SELF-ASSEMBLY OF MOLECULAR SCALE ELECTRONICS BY DNA MOLECULES AND RELATED PROTEINS

Kinneret Keren, Rachel Gilad, Yoav Soen, Michael Krueger, Ilya Baskin, Stav Zaitsev, Doron Lipson, Gidi Ben Yoseph, Yoav Eichen, Erez Braun, Uri Sivan

Technion -Israel Institute of Technology

Molecular Electronics - Gap Between Devices and Circuits
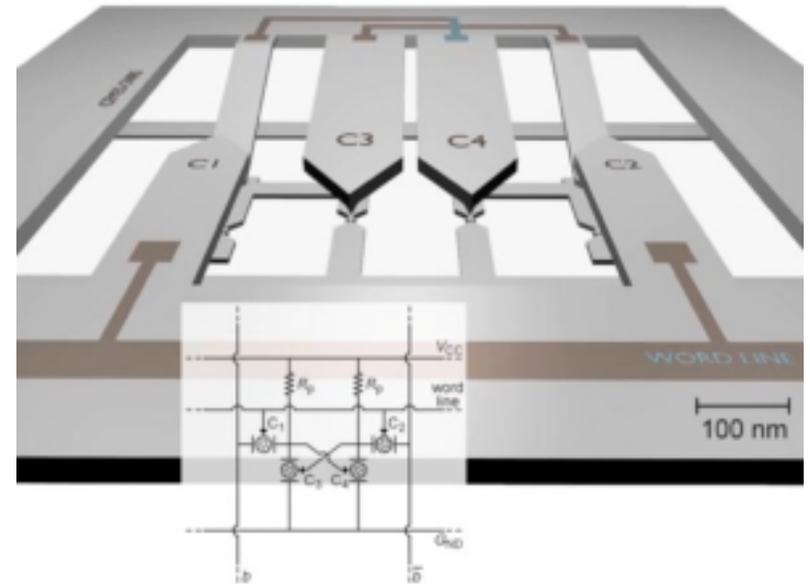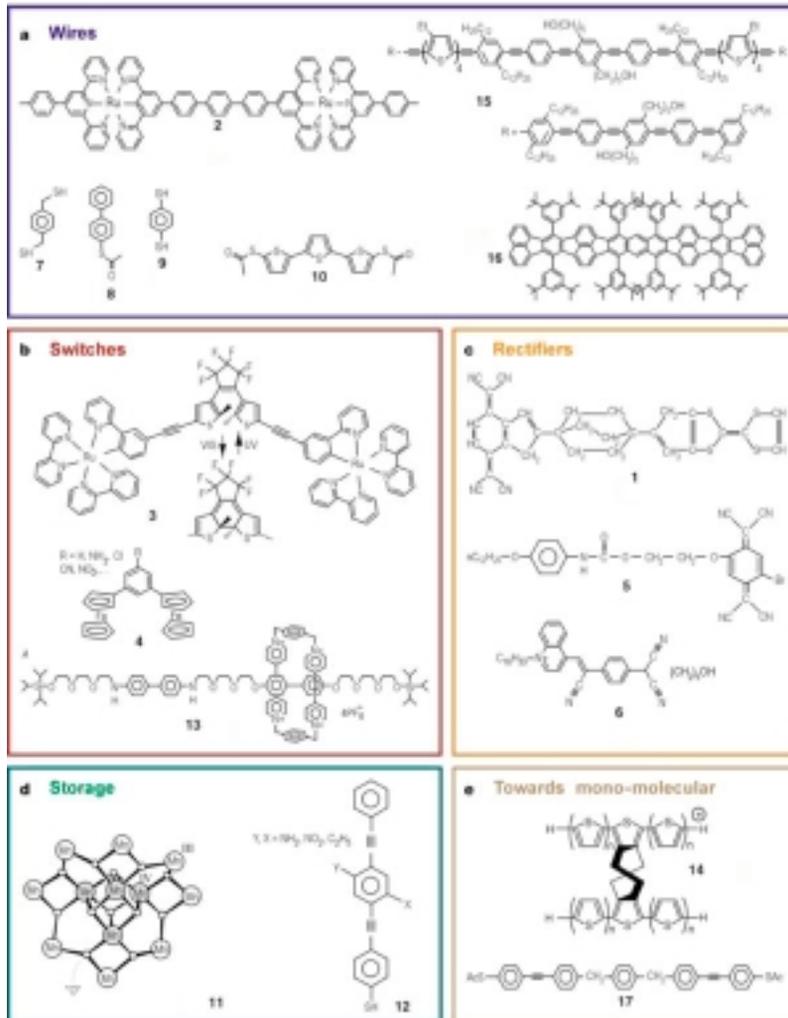
Joachim et al.

DNA Molecular Recognition
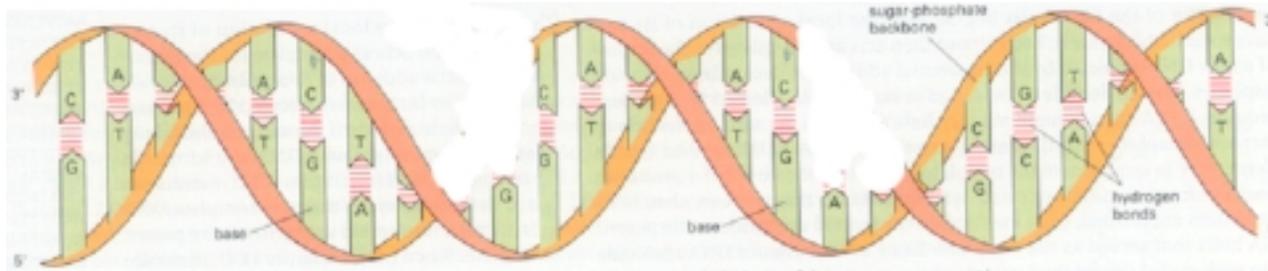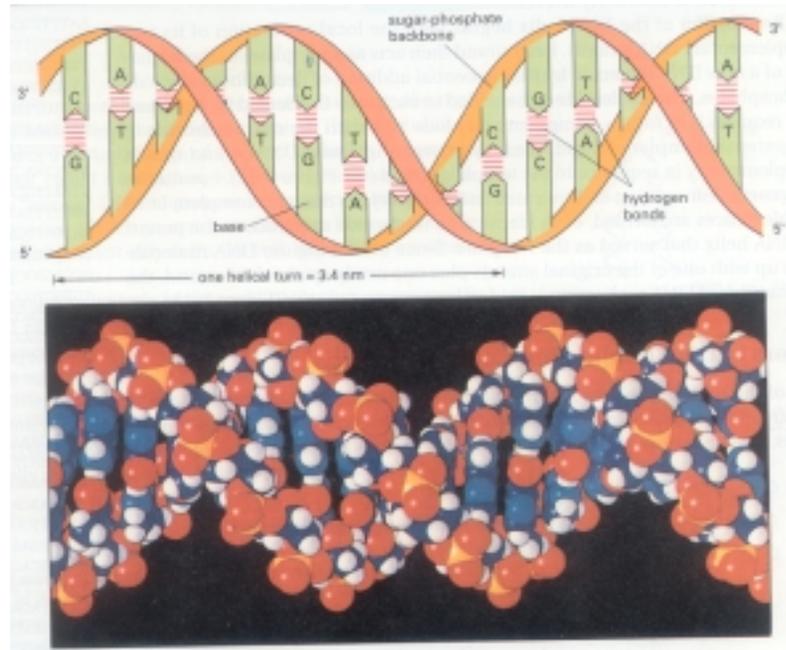
Two Step Self Assembly of an Electronic Circuit Using DNA -Possible Assembly scheme and its Limitations

