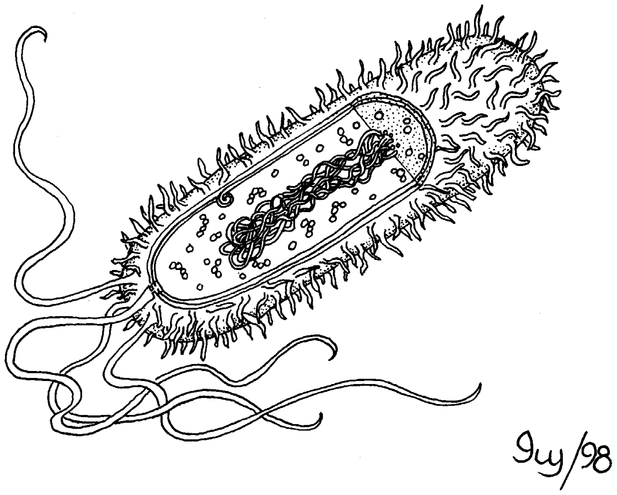
Molecular Programming: Chemistry as a New Information Technology

Erik Winfree

Computer Science & CNS & Bioengineering,
California Institute of Technology
The DNA and Natural Algorithms (DNA) Group

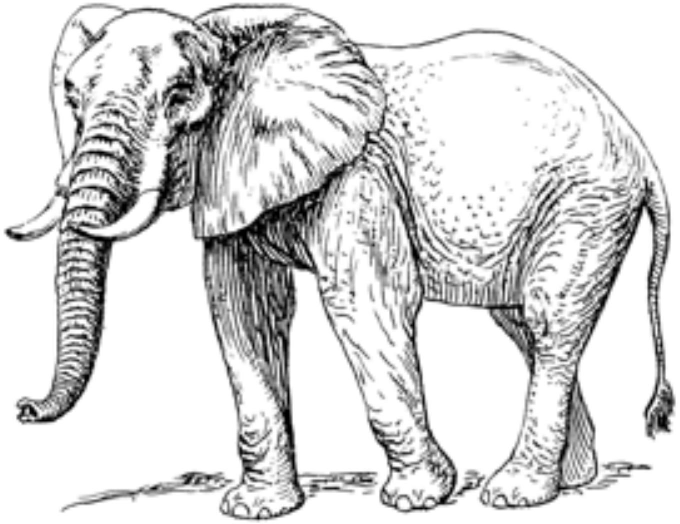
National Science Foundation. Gordon and Betty Moore Foundation.

