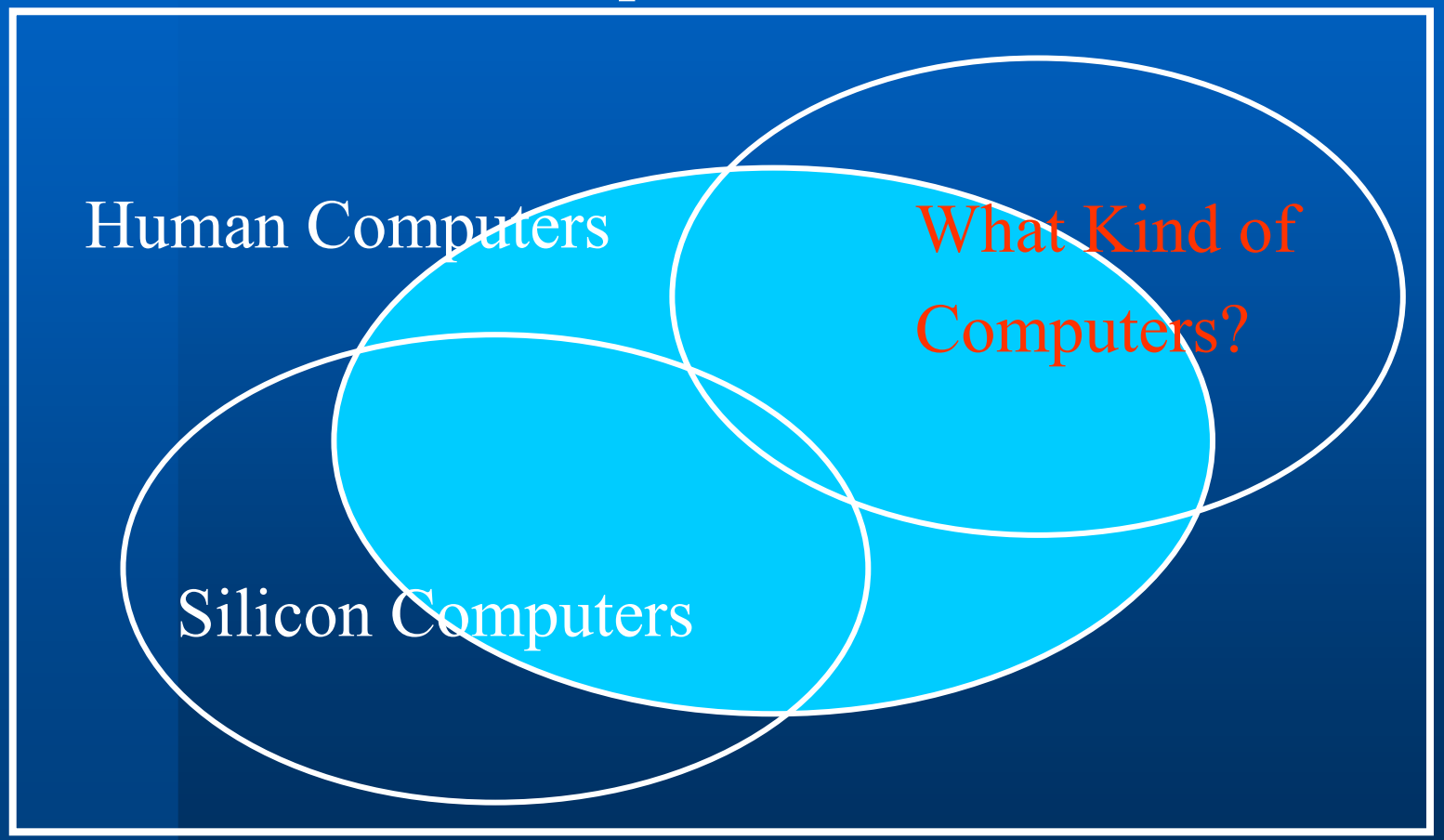
Byoung-Tak Zhang

Center for Bioinformation Technology (CBIT) &
Biointelligence Laboratory
School of Computer Science and Engineering
Seoul National University

btzhang@cse.snu.ac.kr
http://bi.snu.ac.kr/

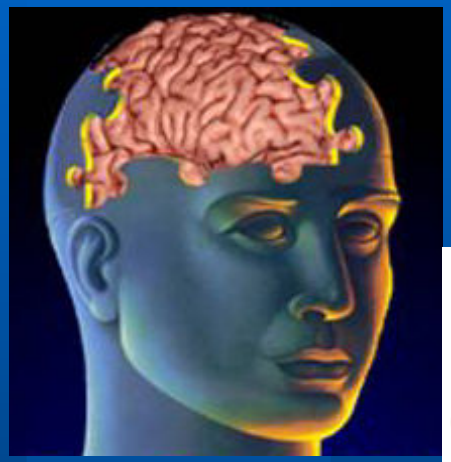
Humans and Computers

The Entire Problem Space



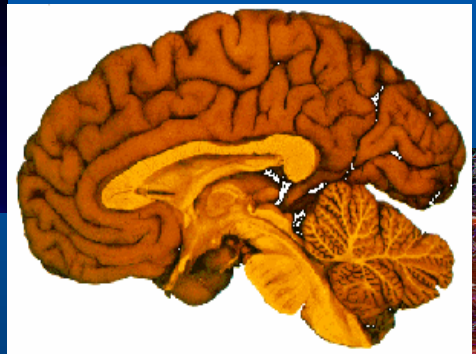
Mind, Brain, Cells, Molecules

Mind



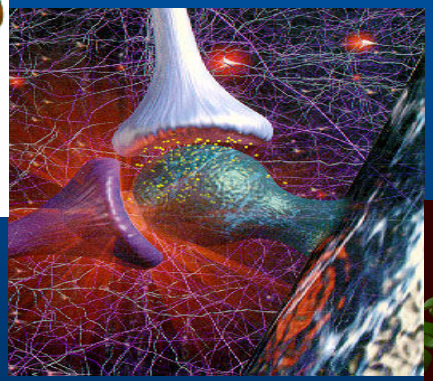
∞ memory

Brain



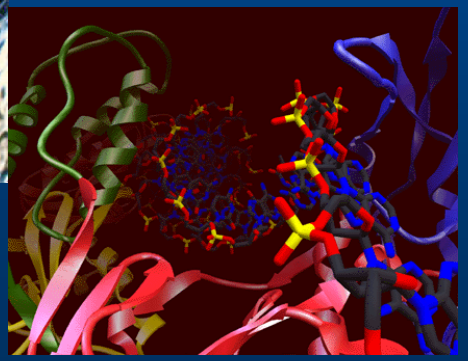
10^{11} cells

Cell

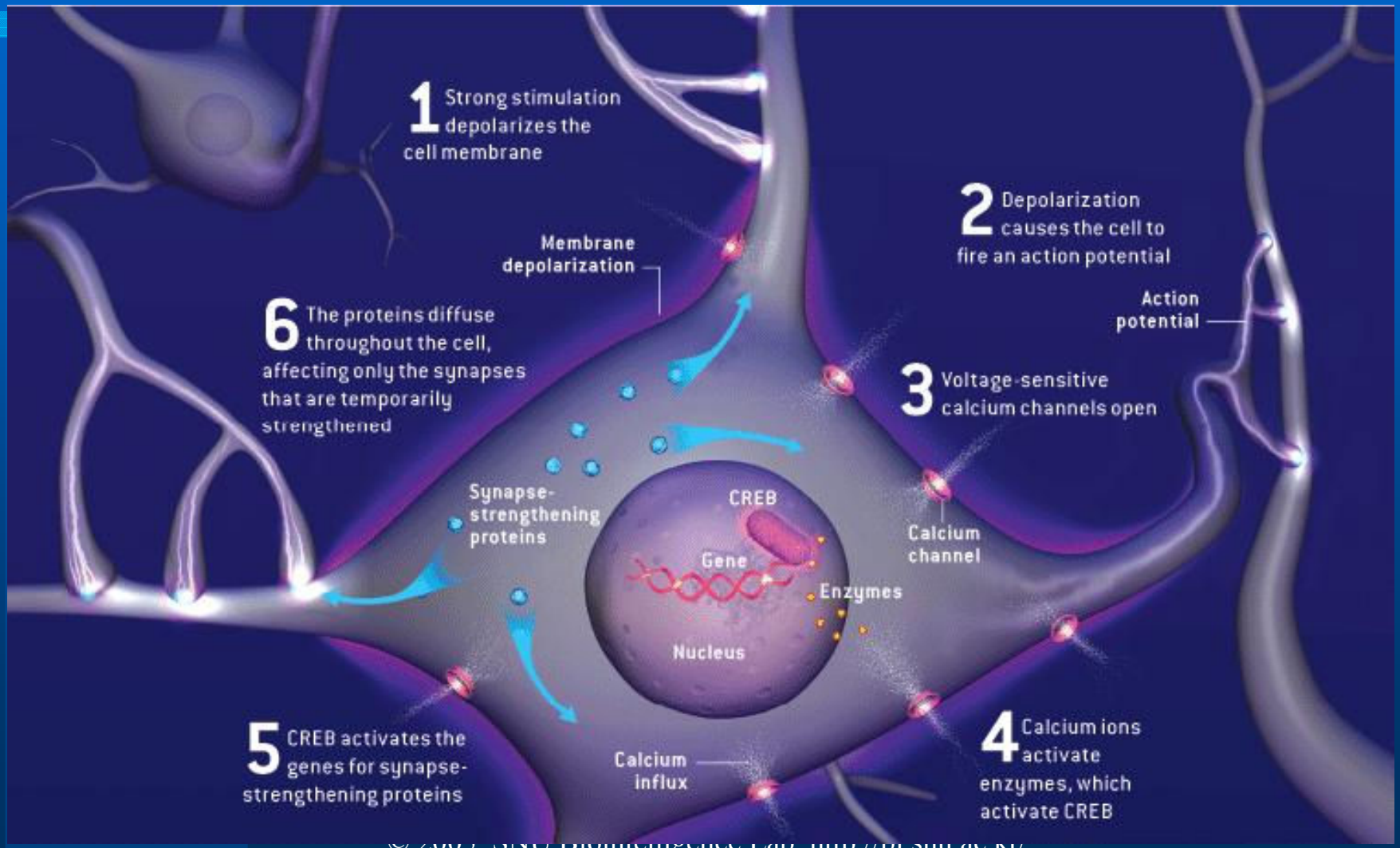


10^{10} mol.

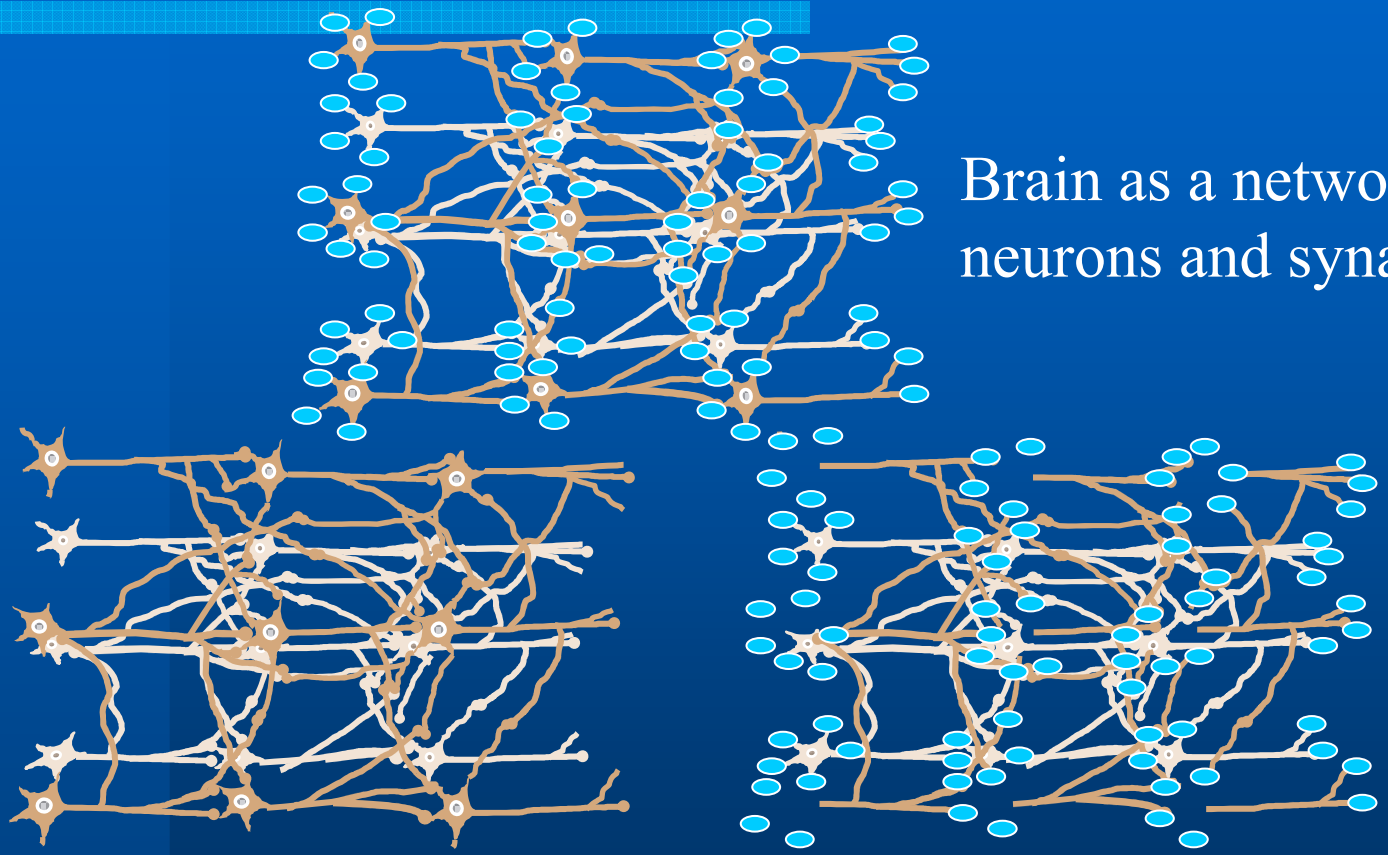
Molecule



Molecular Mechanisms of Synaptic Learning



Two Faces of the Brain: Electrical Waves or Chemical Particles?



Brain as a network of neurons and synapses

(a) Neuron-oriented cellular view ("electrical" waves)

(b) Synapse-oriented molecular view ("chemical" particles)

Large Numbers Count

- **3×10^9 DNA bases in the human genome**
- 3.5×10^9 years since first living cells
- 4.5×10^9 years since origin of Earth
- 1.5×10^{10} years since origin of universe (Big Bang)
- **10^{11} neurons in the human brain**

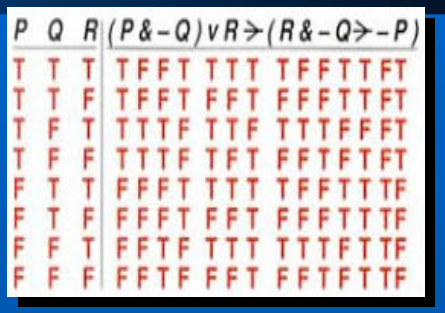
"10^{14} synapses/brain"

- 10^{14} cells in the human body
- 3×10^{23} DNA bases in the human body
  or 10^{14} copies of 3×10^9 bases
- **6×10^{23} molecules/mole or**

"> 10^{14} molecules/nanomole"

Levels of Computation

- Mind = *y*(Symbols) "*Symbolic*"



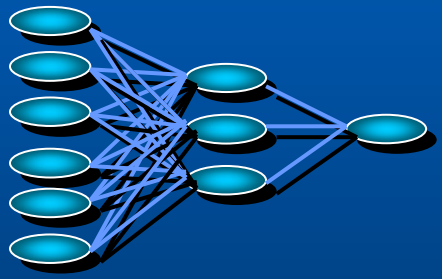
P	Q	R	(P&-Q)vR→(R&-Q→-P)
T	T	T	TFFT TTT TFFTTFT
T	T	F	TFFT FFT FFFTTFT
T	F	T	TTTF TTF TTTFFFT
T	F	F	TTTF TFT FFTFTFT
F	T	T	FFFT TTT TFFTTTF
F	T	F	FFFT FFT FFFTTTF
F	F	T	FFTF TTT TTTFTTF
F	F	F	FFTF FFT FFTFTTF

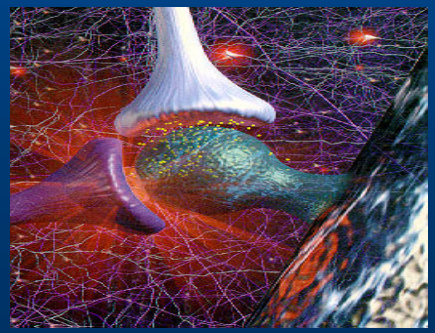
- Mind = *f*(Brain)

  = *f*(*g*(Cells)) "*Connectionist*"

  = *f*(*g*(*h*(Molecules)))

  = *y*(Molecules) "*Interactionist*"





Talk Outline

- Why Nanobiointelligence Computers (NBIC)?
- Molecular Computing Technology for NBIC
- Biomedical Applications
- The Probabilistic Library Model (PLM)
- Future of NBIC

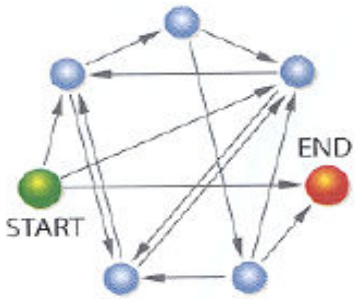
Molecular Computing for NBIC

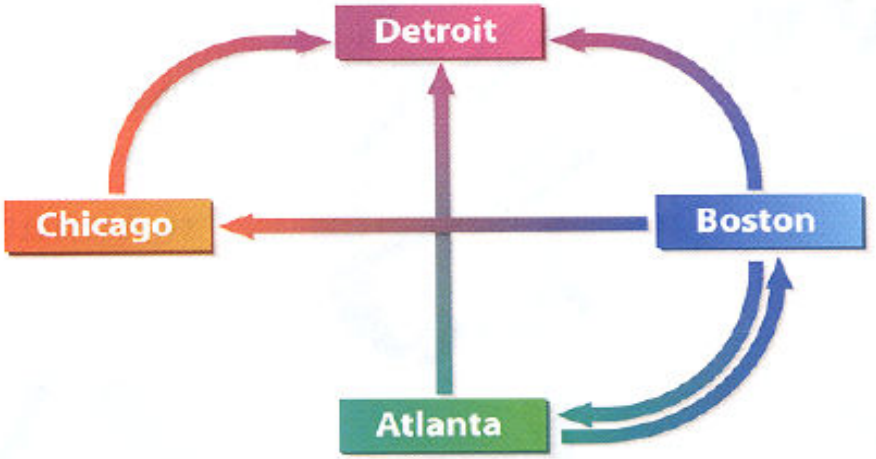
DNA Computation of Hamiltonian Paths

Hamiltonian Path Problem

Consider a map of cities connected by certain nonstop flights (*top right*). For instance, in the example shown here, it is possible to travel directly from Boston to Detroit but not vice versa. The goal is to determine whether a path exists that will commence at the start city (Atlanta), finish at the end city (Detroit) and pass through each of the remaining cities exactly once. In DNA computation, each city is assigned a DNA sequence (*ACTTGCAG* for Atlanta) that can be thought of as a first name (*ACTT*) followed by a last name (*GCAG*). DNA flight numbers can then be defined by concatenating the last name of the city of origin with the first name of the city of destination (*bottom right*).

