united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

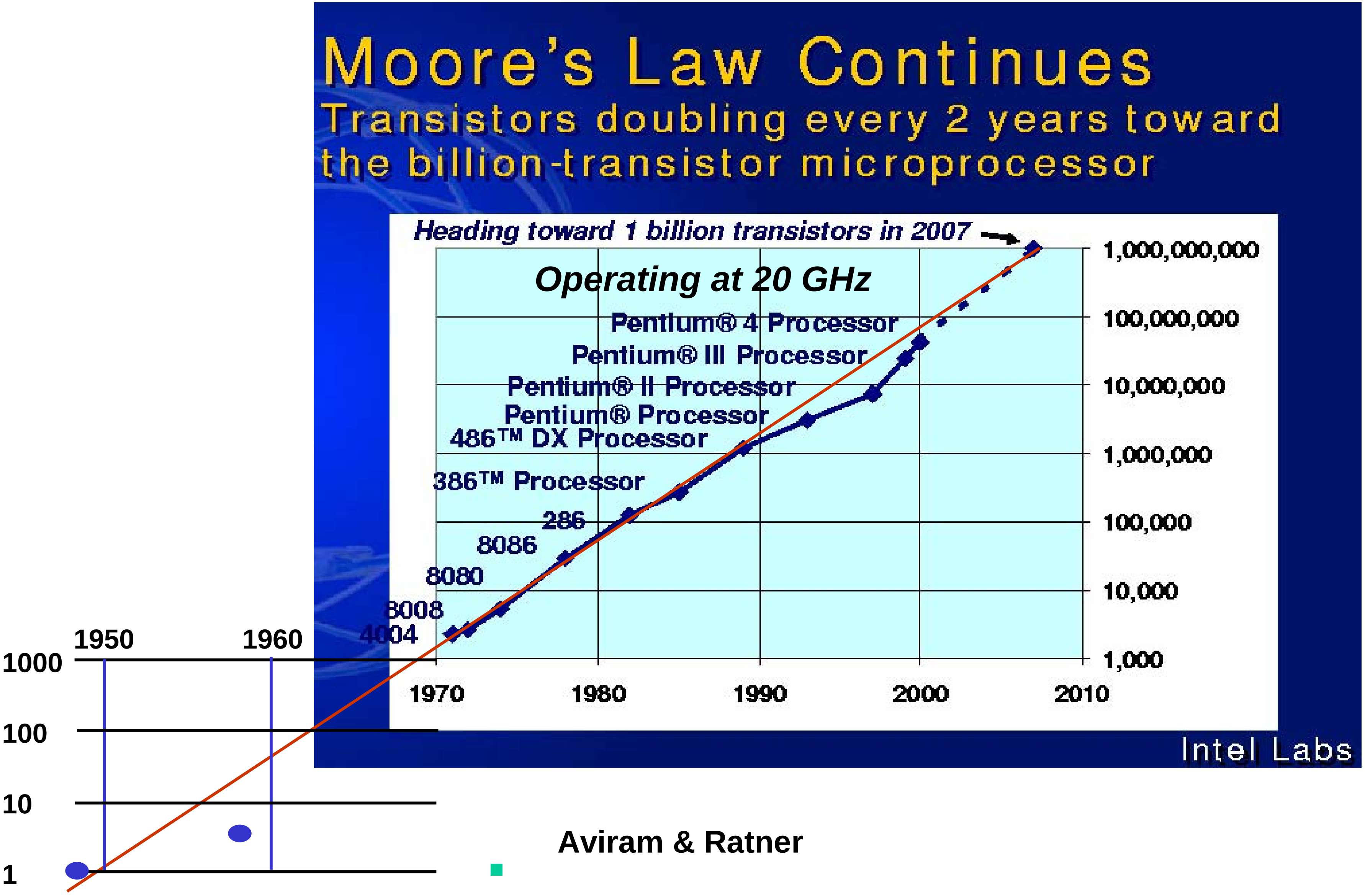
ICTP 40th Anniversary

SMR 1564 - 16

SPRING COLLEGE ON SCIENCE AT THE NANOSCALE
(24 May - 11 June 2004)

DNA self-assembly, molecules, electricity

Uri SIVAN
Israel Institute of Technology (TECHNION)
Institute of Solid State, Haifa, Israel

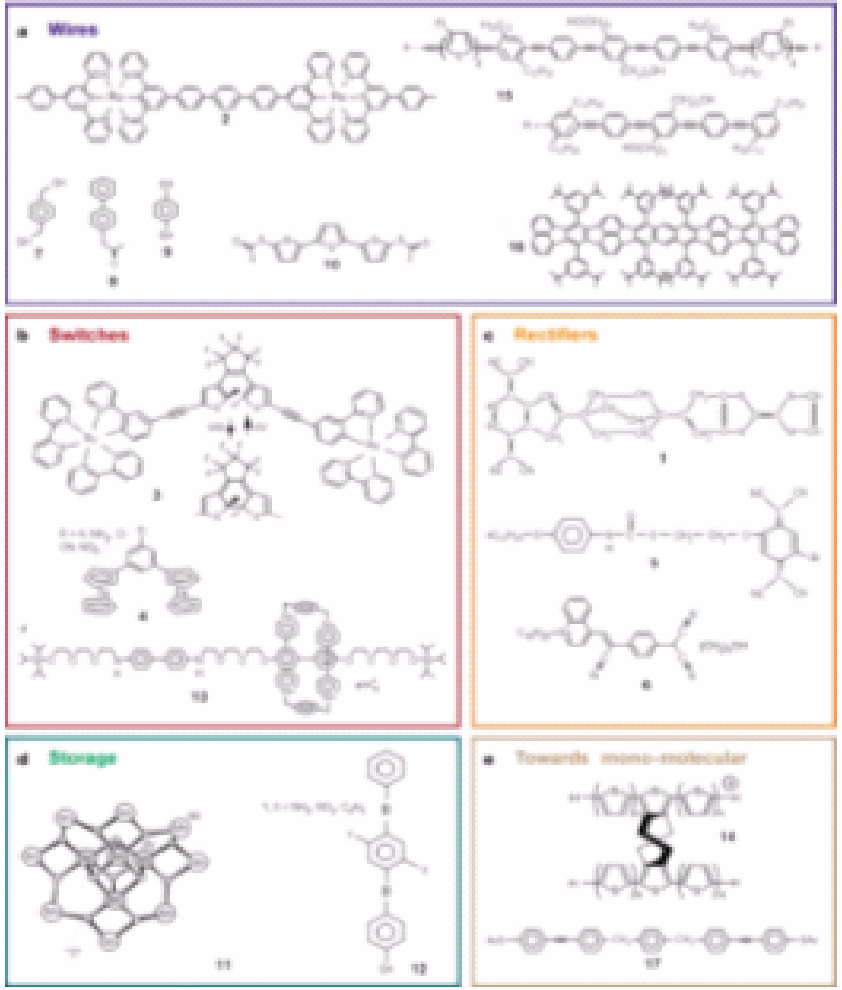


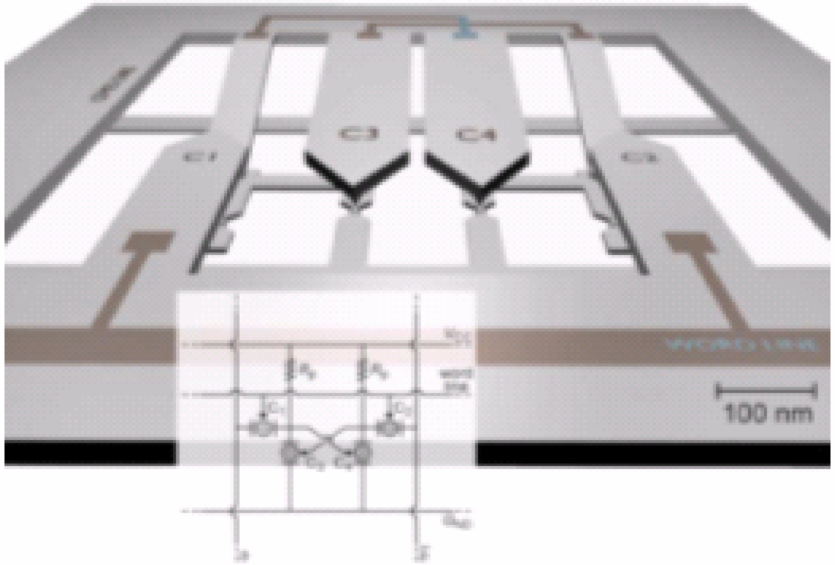
Moore's Law Continues

Transistors doubling every 2 years toward the billion-transistor microprocessor

Aviram & Ratner

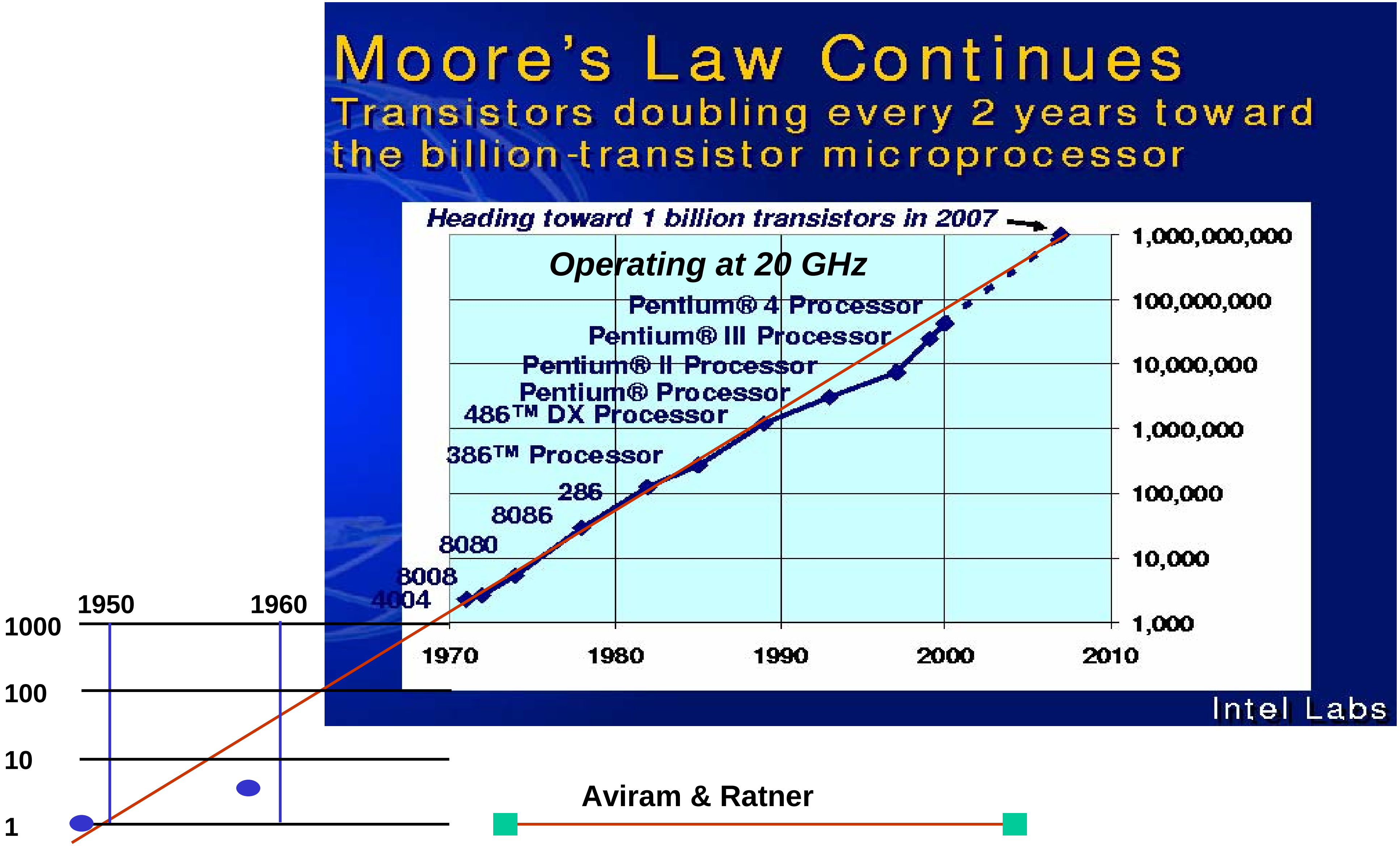
Molecular Electronics - Gap Between Devices and Circuits





Joachim et al.

MOLECULAR ELECTRONICS?



Moore’s Law Continues

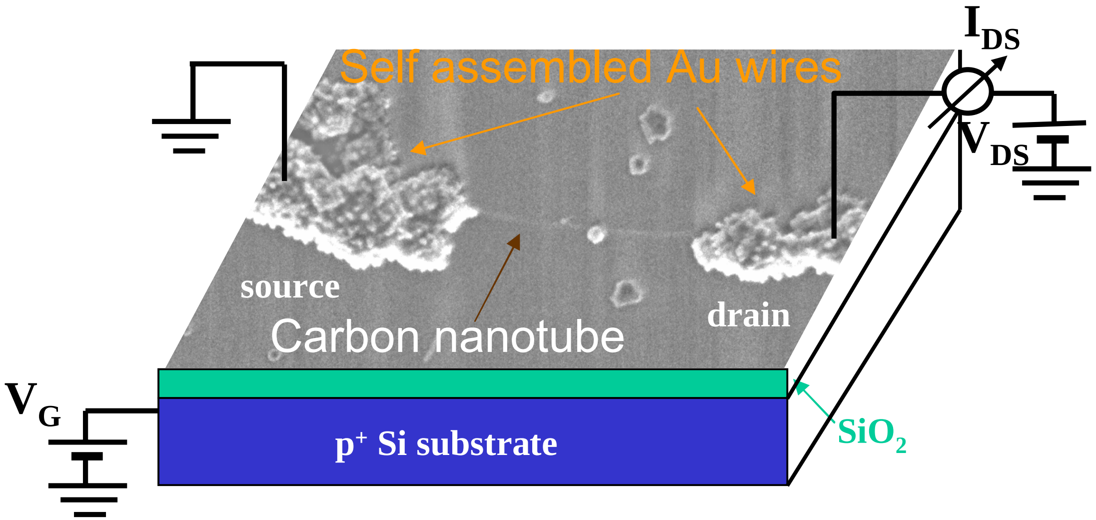
Transistors doubling every 2 years toward the billion-transistor microprocessor

Harnessing Molecular Biology to The Self-Assembly of Molecular Scale Electronics

Uri Sivan

Technion -Israel Institute of Technology

Haifa, Israel



Kinneret Keren, Rotem Berman, Rachel Gilad, Yoav Soen, Michael Krueger, Ilya Baskin, Stav Zaitsev, Doron Lipson, Gidi Ben Yoseph, Evgeny Buchstab, Yoav Eichen, Erez Braun

Two Paradigms for Complex Constructs

✓Microelectronics – a supervised top-down approach

✓Biology – autonomous self-assembly based on the information encoded in the ingredients

FIRST BABY STEPS !

Outline

1. DNA and DNA templated electronics
2. Homologous genetic recombination and “sequence specific molecular lithography” - Self-assembled field effect transistor in a test-tube
3. Approaching complexity
4. Recruiting evolution, antibodies

COMPLEXITY

Is a computer more complex than a toaster?

“The number of instructions in a minimal protocol detailing how to assemble the machine out of elementary building blocks”

Poor definition

- ✓ What is an elementary building block?
- ✓ Non-operative

Can’t tell how complex the machine is by looking at it

Still gives insight

- ✓ Crystals are simple (3 lattice vectors+atom type)
- ✓ So are monolayers
- ✓ Relates structure complexity to information
- ✓ Complex looking objects may, in fact, be simple (DNA templates)

DNA as an Information Carrying Polymer

Components of DNA

•DNA is a double stranded polymer. The monomer units of DNA are nucleotides, and the polymer is known as a "polynucleotide."

•Each nucleotide consists of a 5-carbon sugar (deoxyribose), a nitrogen containing base attached to the sugar, and a phosphate group.

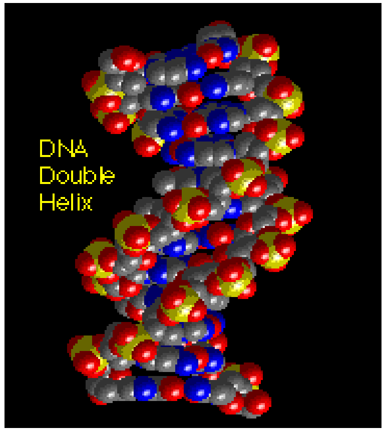
•There are four different types of nucleotides found in DNA, differing only in the nitrogenous base.

A is for adenine

G is for guanine

C is for cytosine

T is for thymine



http://www.blc.arizona.edu/Molecular_Graphics/DNA_Structure/DNA_Tutorial.HTML#Components

DNA is built from Purine and Pyrimidine bases

Adenine and guanine are purines.

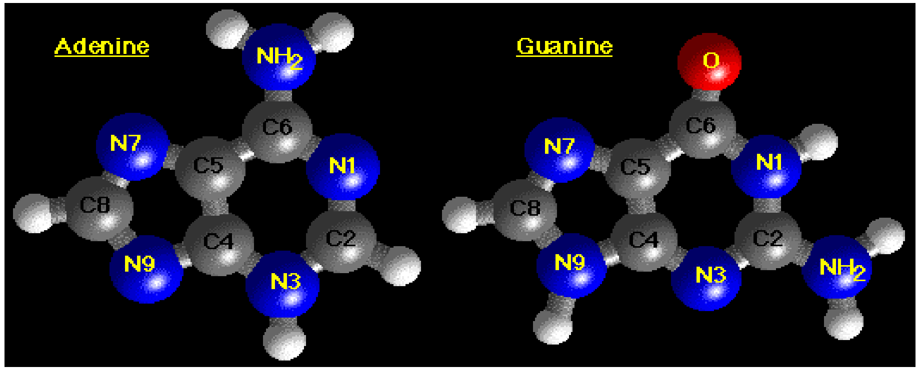
Purines are the larger of the two types of bases found in DNA.

Cytosine and thymine are pyrimidines.

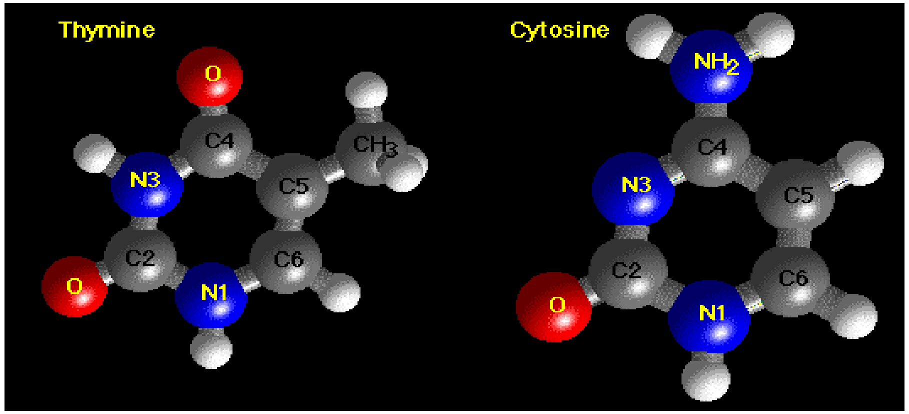
The 6 stoms (4 carbon, 2 nitrogen) are numbered 1-6.

Like purines, all pyrimidine ring atoms lie in the same plane.