I. Electrode definition

100 μm

II. Electrode encoding

3'-TCCAGCGGCGGG
S

III. Scaffold

IV. Device positioning

V. Converting DNA to wires
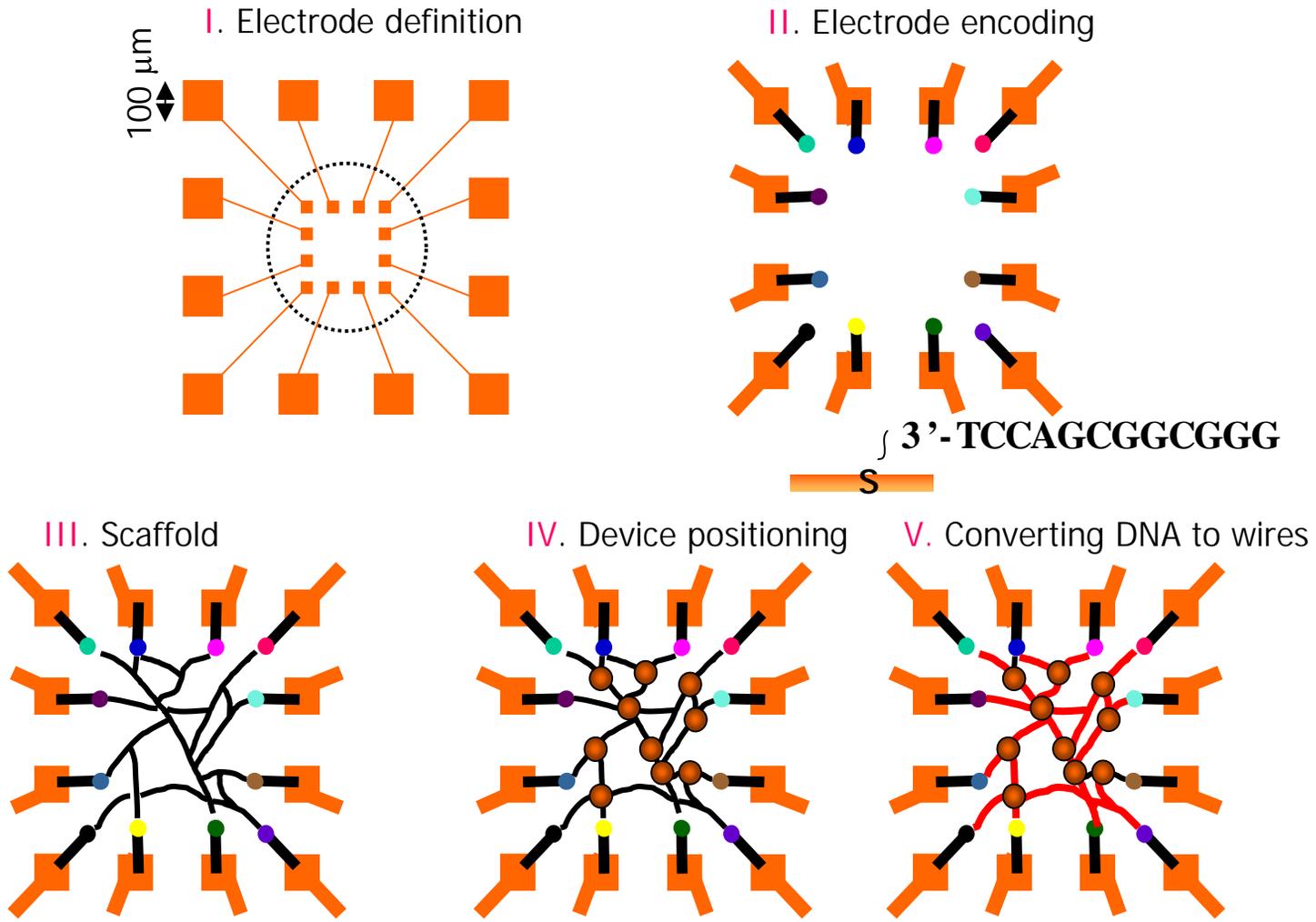
DNA Templated Conductive Wire
E. Braun, Y. Eichen, U. Sivan, G. Ben-Yoseph, Nature 391, 775 (1998)

Conductive silver wire

$$R = 26\Omega$$
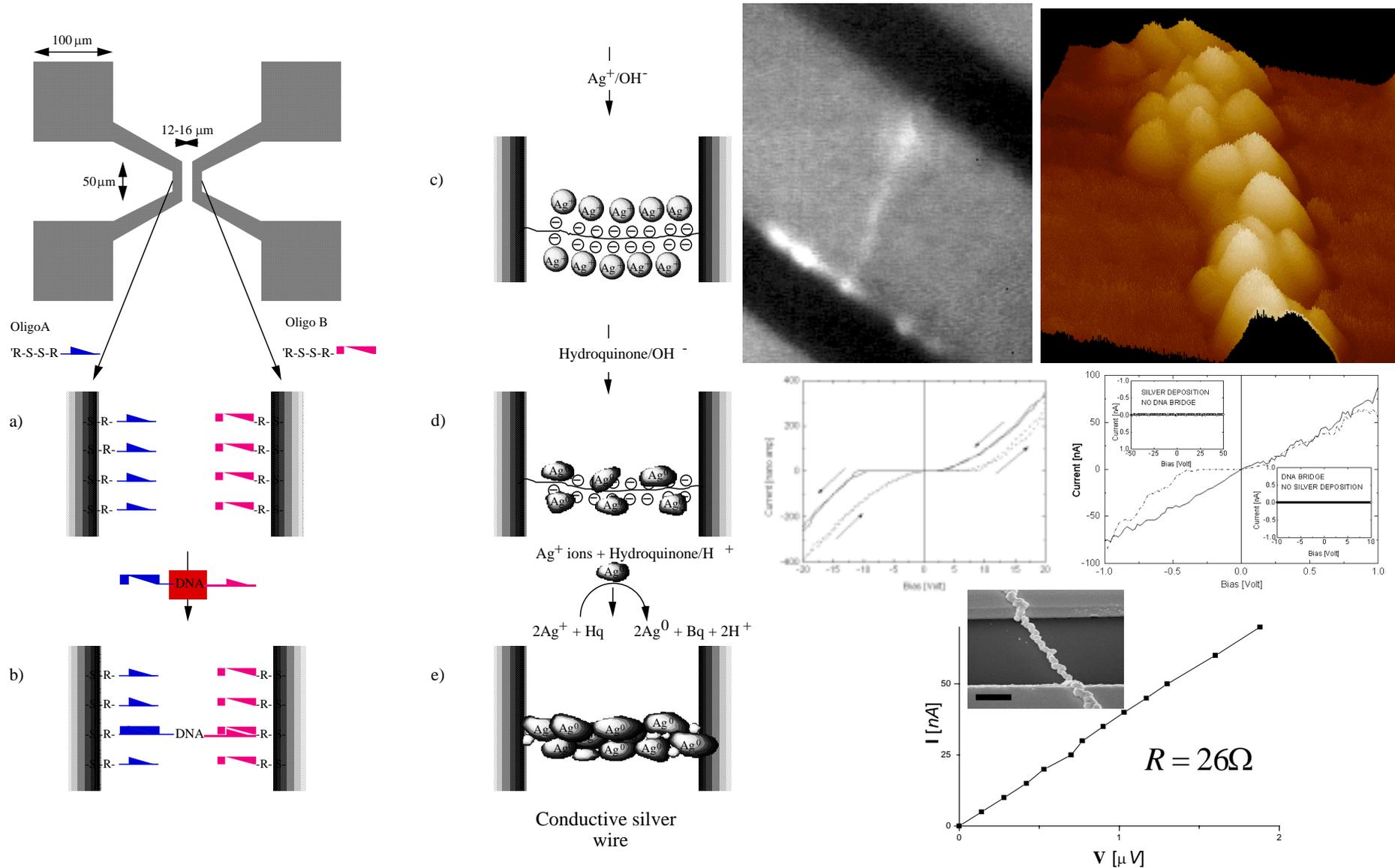
SEQUENCE SPECIFIC LITHOGRAPHY