The complementary DNA city names are the Watson-Crick complements of the DNA city names in which every *C* is replaced by a *G*, every *G* by a *C*, every *A* by a *T*, and every *T* by an *A*. (To simplify the discussion here, details of the 3′ versus 5′ ends of the DNA molecules have been omitted.) For this particular problem, only one Hamiltonian path exists, and it passes through Atlanta, Boston, Chicago and Detroit in that order. In the computation, this path is represented by *GCAGTCGGACTGGGCTATGTCCGA*, a DNA sequence of length 24. Shown at the left is the map with seven cities and 14 nonstop flights used in the actual experiment. —*L.M.A.*





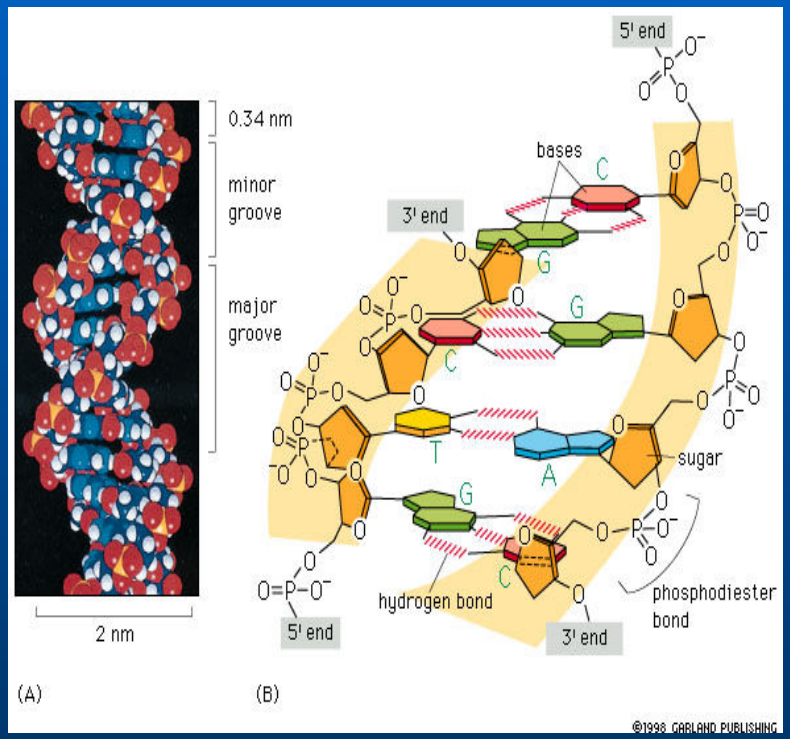
CITY	DNA NAME	COMPLEMENT
ATLANTA	ACTTGCAG	TGAACGTC
BOSTON	TCGGACTG	AGCCTGAC
CHICAGO	GGCTATGT	CCGATACA
DETROIT	CCGAGCAA	GGCTCGTT

FLIGHT	DNA FLIGHT NUMBER
ATLANTA - BOSTON	GCAGTCGG
ATLANTA - DETROIT	GCAGCCGA
BOSTON - CHICAGO	ACTGGGCT
BOSTON - DETROIT	ACTGCCGA
BOSTON - ATLANTA	ACTGACTT
CHICAGO - DETROIT	ATGTCCGA

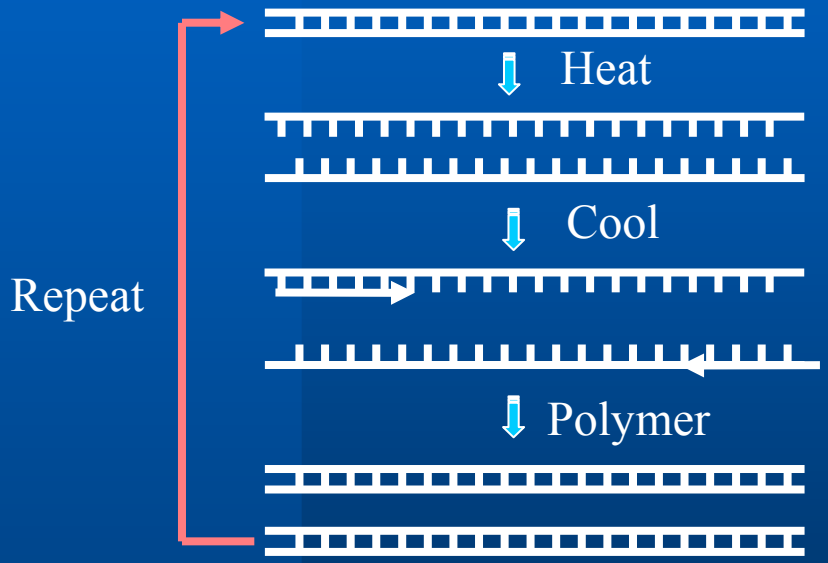
[Adleman, *Science* 1994; *Scientific American* 1998]

DNA as Computing Material

- 초고집적도:
  - 10^6 Gbits per cm^2 (1 bit per nm^3)
  - 반도체 기술: 1 Gbits per cm^2
- 초병렬 탐색:
  - 10^{26} reactions per 1 *m*mol of DNA
  - Desktop: 10^9 operations / sec
  - Supercomputer: 10^{12} operations / sec
- 에너지 효율: 10^{19} operations per Joule
  - 반도체 기술: 10^9 operations per Joule

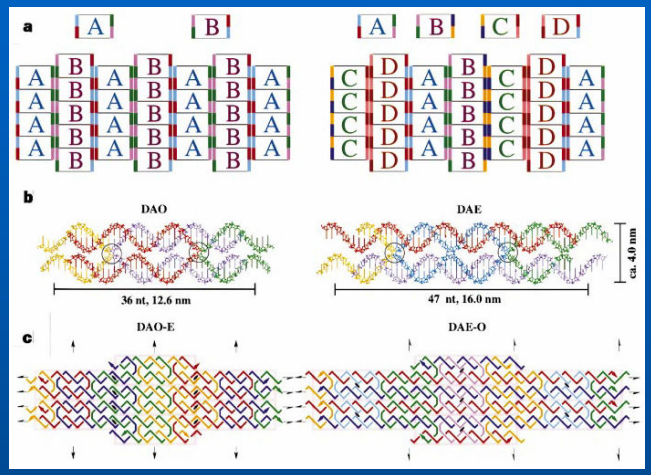


Characteristics of DNA Molecules

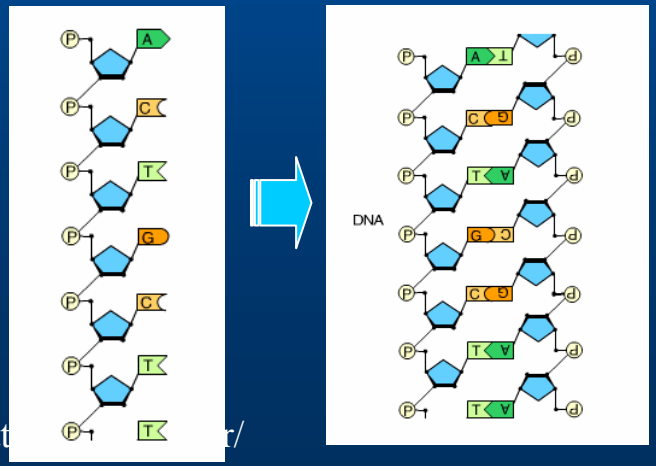


Self-replication

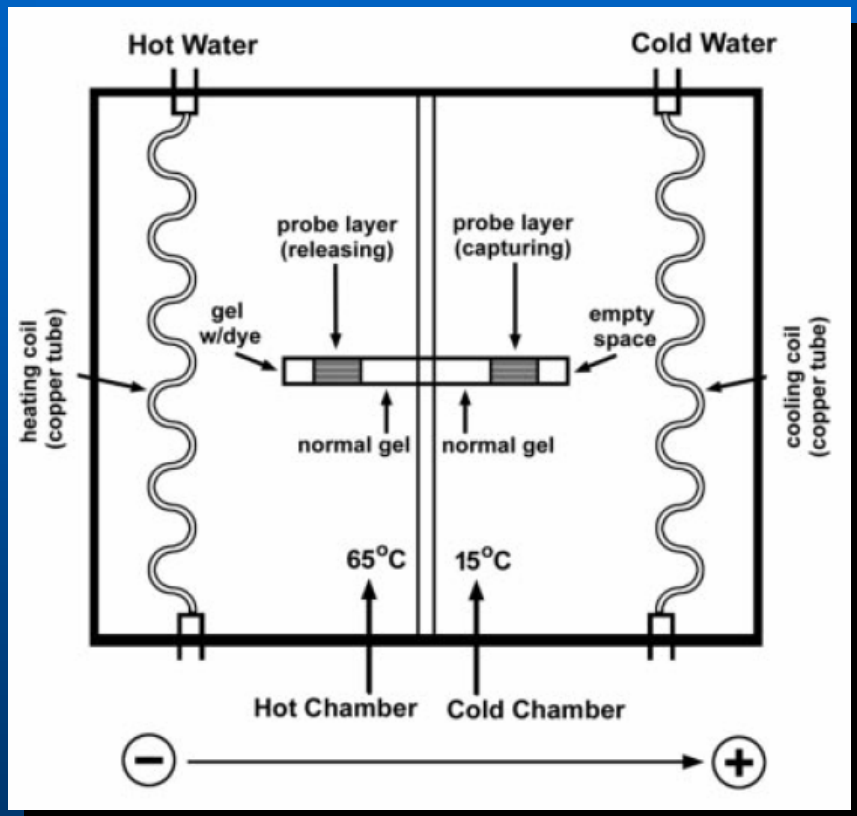
Self-assembly



Molecular recognition



DNA Based Computers



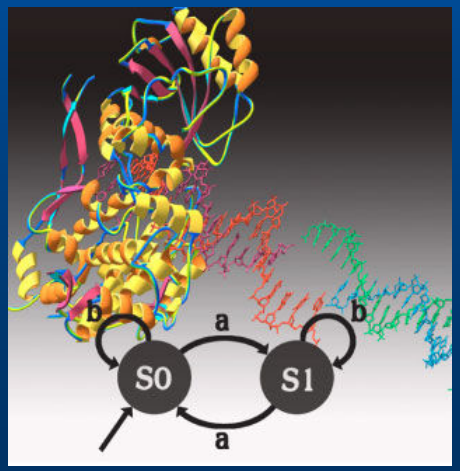
DNA Computer by Adleman's Group

[Braich et al., *Science 2002*]

DNA Computer by Olympus

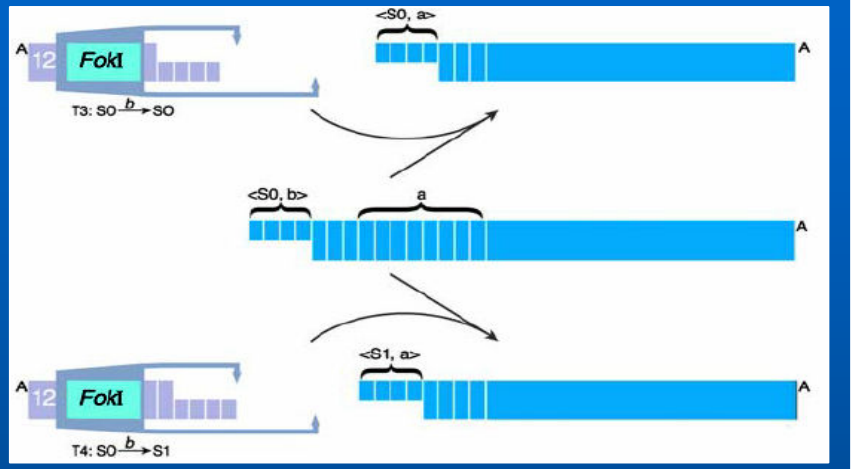
Issues in DNA Computing

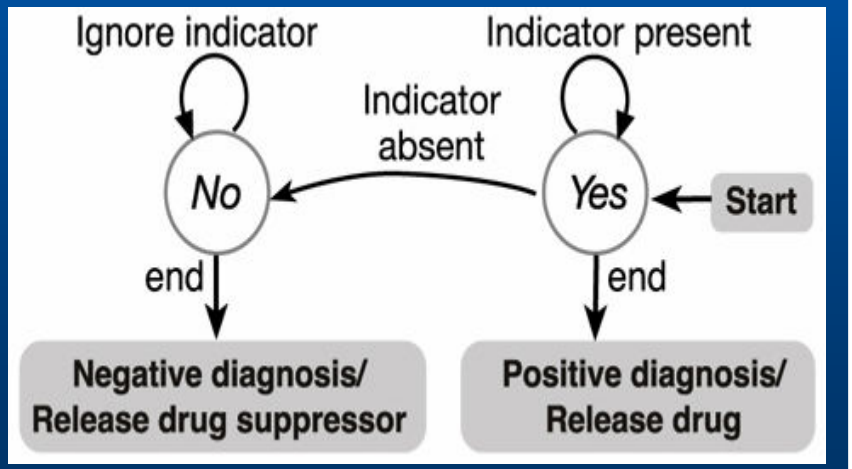
- BT
  - In Vivo Diagnosis
  - Smart Drugs
  - Therapeutics
- IT
  - DNA Processors
  - DNA Memory
  - DNA Electronics
- NT
  - DNA Nanoassembly
  - DNA Nanorobots
  - DNA Motors



DNA as Smart Drugs

[Benenson *et al.*, *Nature* 2001 & *Nature*, 2004]





PPAP2B↓ & GSTP1↓ & PIM1↑ & HEPSIN↑ → Administer GTTGGTATTGCACAT

Solving a 20-var 3-CNF Problem

A

$\Phi = (\sim x_3 \text{ or } \sim x_{16} \text{ or } x_{18})$ *and* $(x_5 \text{ or } x_{12} \text{ or } \sim x_9)$ *and* $(\sim x_{13} \text{ or } \sim x_2 \text{ or } x_{20})$ *and* $(x_{12} \text{ or } \sim x_9 \text{ or } \sim x_5)$ *and* $(x_{19} \text{ or } \sim x_4 \text{ or } x_6)$ *and* $(x_9 \text{ or } x_{12} \text{ or } \sim x_5)$ *and* $(\sim x_1 \text{ or } x_4 \text{ or } \sim x_{11})$ *and* $(x_{13} \text{ or } \sim x_2 \text{ or } \sim x_{19})$ *and* $(x_5 \text{ or } x_{17} \text{ or } x_9)$ *and* $(x_{15} \text{ or } x_9 \text{ or } \sim x_{17})$ *and* $(\sim x_5 \text{ or } \sim x_9 \text{ or } \sim x_{12})$ *and* $(x_6 \text{ or } x_{11} \text{ or } x_4)$ *and* $(\sim x_{15} \text{ or } \sim x_{17} \text{ or } x_7)$ *and* $(\sim x_6 \text{ or } x_{19} \text{ or } x_{13})$ *and* $(\sim x_{12} \text{ or } \sim x_9 \text{ or } x_5)$ *and* $(x_{12} \text{ or } x_1 \text{ or } x_{14})$ *and* $(x_{20} \text{ or } x_3 \text{ or } x_2)$ *and* $(x_{10} \text{ or } \sim x_7 \text{ or } \sim x_8)$ *and* $(\sim x_5 \text{ or } x_9 \text{ or } \sim x_{12})$ *and* $(x_{18} \text{ or } \sim x_{20} \text{ or } x_3)$ *and* $(\sim x_{10} \text{ or } \sim x_{18} \text{ or } \sim x_{16})$ *and* $(x_1 \text{ or } \sim x_{11} \text{ or } \sim x_{14})$ *and* $(x_8 \text{ or } \sim x_7 \text{ or } \sim x_{15})$ *and* $(\sim x_8 \text{ or } x_{16} \text{ or } \sim x_{10})$

B

x_1=F, x_2=T, x_3=F, x_4=F, x_5=F, x_6=F, x_7=T, x_8=T, x_9=F, x_{10}=T, x_{11}=T, x_{12}=T, x_{13}=F, x_{14}=F, x_{15}=T, x_{16}=T, x_{17}=T, x_{18}=F, x_{19}=F, x_{20}=F

Fig. 1. The computational problem. (**A**) 20-variable 3-CNF Boolean formula Φ. The symbol "$\sim$" indicates "not." (**B**) The unique truth assignment satisfying Φ.