Structure of A and G



Structure of T and C



http://www.blc.arizona.edu/Molecular_Graphics/DNA_Structure/DNA_Tutorial.HTML#Components

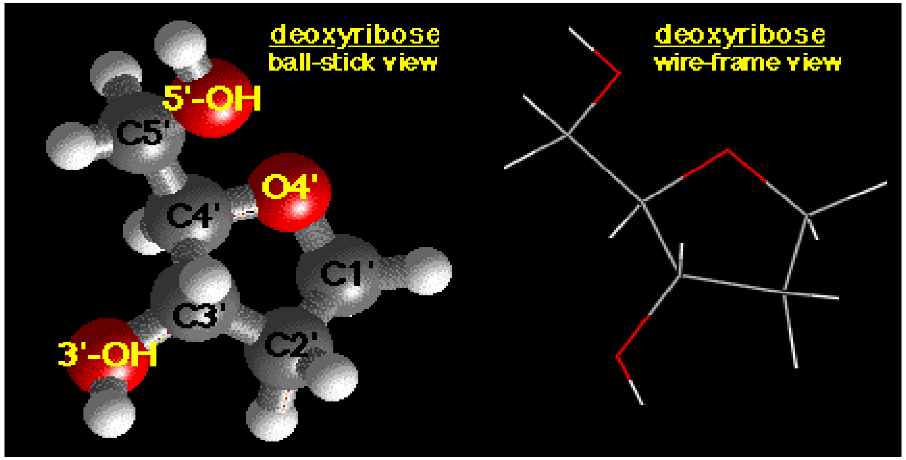
Deoxyribose Sugar

The deoxyribose sugar of the DNA backbone has 5 carbons and 3 oxygens.

The carbon atoms are numbered 1', 2', 3', 4', and 5' to distinguish from the numbering of the atoms of the purine and pyrmidine rings.

The hydroxyl groups on the 5'- and 3'- carbons link to the phosphate groups to form the DNA backbone.

Deoxyribose lacks an hydroxyl group at the 2'-position when compared to ribose, the sugar component of RNA.



http://www.blc.arizona.edu/Molecular_Graphics/DNA_Structure/DNA_Tutorial.HTML#Components

Nucleosides

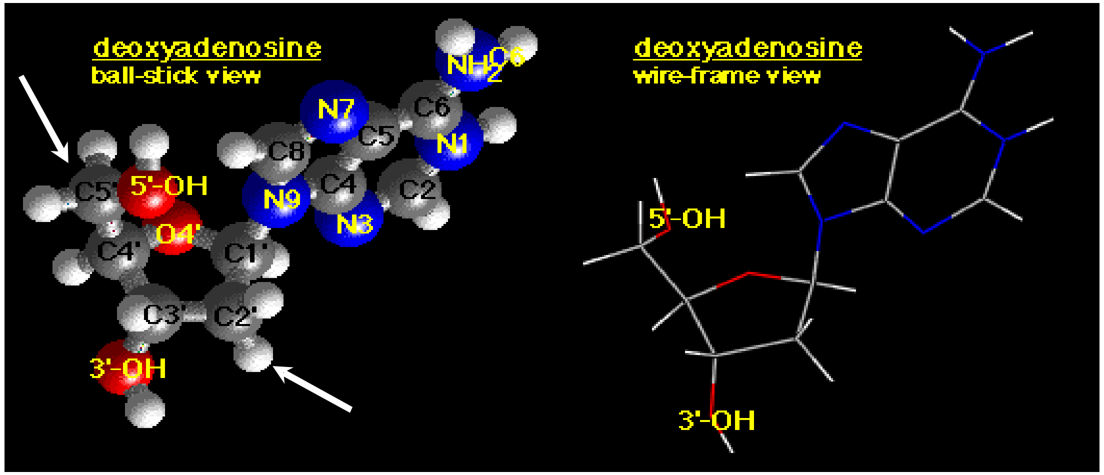
A nucleoside is one of the four DNA bases covalently attached to the C1' position of a sugar.

The sugar in deoxynucleosides is 2'-deoxyribose.

The sugar in ribonucleosides is ribose.

Nucleosides differ from nucleotides in that they lack phosphate groups.

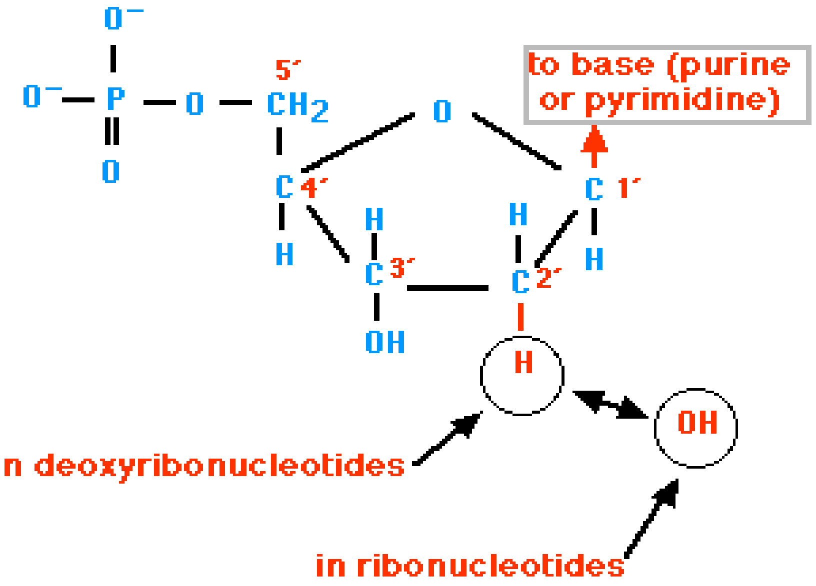
The four different nucleosides of DNA are deoxyadenosine (dA), deoxyguanosine (dG), deoxycytosine (dC), and (deoxy)thymidine (dT, or T).



http://www.blc.arizona.edu/Molecular_Graphics/DNA_Structure/DNA_Tutorial.HTML#Components

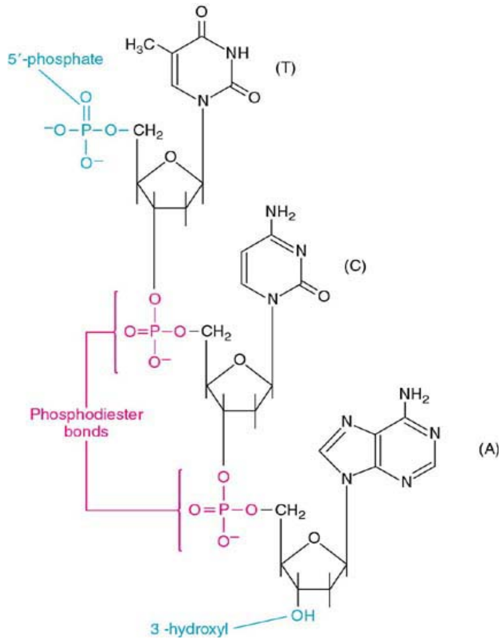
Nucleotides

A nucleotide is a nucleoside with one or more phosphate groups covalently attached to the 3'- and/or 5'-hydroxyl group(s).



http://www.blc.arizona.edu/Molecular_Graphics/DNA_Structure/DNA_Tutorial.HTML#Components

DNA Backbone



The DNA backbone is a polymer with an alternating sugar-phosphate sequence.

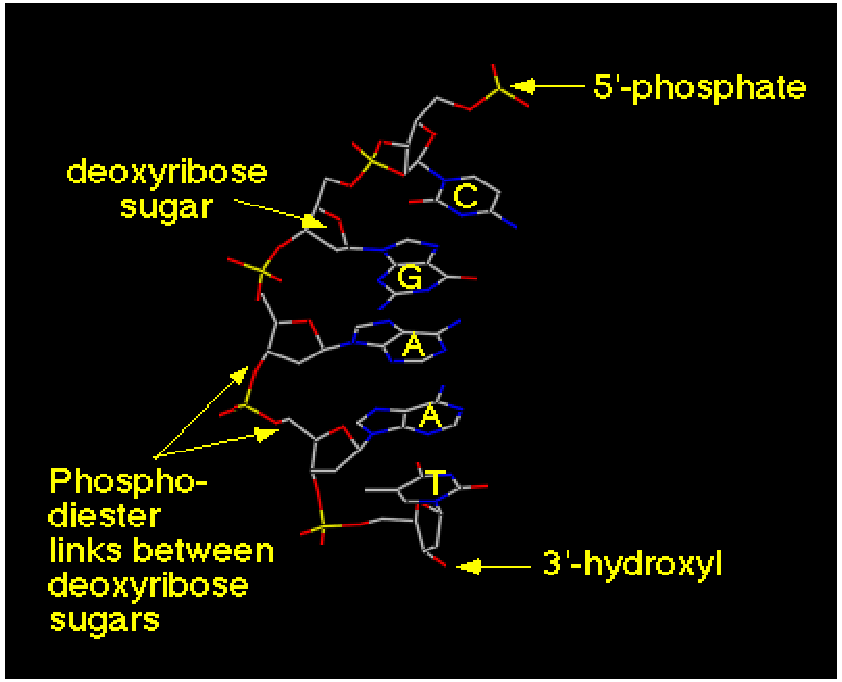
The deoxyribose sugars are joined at both the 3'-hydroxyl and 5'-hydroxyl groups to phosphate groups in "phosphodiester" bonds.

Chain has a direction (known as polarity), 5'- to 3'- from top to bottom

Oxygens (red atoms) of phosphates are polar and negatively charged

A, G, C, and T bases can extend away from chain, and stack atop each other

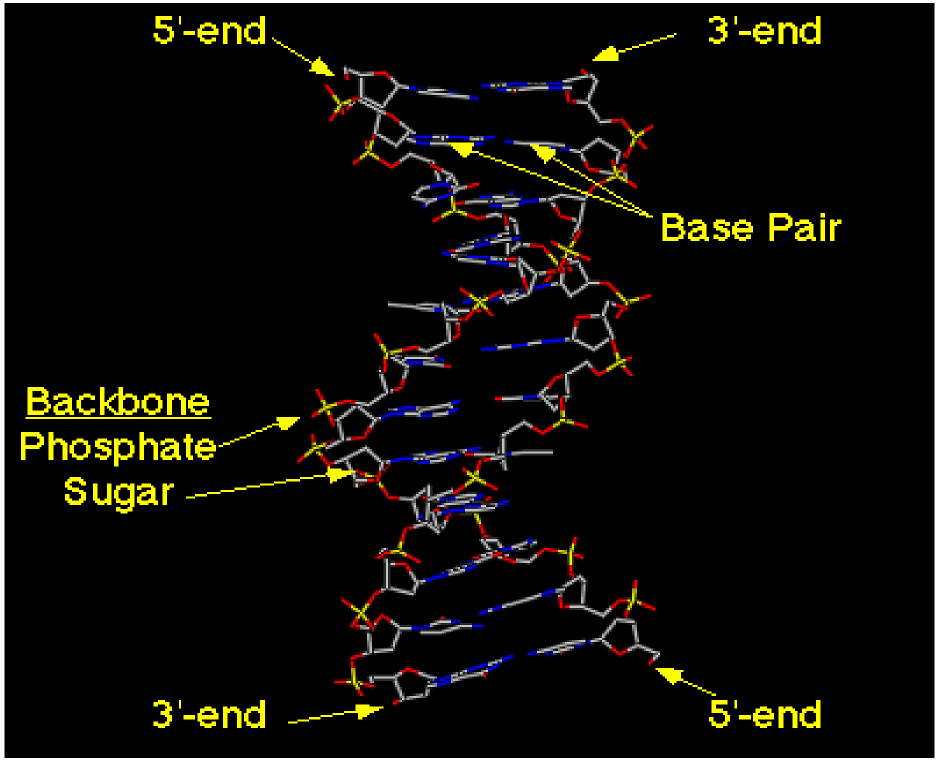
Bases are hydrophobic



DNA Double Helix

DNA is a normally double stranded macromolecule. Two polynucleotide chains, held together by weak thermodynamic forces, form a DNA molecule.

- •Two DNA strands form a helical spiral, winding around a helix axis in a right-handed spiral
- •The two polynucleotide chains run in opposite directions
- •The sugar-phosphate backbones of the two DNA strands wind around the helix axis like the railing of a spiral staircase
- •The bases of the individual nucleotides are on the inside of the helix, stacked on top of each other like the steps of a spiral staircase.

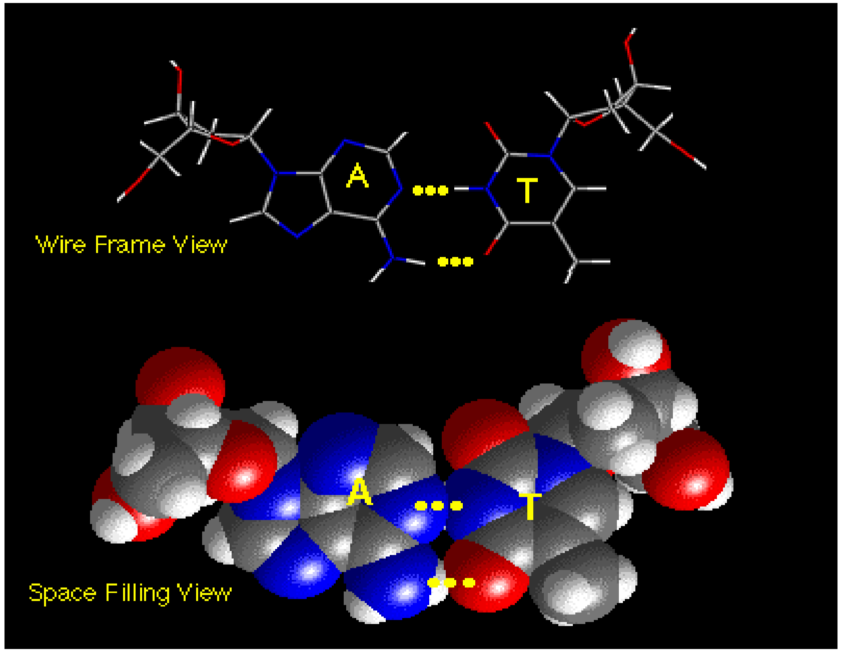


http://www.blc.arizona.edu/Molecular_Graphics/DNA_Structure/DNA_Tutorial.HTML#Components

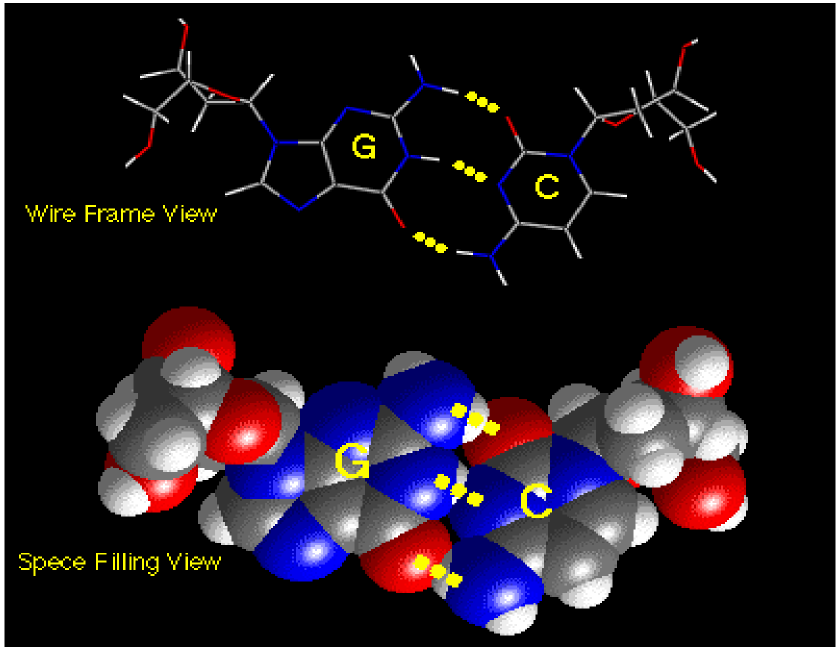
Base Pairs

Within the DNA double helix, A forms 2 hydrogen bonds with T on the opposite strand, and G forms 3 hyrdorgen bonds with C on the opposite strand.

dA-dT base pair as found within DNA double helix



dG-dC base pair as found within DNA double helix

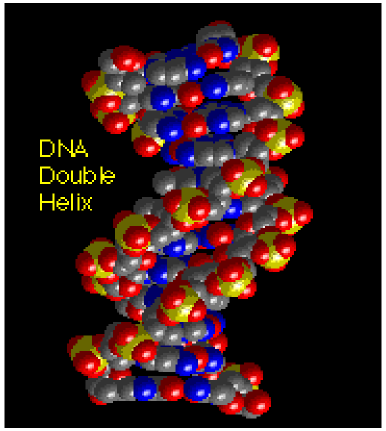


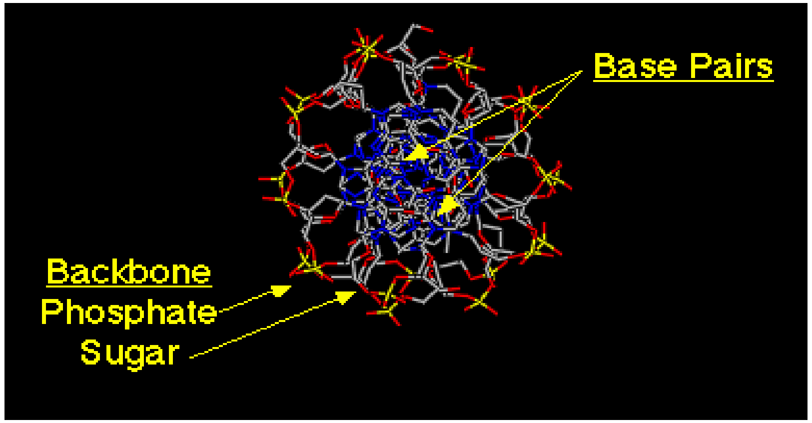
dA-dT and dG-dC base pairs are the same length, and occupy the same space within a DNA double helix. Therefore the DNA molecule has a uniform diameter.

http://www.blc.arizona.edu/Molecular_Graphics/DNA_Structure/DNA_Tutorial.HTML#Components

DNA Helix Axis

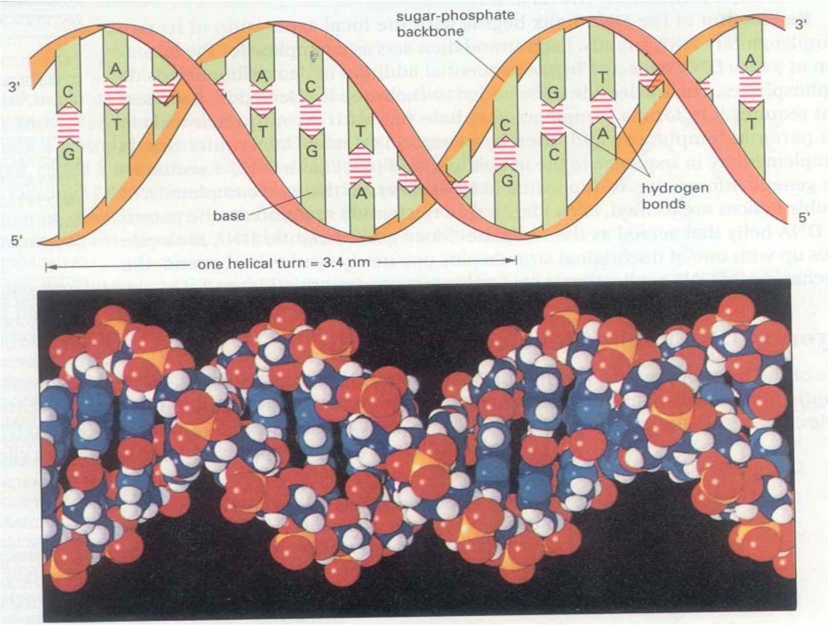
The helix axis is most apparent from a view directly down the axis.
The sugar-phosphate backbone is on the outside of the helix where the polar phosphate groups (red and yellow atoms) can interact with the polar environment.
The nitrogen (blue atoms) containing bases are inside, stacking perpendicular to the helix axis.