- Microelectronics relies on lithography
- Not merely a technology - It 's a concept how to handle complexity

Homologous Recombination by RecA

DS DNA   **SS DNA**   **RecA**
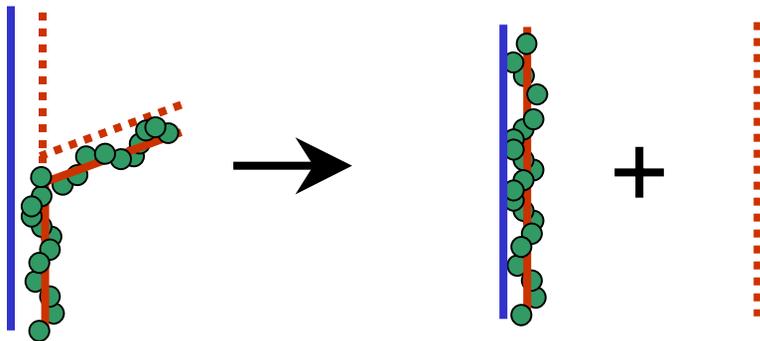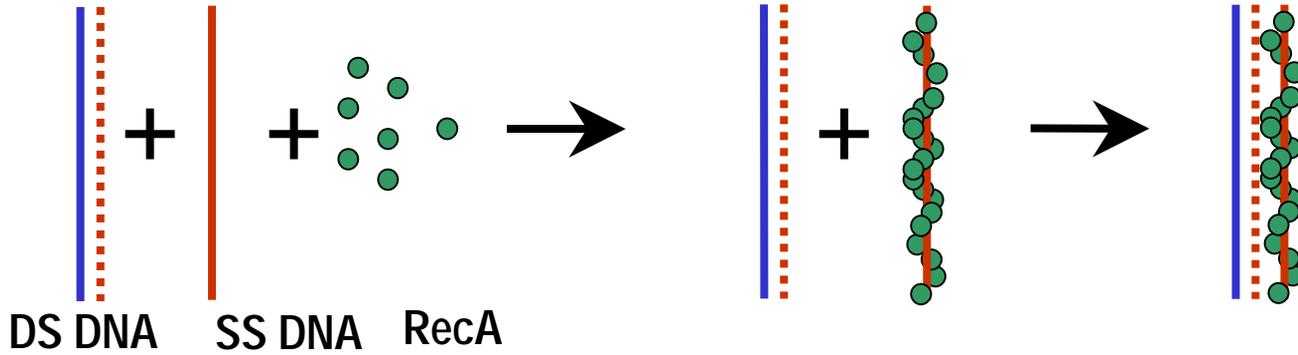
Sequence Specific Molecular Lithography Using RecA Protein

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

RecA protects the DNA against certain operations. Can be used as sequence specific resist
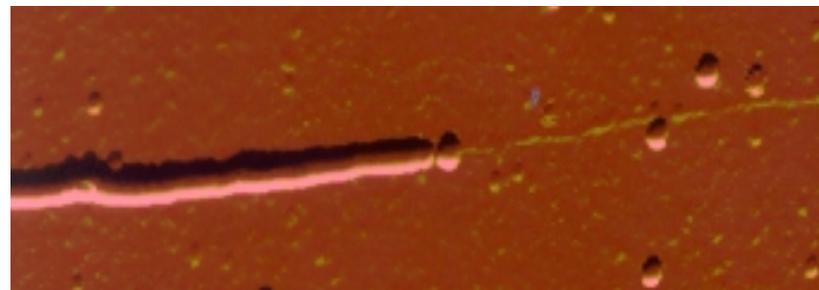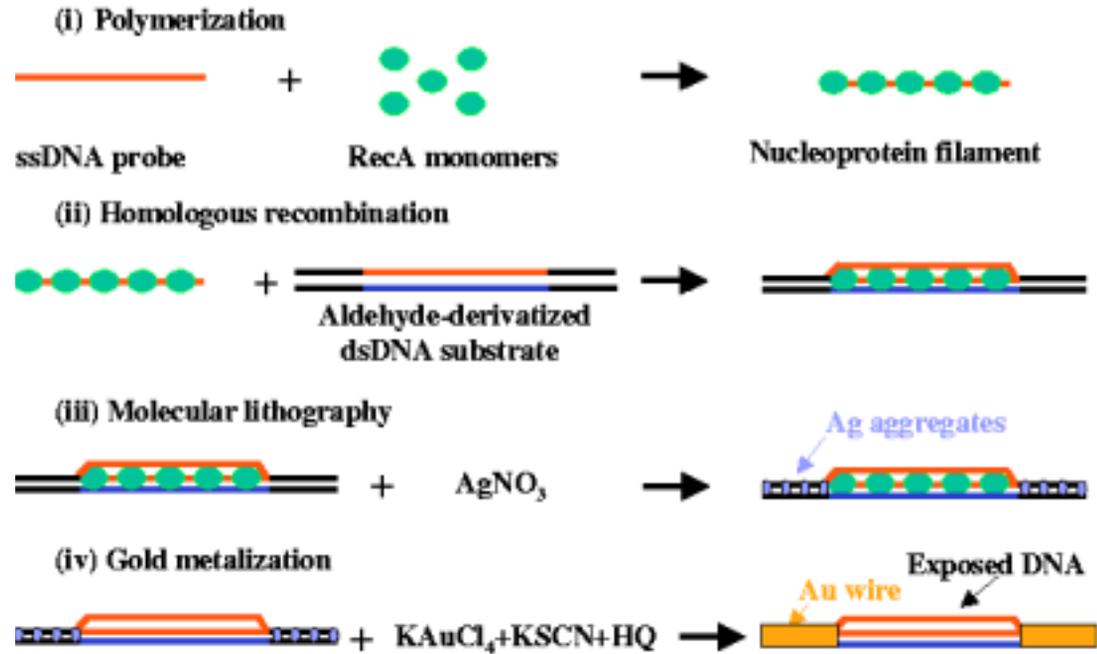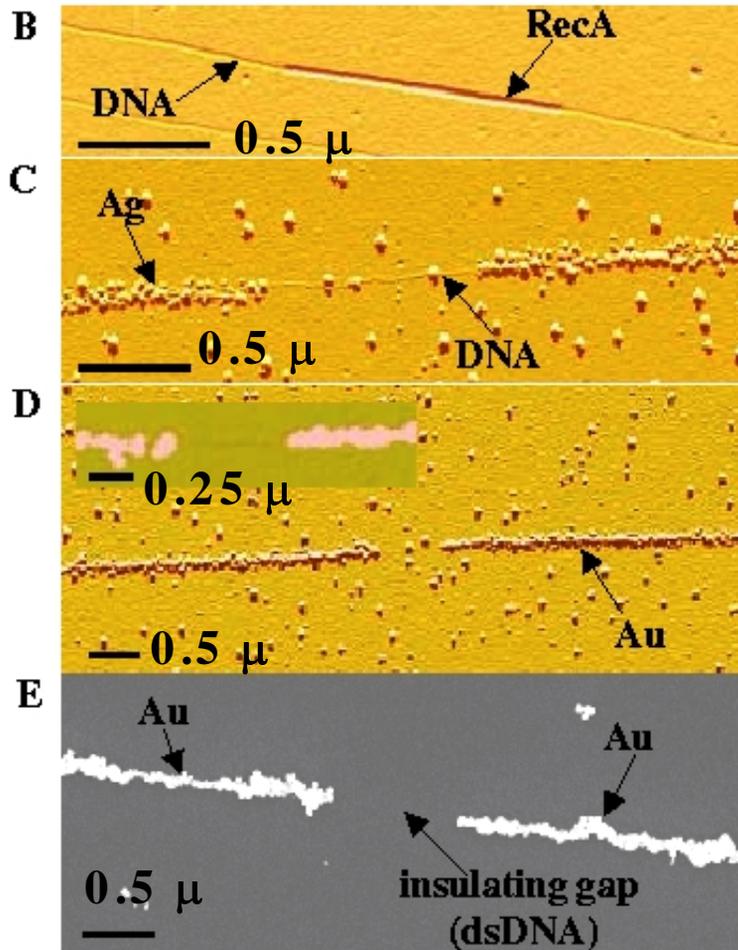
RecA as a Sequence Specific Junction Generator