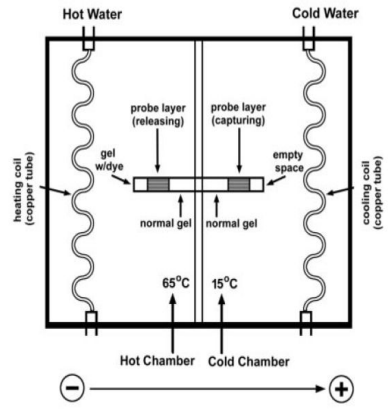


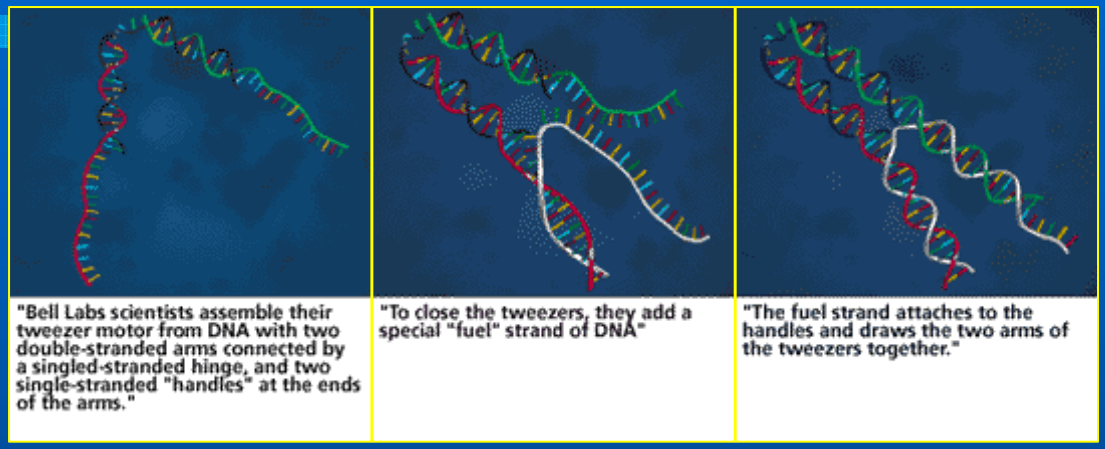
Fig. 4. The computer. An electrophoresis box 30 cm long, 15 cm wide, and 8 cm high was constructed from 0.5-cm-thick plexiglass. The box was partitioned into a hot chamber and a cold chamber of equal volumes, by a plexiglass divider. Each chamber was attached via plastic tubing to a circulating water bath. Water from the bath was transported across the chamber through copper tubing, acting as a cooling/heating coil. A platinum wire electrode was inserted into each chamber. For each of the 24 clauses of Φ, a 100-μl clause solution was prepared containing 15 μM of each of three Acrydite-modified probes, one for each literal in the clause (if x_k appeared in the clause, the probe with sequence $\bar{X}^T_k$ was added, if $\sim x_k$ appeared, the probe with sequence for $\bar{X}^F_k$ was added). For example, for the first clause, $(\sim x_3 \text{ or } \sim x_{16} \text{ or } x_{18})$, probes with sequences $\bar{X}^F_3$, $\bar{X}^F_{16}$, and X^T_{18} were added. For each clause solution, a clause module was created in a 4.5-cm-long glass tube with outer diameter of 0.5 cm and inner diameter of 0.3 cm. A base layer of 5% polyacrylamide gel was polymerized in the bottom 3.2 cm of the tube. A probe layer of 5% polyacrylamide gel containing the clause solution was polymerized on top of the base layer. A library module was prepared as above, but instead of a clause solution, 500 pmol of the full library in 100 μl was used (*28*). The end of the module destined for the hot chamber was plugged with 5% polyacrylamide gel containing bromophenol blue and xylene cyanol dyes. The hot chamber was brought to 65°C and the cold chamber to 15°C. Electrophoresis was performed at 12 V/cm. Movement of the dyes through the gel provided a means of monitoring progress and of detecting possible leaks at the interface between the modules. Electrophoresis was stopped when the Xylene Cyanol dye had gone through the modules and into the cold buffer, approximately 4 hours.

[Braich et al., *Science* 2002]

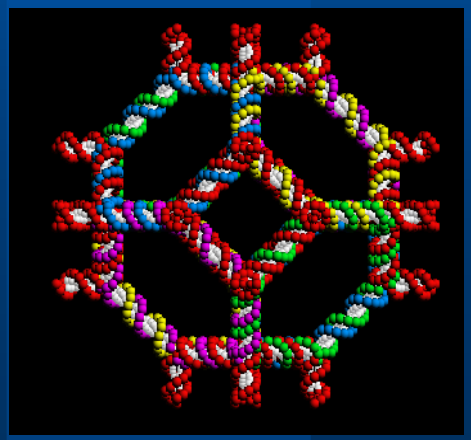
DNA Nanostructures

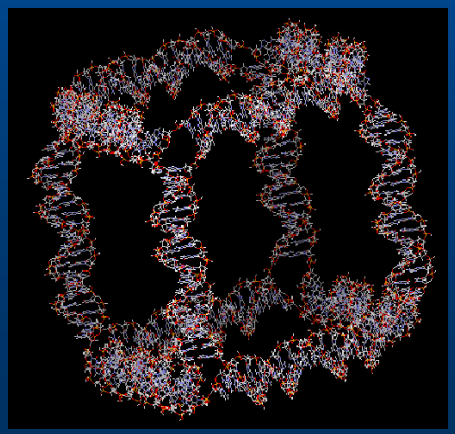
- Molecular Tweezers

  [Yurke et al., *Nature 2000*]



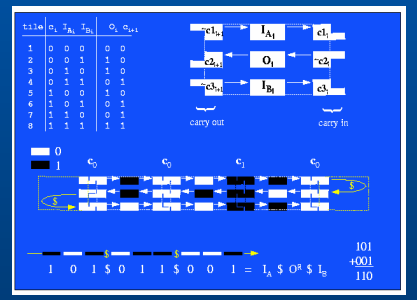
- DNA nanostructure

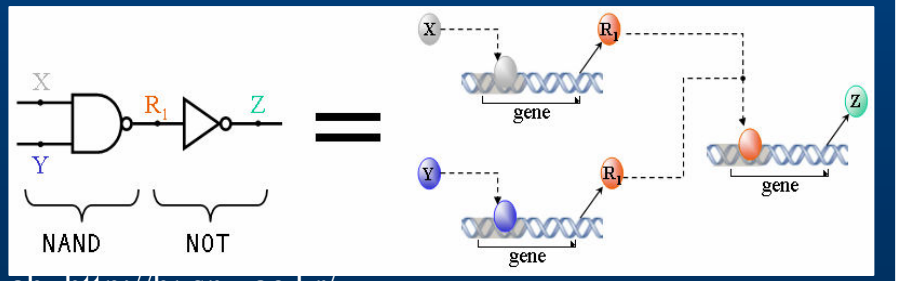




[Chen and Seeman, *Nature* 1991]

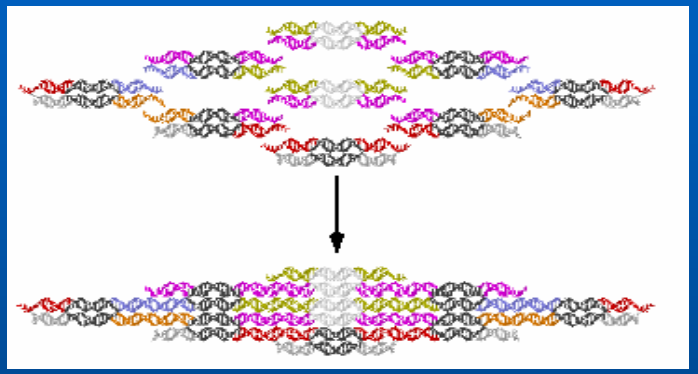
- Information processing methods

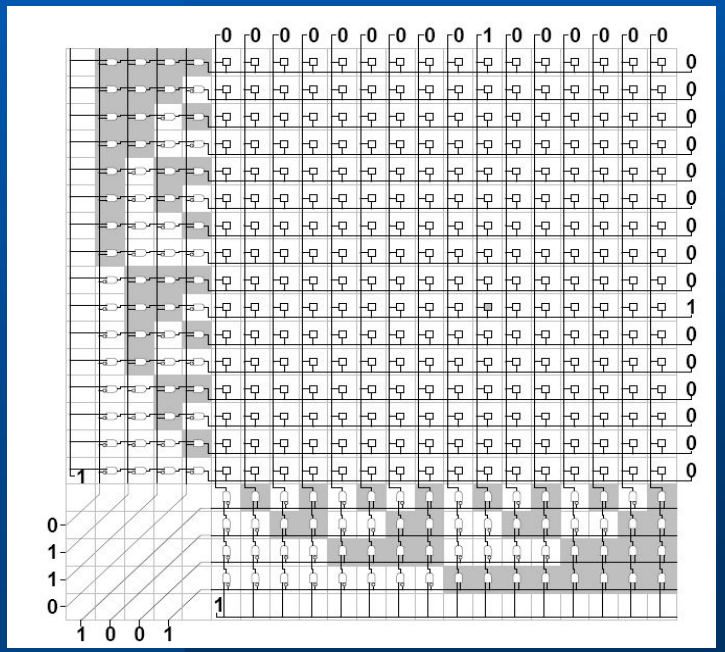


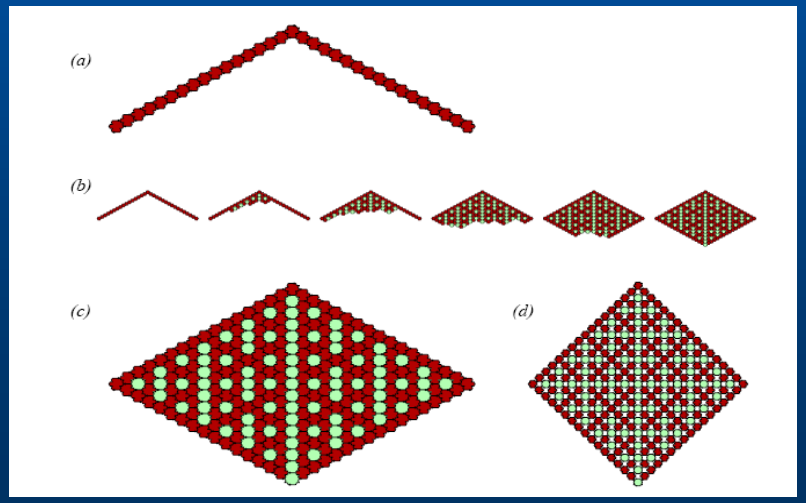


Molecular Memory and Self-Assembly

- DNA self-assembly as information processing utilizing
  - Parallel-interaction
  - Molecular recognition
  - Self-organization

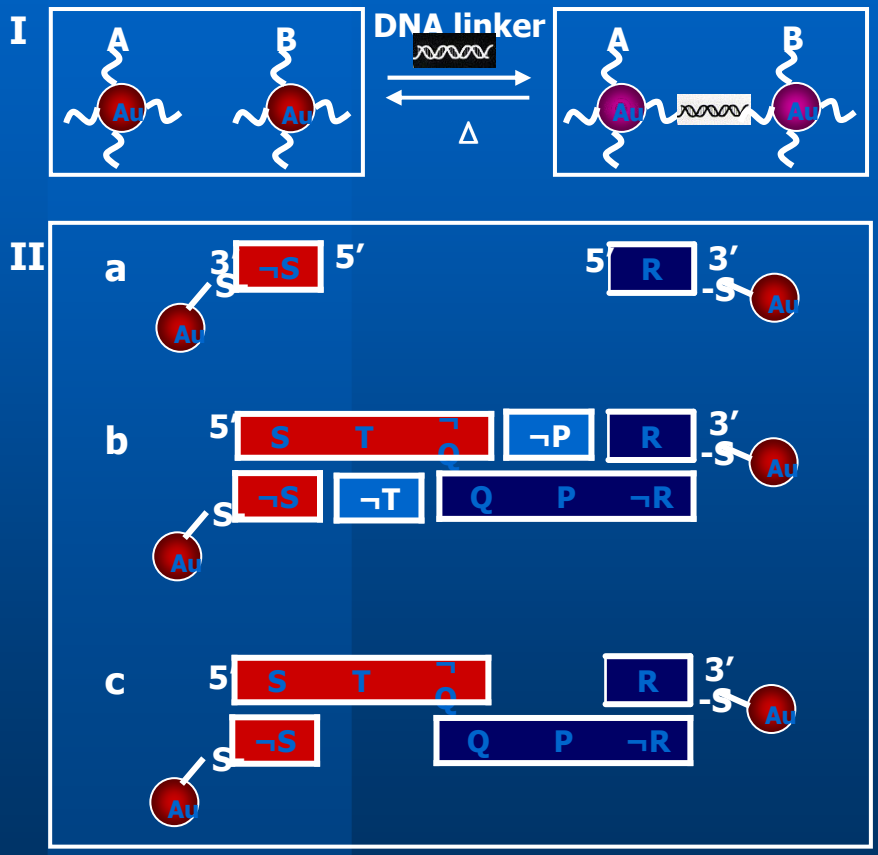


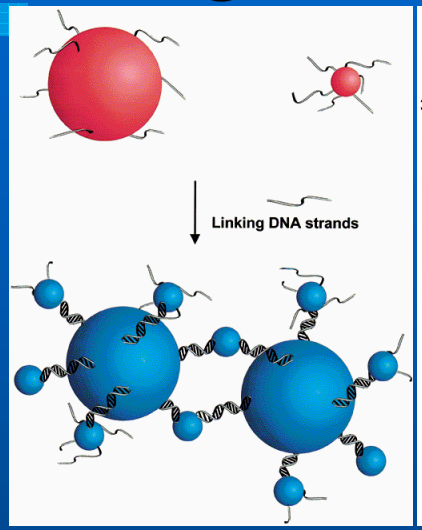


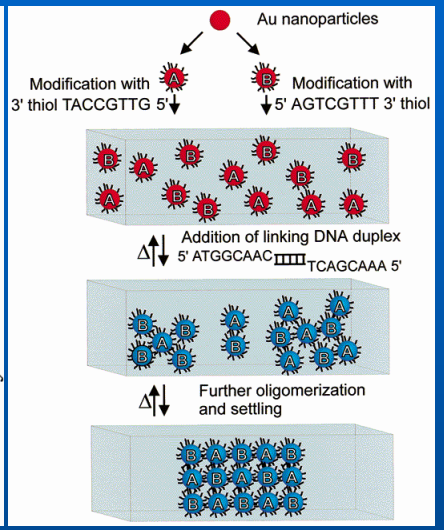


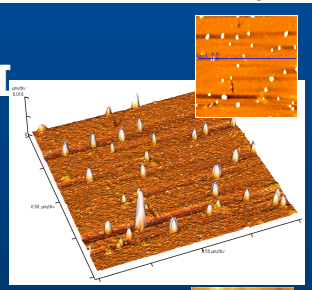
[Caltech and Duke]

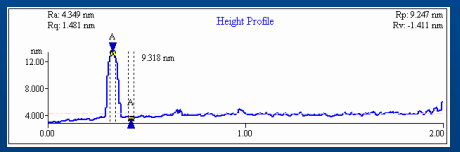
Nanoparticle-Based Theorem Proving for Medical Diagnosis

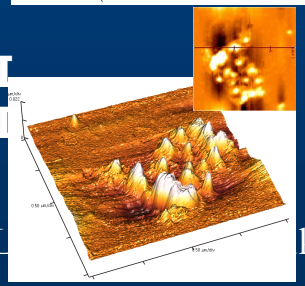


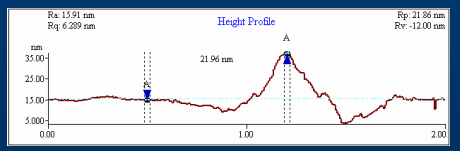




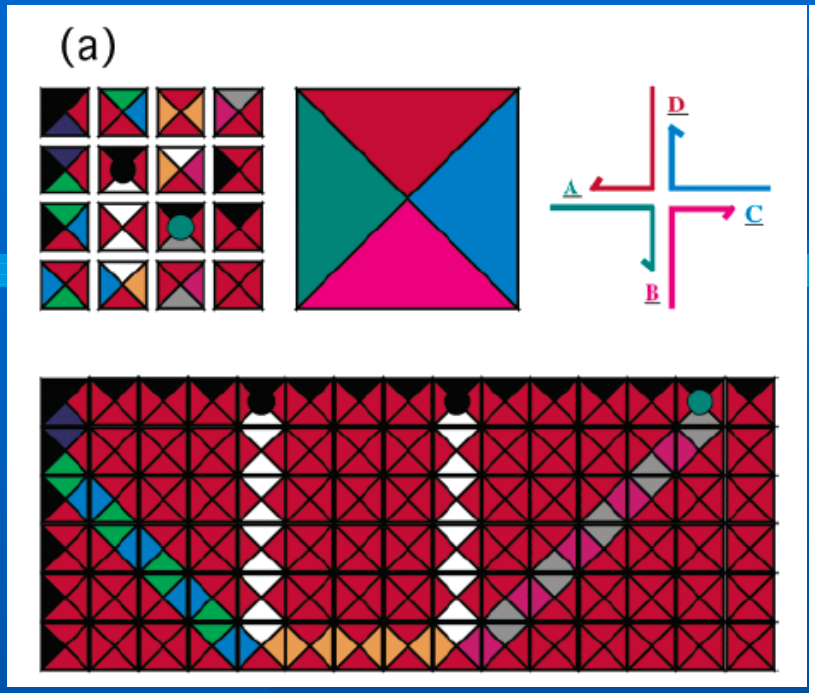








[Park et al., in preparation]