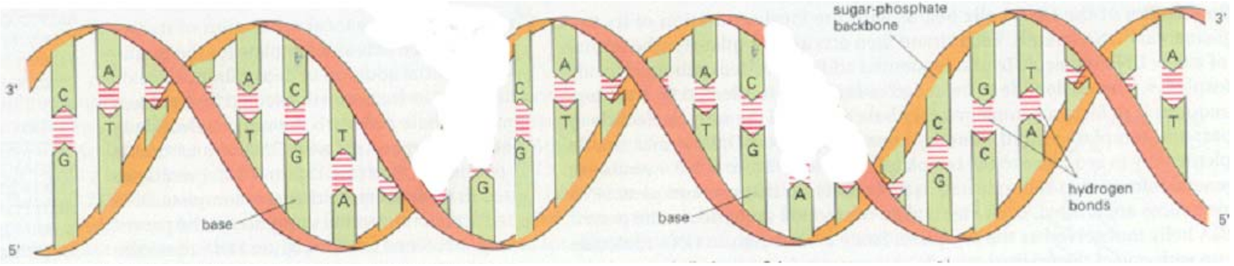




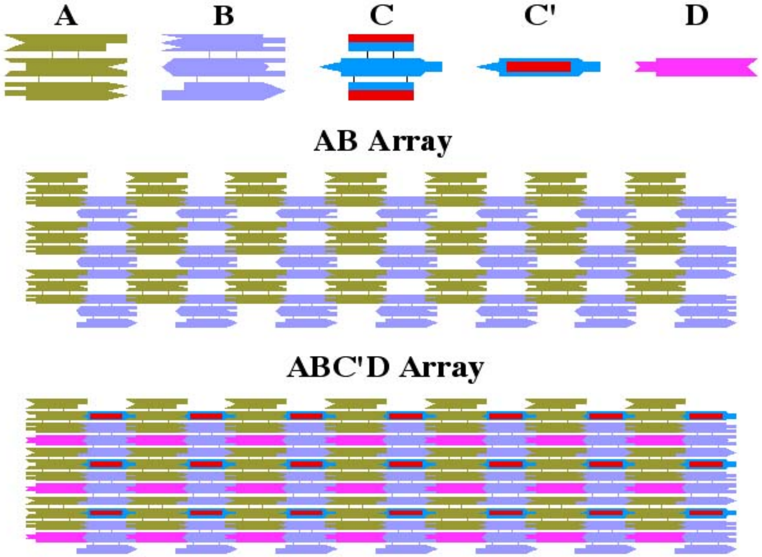
One turn extends 3.4 nm. The bare charge is -1e/1.7A. The dressed charge under physiological conditions is -1e/7A.
The persistence length of double stranded DNA is ~500A

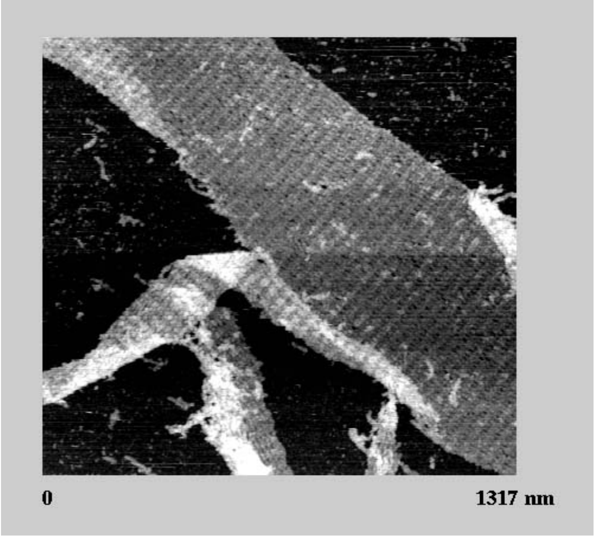
DNA Molecular Recognition

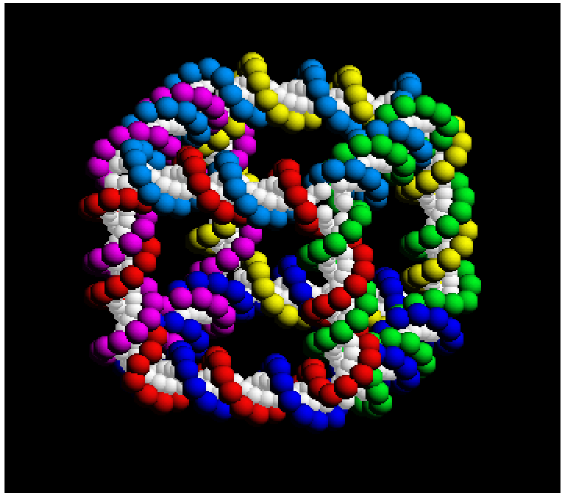


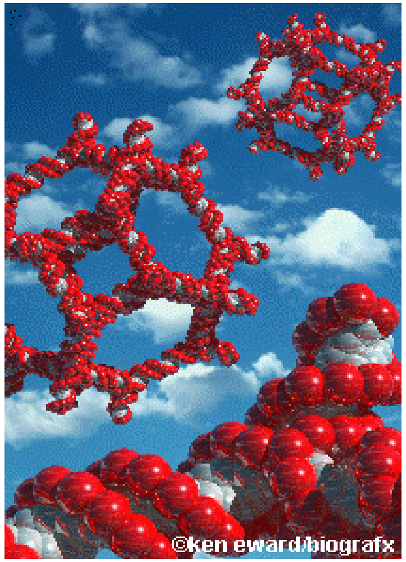


DNA Tiling and Other Gismos
Ned Seeman & collaborators








©ken eward/biografx

http://seemanlab4.chem.nyu.edu

SCALES OF SELF ASSEMBLY

- 1 k_BT <Energy scale<10 k_BT. Provided by screened Coulomb interaction (Ry/dielectric constant of water)

$$E \approx \frac{e^2}{\kappa a_0} \approx \frac{10\,eV}{100} = 100\,meV$$

$$k_B T = 25\,meV$$

- Length scale 0.1-1 nm
- Force=energy/length scale = 10-100 pN
- Self Assembly is hence synonymous to errors!
- Can self assembly and electronic conduction coexist in the same place?
- Self Assembly with a single energy scale (Lego bricks analogy)? Enzymes?

MINIATURIZATION AND COMPLEXITY

MINIATURIZATION

- ✓ 100 Atoms per bit of information-1nm sphere
- ✓ Working example (Feynman) - DNA.

  ROM, ~100 atoms per base pair.

  $3x10^9$ bits in a tiny fraction of a cell.

- ✓ ~10^{19} bits of information are generated worldwide every year.
- ✓ Approximately 10^8 hard disks or 10^9 CDs (1000Km tall stack!).
- ✓ $10^{19}x100 = 10^{21}$ atoms. 10^{-3} Avogadro number, A cube, 3mm on the side.

COMPLEXITY

- ✓ Microelectronics - 10^8 elements in a CPU
- ✓ Molecular electronics - 10^{14}-10^{17} elements in a circuit

Main Challenges in Molecular Scale Electronics

- ✓ Operation Principles of molecular scale devices - single electron transistors?
- ✓ Circuit Organization - positioning a tremendous number of nanometer scale objects with nanometer accuracy
- ✓ Inter device wiring
- ✓ Wiring molecular scale objects to the external macroscopic world

Barriers in the Self Assembly of Molecular Scale Electronics

- Electronic materials like metals, semiconductors, and polymers - superb electronic properties but only trivial self assembly capabilities
- Biological Molecules - poor electronic properties but superb self assembly capabilities

Combine electronic materials with biological molecules

Assembly in a three step process

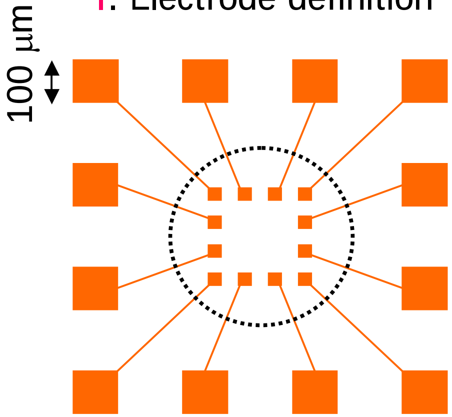
•Build a DNA template

•Localize electronic devices at molecular addresses

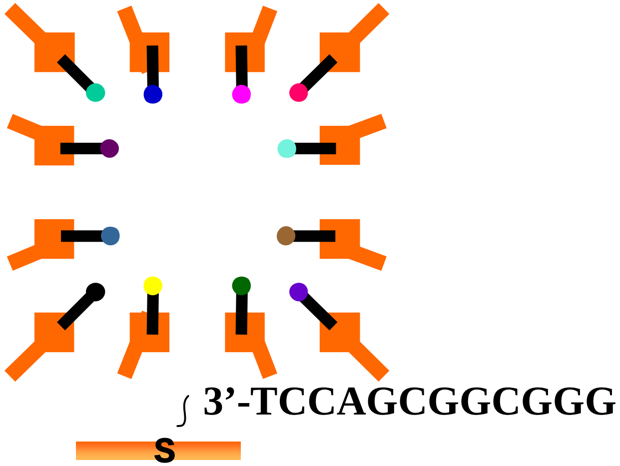
•Wire by DNA metallization

Self Assembly of an Electronic Circuit Using DNA - Possible Assembly Scheme and its Limitations

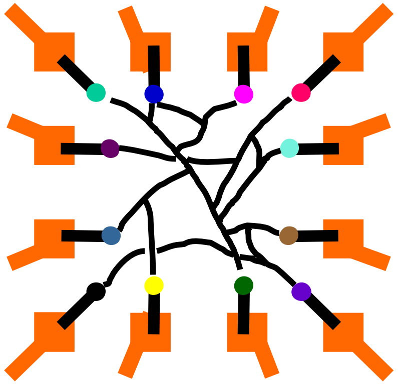
I. Electrode definition



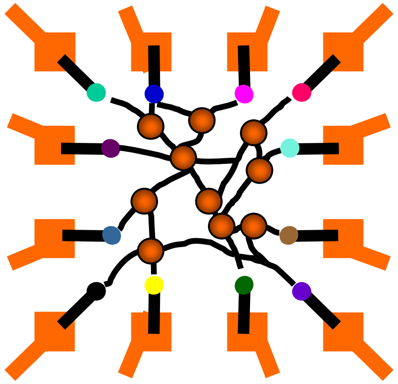
II. Electrode encoding



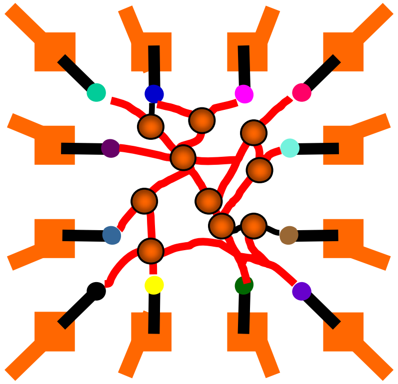
III. Scaffold



IV. Device positioning

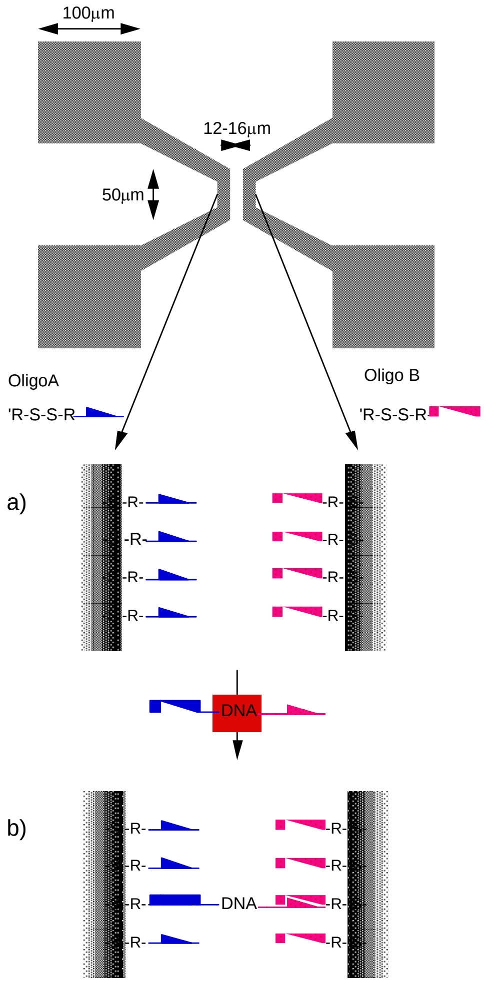


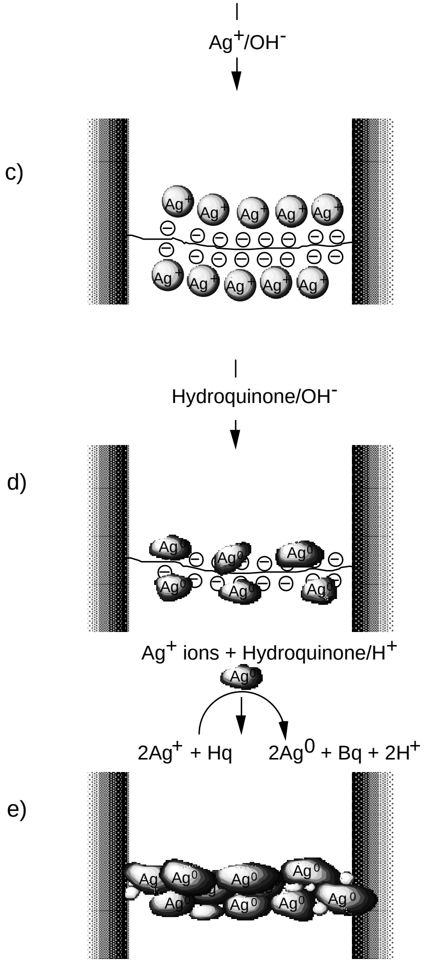
V. Converting DNA to wires



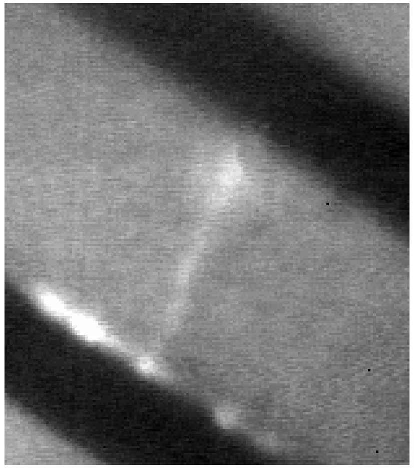
DNA Templated Conductive Wire

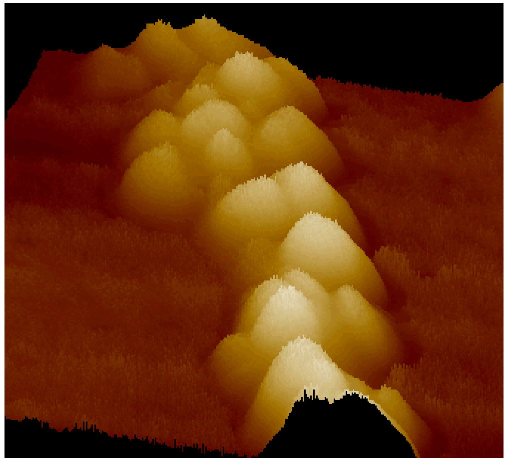
E. Braun, Y. Eichen, U. Sivan, G. Ben-Yoseph, Nature 391, 775 (1998)

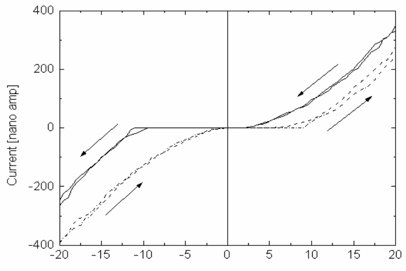


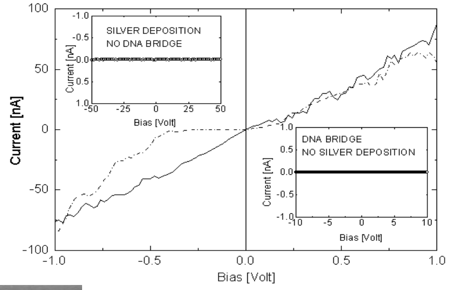


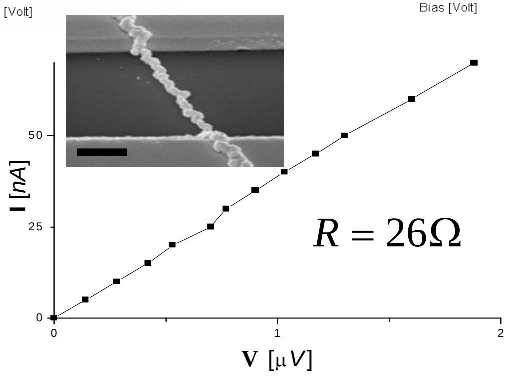
Conductive silver
wire







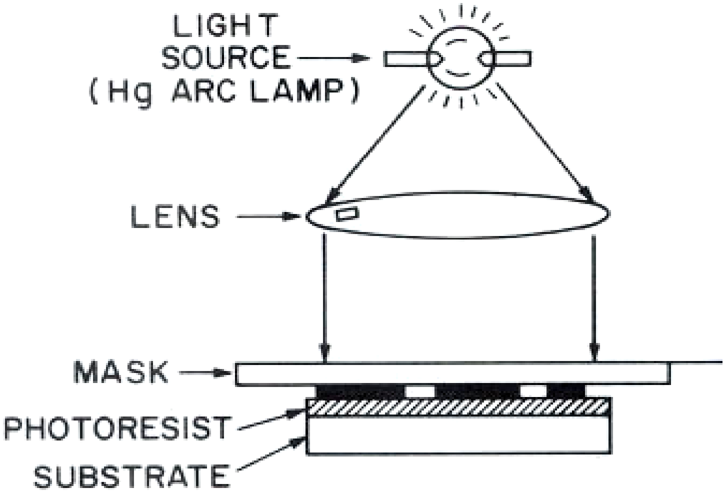




$R = 26\Omega$

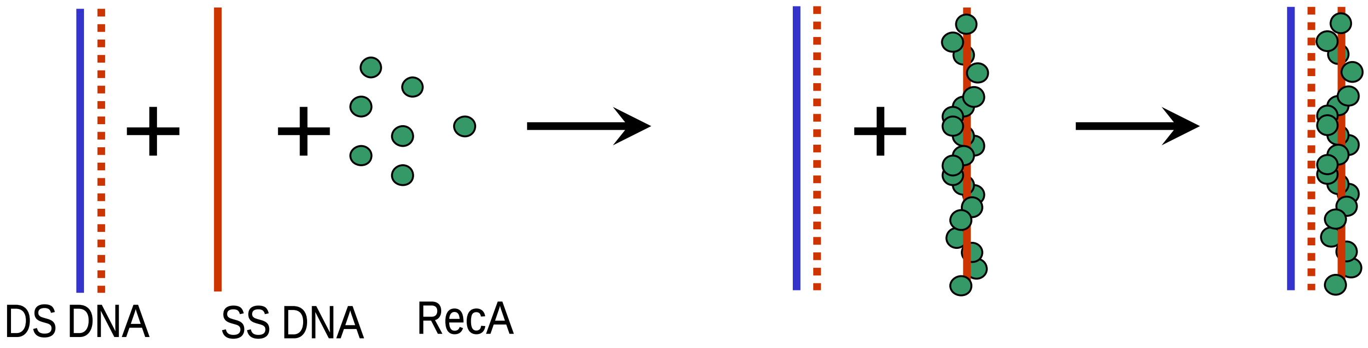
MICROELECTRONICS RELIES ON LITHOGRAPHY

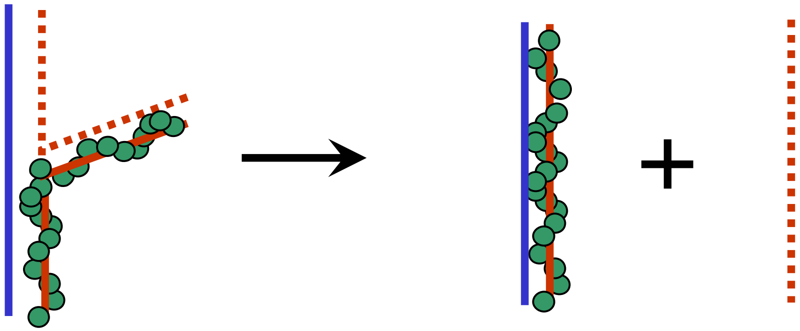
- Not merely a technology - It’s a concept how to start with a virgin silicon wafer and embed in it tremendous amounts of information

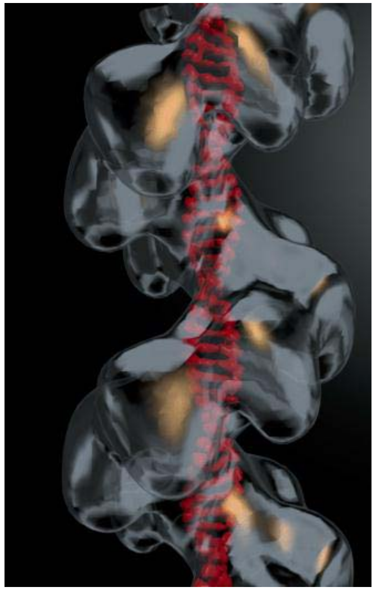


- Can we invent an equivalent concept for molecular scale electronics with the information embedded in molecules rather than in glass masks?