ATP		+	+	+	+	+	+	-
RecA		-	+	+	+	+	-	+
Time	M	0	30	60	120	210	210	210
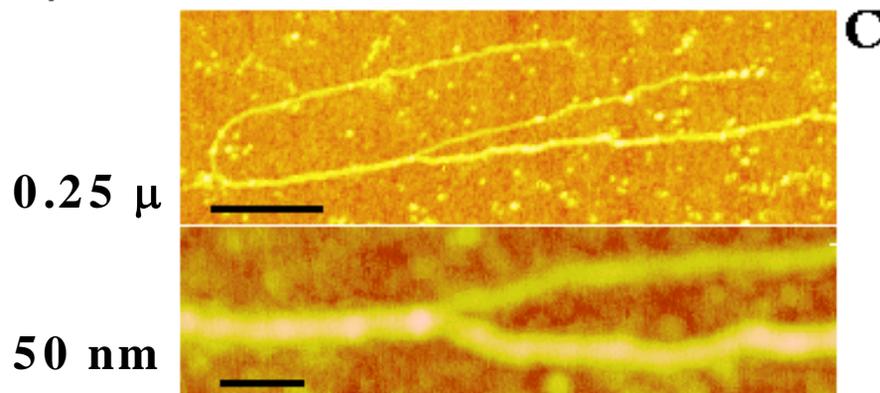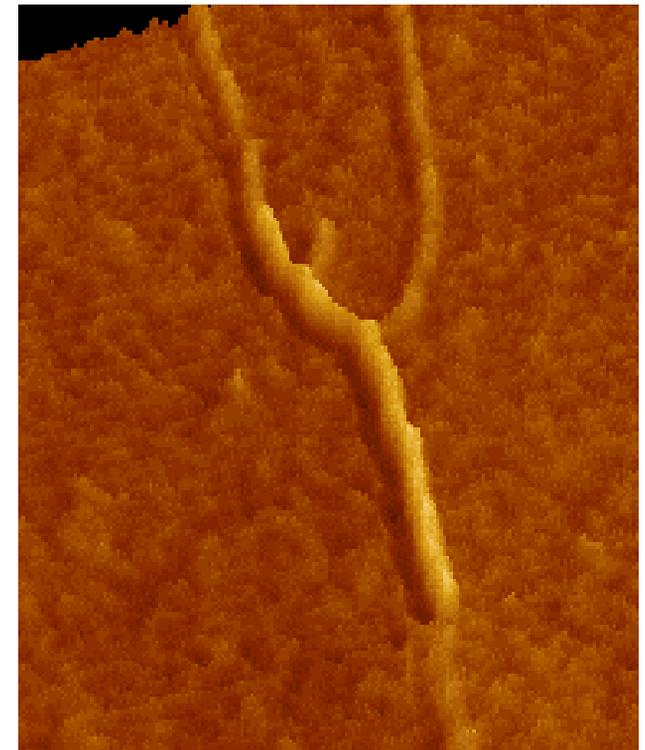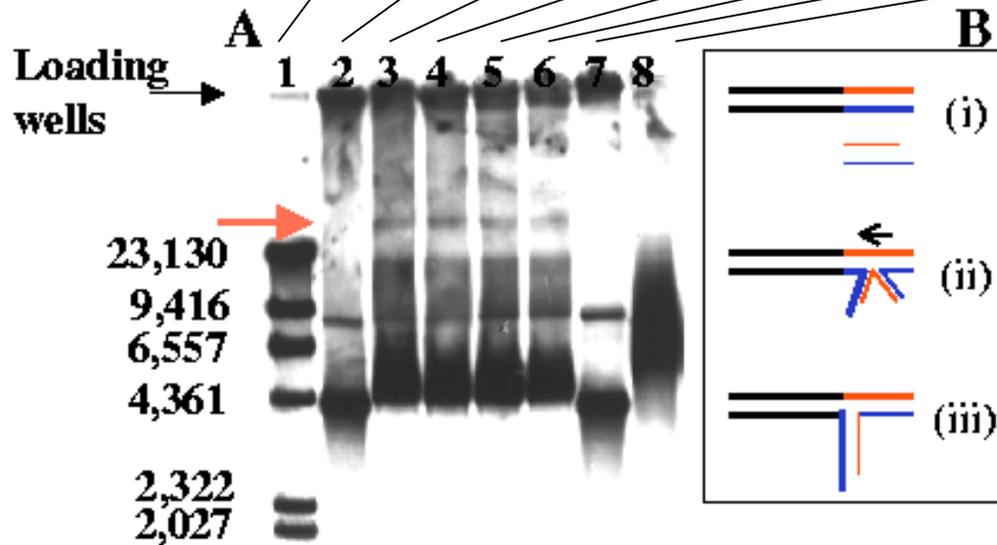
Homologous Recombination
RecA - a Universal Molecular Assembler