Biology is an information technology



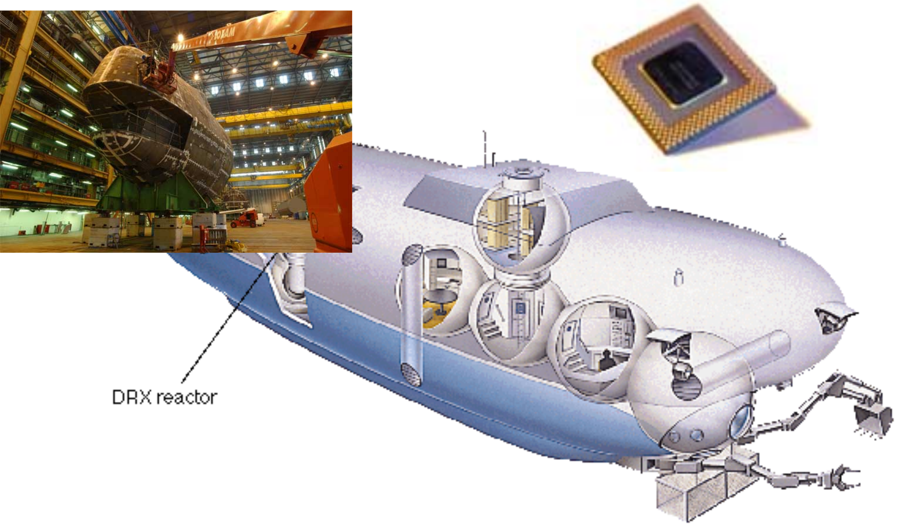
I. Livingstone © BIODIDAC

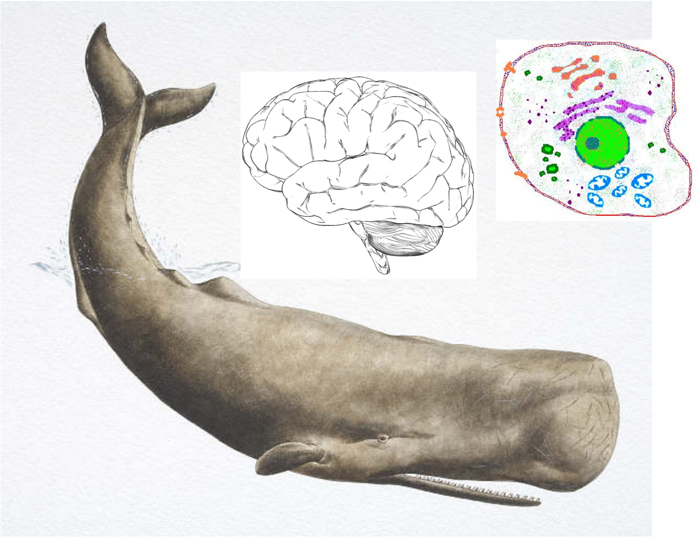
1 micron^3 volume
4 megabit genome
biochemical circuitry
manufacturing plant
atomic-level design
➔ deep physics & chemistry



grows from a single cell
contains 10^{15} cells
10^{27} macromolecules
spans atomic to macro scales
intelligent behavior
➔ deep algorithmic issues