Homologous Recombination by RecA





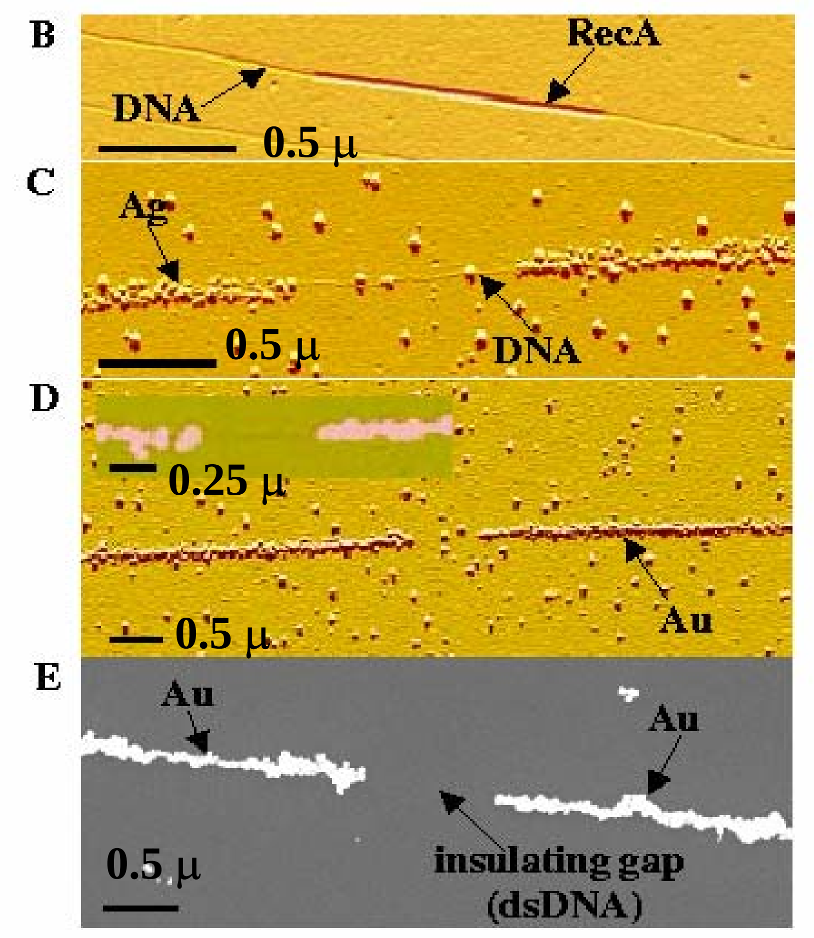


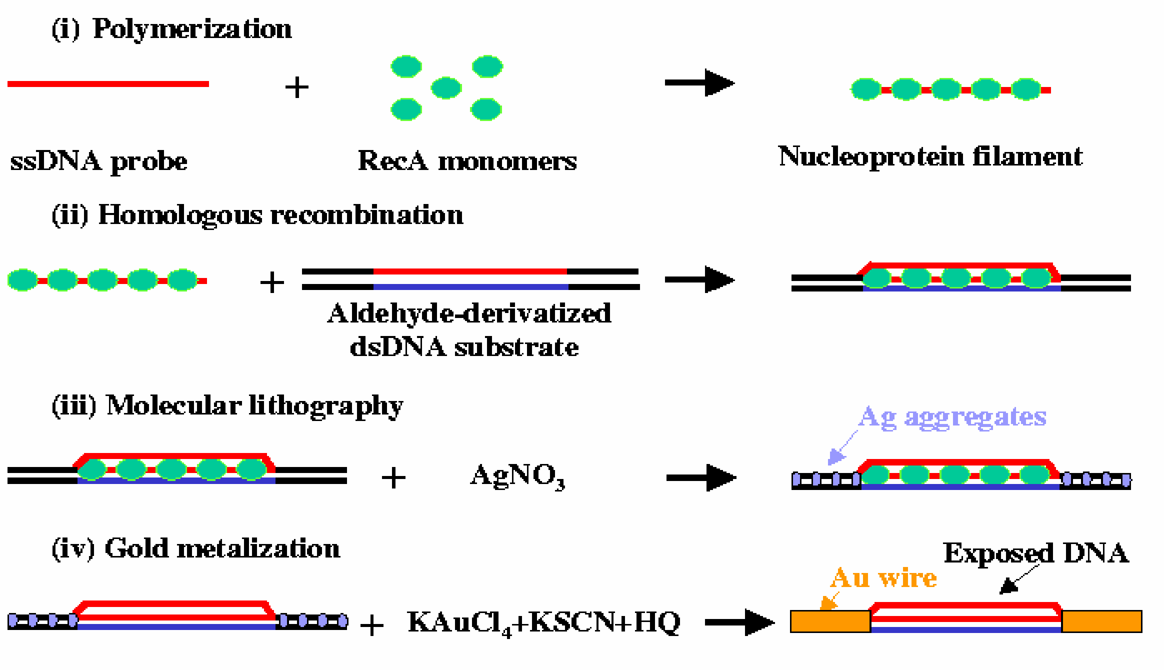
Sequence Specific Molecular Lithography Using RecA Protein

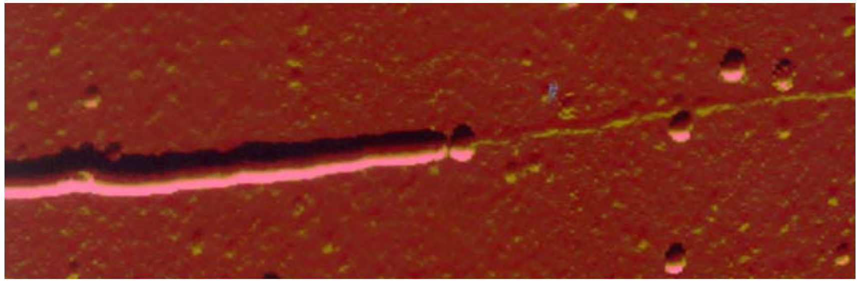
K.Keren, M.Krueger, R.Gilad,
G. Ben-Yoseph, U.Sivan &
E.Braun, *Science* 2002

Photo resist ↔ RecA Protein

Masks ↔ DNA sequence

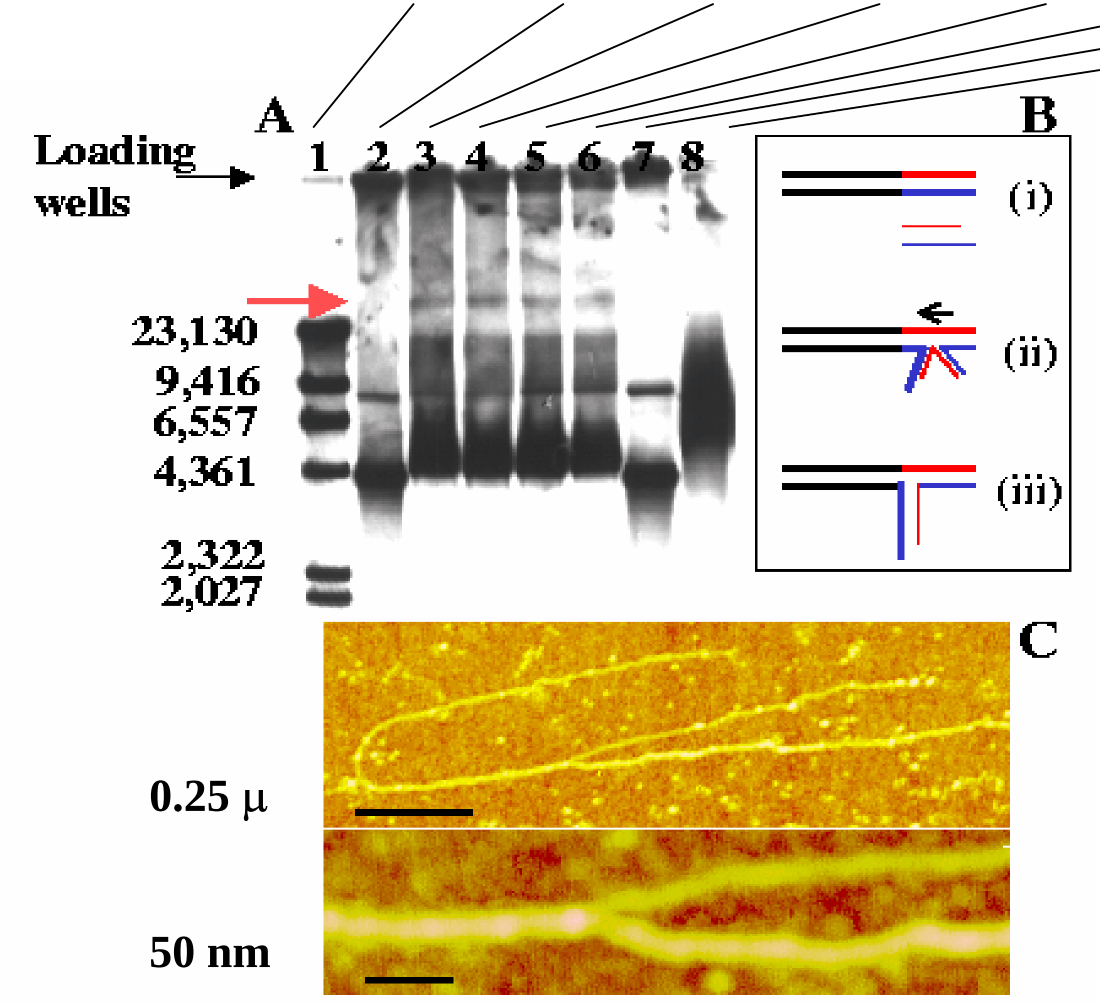


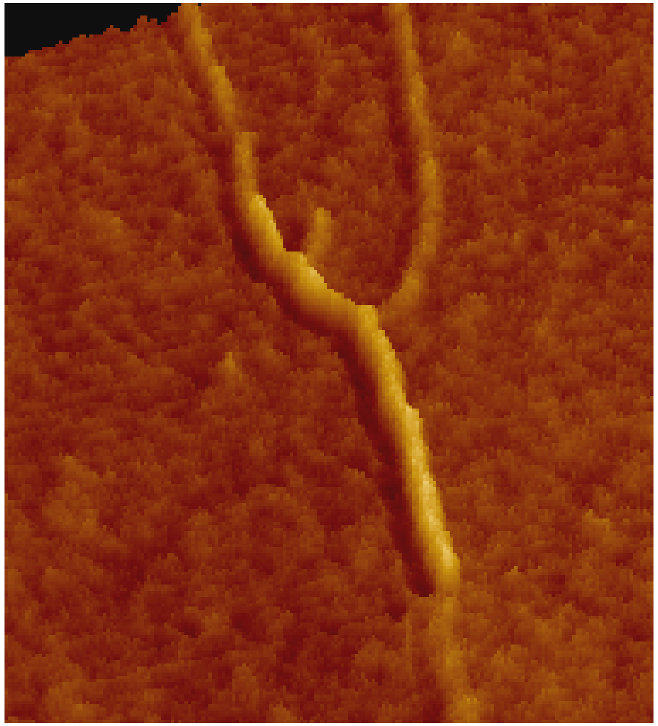




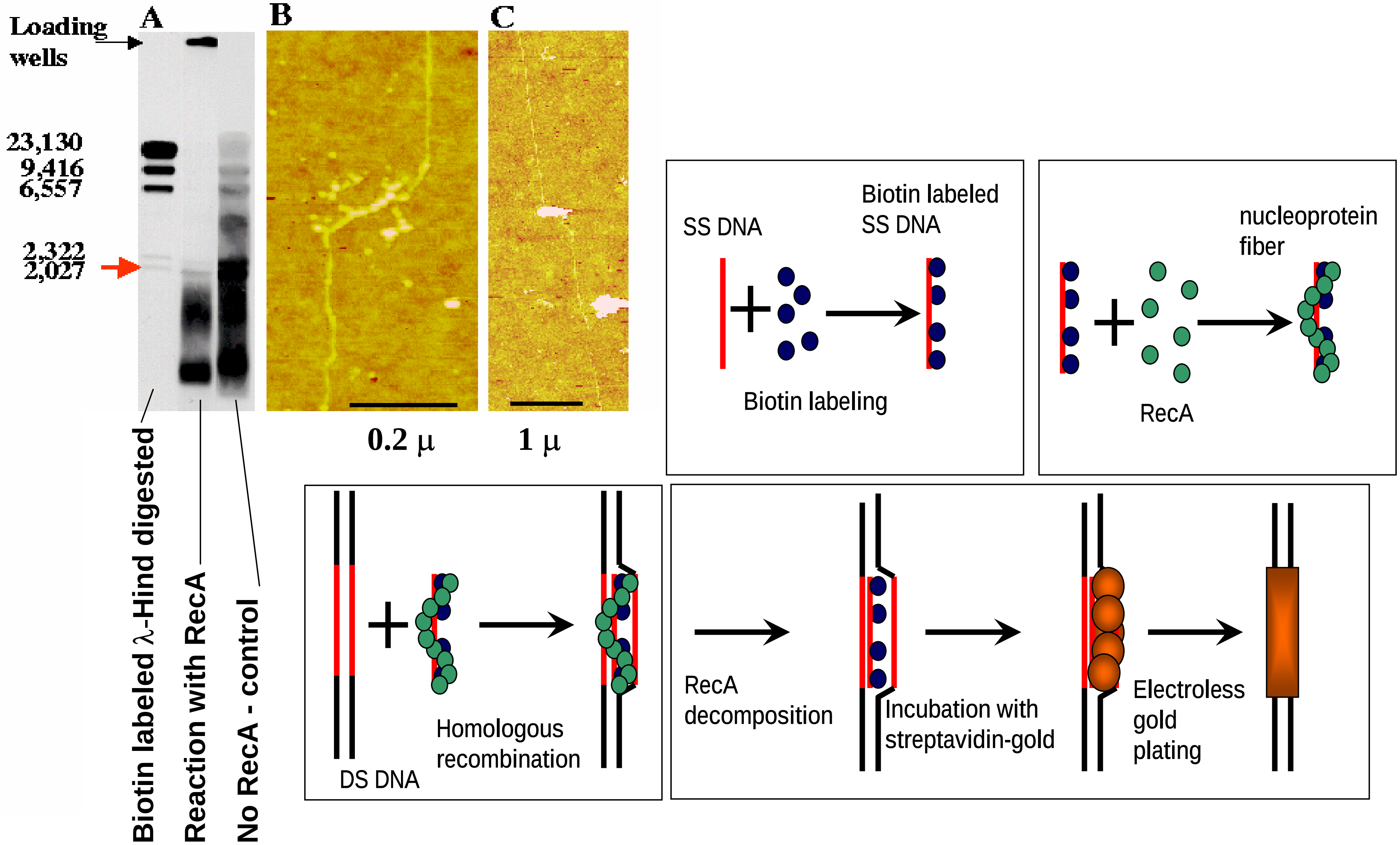
RecA as a Sequence Specific Junction Generator

ATP		+	+	+	+	+	+	-
RecA		-	+	+	+	+	-	+
Time	M	0	30	60	120	210	210	210





RecA as a Universal Molecular Assembler



SS DNA

Biotin labeled SS DNA

Biotin labeling

nucleoprotein fiber

RecA

DS DNA

Homologous recombination

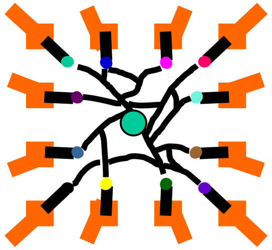
RecA decomposition

Incubation with streptavidin-gold

Electroless gold plating

Self-assembled Carbon Nanotube FET

K.Keren, R. Berman, E. Buchstab, U.Sivan, & E.Braun, *Science* 2003



(i) Choose a site for the FET on dsDNA, synthesize ssDNA with the same sequence, and polymerize RecA on the ssDNA

ssDNA probe + **RecA monomers** → **Nucleoprotein filament**

(ii) Add to dsDNA scaffold molecules

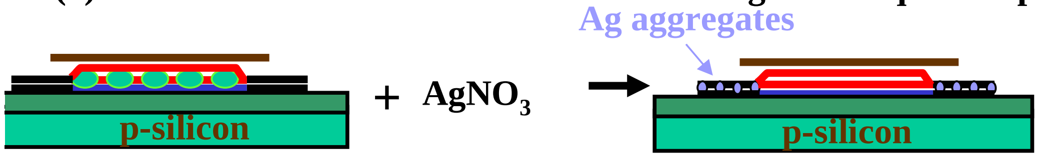
Aldehyde-derivatized dsDNA substrate

(iii) Localize a carbon nanotube using antibodies

Streptavidin coated carbon nanotube

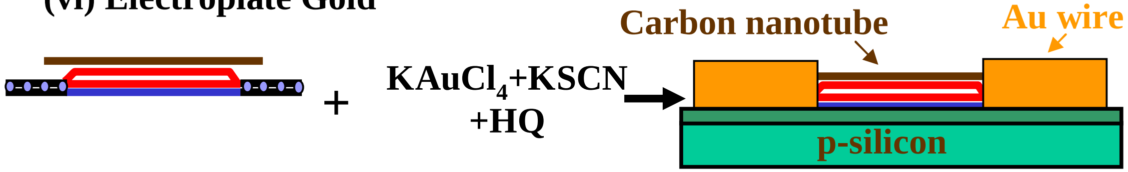
(iv) Stretch the complex on oxidized p-silicon substrate