DS DNA **SS DNA** **RecA**

0.2 μ   **1 μ**
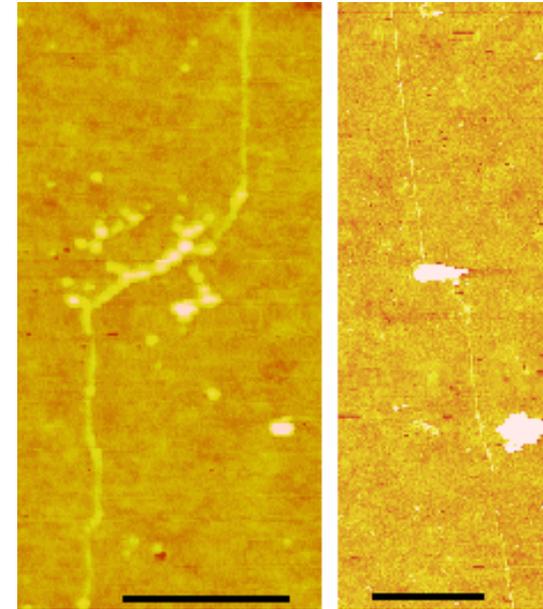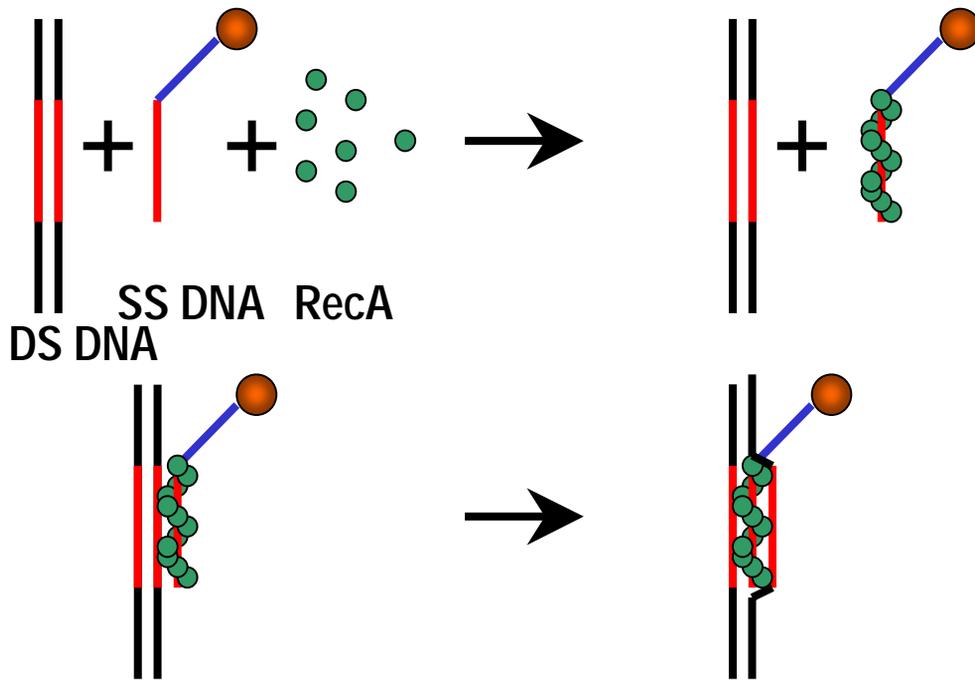
- Operates on arbitrary double stranded sequences
- Facilitates positioning of arbitrary molecular scale objects

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.
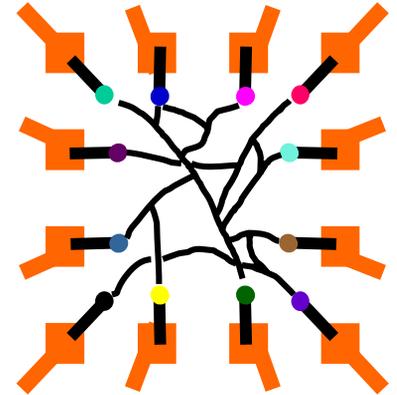
Reminds "random" number generator on a computer

COPYING DNA

Template has free 3'-OH end

Incoming nucleotide has 5'-triphosphate

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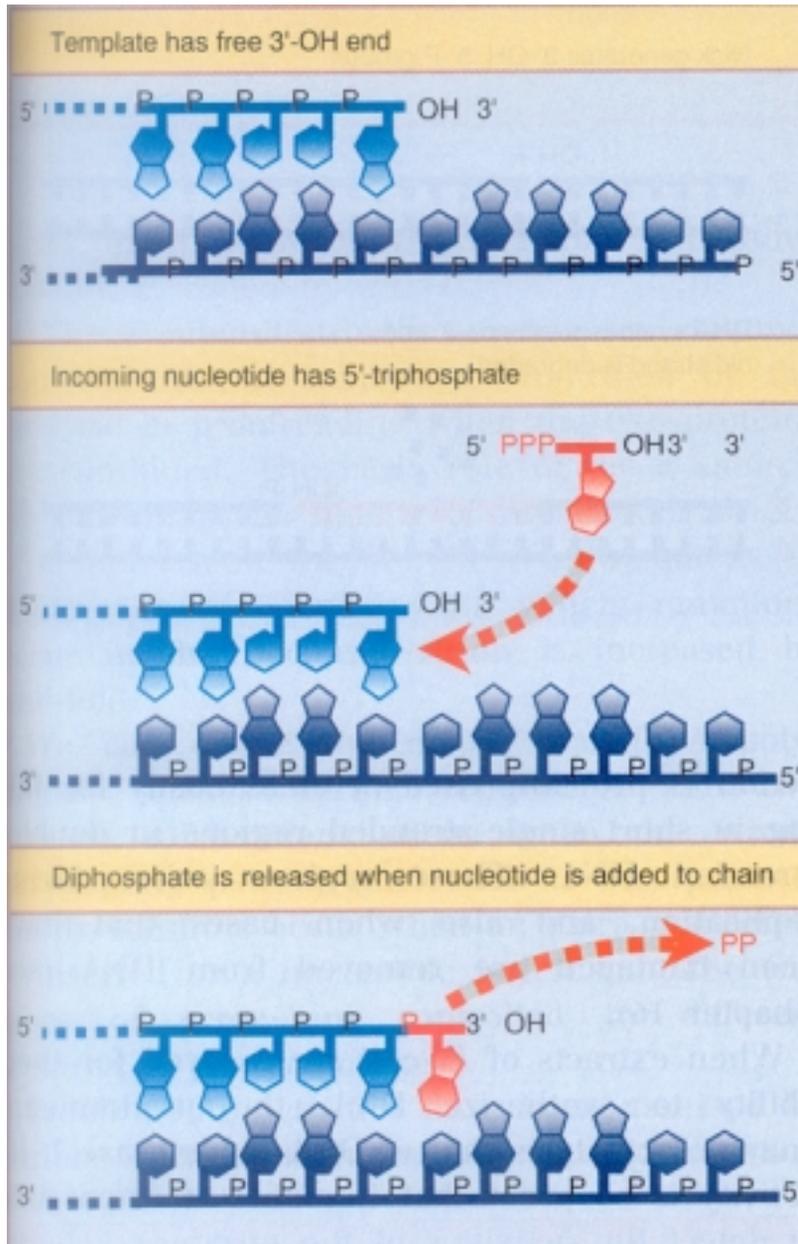
- DNA is copied with the help of an enzyme - DNA polymerase
- Complementary nucleotide is added to the 3'-OH end of the growing chain, so that the new chain is synthesized in the 5' to 3' direction
- The precursor for DNA synthesis is a nucleoside triphosphate, which looses the terminal two phosphate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