Biology as an information technology

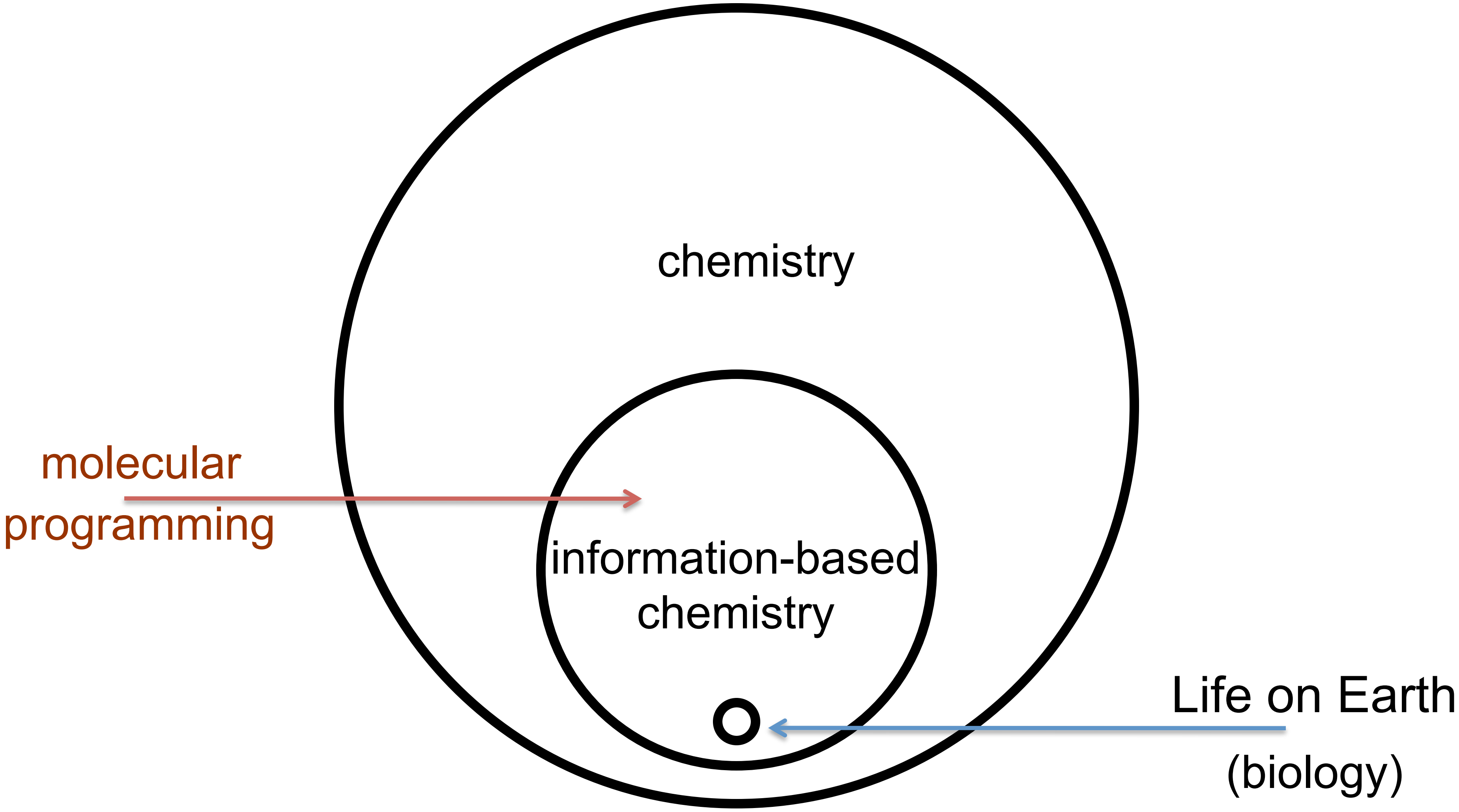




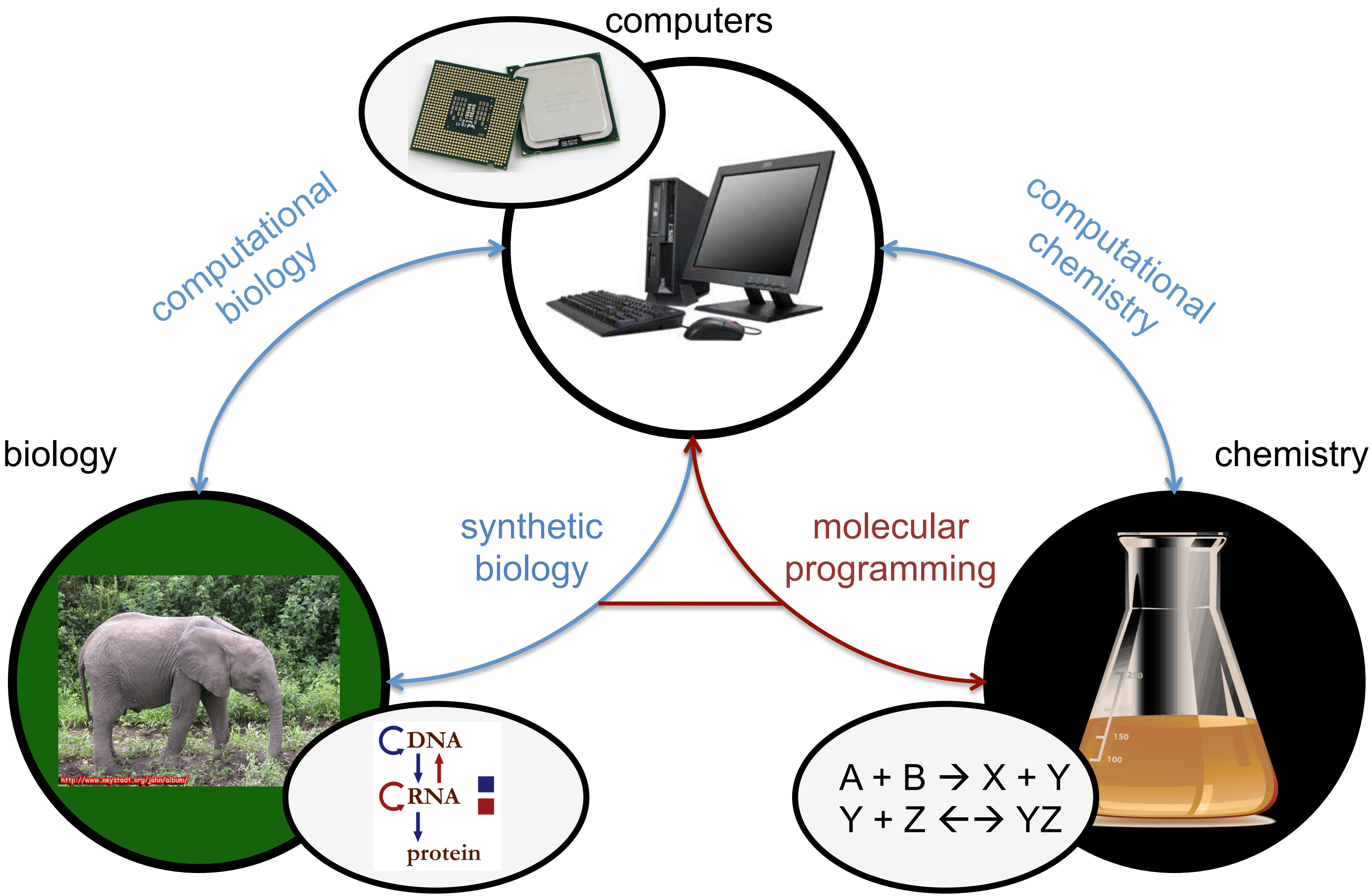
Size:	25 m, 30 tons	20 m, 50 tons.
Smarts:	multiple CPUs	7 kg brain
Resolution:	22 nm in chips	0.3 nm everywhere
Complexity:	10^6 parts, 10^{10} transistors	10^{15} cells, 10^{27} proteins
Construction:	built in factory	growth algorithm
Specification:	CAD files	genetic program