(v) Grow silver on DNA with the RecA servings as a sequence specific resist

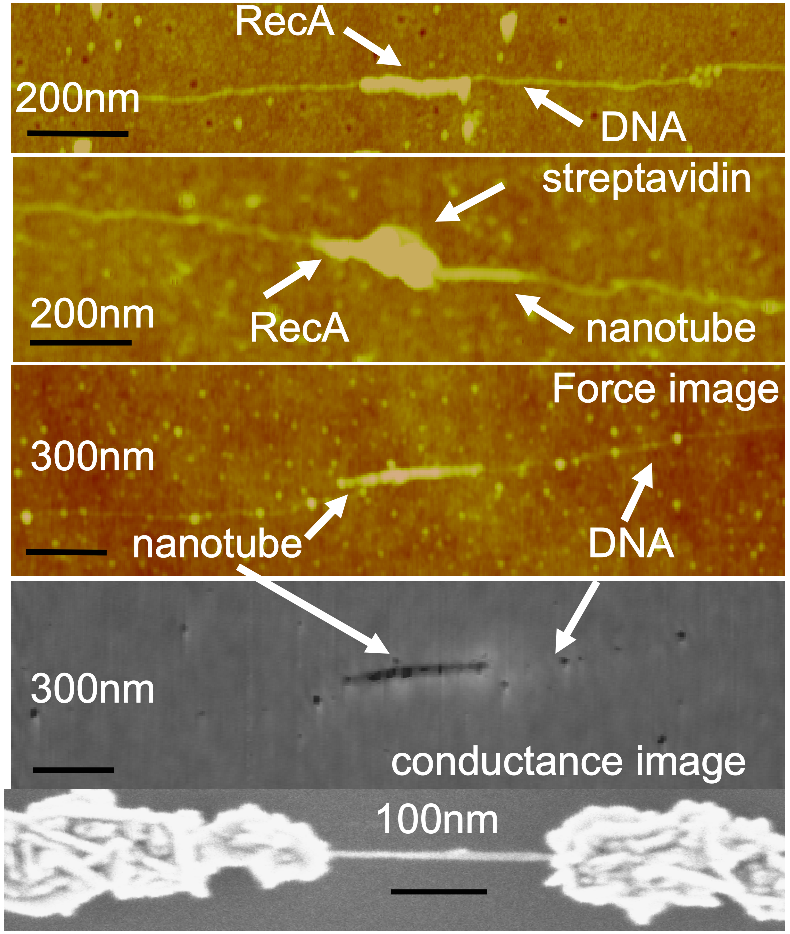


+ $AgNO_3$

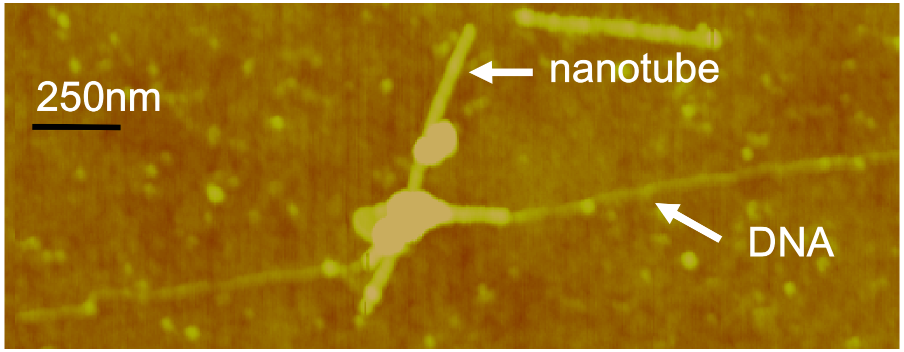
(vi) Electroplate Gold



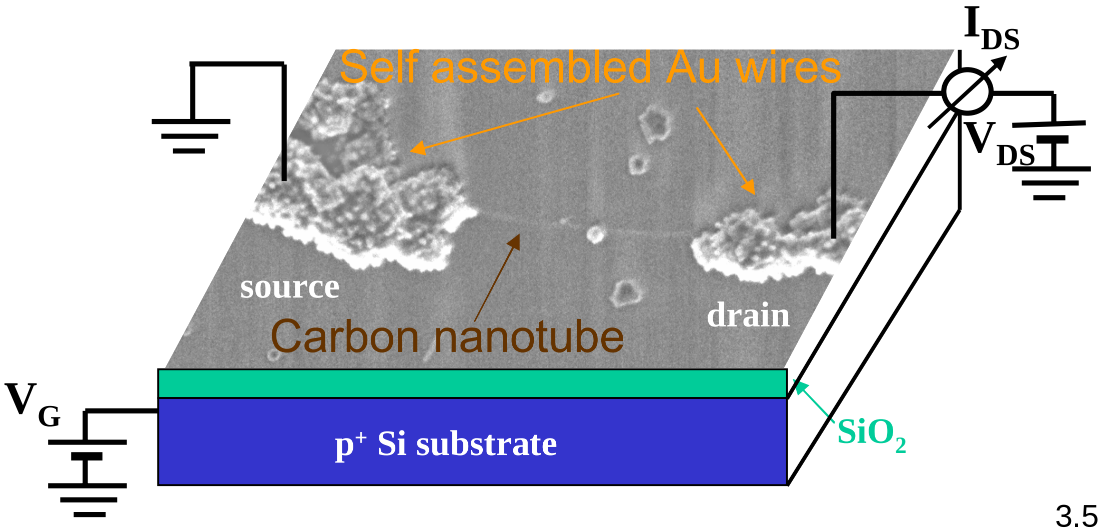
+ $KAuCl_4$+KSCN +HQ

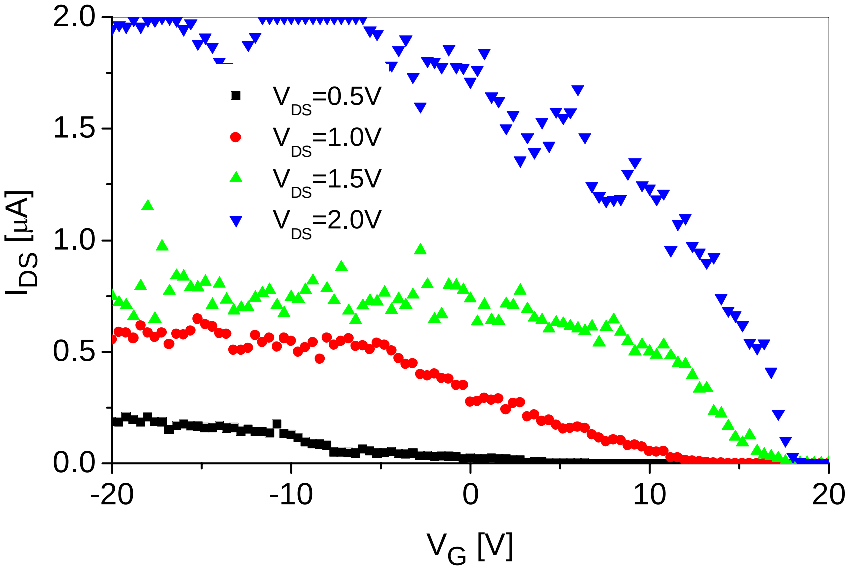


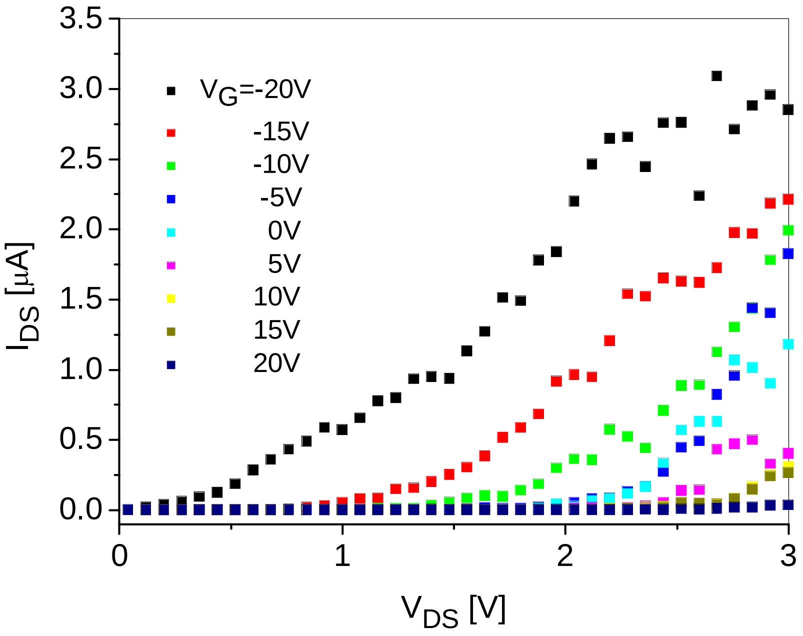
Sometimes…



Electrical Characteristics of Self-assembled Field Effect Transistors

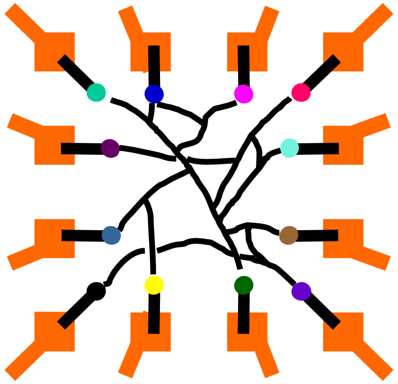






HIGHER COMPLEXITY REQUIRES ELABORATE DNA SCAFOLDS

CHALLENGE - DEVICE AN AUTONOMOUS DNA SYNTHESIZER SUCH THAT



(a) The synthesizer lends itself to the generation of a large variety of sequences.

(b) The number of distinct addresses along each generated sequence is large.

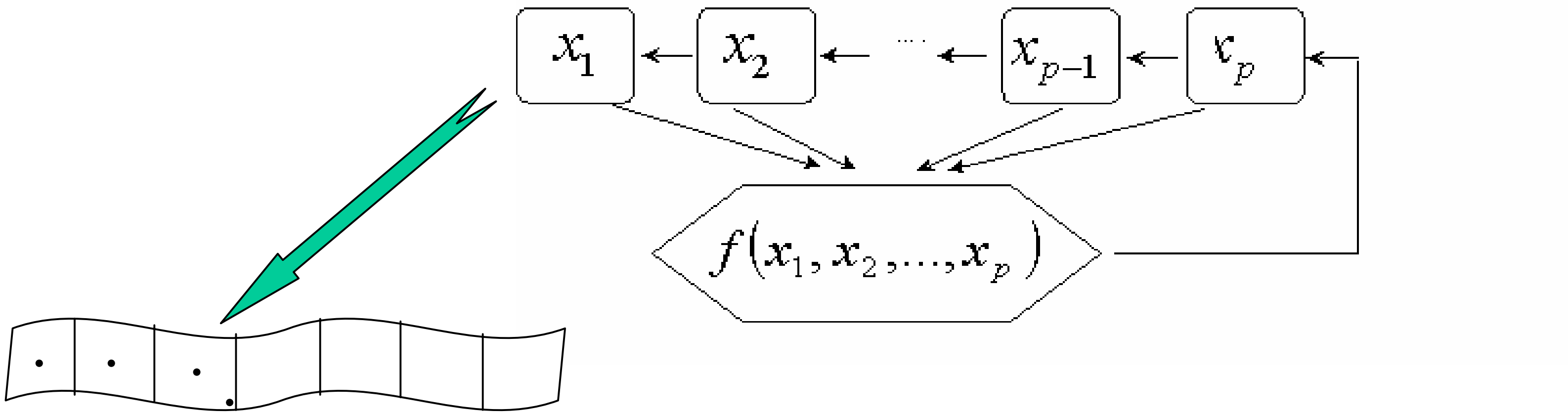
(c) The sequence is fully known

(d) Each address longer than a given length appears only once per certain DNA length.

(e) The synthesis effort is exponentially small compared with direct synthesis of all addresses.

- At first sight sounds impossible from the informational point of view
- Not true – just pseudo random number generator

Autonomous Binary *p*-Shift Register



- A computing machine with 2^p internal states represented by an array of *p* cells, each occupying one bit.
- In each step a binary function, *f*, is computed and its value is inserted into cell *p*.
- Simultaneously, the content of all cells is shifted one cell to the left.
- On printing x_1 to a tape, a long periodic binary sequence is generated.
- The generated sequence is uniquely determined by *f* and the seed.

Maximal Linear *p*-Shift Register

Example - 3-shift register following the rule $f(x_1, x_2, x_3) = x_1 \oplus x_3$

0011101001110100…..

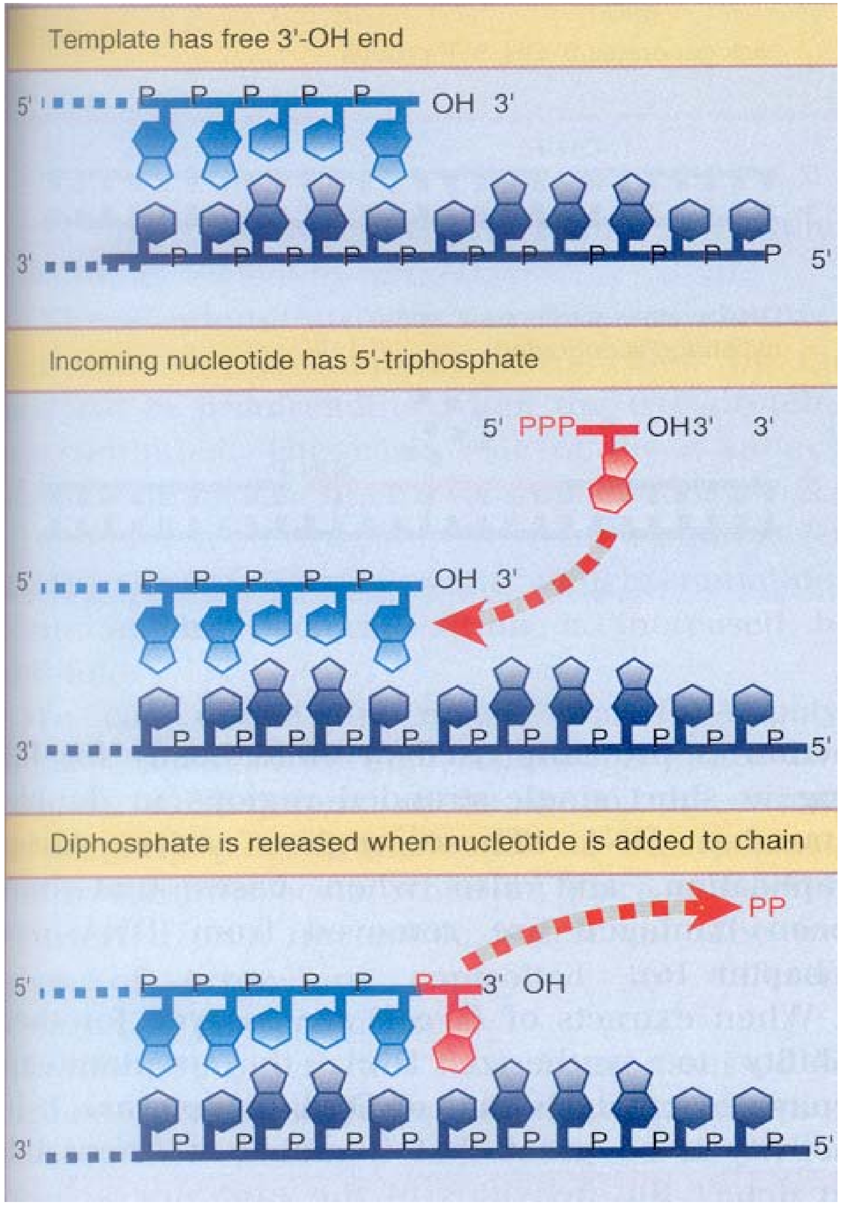
- 7 bit period
- Each string longer than 3 bits appears exactly once per period

x_1	x_3	f	Rule Strand
0	0	0	$\overline{0}\,\overline{1}\,\overline{0}\,\overline{0}$
0	1	1	$\overline{0}\,\overline{0}\,\overline{1}\,\overline{1}$ $\overline{0}\,\overline{1}\,\overline{1}\,\overline{1}$
1	0	1	$\overline{1}\,\overline{0}\,\overline{0}\,\overline{1}$ $\overline{1}\,\overline{1}\,\overline{0}\,\overline{1}$
1	1	0	$\overline{1}\,\overline{0}\,\overline{1}\,\overline{0}$ $\overline{1}\,\overline{1}\,\overline{1}\,\overline{0}$

Generally - for a linear *p*-shift register $x_{p+1} = \sum_1^p \alpha_j x_j \quad \alpha_j \in \{0,1\}$

- 2^p-1 bit period
- Each string longer than *p* bits appears exactly once per period
- Rules can be found such that the number of non-vanishing α_j is significantly smaller than *p* (truth table dimension << *p*)
- Consequently, the number of rules is exponentially smaller than the number of generated addresses !

COPYING DNA



- •DNA is copied with the help of an enzyme - DNA polymerase
- •Complementary nucleotide is added to the 3’-OH end of the growing chain, so that the new chain is synthesized in the 5’ to 3’ direction
- •The precursor for DNA synthesis is a nucleoside triphosphate, which looses the terminal two phosphate groups in the reaction

DNA Based Molecular *p*-Shift Register

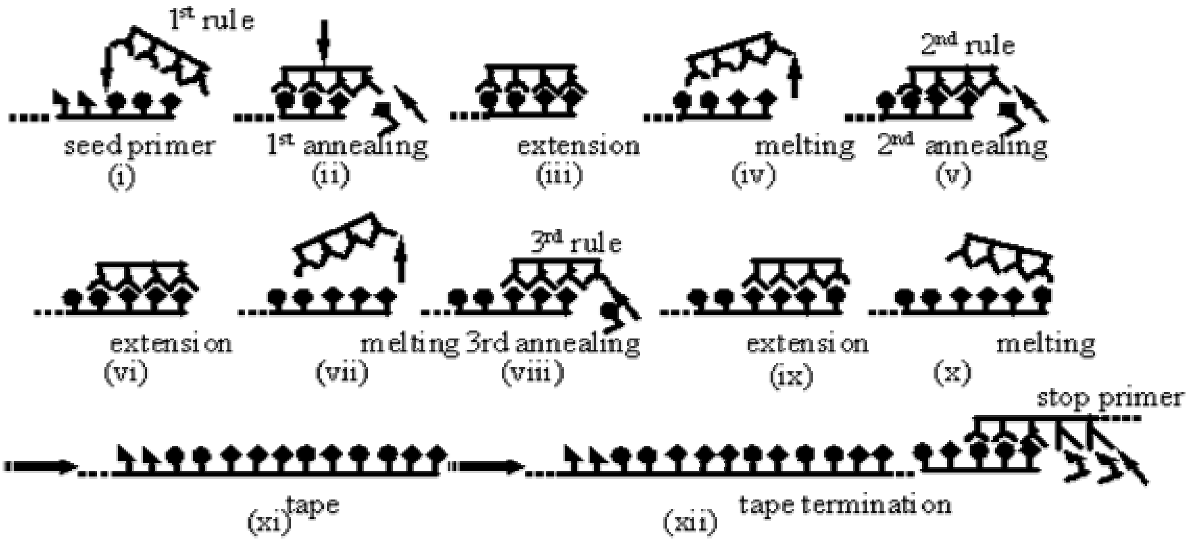
Consider a Boolean DNA with 4 “bases” $1, \overline{1}, 0, \overline{0}$

1 binds $\overline{1}$ but not $0, \overline{0}$

0 binds $\overline{0}$ but not $1, \overline{1}$

Realize the function *f* with 7 rule strands. Add a seed strand and polymerase. Cycle thermally. Terminate with a stop strand.