Autonomous Binary *p*-Shift Register
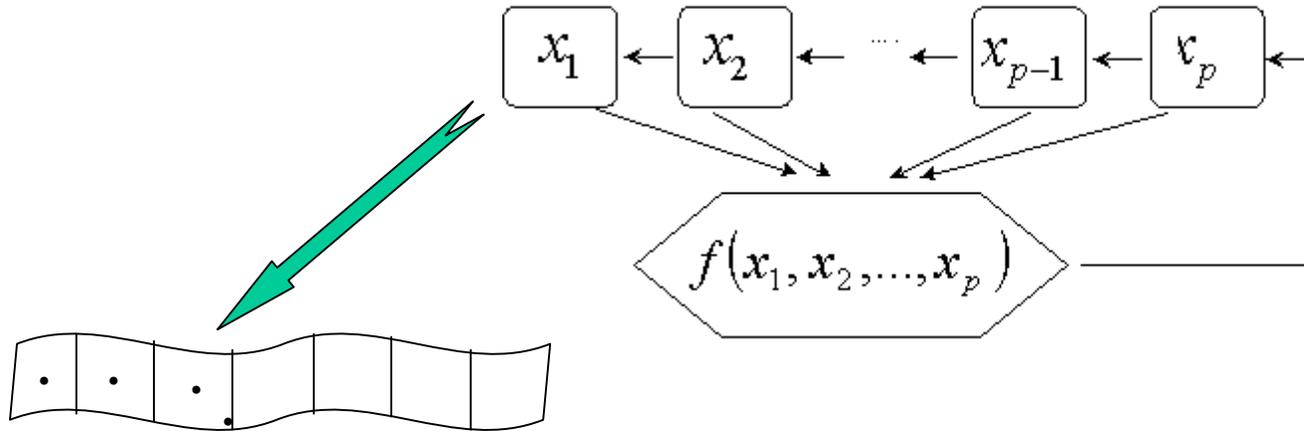
- A computing machine with 2^p internal states represented by an array of *p* cells, each occupying one bit.

- 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 !

DNA Based Molecular *p*-Shift Register

Consider a Boolean DNA with 4 "bases" $1, \bar{1}, 0, \bar{0}$

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

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

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	$\bar{0}\,\bar{1}\,0\,0$
0	1	1	$\bar{0}\,0\,\bar{1}\,\bar{1}$
			$\bar{0}\,1\,\bar{1}\,\bar{1}$
1	0	1	$1\,\bar{0}\,\bar{0}\,\bar{1}$
			$1\,1\,\bar{0}\,\bar{1}$
1	1	0	$1\,\bar{0}\,1\,\bar{0}$
			$1\,1\,1\,\bar{0}$

1st rule
seed primer (i) — 1st annealing (ii) — extension (iii) — melting (iv) — 2nd rule 2nd annealing (v)

extension (vi) — melting (vii) — 3rd rule 3rd annealing (viii) — extension (ix) — melting (x)

stop primer

tape (xi) — tape termination (xii)

◆ 1
∧ $\bar{1}$
● 0
⌢ $\bar{0}$
◣ sequences other than $0, 1$
◥ sequences other than $\bar{0}, \bar{1}$
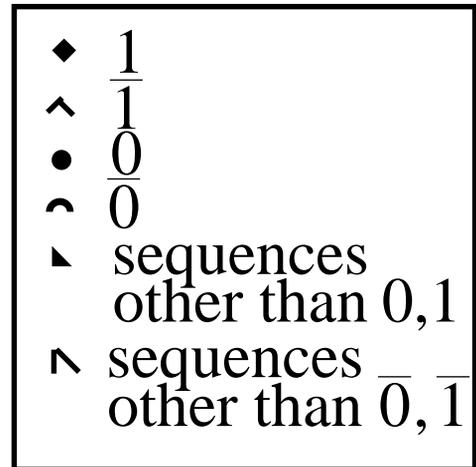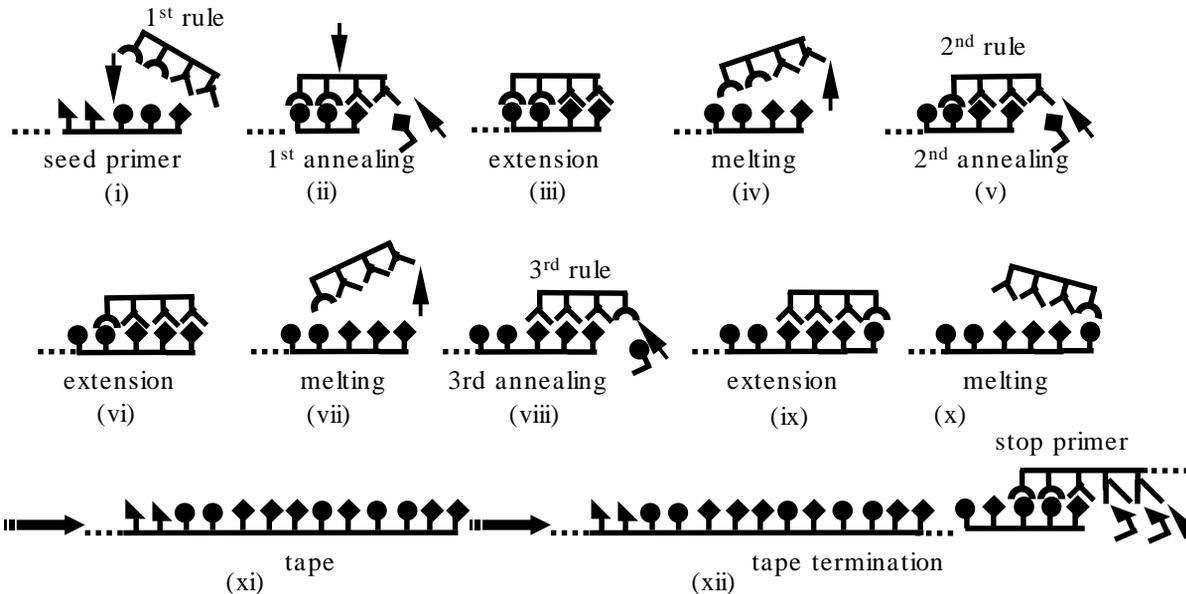
DNA Based Molecular *p*-Shift Register

•Works also in a thermal ratchet mode at a fixed temperature

•Rule strands function as enzymes. They direct the reaction but not consumed

Initiation

3-shift register realized in 5-bit space

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

5'GCATGCGCCCGTCAGGCG 00111   Seed strand

3' $\overline{0}\,\overline{0}\,1\,1\,\overline{1}\,\overline{0}$
3' $\overline{0}\,1\,1\,1\,\overline{0}\,\overline{1}$
3' $\overline{1}\,1\,1\,\overline{0}\,1\,\overline{0}$
3' $\overline{1}\,1\,\overline{0}\,1\,\overline{0}\,\overline{0}$ } 7 rule strands
3' $\overline{1}\,\overline{0}\,1\,\overline{0}\,\overline{0}\,\overline{1}$
3' $\overline{0}\,1\,\overline{0}\,\overline{0}\,1\,\overline{1}$
3' $\overline{1}\,\overline{0}\,\overline{0}\,1\,1\,\overline{1}$