How can we engineer molecular systems of comparable sophistication? Molecular programming!

Chemistry as an information technology



Information technologies



computers

computational biology

computational chemistry

biology

chemistry

synthetic biology

molecular programming

DNA

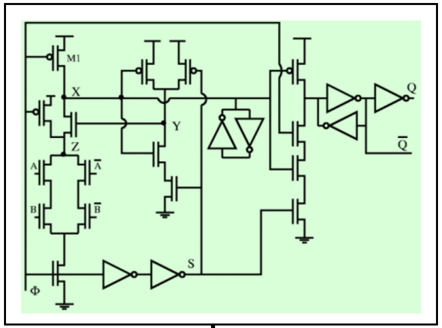
RNA

protein

$A + B \rightarrow X + Y$

$Y + Z \leftrightarrow YZ$

Embedded computation in electro-mechanical systems



Computers

Appliances

Vehicles

Embedded computation in chemical and biological systems

$$A \xrightarrow{0.03} 2A$$

$$2A \xrightarrow{5\times10^4} A$$

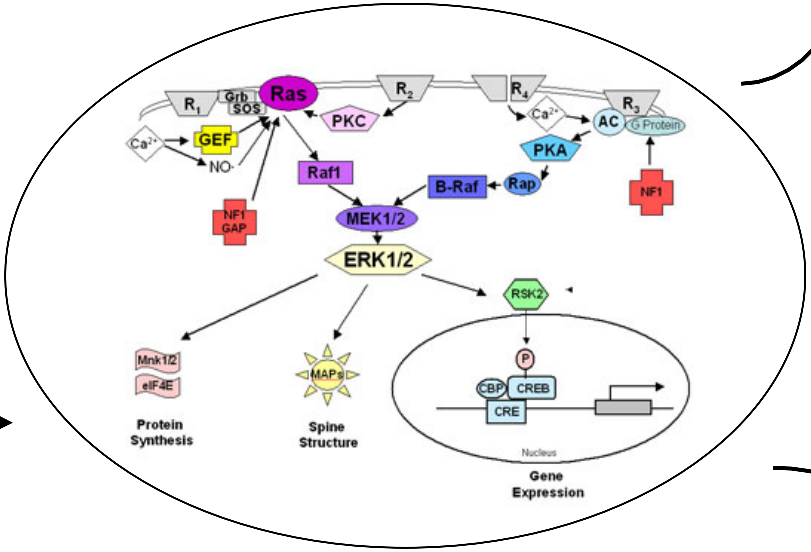
$$B + A \xrightarrow{10^5} 2B$$

$$B \xrightarrow{0.01}$$

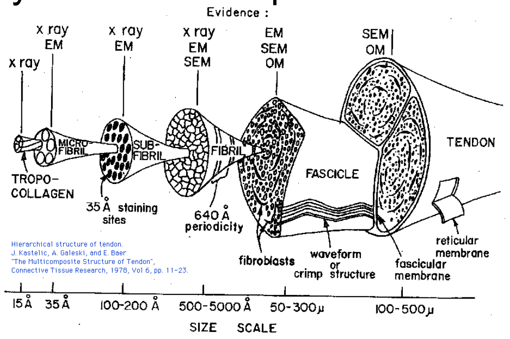
$$A + C \xrightarrow{10^5}$$

$$C \xrightarrow{0.0165} 2C$$

$$2C \xrightarrow{5\times10^4} C$$

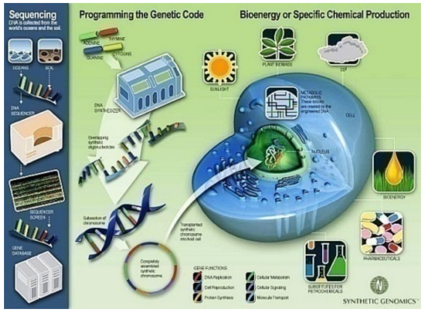


Synthesis of complex materials

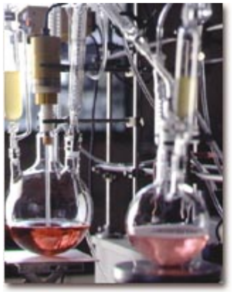


Hierarchical structure of tendon.
J. Kastelic, A. Galeski, and E. Baer
"The Multicomposite Structure of Tendon",
Connective Tissue Research, 1978, Vol 6, pp. 11–23.

Medical diagnostics & therapeutics


Sequencing
Programming the Genetic Code
Bioenergy or Specific Chemical Production
SYNTHETIC GENOMICS

Controlling complex chemical synthesis



Natural and artificial technology

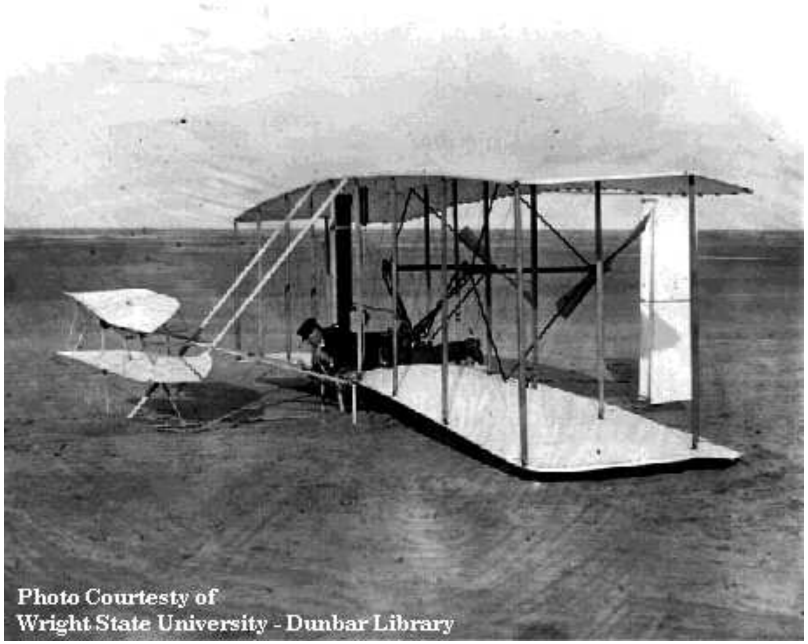
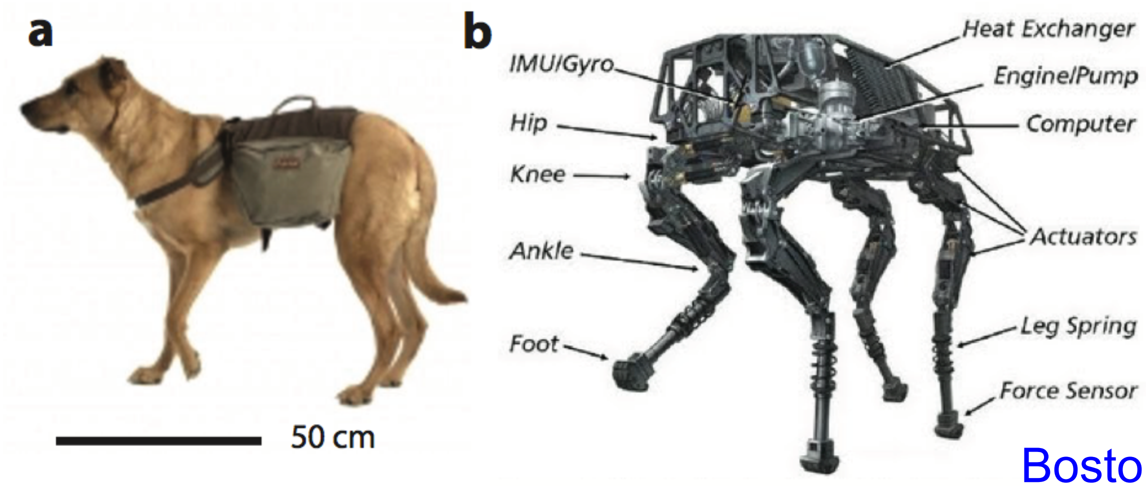