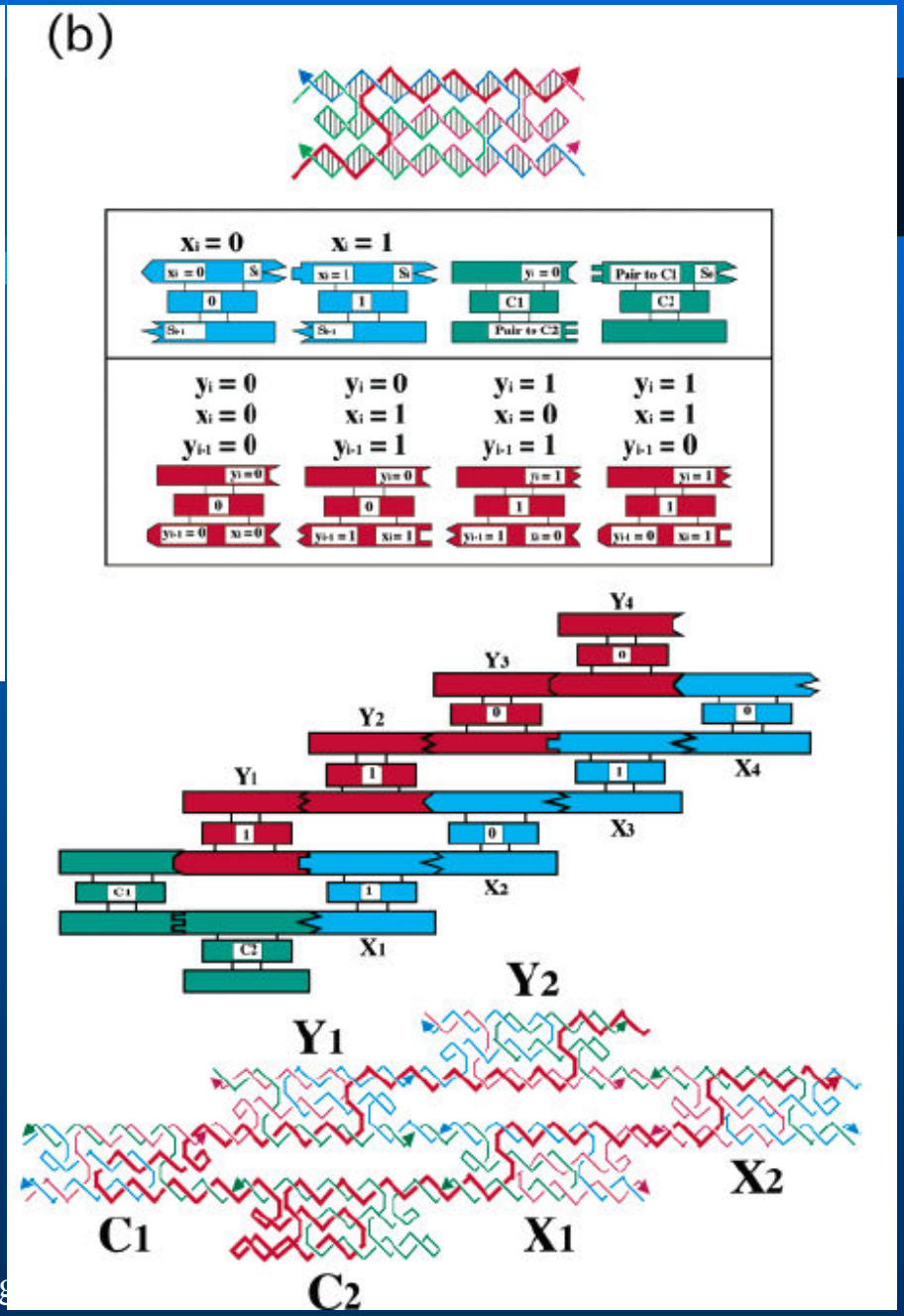
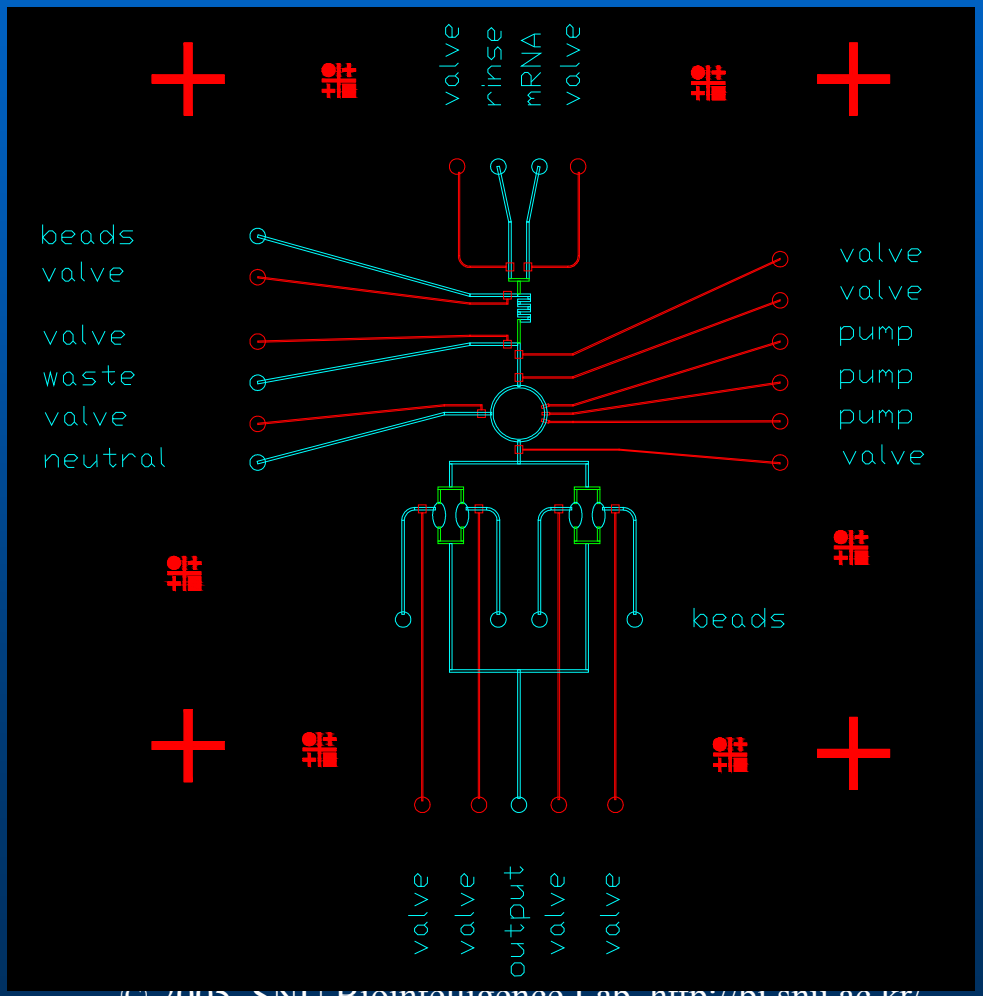


FIGURE 7: Wang tiles and algorithmic assembly. (a) Wang tiles. The upper left shows 16 wang tiles that fit together in the mosaic below according to the rule that each edge is flanked by the same color (modified from ref *40*). The relationship between Wang tiles and a branched junction with a variety of sticky ends is illustrated at the upper center and right. (b) A four-bit XOR self-assembly. At the top of the panel is a TX tiles with its reporter strand emphasized in red. Below this are schematics of the input tiles (blue), initiator tiles (green), and gating tiles (red). The four possible inputs to the XOR gate correspond to sticky ends on the bottom domains of the red tiles. The schematic tiles are shown to self-assemble to produce the output arrangement of red tiles in the schematic below this. At the bottom, the answer is extracted by ligation of the reporter strands, which are later subjected to partial restriction.

DNA Computing Chip for Medical Diagnosis



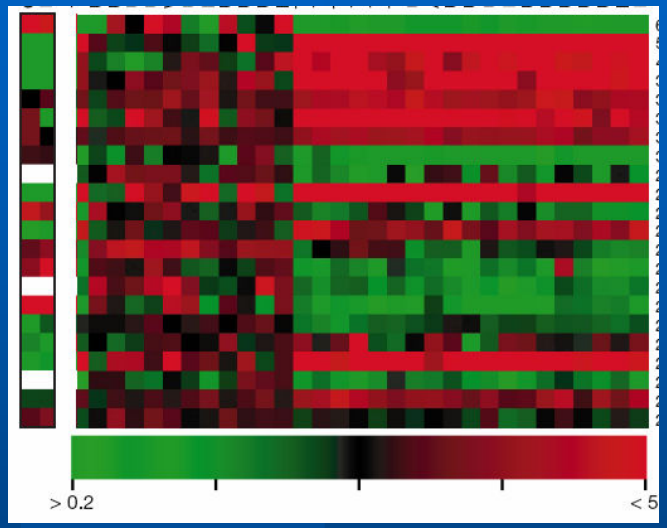
[Lee et al., in preparation]

Molecular Computers vs. Silicon Computers

	Molecular Computers	Silicon Computers
Processing	Ballistic	Hardwired
Medium	Liquid (wet) or Gaseous (dry)	Solid (dry)
Communication	3D collision	2D switching
Configuration	Amorphous (asynchronous)	Fixed (synchronous)
Parallelism	Massively parallel	Sequential
Speed	Fast (millisec)	Ultra-fast (nanosec)
Reliability	Low	High
Density	Ultrahigh	Very high
Reproducibility	Probabilistic	Deterministic

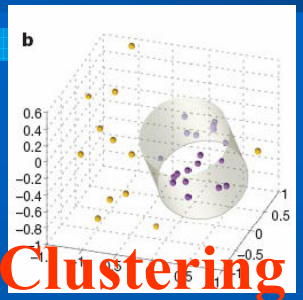
Diagnosis by DNA-Based Theorem Proving

Diagnosis Scheme

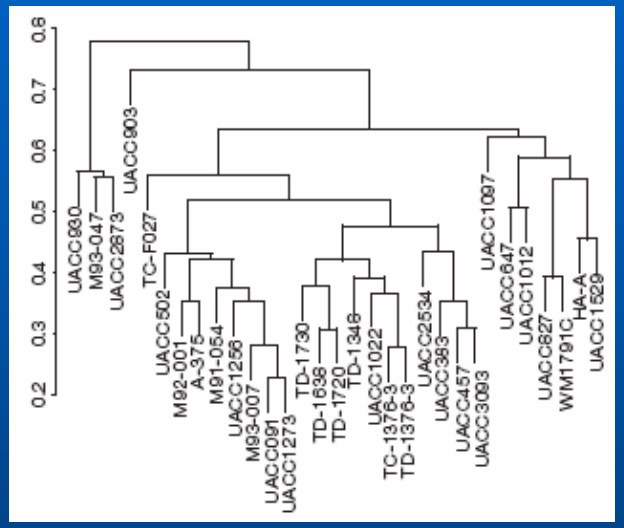


Gene expression data

[Bittner et al., *Nature*, 406, 536-540, 2000]



Clustering



- Refine logical rules from clustered data
- Implement logical inference by DNA computing

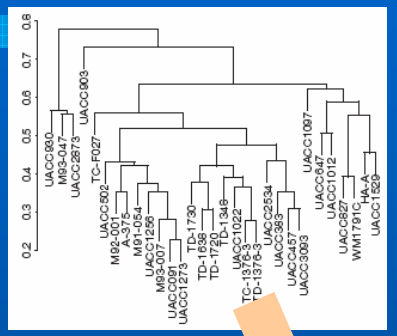
[Zhang et al., *Private discussion*, 2004]

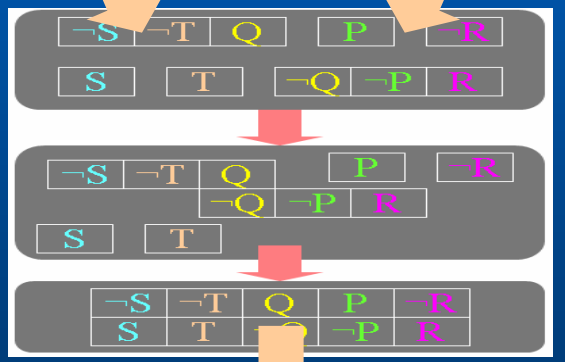
Diagnosis by DNA Computing

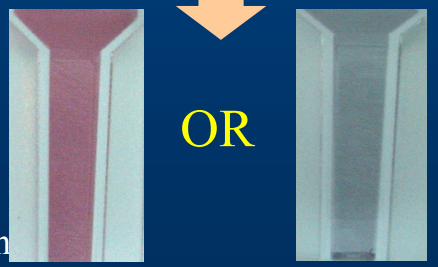
1. Transformation of Gene Expression Information into DNA Signal

2. Autonomous Logical Inference from DNA Signal

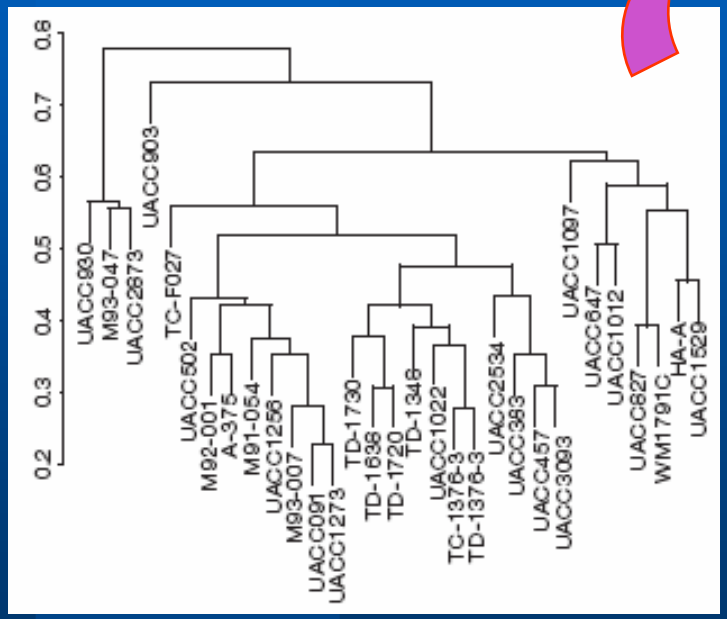
3. Detection of Inference Results







Diagnosis by DNA Computing: Logical Inference



If gene A is expressed$_A$ and gene C is expressed$_C$, he (she) has a lung cancer$_L$.

If gene H is expressed$_H$, then gene A is expressed$_A$.

In sample, gene H$_H$ and C$_C$ are expressed.