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

$0 = 5'\text{TGC}$	$1 = 5'\text{GCT}$
$\overline{0} = 3'\text{ACG}$	$\overline{1} = 3'\text{CGA}$

5' GCATGCGCCCGTCAGGCG00111$(0100111)_n$01001CTGCAG   with $n = 0,1,...$
seed primer $\hookleftarrow$ $\mapsto$ complementary to stop primer

Confirmed by Sequencing!
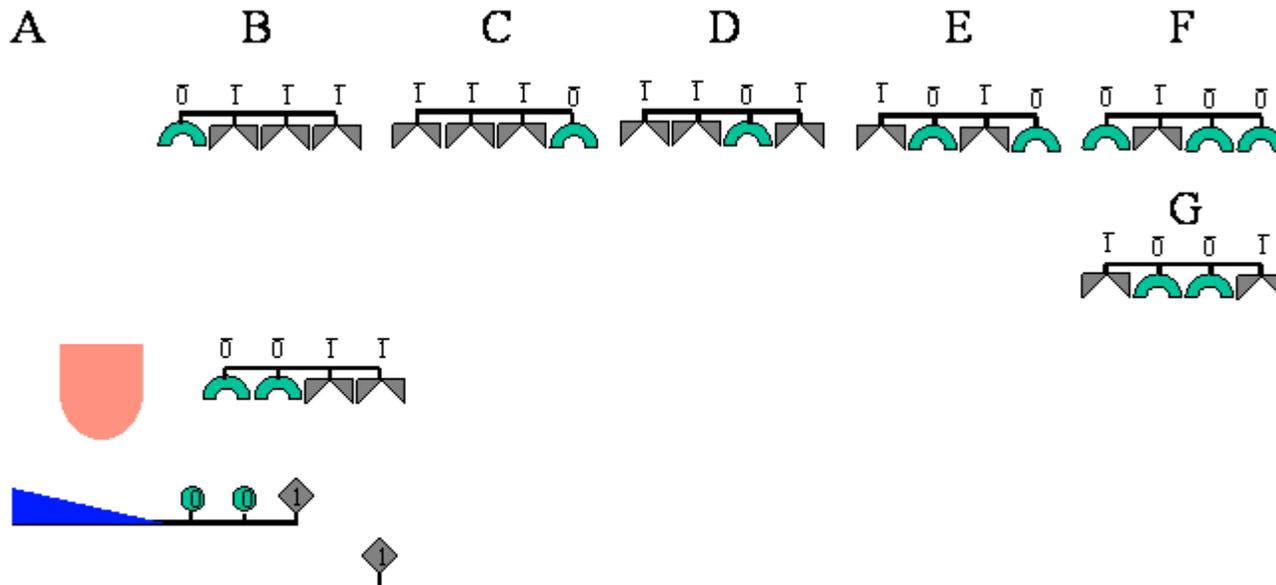
4-shift register realized in 6-bit space

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

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{00}$1 1 1 1$\overline{0}$ 3'$\overline{0}$1 1$\overline{00}$1$\overline{0}$
3'$\overline{0}$1 1 1 1$\overline{0}$1 3'1 1$\overline{00}$1$\overline{00}$
3'1 1 1 1$\overline{0}$1$\overline{0}$ 3'1$\overline{0}$ 0 1$\overline{000}$
3'1 1 1$\overline{0}$1$\overline{0}$1 3'$\overline{0}$ 0 1$\overline{0}$0$\overline{0}$1
3'1 1$\overline{0}$1$\overline{0}$1 1 3'$\overline{0}$1$\overline{0}$0 0 1 1
3'1$\overline{0}$1$\overline{0}$1 1$\overline{0}$ 3'1$\overline{000}$1 1 1
3'$\overline{0}$1$\overline{0}$ 1 1$\overline{00}$ 3'$\overline{000}$1 1 1 1

15 rule

strands

3' 1$\overline{0}$1 1$\overline{00}$1

3' 1$\overline{0}$1 1$\overline{00}$ GCGC CAG GAC GCG GAC GTC stop strand
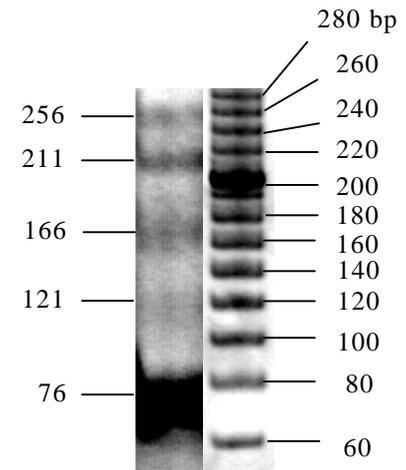
$$0011110 \left(1011001000\ 11110\right)_n 1011; n = 0,1,...$$

280 bp
260
256 — 240
211 — 220
200
180
166 — 160
140
121 — 120
100
80
76 — 60

Confirmed by Sequencing!

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

240 — — 254
200 — — 212
160 — — 170
120 — — 128
80 — — 86
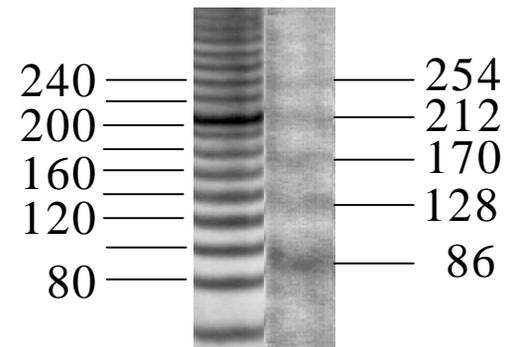
$$03110(12231323203110)_n 1223132320$$

Number of shift register sequences that can be generated by rule strands of a given length

For *k* letter alphabet and *p* cells the number of maximal shift-register sequences of length k^p is

$$[(k-1)!]^{k^{p-1}} k^{k^{p-1}-p}$$

For *k*=3 and *p*=5, for instance, combinations of the 3^6=729 rule strands, which can be synthesized in a reasonable effort, yield more than 10^{60} different maximal shift register sequences!