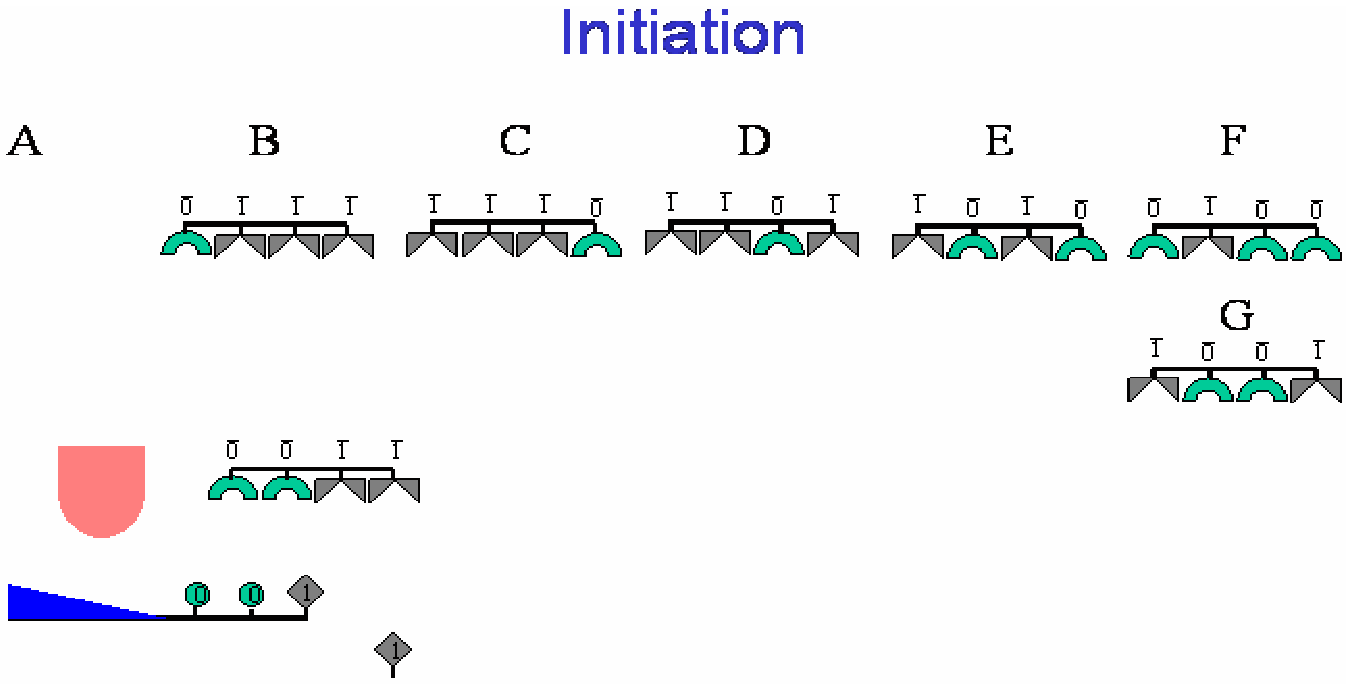
x_1	x_3	f	Rule Strand
0	0	0	$\overline{0}\,\overline{1}\,\overline{0}\,\overline{0}$
0	1	1	$\overline{0}\,\overline{0}\,\overline{1}\,\overline{1}$ $\overline{0}\,\overline{1}\,\overline{1}\,\overline{1}$
1	0	1	$\overline{1}\,\overline{0}\,\overline{0}\,\overline{1}$ $\overline{1}\,\overline{1}\,\overline{0}\,\overline{1}$
1	1	0	$\overline{1}\,\overline{0}\,\overline{1}\,\overline{0}$ $\overline{1}\,\overline{1}\,\overline{1}\,\overline{0}$



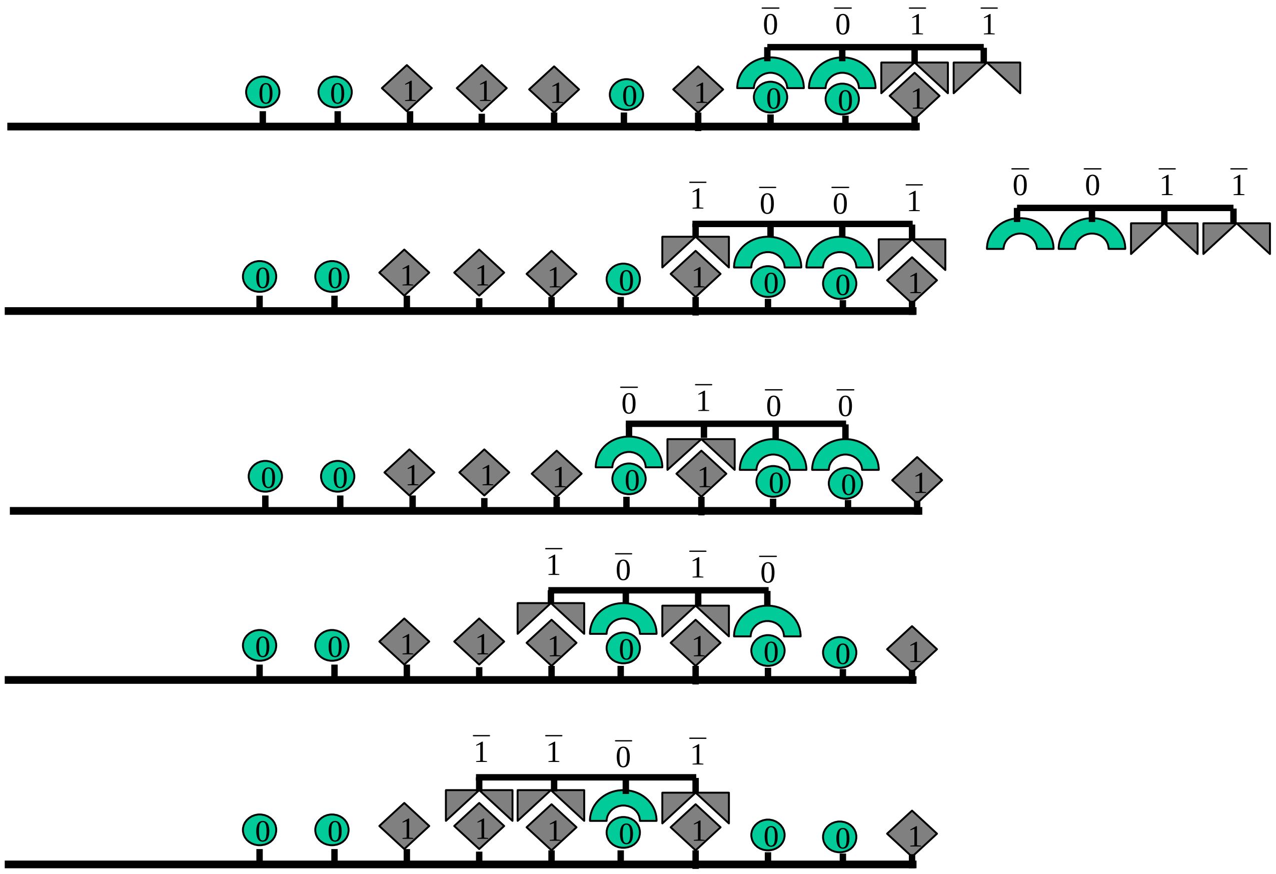
◆ 1
⌃ $\overline{1}$
● 0
⌒ $\overline{0}$
◣ sequences other than $0,1$
↖ sequences other than $\overline{0}, \overline{1}$

DNA Based Molecular *p*-Shift Register

- Works also in a thermal ratchet mode at a fixed temperature
- Rule strands function as enzymes. They direct the reaction but not consumed



Competing Blocking Processes



Correct synthesis

blocking

synthesis

Reaction proceeds through thermal fluctuations (ratchet)

3-shift register realized in 5-bit space

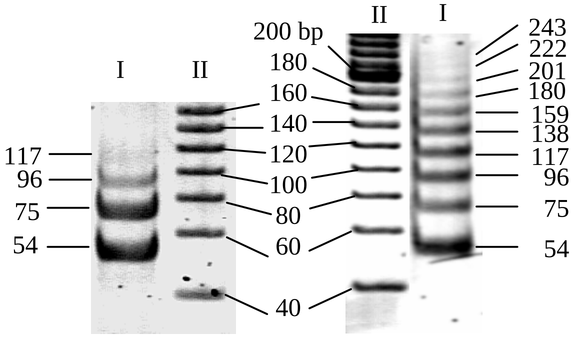
$x_{n+1} = x_n \oplus x_{n-2}$ 7 bit=21 base period

5'GCATGCGCCC GTCAGGCG 00111 Seed strand

$$\left.\begin{array}{l} 3'\,\bar{0}\,\bar{0}\,\bar{1}\,\bar{1}\,\bar{1}\,\bar{0} \\ 3'\,\bar{0}\,\bar{1}\,\bar{1}\,\bar{1}\,\bar{0}\,\bar{1} \\ 3'\,\bar{1}\,\bar{1}\,\bar{1}\,\bar{0}\,\bar{1}\,\bar{0} \\ 3'\,\bar{1}\,\bar{1}\,\bar{0}\,\bar{1}\,\bar{0}\,\bar{0} \\ 3'\,\bar{1}\,\bar{0}\,\bar{1}\,\bar{0}\,\bar{0}\,\bar{1} \\ 3'\,\bar{0}\,\bar{1}\,\bar{0}\,\bar{0}\,\bar{1}\,\bar{1} \\ 3'\,\bar{1}\,\bar{0}\,\bar{0}\,\bar{1}\,\bar{1}\,\bar{1} \end{array}\right\}\text{7 rule strands}$$

$3'\,\bar{0}\,\bar{1}\,\bar{0}\,\bar{0}\,\bar{1}\,\mathrm{GACGTC}$ stop strand

$$\boxed{\begin{array}{ll} 0 = 5'\mathrm{TGC} & 1 = 5'\mathrm{GCT} \\ \bar{0} = 3'\mathrm{ACG} & \bar{1} = 3'\mathrm{CGA} \end{array}}$$



$$5'\,\mathrm{GCATGCGCCC}\;\;\underbrace{\mathrm{GTCAGGCG00}}_{\text{seed primer}}\;\;\underset{\hookleftarrow}{111}\left(0100111\right)_n\underbrace{01001\mathrm{CTGCA}\;\;\mathrm{G}}_{\mapsto\ \text{complementary to stop primer}}\quad\text{with } n = 0,1,\ldots$$

Confirmed by Sequencing!

4-shift register realized in 6-bit space

$$x_{n+1} = x_n \oplus x_{n-3}$$

15 bit=45 base period

Exponentially more addresses for the same synthesis effort!

5'GCA TGC GCC CGT CAG GCG 001111 seed strand

$3'\,\overline{0}\,\overline{0}\,\overline{1}\,\overline{1}\,\overline{1}\,\overline{1}\,\overline{0}$

$3'\,\overline{0}\,\overline{1}\,\overline{1}\,\overline{1}\,\overline{1}\,\overline{0}\,\overline{1}$

$3'\,\overline{1}\,\overline{1}\,\overline{1}\,\overline{1}\,\overline{0}\,\overline{1}\,\overline{0}$

$3'\,\overline{1}\,\overline{1}\,\overline{1}\,\overline{0}\,\overline{1}\,\overline{0}\,\overline{1}$

$3'\,\overline{1}\,\overline{1}\,\overline{0}\,\overline{1}\,\overline{0}\,\overline{1}\,\overline{1}$

$3'\,\overline{1}\,\overline{0}\,\overline{1}\,\overline{0}\,\overline{1}\,\overline{1}\,\overline{0}$

$3'\,\overline{0}\,\overline{1}\,\overline{0}\,\overline{1}\,\overline{1}\,\overline{0}\,\overline{0}$

$3'\,\overline{1}\,\overline{0}\,\overline{1}\,\overline{1}\,\overline{0}\,\overline{0}\,\overline{1}$

$3'\,\overline{0}\,\overline{1}\,\overline{1}\,\overline{0}\,\overline{0}\,\overline{1}\,\overline{0}$

$3'\,\overline{1}\,\overline{1}\,\overline{0}\,\overline{0}\,\overline{1}\,\overline{0}\,\overline{0}$

$3'\,\overline{1}\,\overline{0}\,\overline{0}\,\overline{1}\,\overline{0}\,\overline{0}\,\overline{0}$

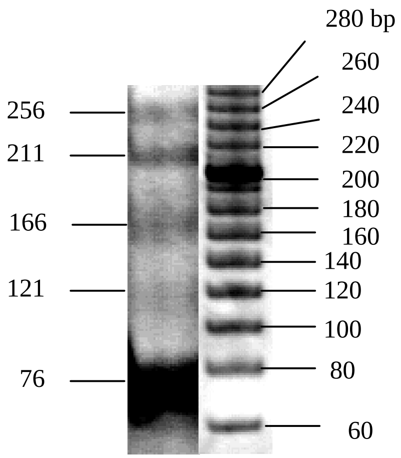
$3'\,\overline{0}\,\overline{0}\,\overline{1}\,\overline{0}\,\overline{0}\,\overline{0}\,\overline{1}$

$3'\,\overline{0}\,\overline{1}\,\overline{0}\,\overline{0}\,\overline{0}\,\overline{1}\,\overline{1}$

$3'\,\overline{1}\,\overline{0}\,\overline{0}\,\overline{0}\,\overline{1}\,\overline{1}\,\overline{1}$

$3'\,\overline{0}\,\overline{0}\,\overline{0}\,\overline{1}\,\overline{1}\,\overline{1}\,\overline{1}$

15 rule strands



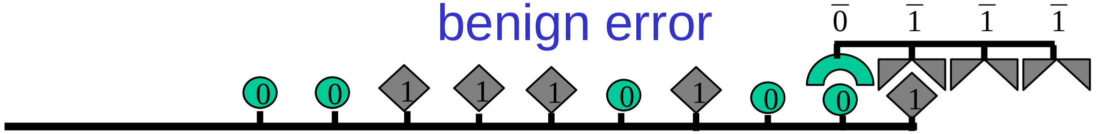
$3'\,\overline{1}\,\overline{0}\,\overline{1}\,\overline{1}\,\overline{0}\,\overline{0}$ GCGC CAG GAC GCG GAC GTC stop strand

$$0011110(1011001000\,11110)_n\,1011;\ n = 0,1,\ldots$$

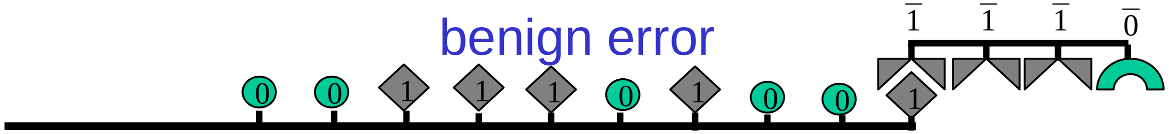
Confirmed by Sequencing!

Errors

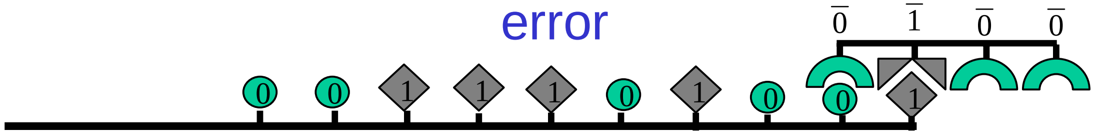
benign error



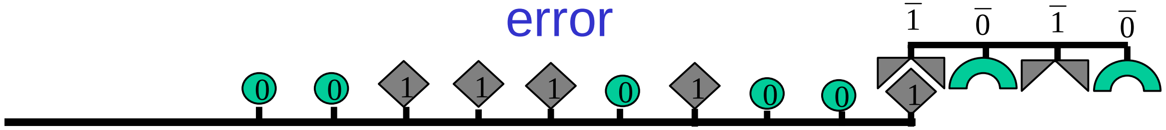
benign error



error



error

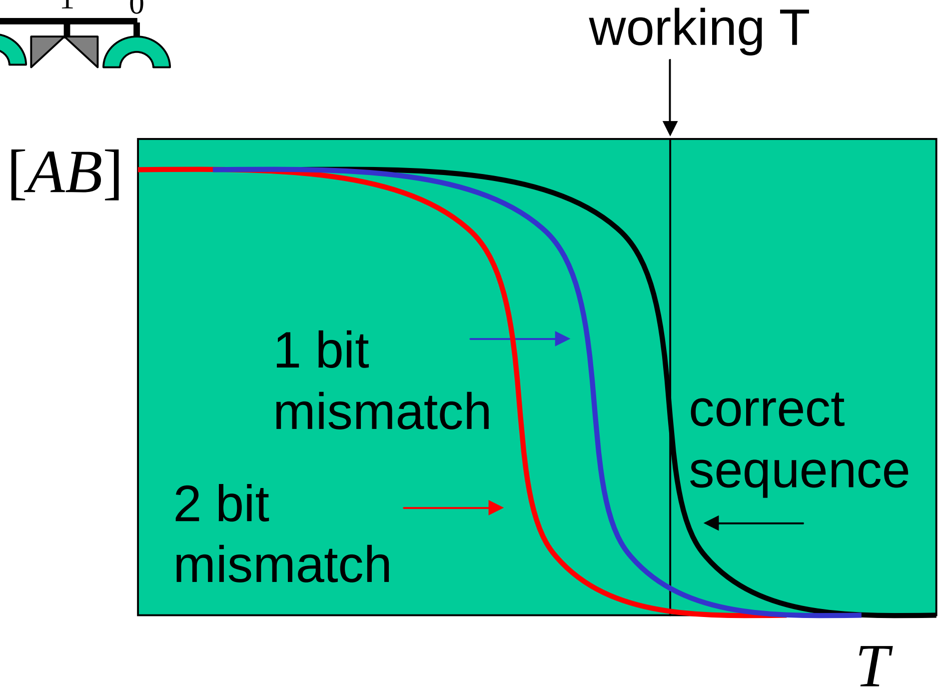


$$A + B \Leftrightarrow AB$$

$$\exp(-\Delta G/T) = \frac{[AB]}{[A][B]}$$

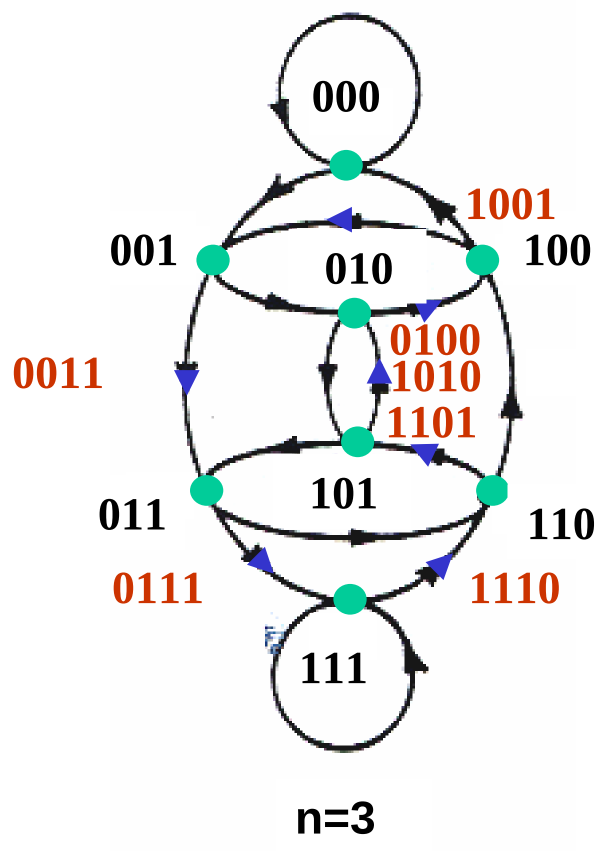
$$\text{if } [B_0] >> [A_0];\ \text{and } [AB] = [A_0] - [A]$$

$$[AB] = \frac{[A_0]}{1 + [B_0]^{-1}\exp(\Delta G/T)}$$



Shift Register Sequence is a Path on a de-Bruijn Graph

00111010



00111010011101 00…..