Does he (she) have a Lung cancer$_L$?

$$A \wedge C \rightarrow L,\ H \rightarrow A,\ H,\ C$$

$$L\,?$$

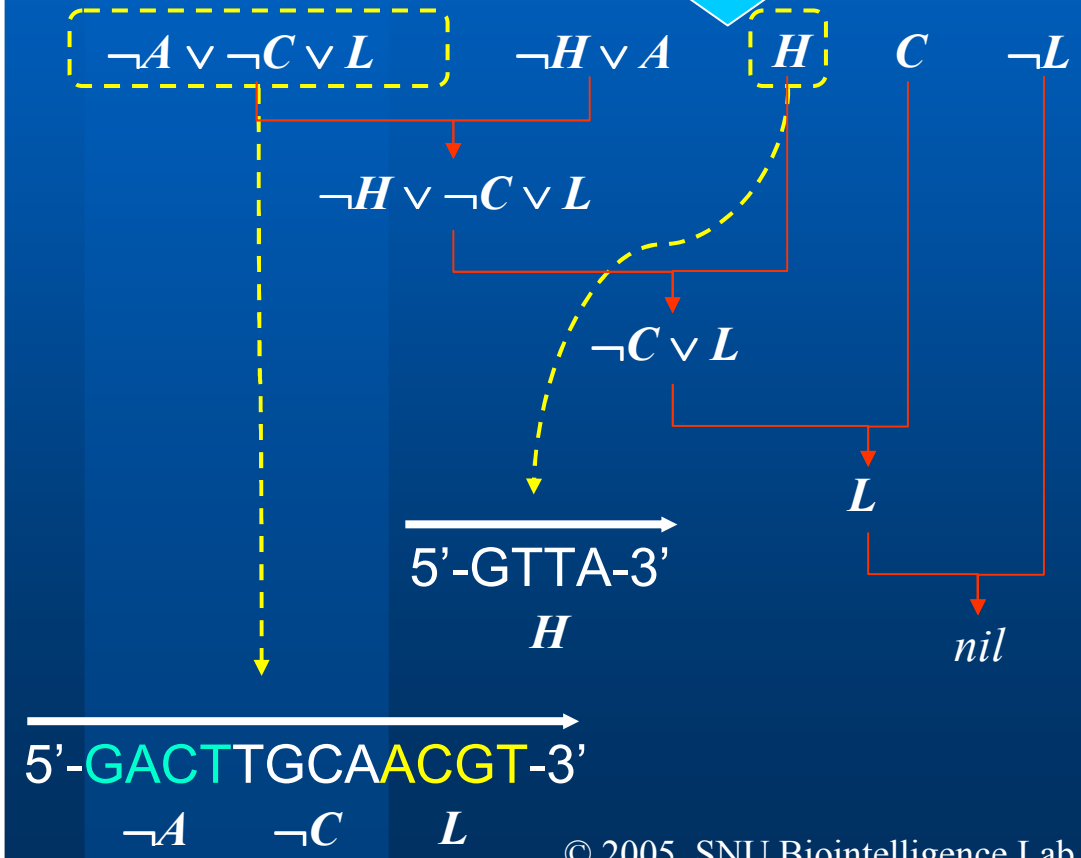
$$(\neg A \vee \neg C \vee L) \wedge (\neg H \vee A) \wedge H \wedge C \wedge \neg L$$

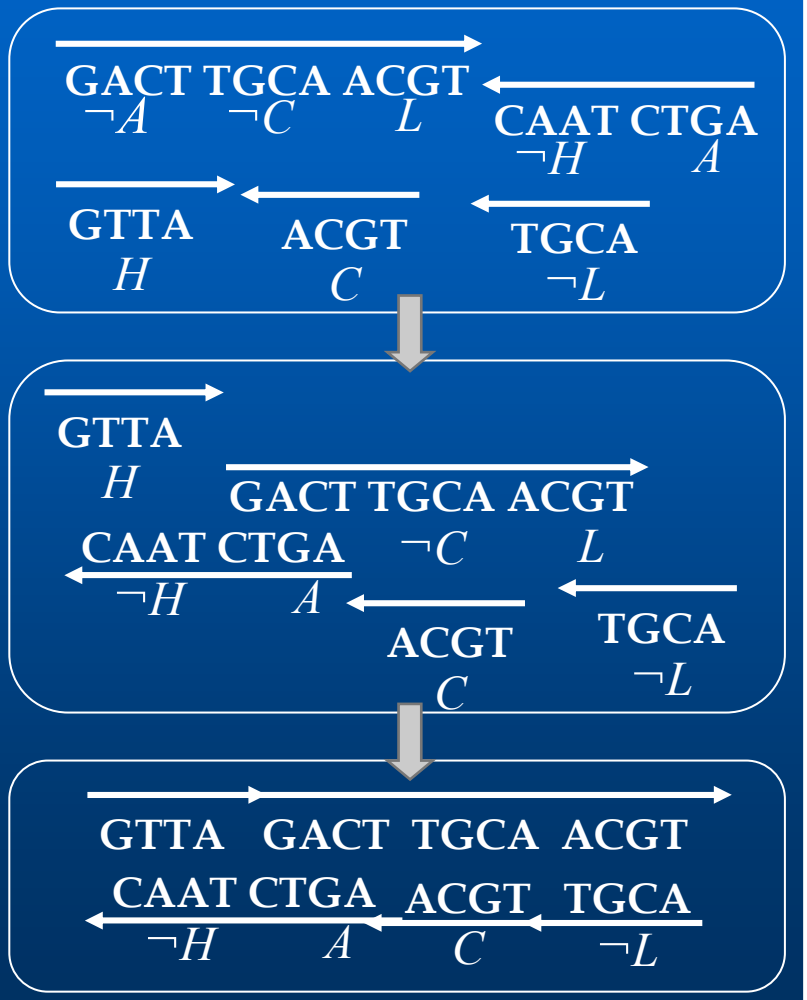
Transform into CNF

Diagnosis by DNA Computing: Logical Inference

$$(\neg A \vee \neg C \vee L) \wedge (\neg H \vee A) \wedge H \wedge C \wedge \neg L$$

Resolution

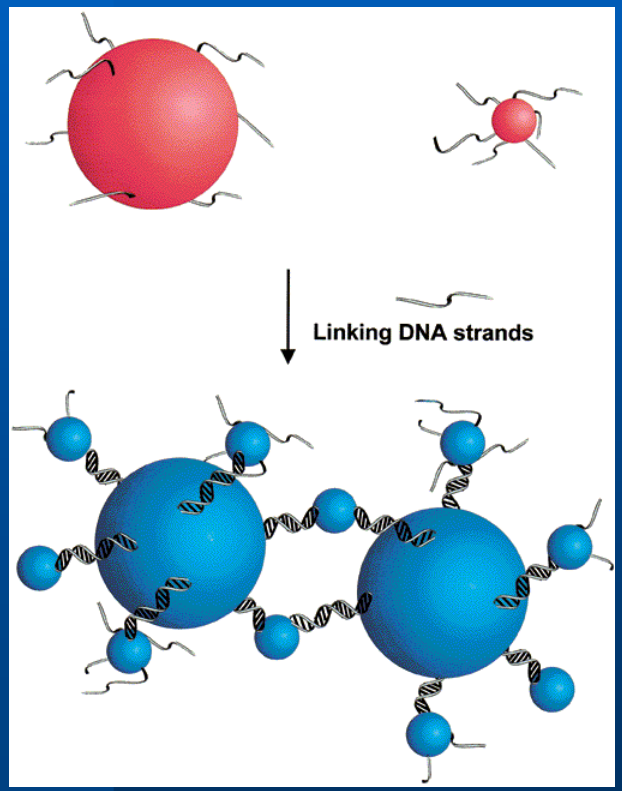


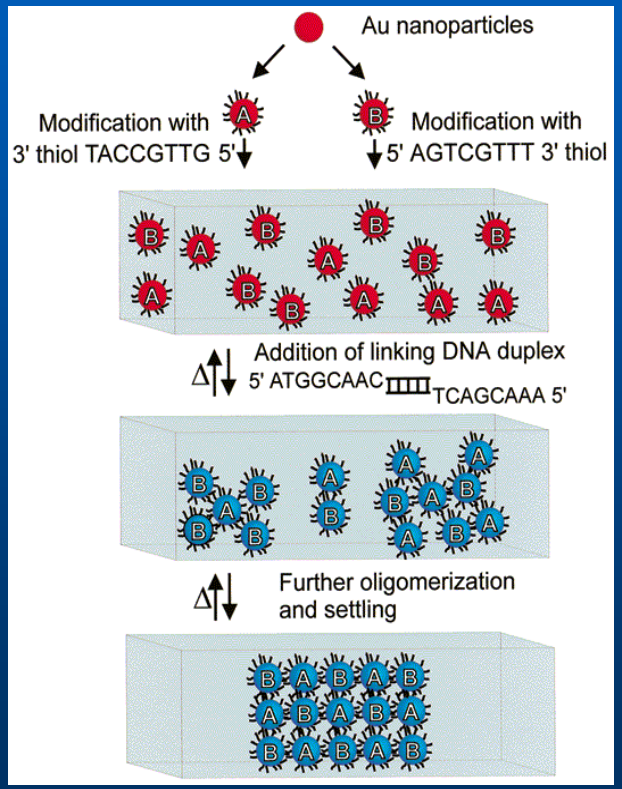


[Lee et al., *Lecture Notes in Computer Science*, 2003]

Diagnosis by DNA Computing: Detection

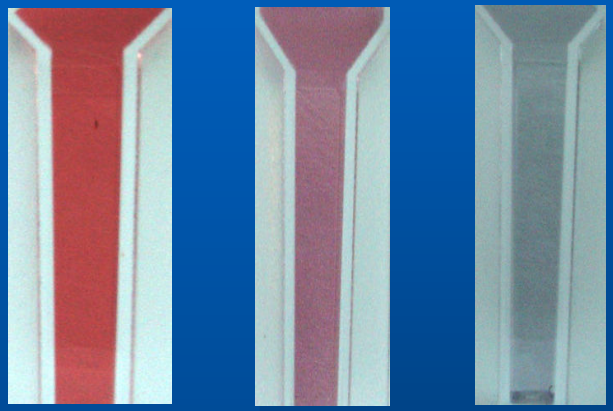
- DNA-Based nanoparticle assembly strategy





Color Change of DNA-Induced Assembly

(a) (b) (c)

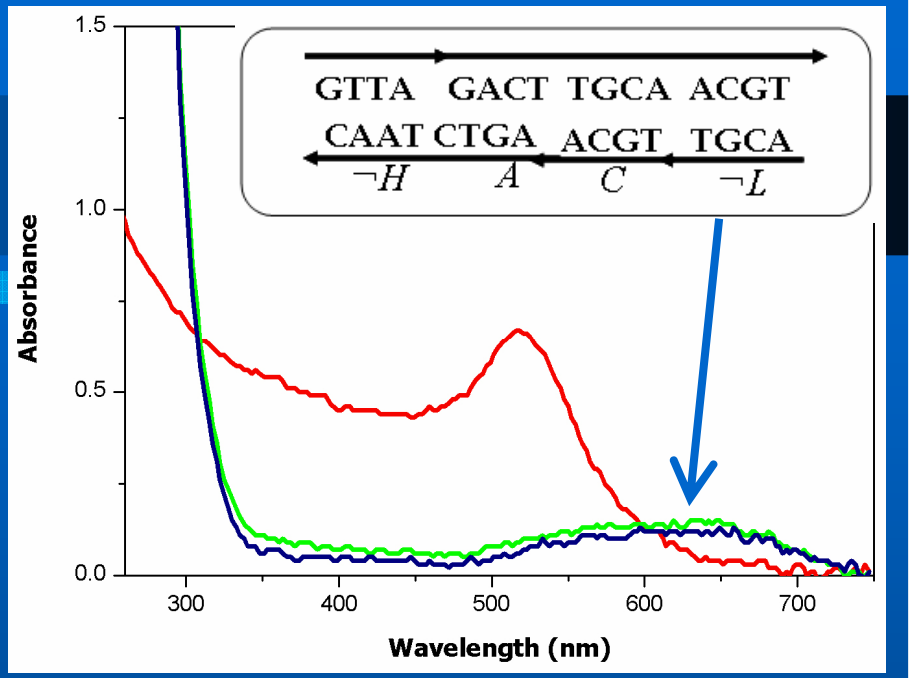


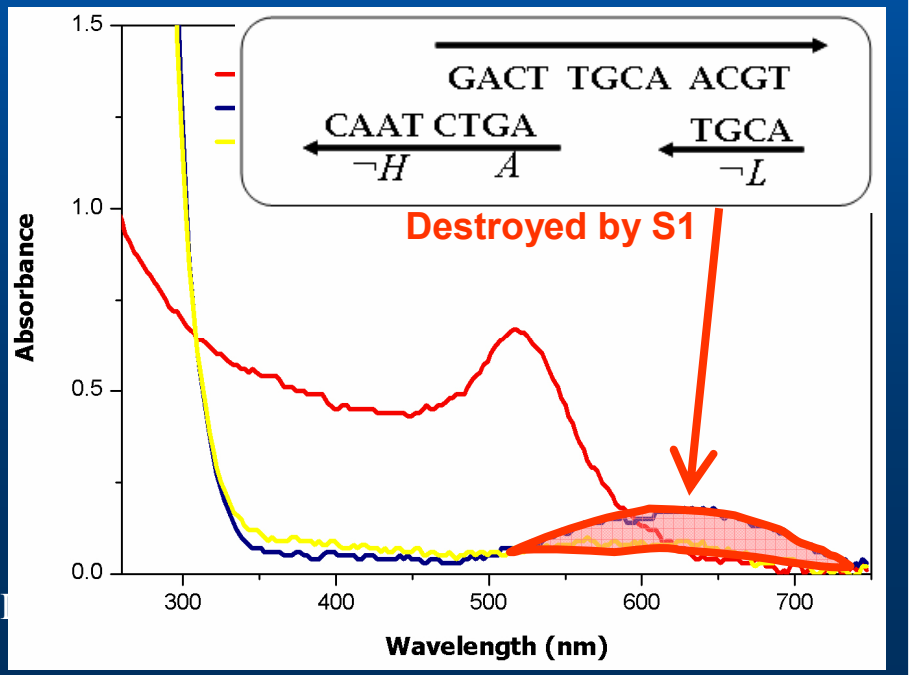
(a) The 10-nm gold particles

(b) The solutions of DNA-linked full assembly

(c) Aqueous solution of the addition of NaCl

[J.-Y. Park, *Ph.D. Thesis*, 2004]





The Probabilistic Library Model (PLM)

Probabilistic Library Model (PLM)

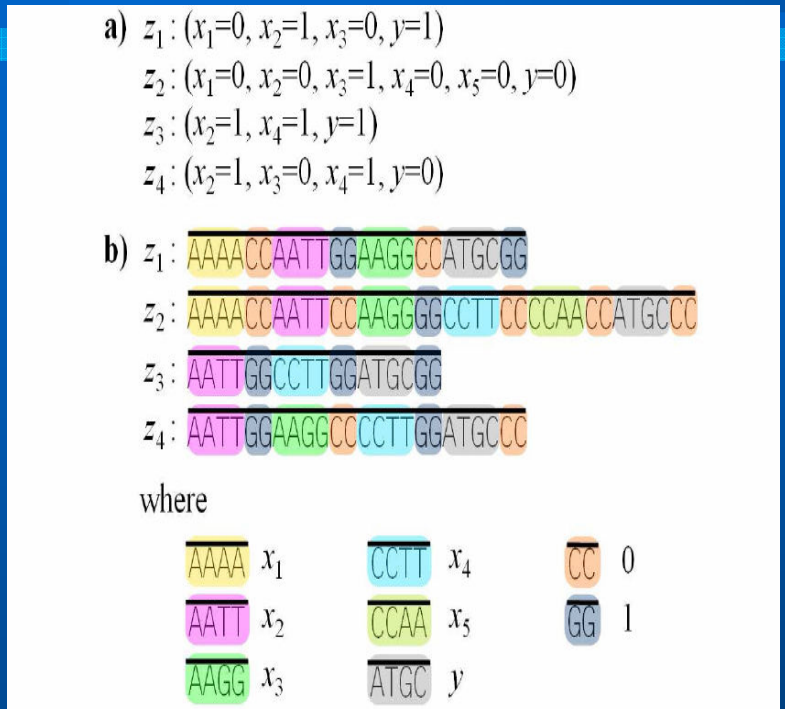


Figure 1: Population of genetic programs in two different representations: (a) set of decision lists, (b) library of DNA molecules corresponding to (a). The DNA code shown are illustration-purposes only and this design does not fully reflect the biochemical properties of the sequences.