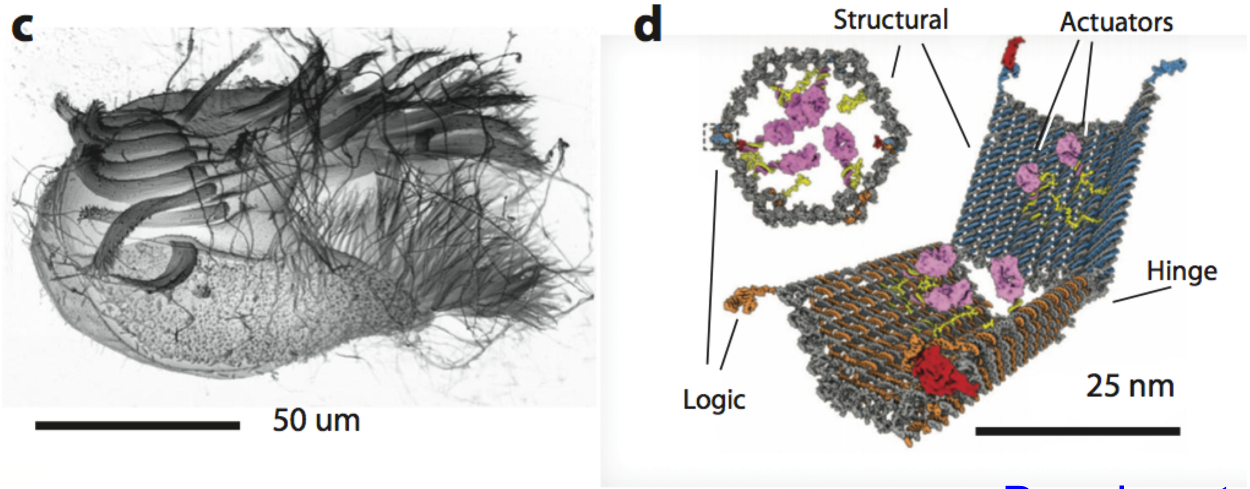
Photo Courtesty of
Wright State University - Dunbar Library

Biomimetic technologies



Electro-mechanical technologies make it possible to create macroscopic robots with life-like autonomous behaviors.

Boston Dynamics Big Dog



Molecular technologies will make it possible to create molecular robotic systems with life-like autonomous behaviors.