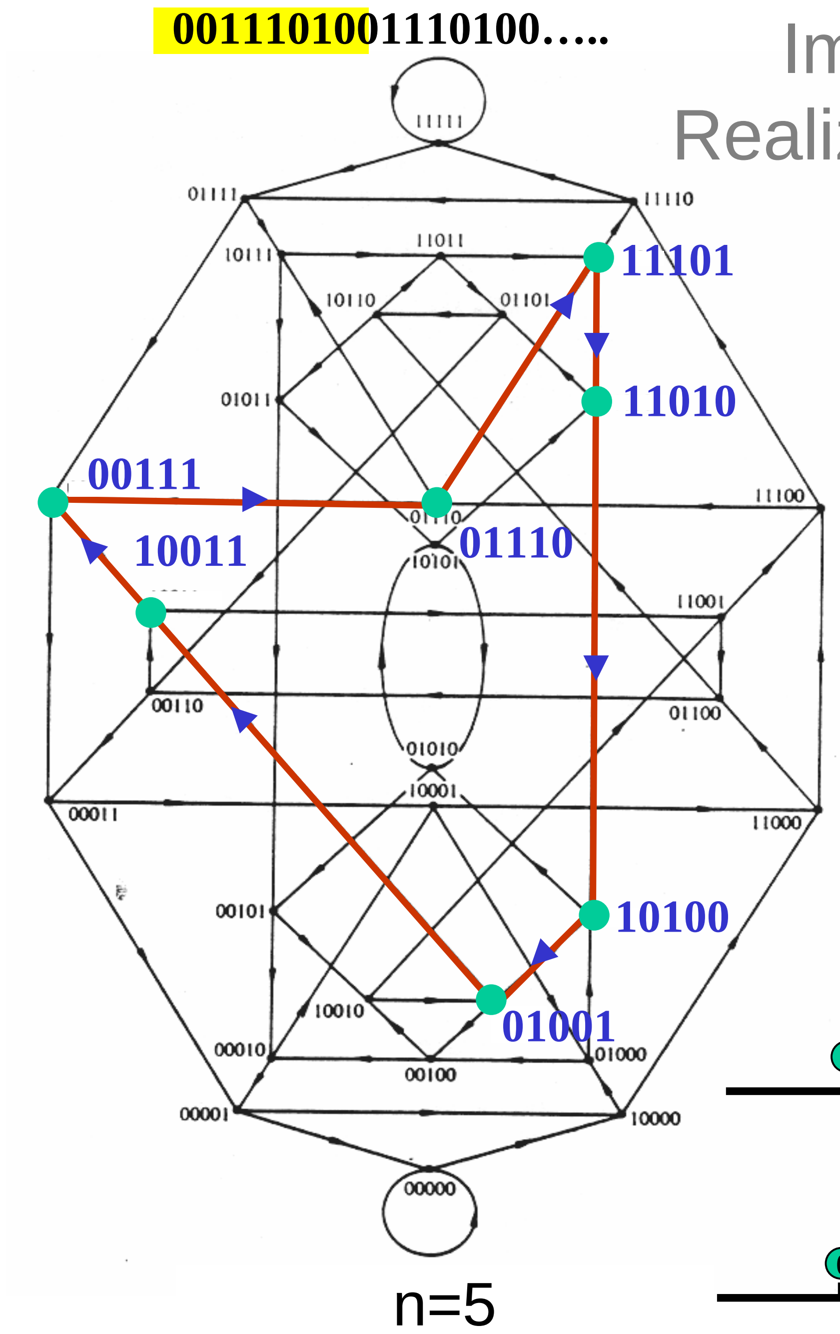
- Nodes correspond to machine states
- Lines correspond to transition rules

Prone to errors since

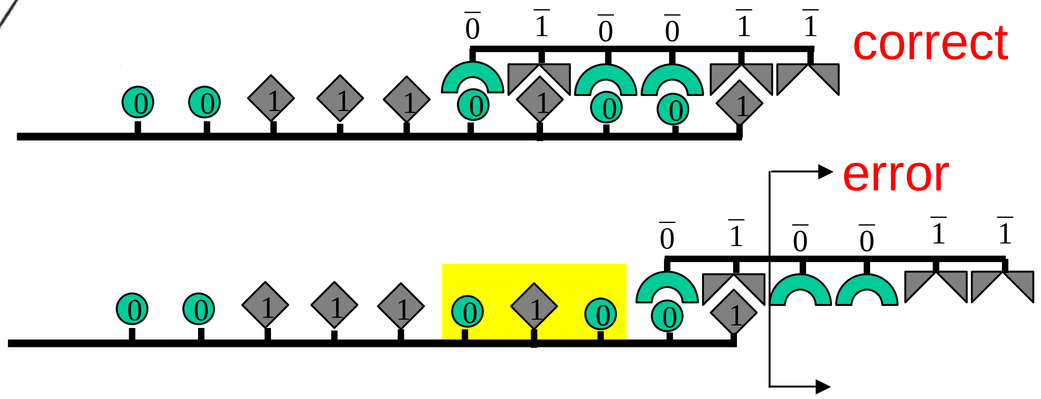
all nodes are legal, namely,

recognized by rule strands

Immunity Against Errors by Realization in a Higher Dimension



- Errors usually lead to a node which is not recognizable by any rule strand. Consequently further elongation requires a second error
- When a *p*-SR is realized with (*p*+*q*)-SR error requires *q*+1 mismatches. Consequently the errors are suppressed by $\exp\left[-(q+1)\Delta G/k_BT\right]$ where ΔG is the free energy associated with one base mismatch
- $\Delta G \approx 8.5 \div 10.5 k_BT$



Richer Alphabets

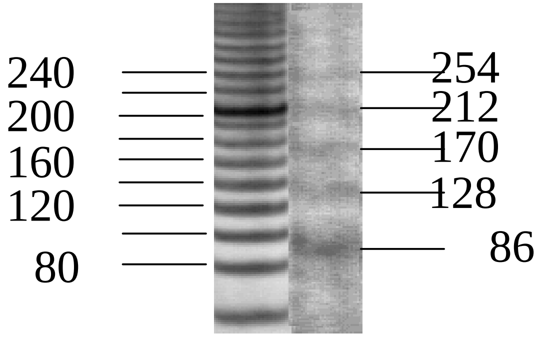
- Using 3 nucleotides for two letter alphabet is very inefficient
- Maximal alphabet includes $4^3=64$ letters
- Probably can't use such a large alphabet due to interference
- Optimal alphabet is probably in between
- Nature uses 3 nucleotide codons (albeit with a reading frame) to code 20 amino acids plus stop codons
- With k-letter alphabet the maximal shift register sequence measures k^p bits!

4-letter alphabet

3-shift register realized with 5+1 bits

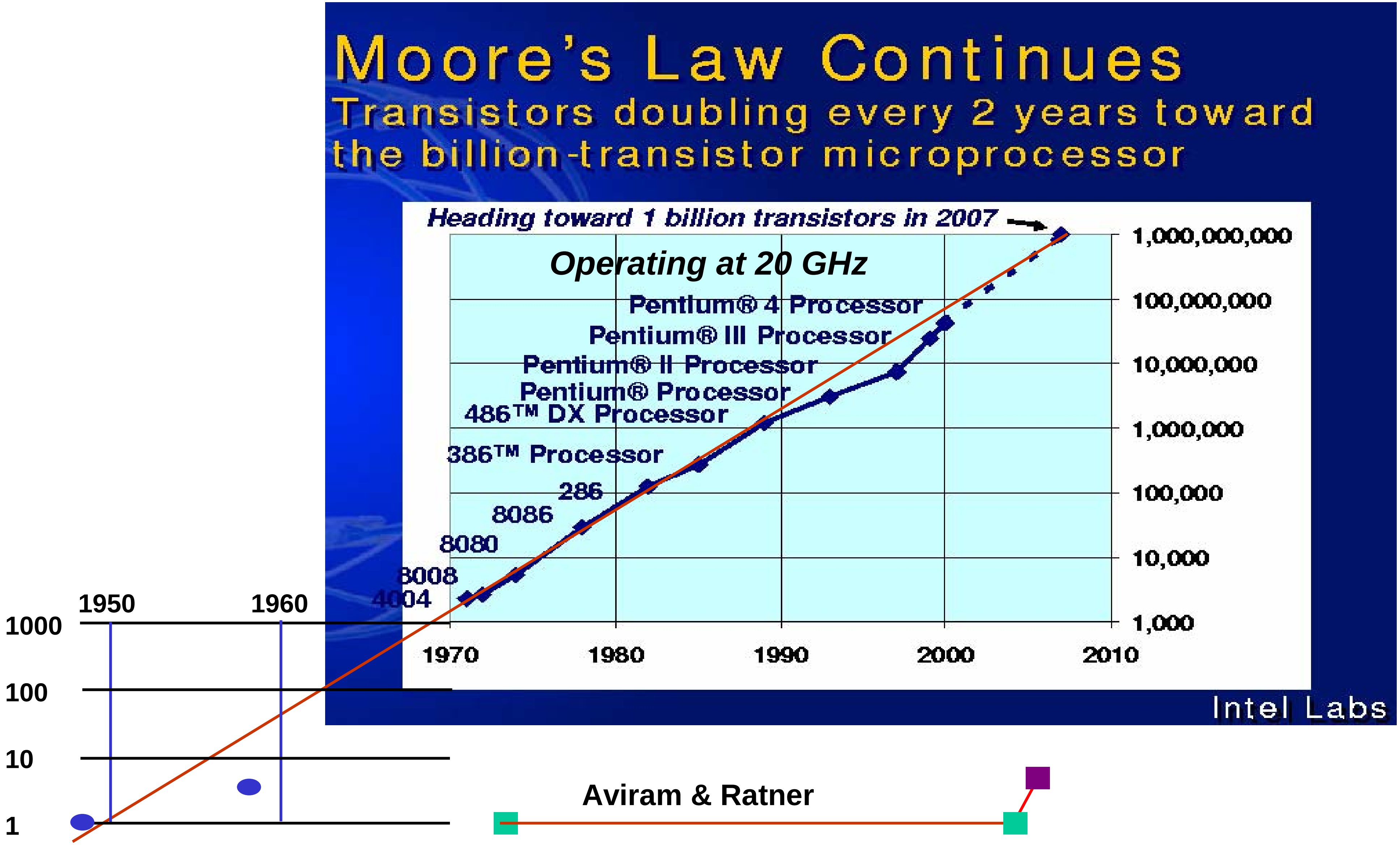
14 bit period (42 nucleotides)

0=ACC 1=CAG 2=CGA 3=GGA



$$03110\,(1223132320\ \ 3110\,)_n\,1223132320$$

CAN PROBABLY TAKE OUR TECHNOLOGY TO ~5 TRANSISTORS



Moore’s Law Continues

Transistors doubling every 2 years toward the billion-transistor microprocessor

COMPLEX SELF ASSEMBLY

✓ Requires hierarchical assembly

✓ Breaking to modules - what’s the optimal module size? Can’t be too large or too small

✓ Error reduction and correction on structural level

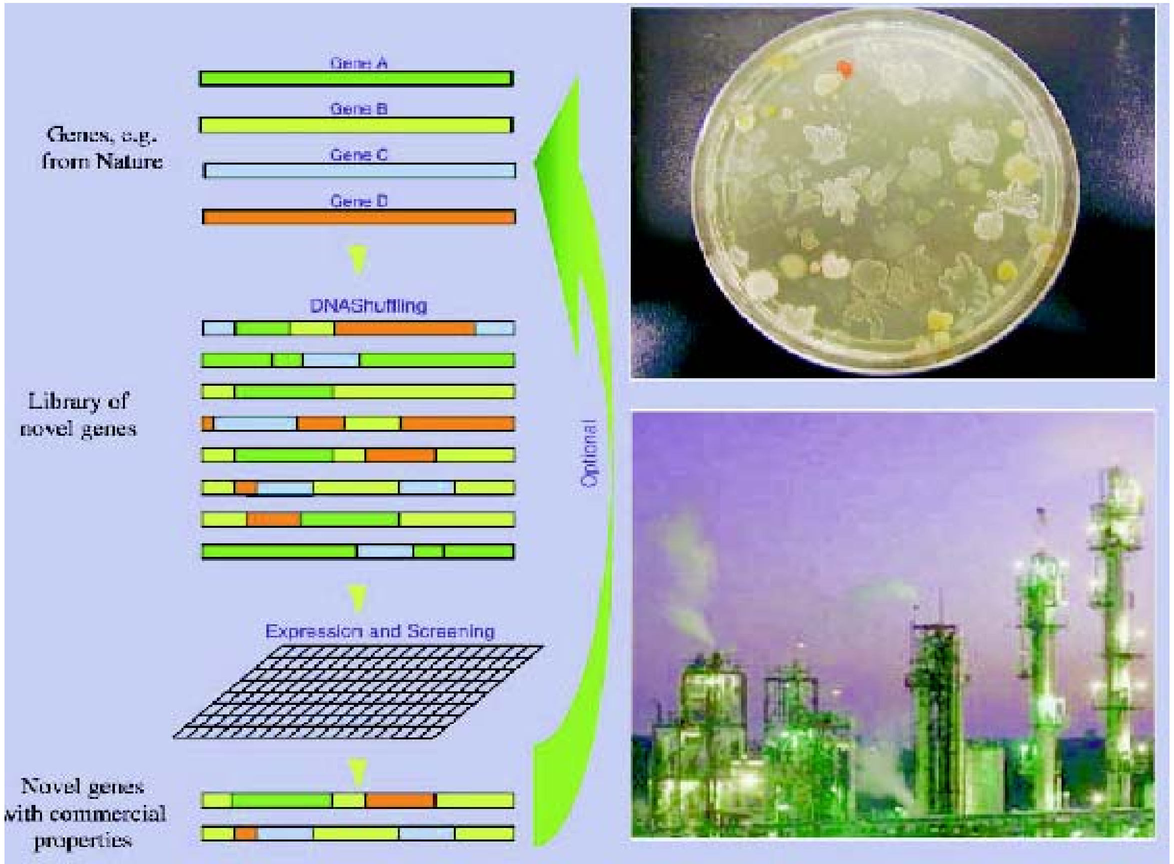
✓ Feedback loops and molecular switches from function to structure and from function to function

✓ Need further tools. How about recruiting the immune system?

DIRECTED EVOLUTION OF BIOCATALYSTS

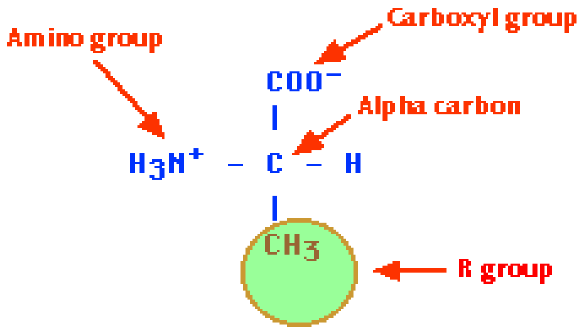
Powel et al. , Angew. Chem. Int. Ed. 40, 3948 (2001)

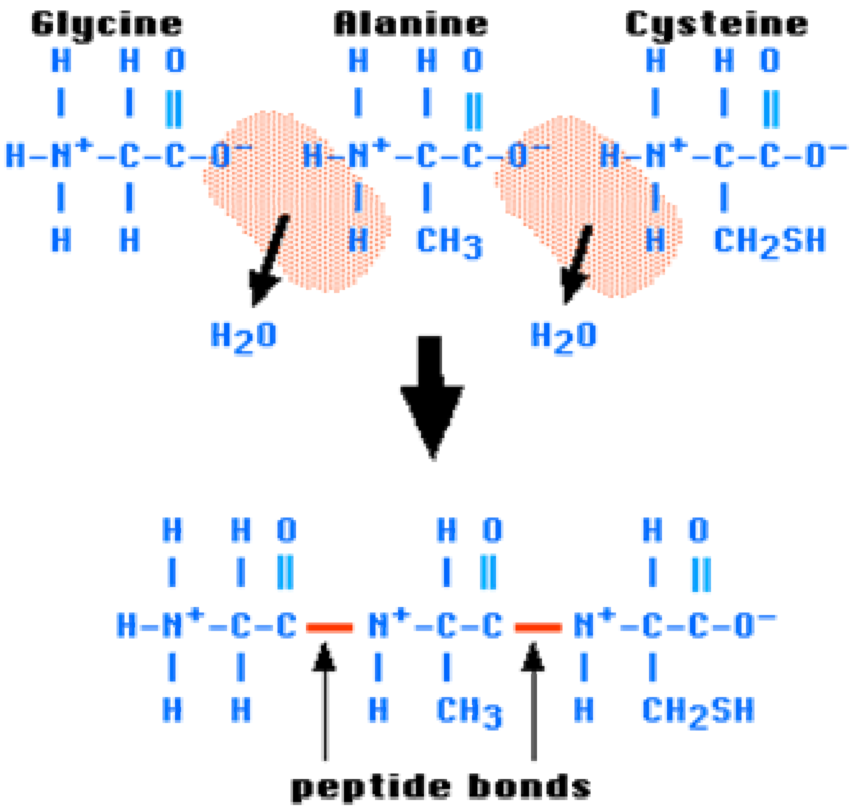
Directed evolution produces better enzymes for industrial applications without really understanding the microscopics.