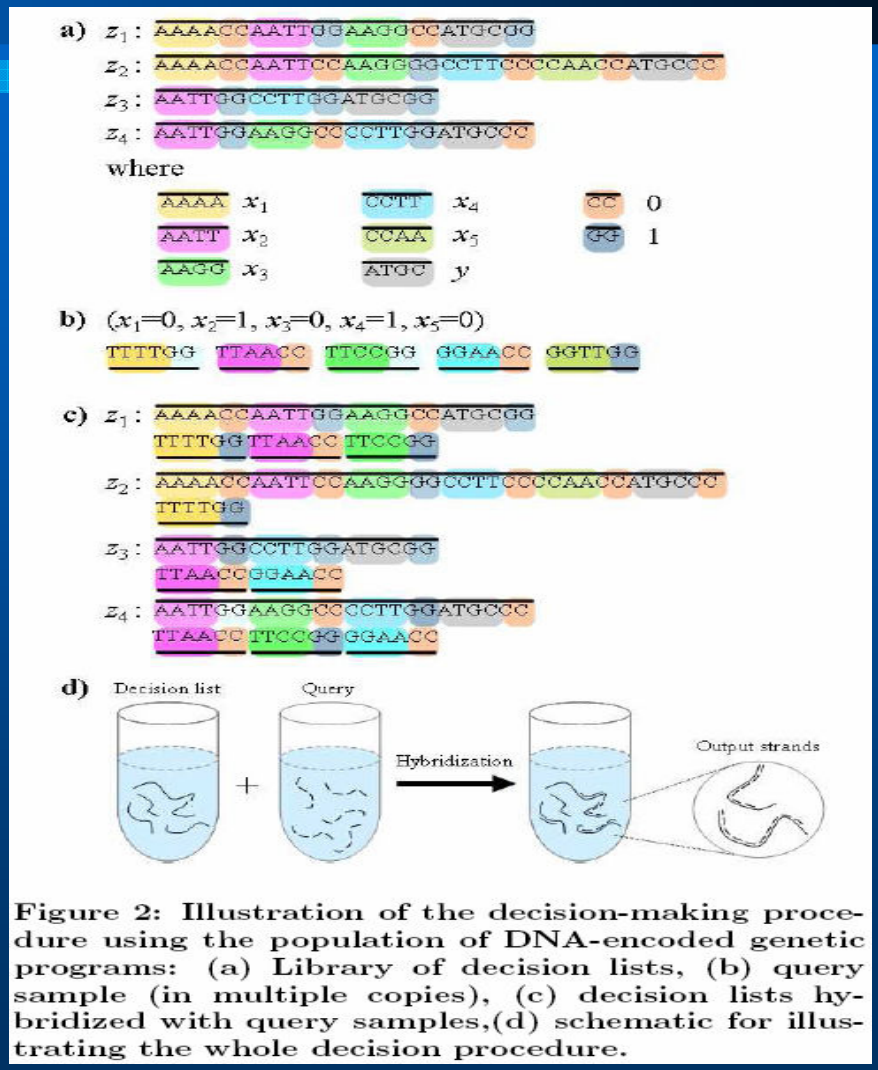
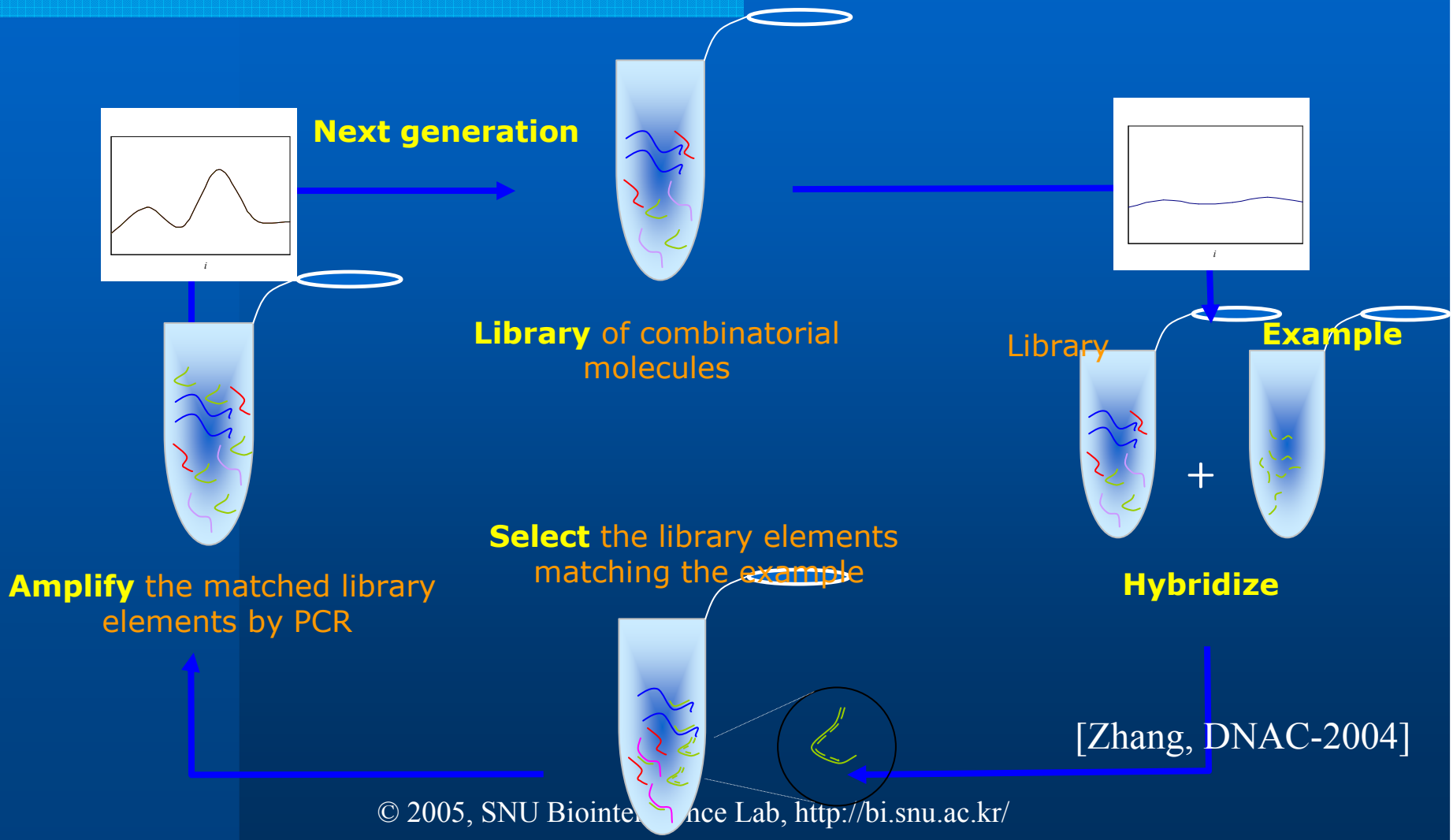


Figure 2: Illustration of the decision-making procedure using the population of DNA-encoded genetic programs: (a) Library of decision lists, (b) query sample (in multiple copies), (c) decision lists hybridized with query samples,(d) schematic for illustrating the whole decision procedure.

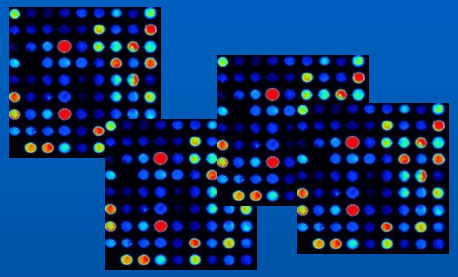
PLM (Probabilistic Library Model):
Learning Probability Distributions with DNA



[Zhang, DNAC-2004]

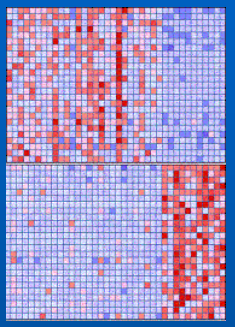
Application to Leukemia Diagnosis

120 samples from
60 leukemia patients

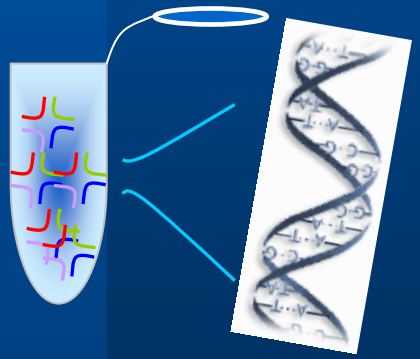


Gene expression data

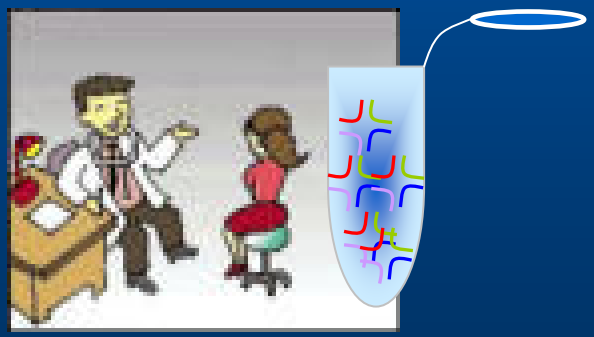
&



Class: ALL/AML



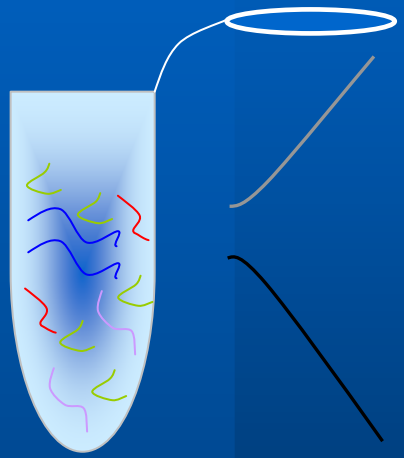
Training with
6-fold validation

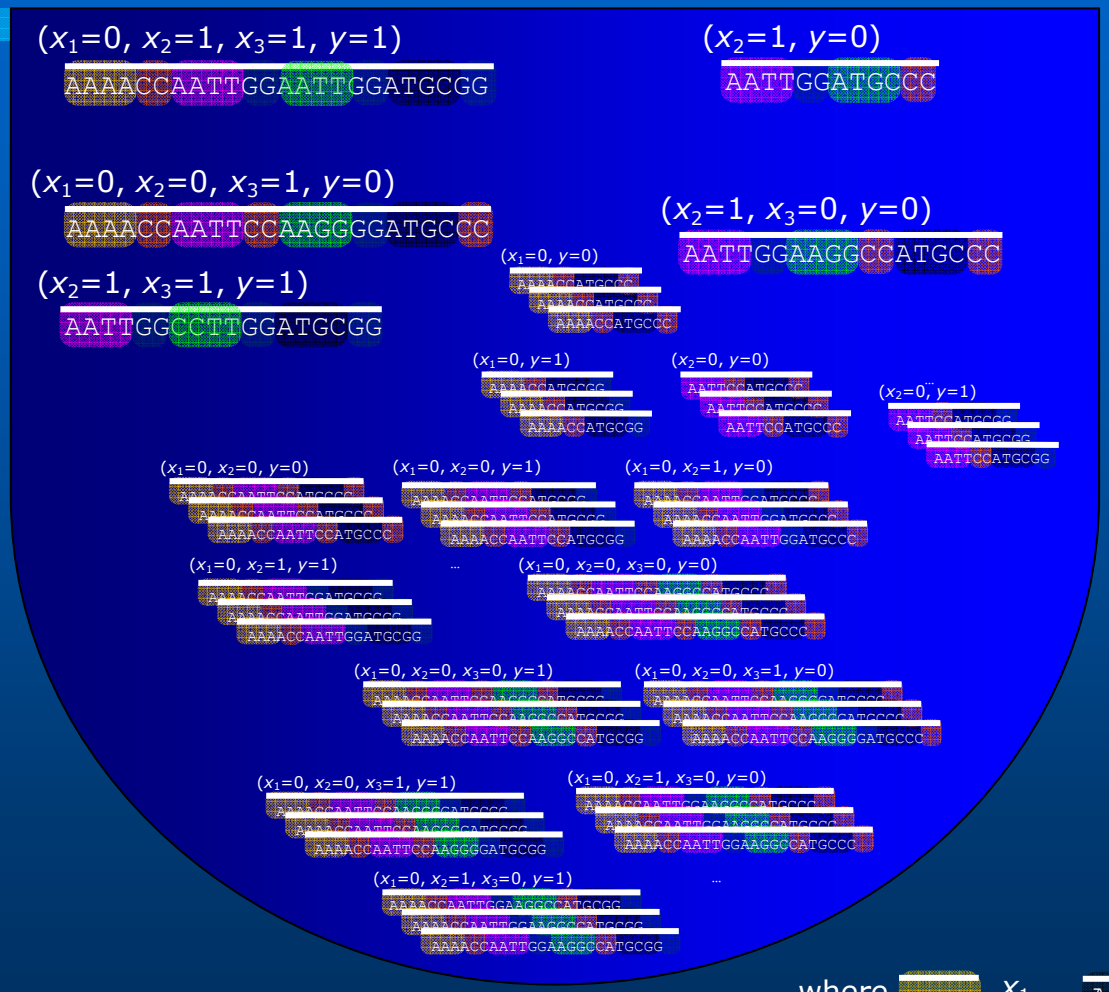


Diagnosis

[Cheok *et al*., *Nature Genetics*, 2003]