Put it differently. For synthesis of *s* strands with *k* letter alphabet

one can generate $[k!]^{s/k^2} \dfrac{k}{s}$ different maximal shift register sequences

Competing Blocking Processes
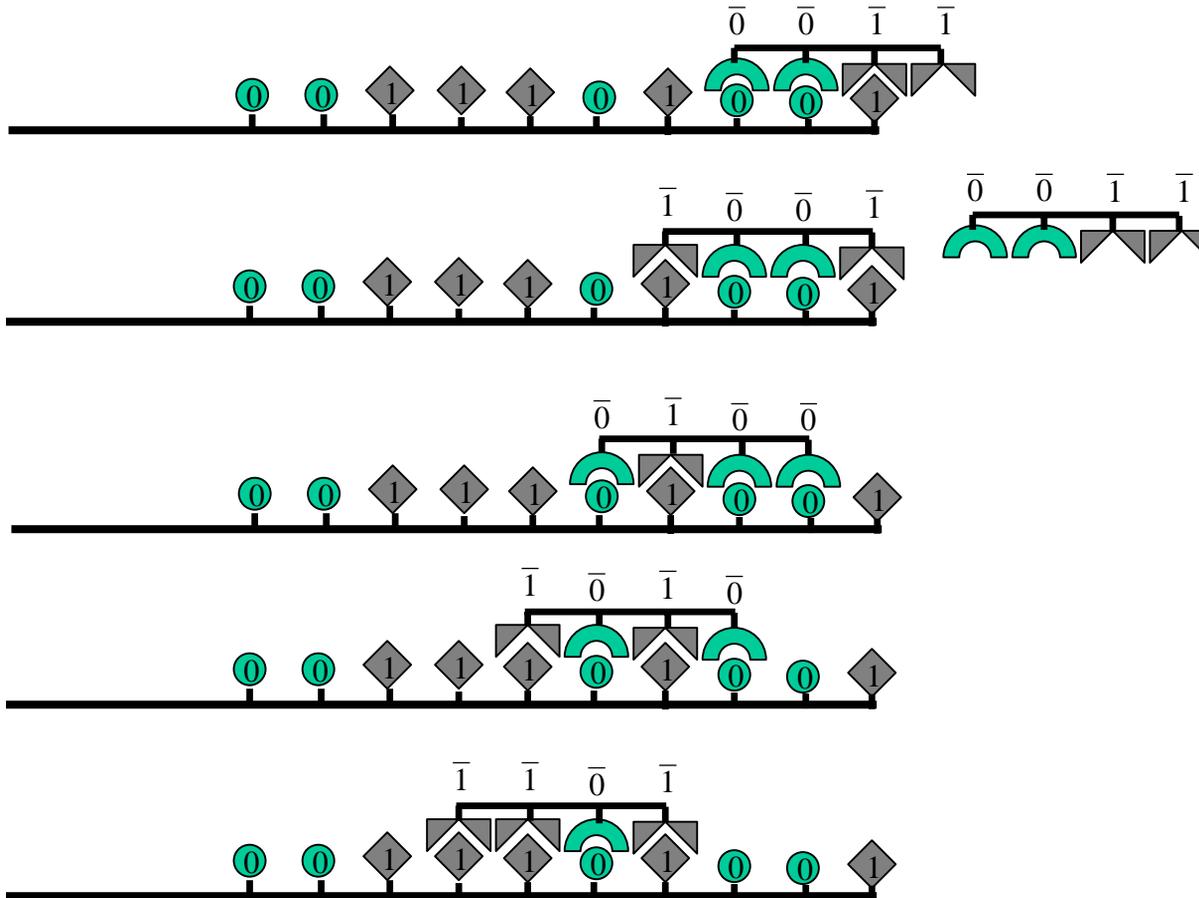
Correct synthesis

blocking

synthesis

Reaction proceeds through thermal fluctuations (ratchet)

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

00111010

n=3

0011**11010001**110100.....

- Nodes correspond to machine states

- Lines correspond to transition rules

Prone to errors since

all nodes are legal, namely,

recognized by rule strands

00111101001110100.....

n=5
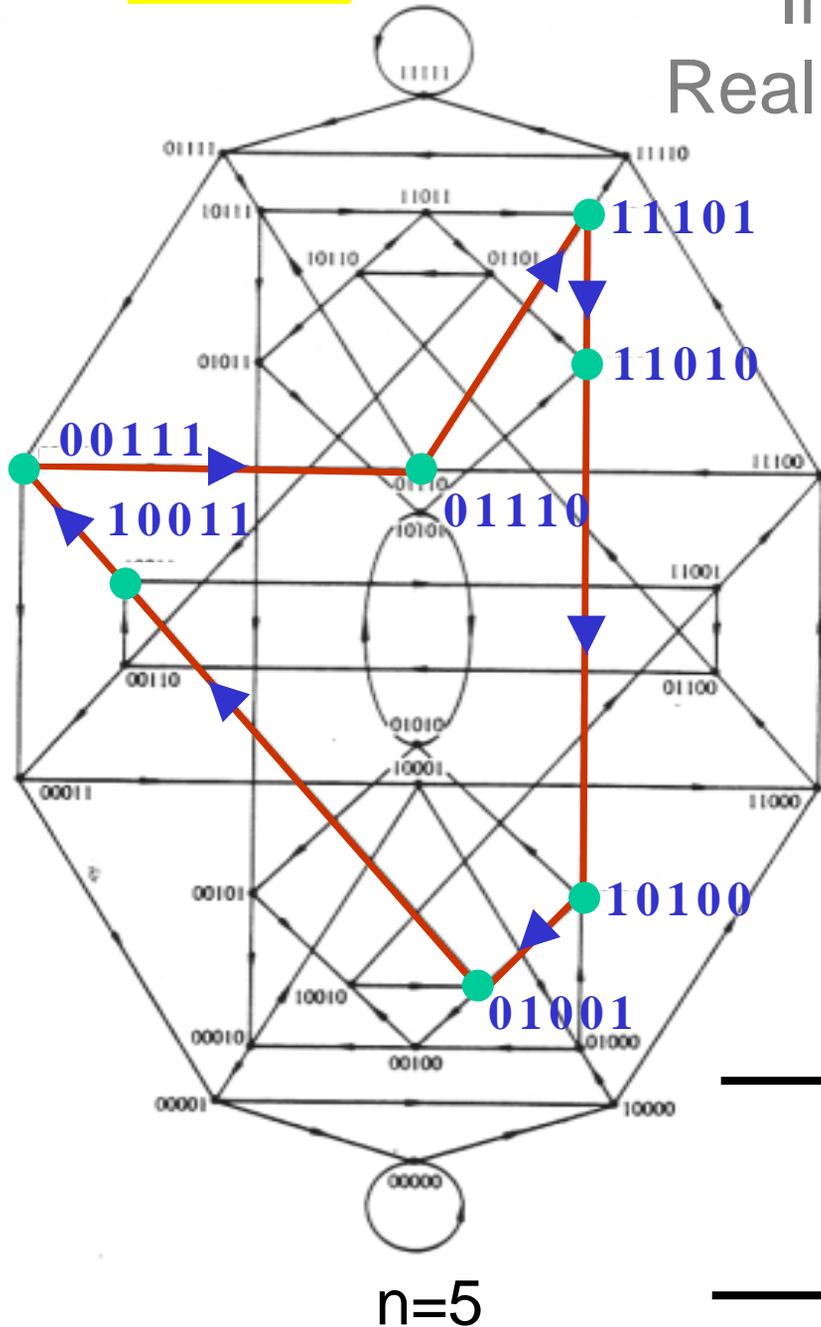
• Errors usually lead to a node which is not recognizable by any rule strand. Consequently further elongation requires a second error

• When a p-SR is realized with ($p+q$)-SR error requires $q+1$ mismatches. Consequently the errors are suppressed by $\exp\left[-(q+1)\Delta G / k_B T\right]$ where ΔG is the free energy associated with one base mismatch

• $\Delta G \approx 8.5 \div 10.5 k_B T$

correct

error