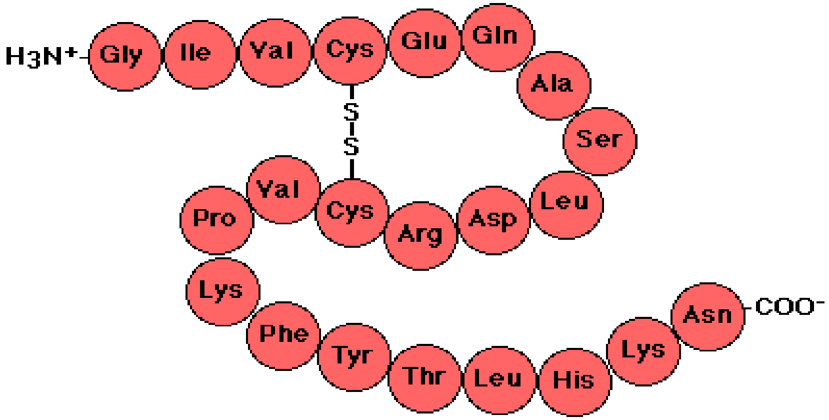
Amino Acids, Peptides, Proteins

- ✓ 20 amino acids in proteins
- ✓ More natural and artificial amino acids

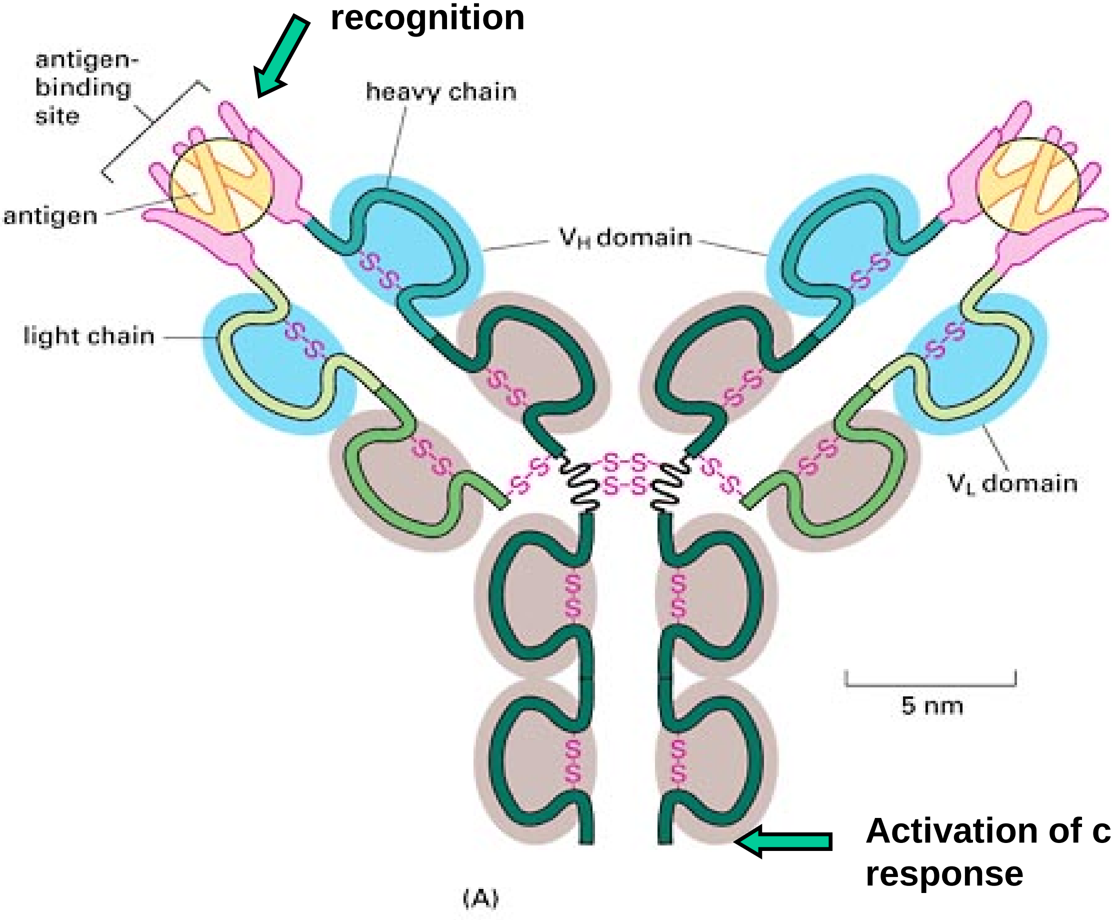




Unlike DNA, rich secondary and tertiary structure due to interaction between side groups


H_3N^+
Gly
Ile
Val
Cys
Glu
Gln
Ala
Ser
Leu
Asp
Arg
Cys
Val
Pro
Lys
Phe
Tyr
Thr
Leu
His
Lys
Asn
COO^-
S
S

Antibody Molecules

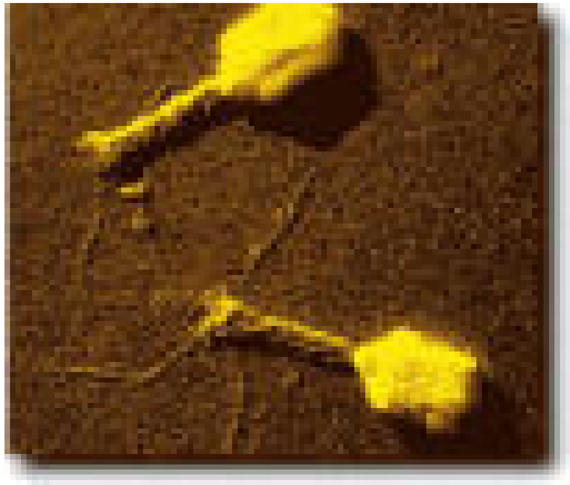


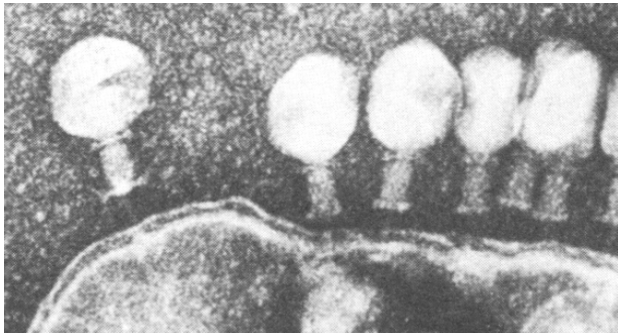
recognition

Activation of cell response

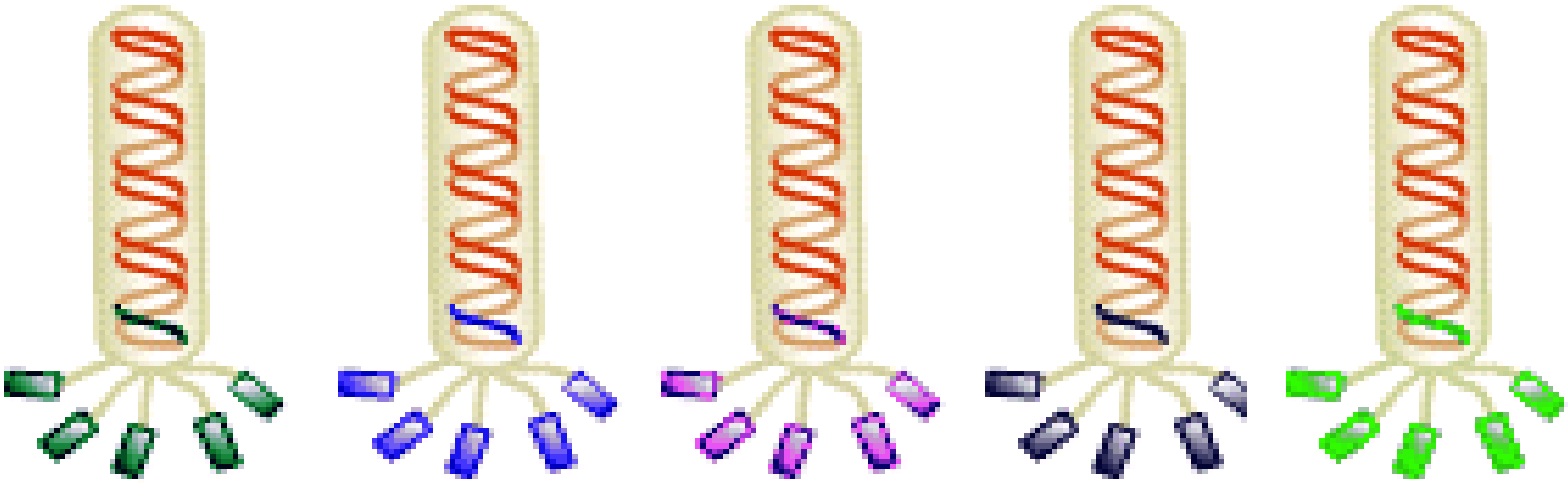
Phage

- Phage (bacteriophage) is a virus that infects bacteria.
- A phage is made out of two components: genetic material and a protein coat (capsid).
- The phage attaches to a specific receptor in the bacterial cell wall and injects its genome into the cell.
- After injection the virus proteins are expressed in the bacterium.
- In nature these proteins make new viruses. Eventually the bacterium is disrupted and the viruses disperse and attack further bacteria.
- If genes are added to the viral genome, any proteins can be added to e.g. the viral coat proteins.
- This is the basis of phage display.
- The proteins can be antibody fragments.

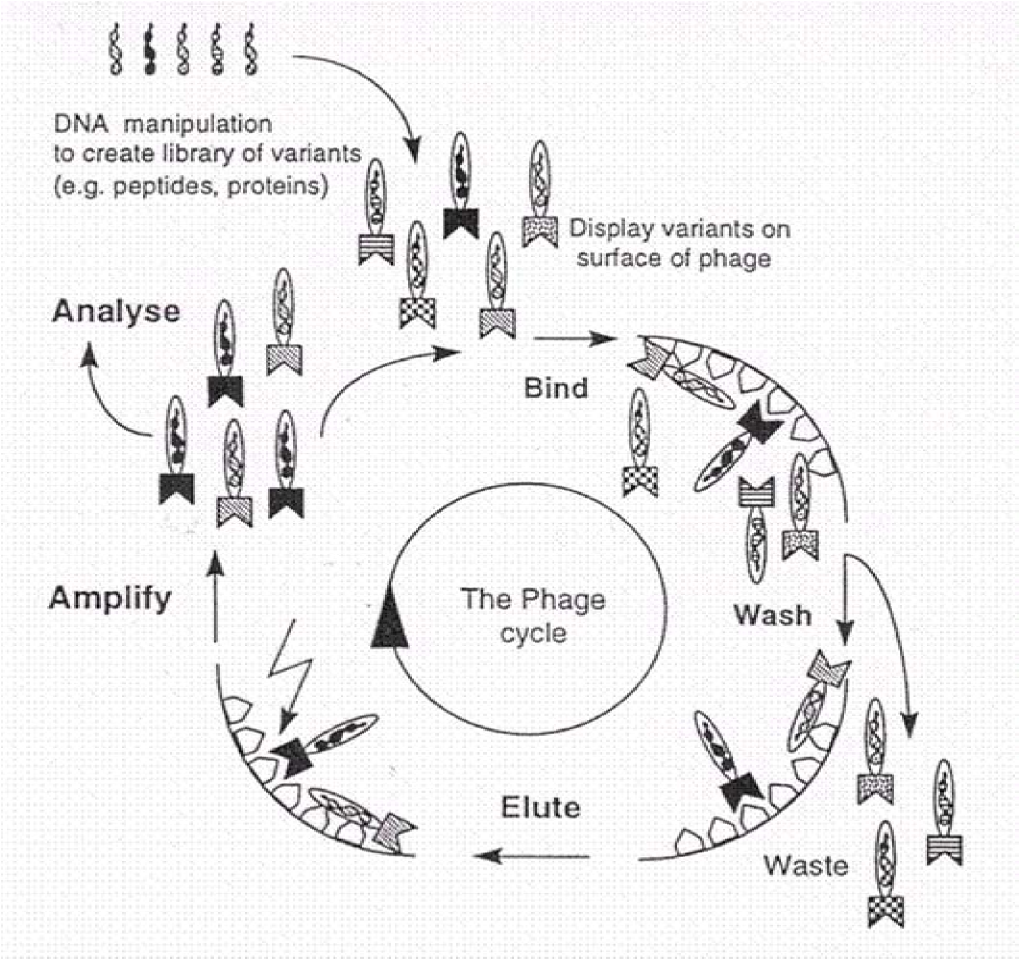


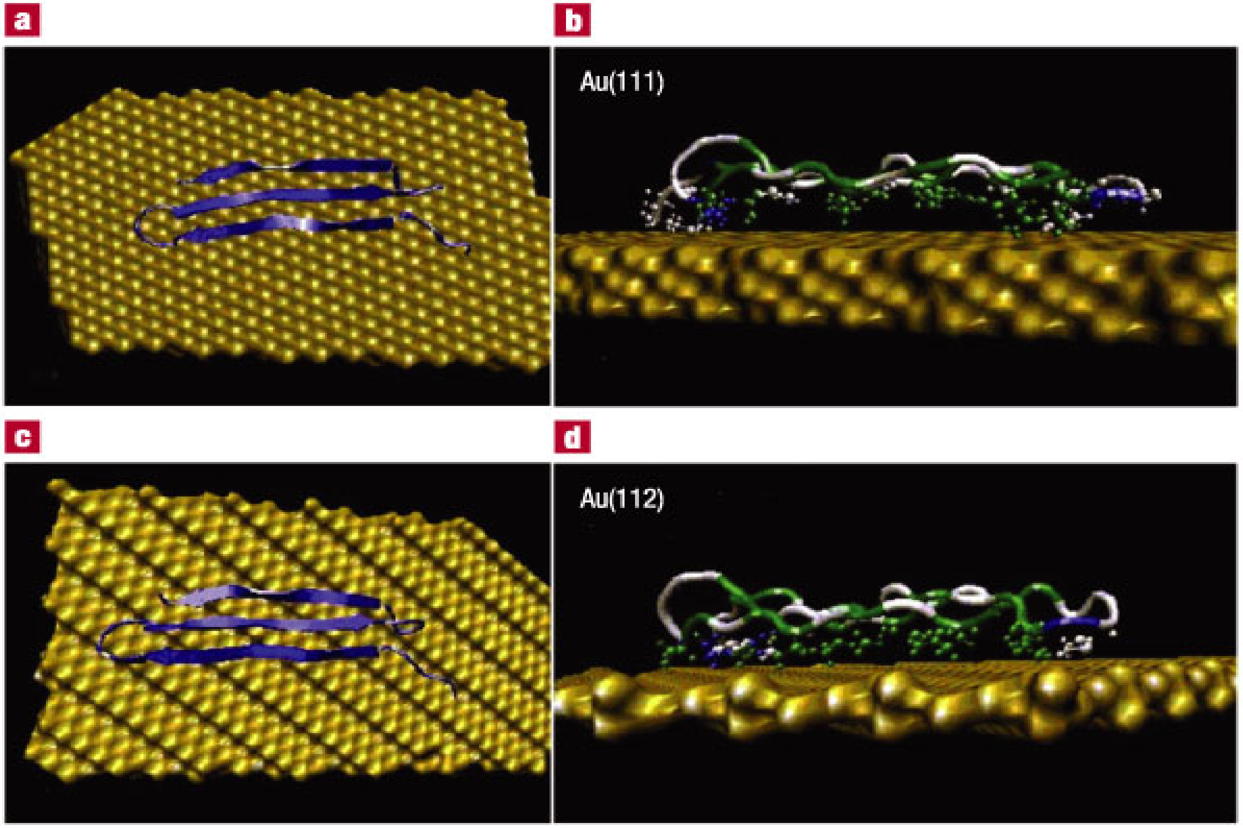


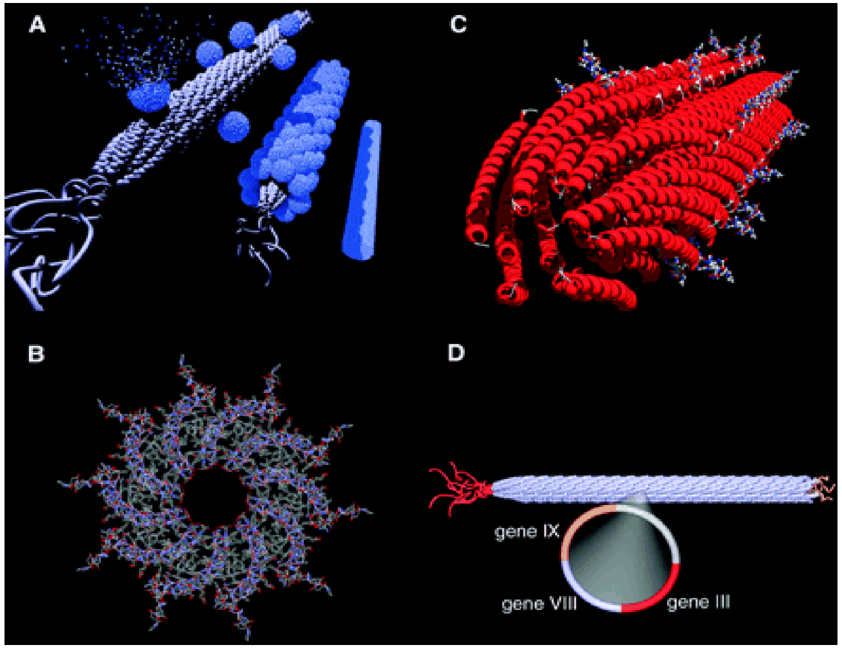
Antibody Library displayed on Phages



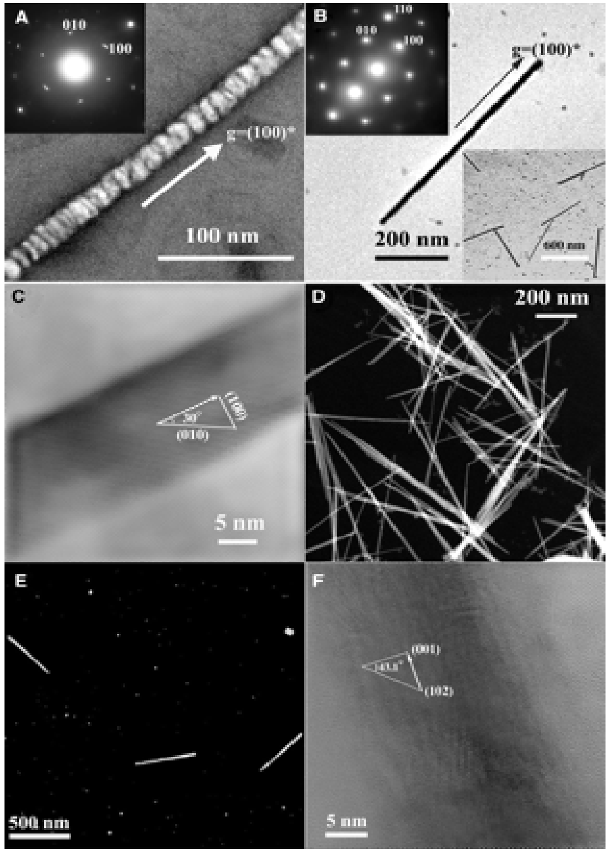
Antibody Selection by Phage Display




a
b
Au(111)
c
d
Au(112)



Belcher and collaborators 2003, 2004



The Challenge

GENERATE ANTIBODIES THAT RECOGNIZE VARIOUS ELECTRONIC COMPONENTS. USE THEM AS MOLECULAR SWITCHES TO GO FROM ONE CONSTRUCTION HIERARCHY TO THE NEXT

GRAND CHALLENGES

✓ Metallization must be replaced with biologically compatible process

✓ Error reduction and correction on structural level

✓ Need more than one energy scale - fixation

✓ Breaking to modules - what’s the optimal module size? Can’t be too large or too small

✓ Feedback loops and molecular switches from function to structure and from function to function

References

1. E. Braun, Y. Eichen, U. Sivan, G. Ben-Yoseph, *Nature* **391**, 775 (1998)

2. K.Keren, M.Krueger, R.Gilad, G. Ben-Yoseph, U.Sivan & E.Braun, *Science* **297**, 72 (2002).

3. K.Keren, R. Berman, E. Buchstab, U.Sivan, E.Braun, *Science* (2003) in press