Initial Library L_0

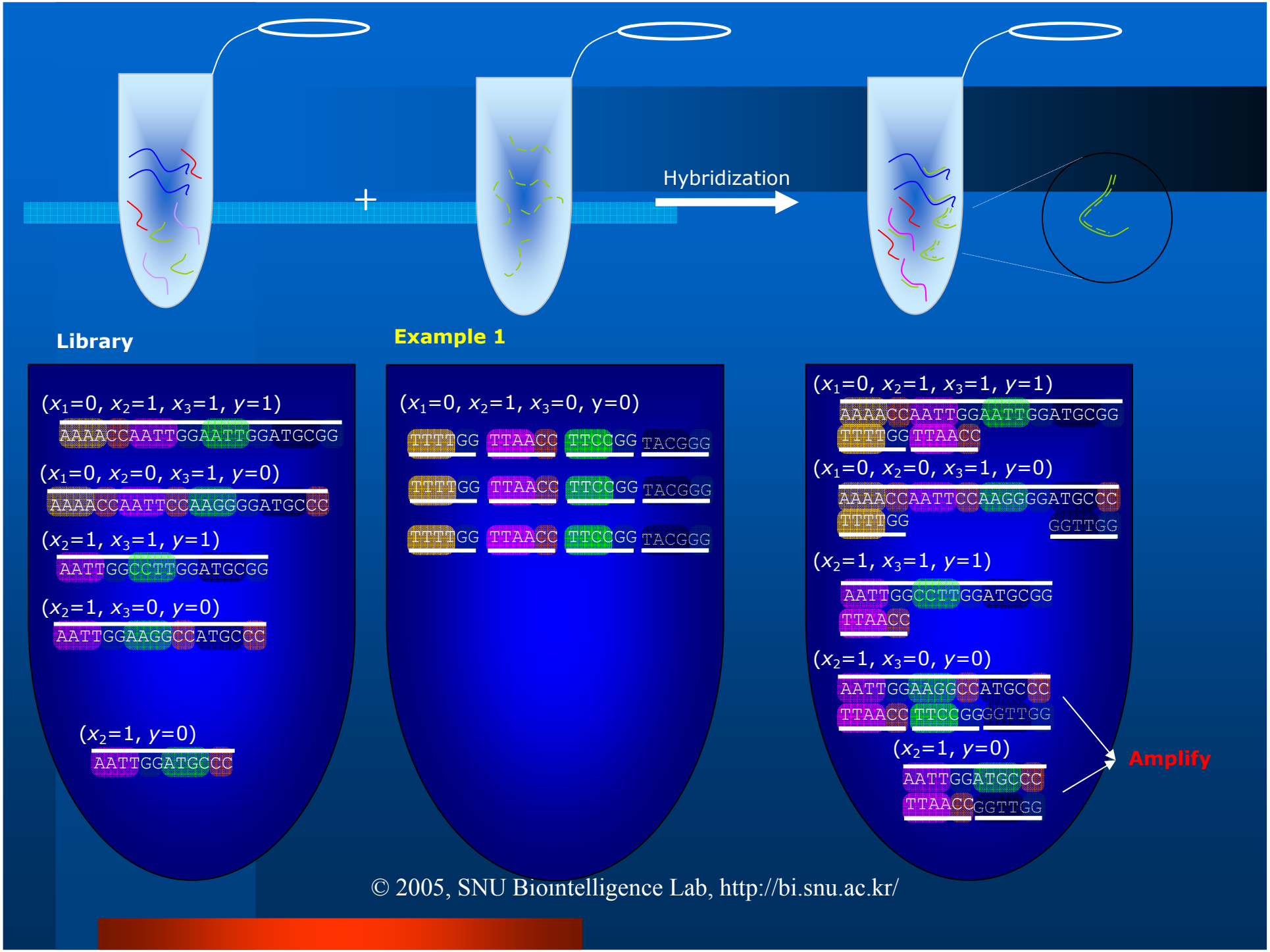




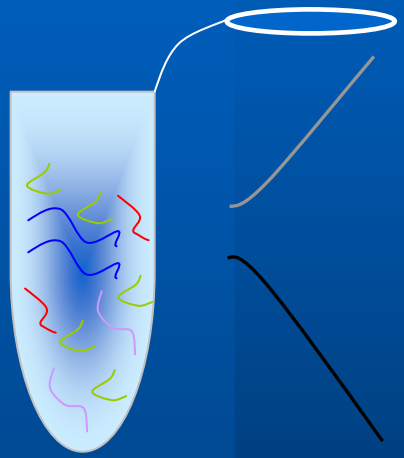
where AAAA x_1 ATGC y

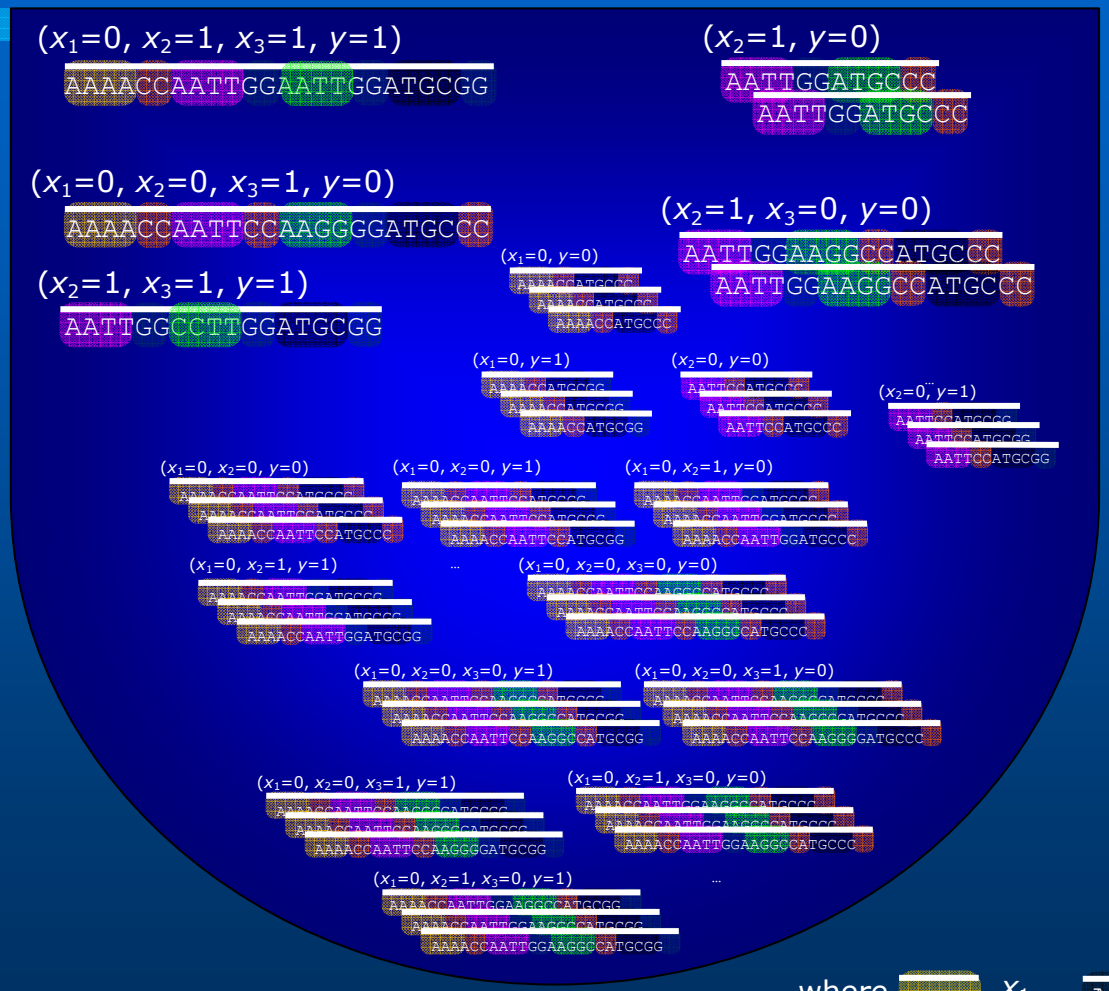
AATT x_2 CC 0

AAGG x_3 GG 1



Updated Library L_1

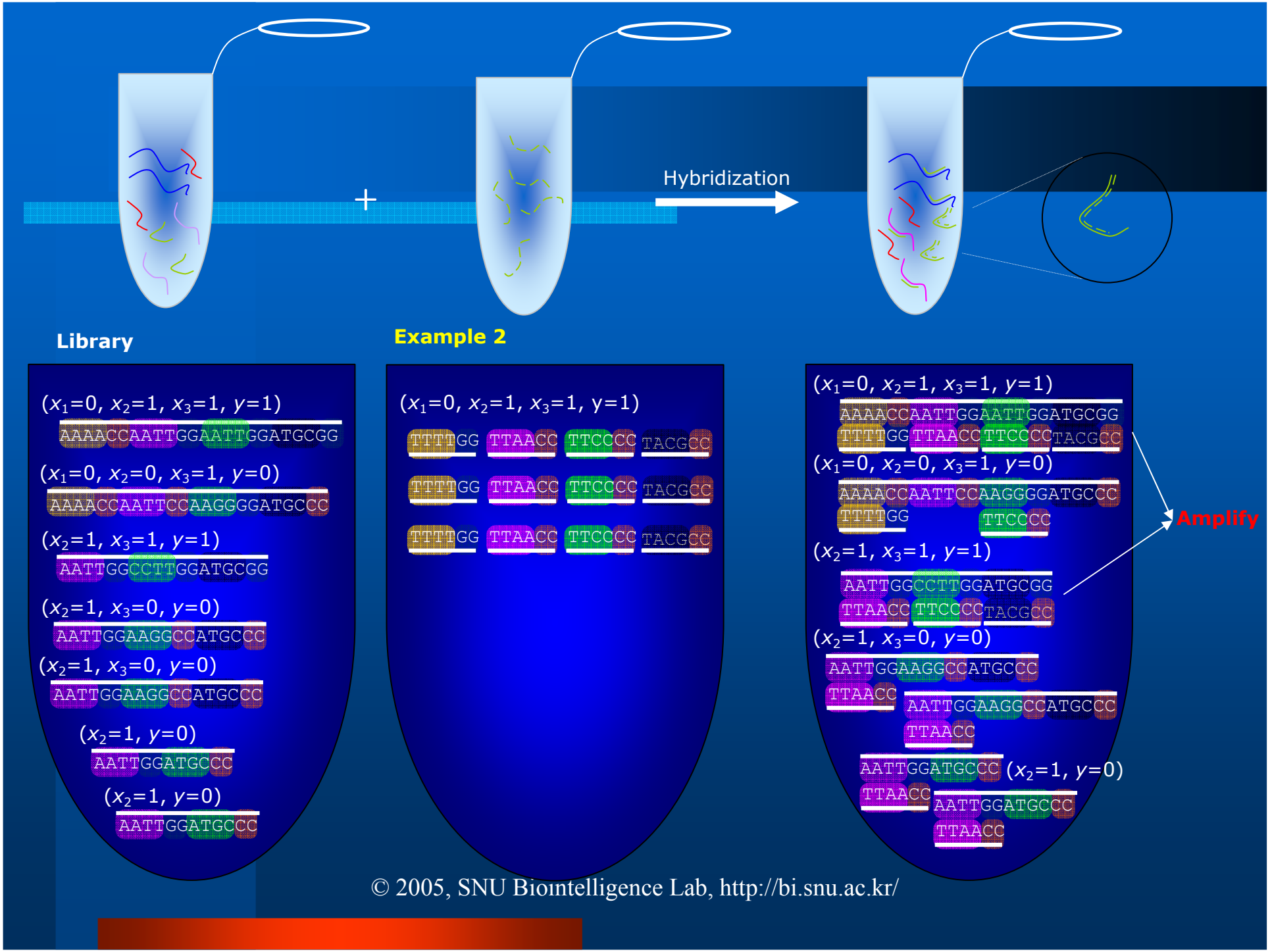




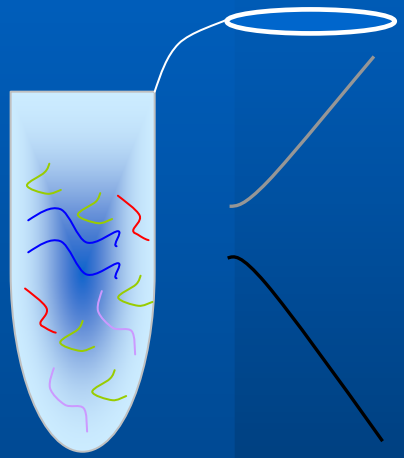
where AAAA x_1 ATGC y

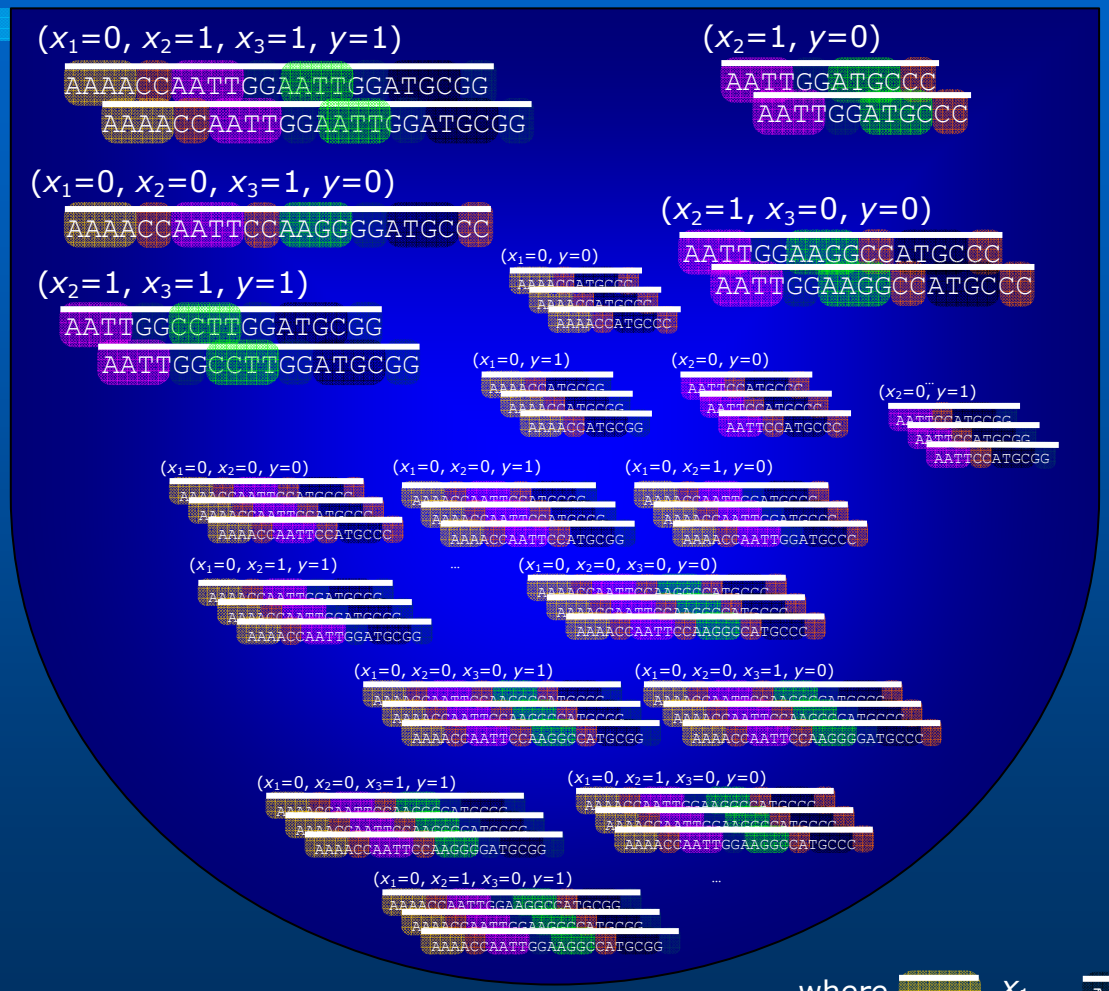
AATT x_2 CC 0

AAGG x_3 GG 1

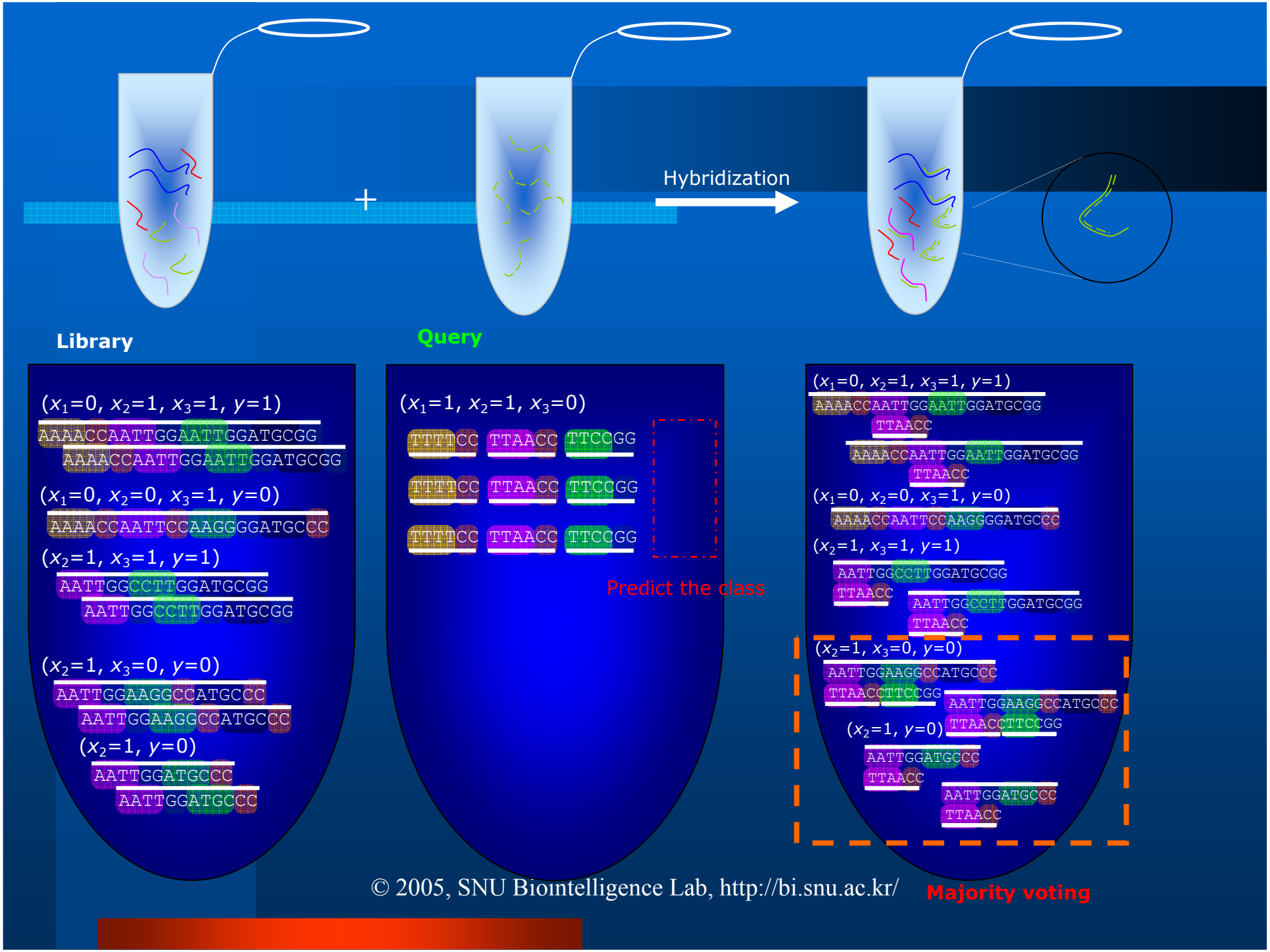


Updated Library L_2

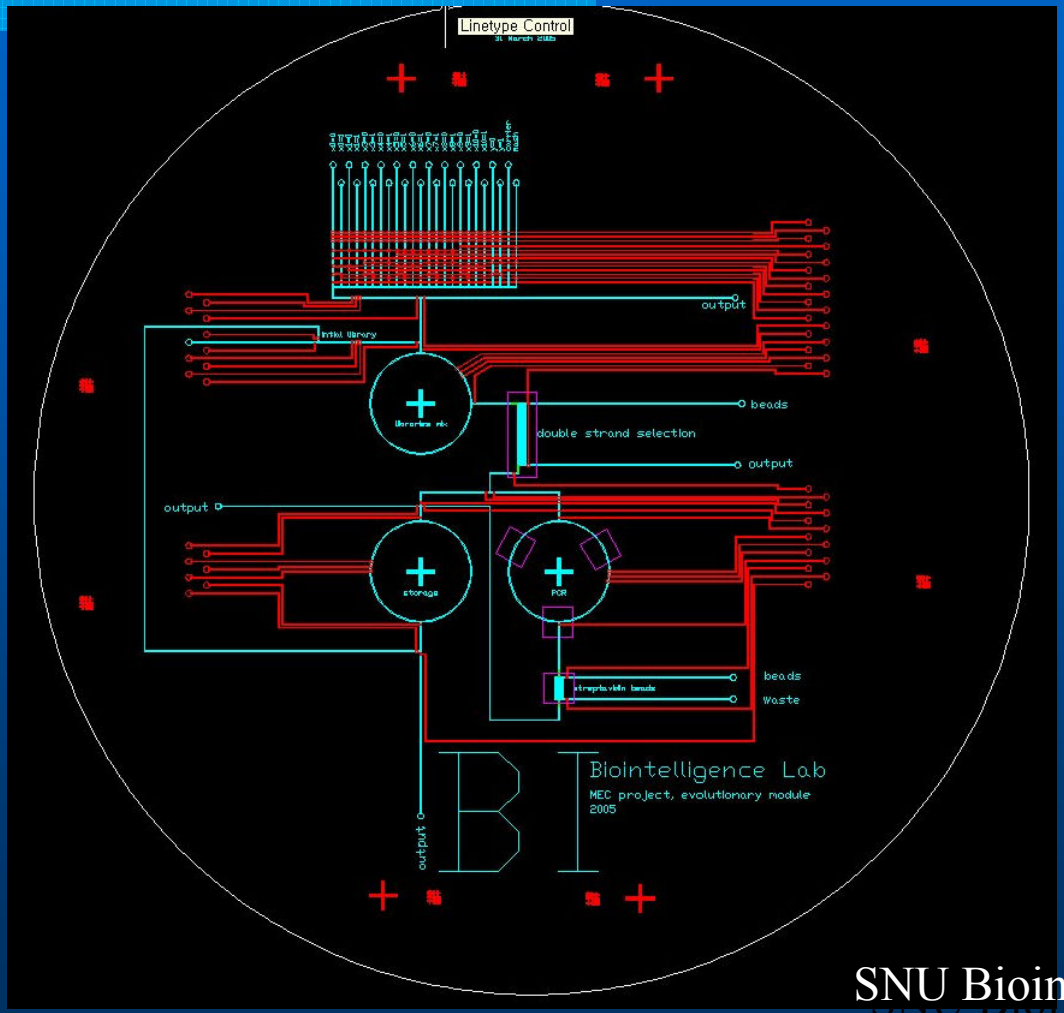




where AAAA x_1 AATT x_2 AAGG x_3 ATGC y CC 0 GG 1

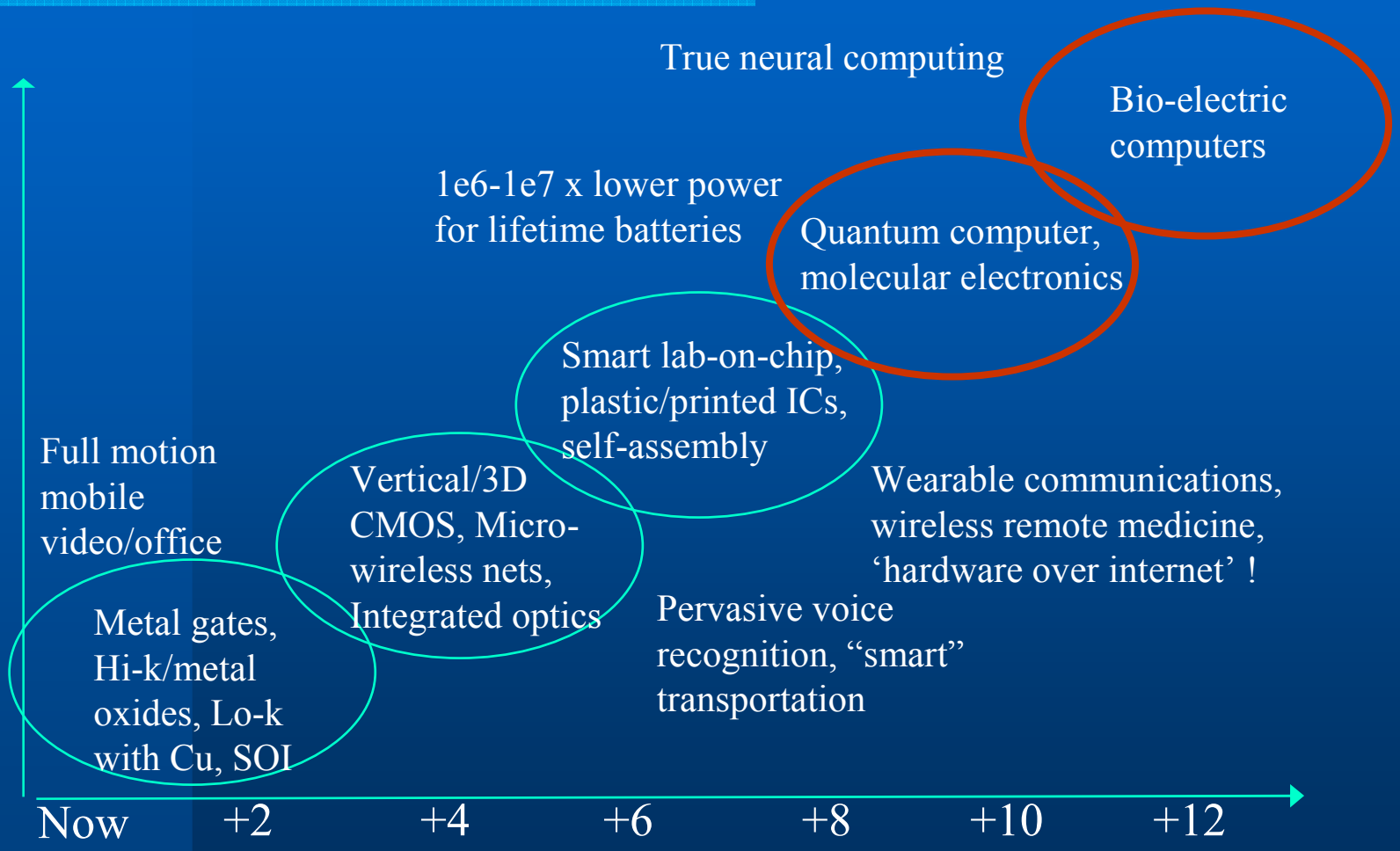


PLM Chip



Future of Molecular Nanobiointelligence Computers

Future Technology Enablers



Source: Motorola, Inc, 2000

Molecular Biocomputers

	하드웨어	소프트웨어
초소형 초대용량 정보검색 시스템	분자 기반의 대규모 연상 메모리	분자 기반의 대규모 데이터베이스
	In Vitro Wet 데이터 뱅크를 이용한 정보검색	
Molecular electronic components & circuits	자기조립에 기반한 나노구조 생성	DNA 구조 설계 지원 소프트웨어
	DNA 나노구조를 이용한 Patterning	

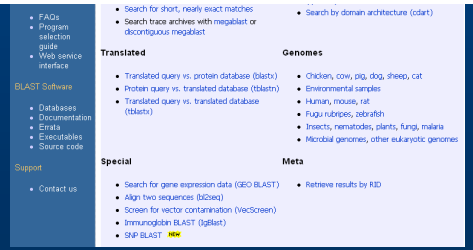
H/W & S/W Technology for Wet Information Retrieval

DNA
Processor

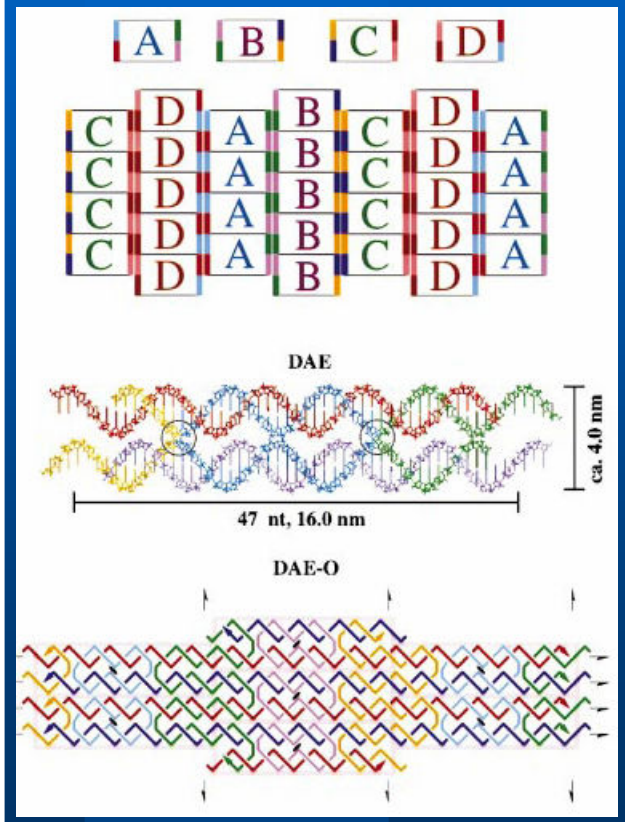
DNA Computer

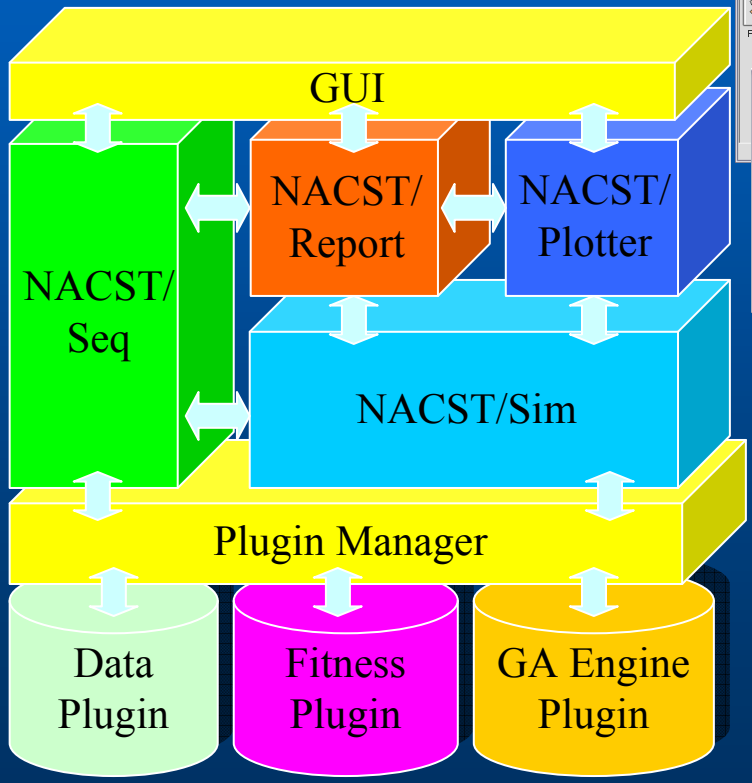
Silicon
Processor

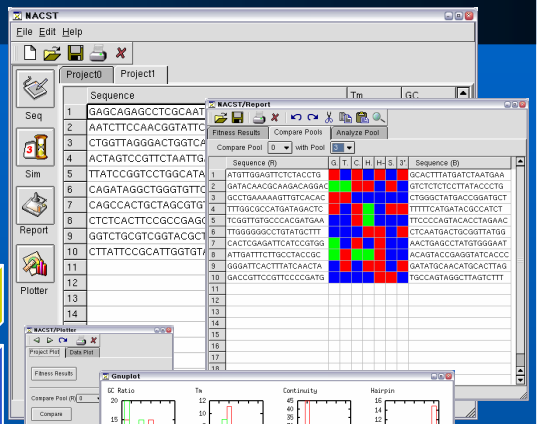
Wet Blast

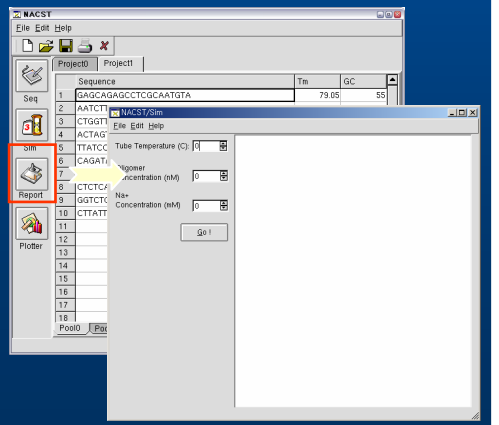


Design Support and Programming Software for DNA Computers and Nano-Machines



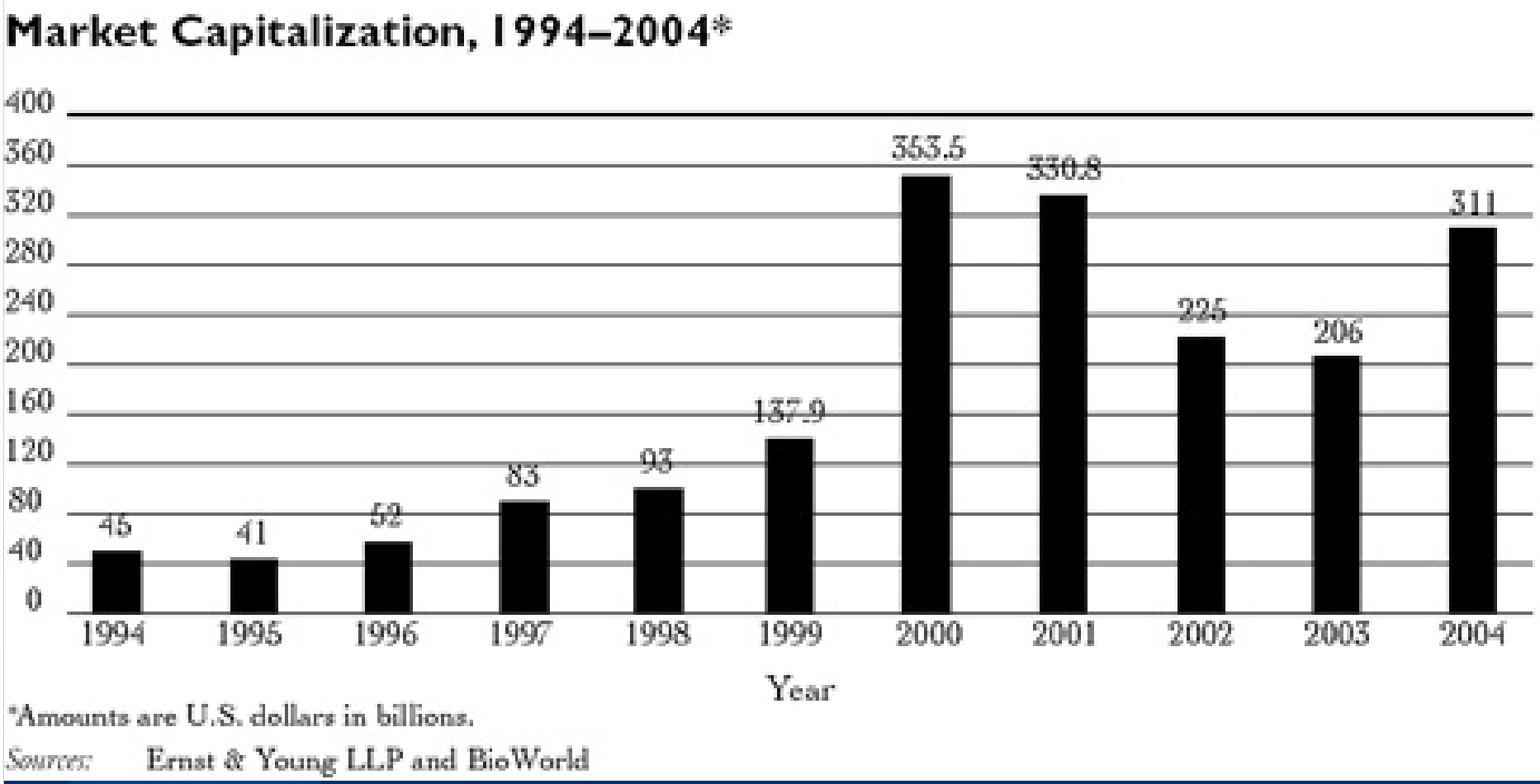






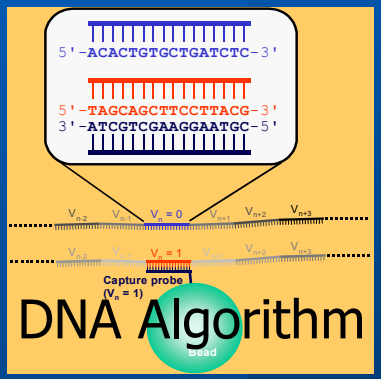
[Shin et al., *IEEE TEC* 2005]

BIT 시장 전망

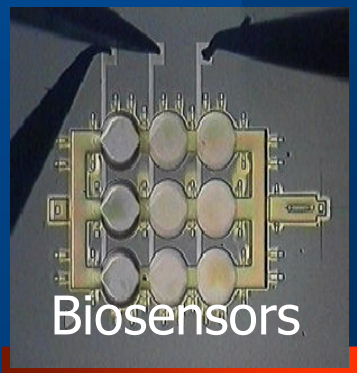


Biosensors (Bio Data) + Biocomputers (Bio H/W) + Bioinformatics (Bio S/W)

Bioinformatics S/W

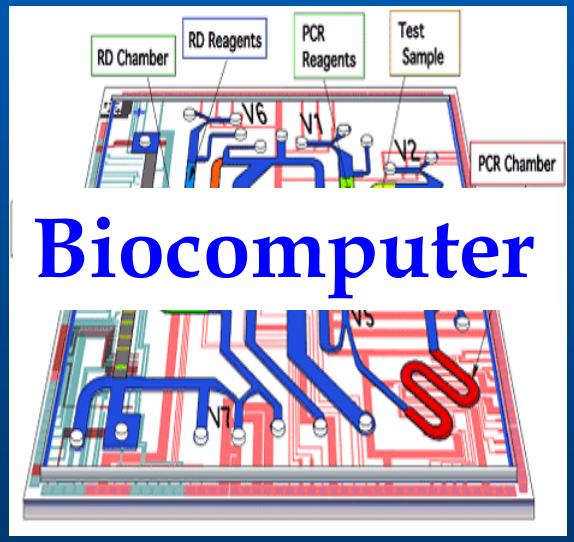


DNA Algorithm



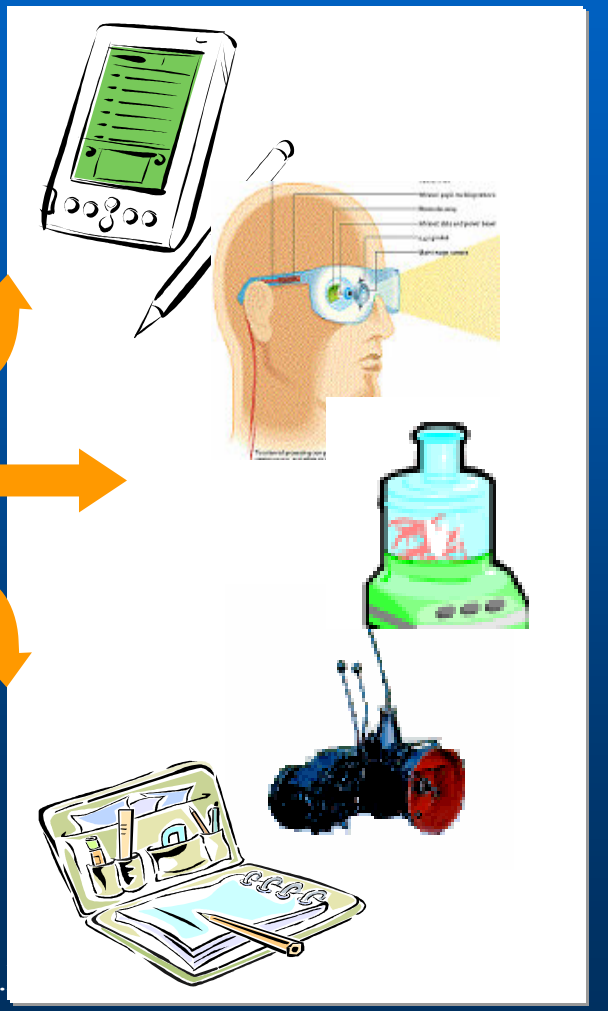
Biosensors

Design Support Software



Biocomputer

Bio-MEMS Technology



Acknowledgements

Collaborating labs

서울대 바이오지능 & 인공지능 연구실
서울대 의대 생화학교실
서울대 세포 및 미생물공학 연구실
한양대 프로테오믹스 연구실
㈜바이오니아 & ㈜바이오인포메틱스

Supported by

과기부 국가지정연구실사업

산자부 차세대신기술연구개발사업

More information at

http://bi.snu.ac.kr/

http://cbit.snu.ac.kr/