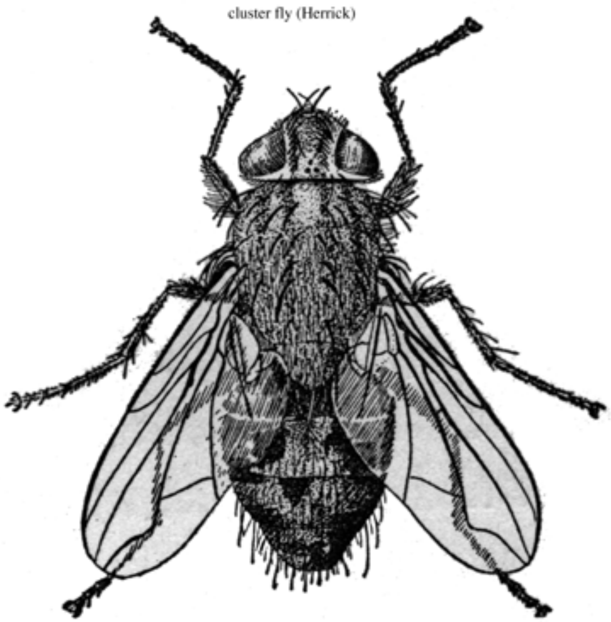
Douglas et al, Science, 2012

The challenge of programming chemical systems

- **What kind of programming language do we need to build a “fly”?**
  - **What are the programmable molecular and biochemical building blocks?**
  - **New concepts for programming and analyzing such systems?**
  - **Intrinsic fault-tolerance, adaptation, and learning conceptually built in?**
  - **How to incorporate geometric and mechanical factors?**

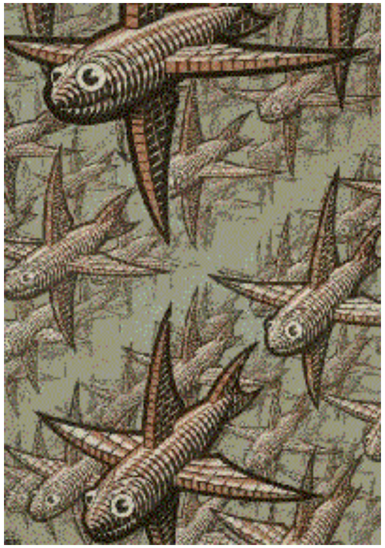
```
include gro
alpha := 0.75;
program p() := {
  gfp := 0;
  r := [ t := 0 ];
  selected & just_divided : {
    print ( "At time ", r.t, ":
        After division, cell ", id,
        " has ", gfp, " gfp molecules" )
  }
  rate ( alpha * volume ) : {
    gfp := gfp + 1
  }
  true : {
    r.t := r.t + dt
  }
};
stemcell ( [], program p() );
```

?



Nucleic acid nanotechnology as a first step toward chemical information technology

- Information-based
- Programmable
- Algorithmic
- Computation and construction

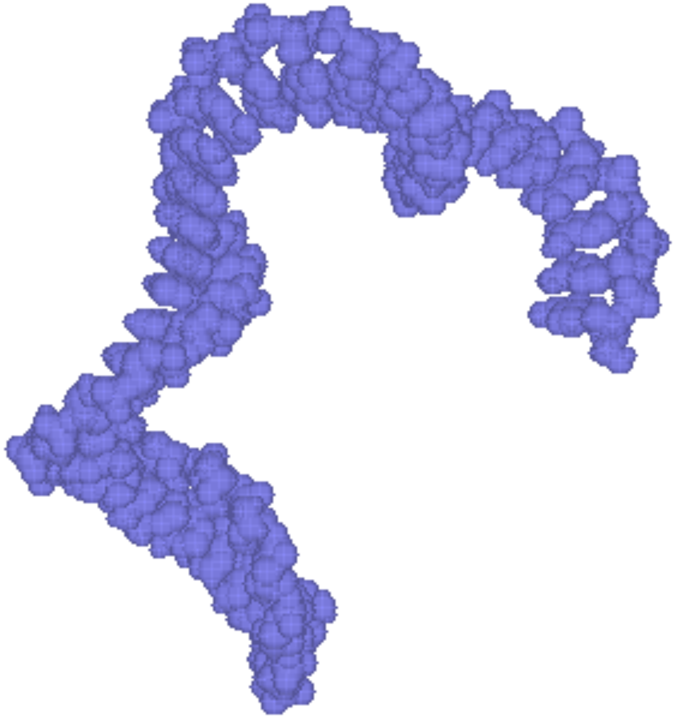


The Brewery (by Ann Erpino)

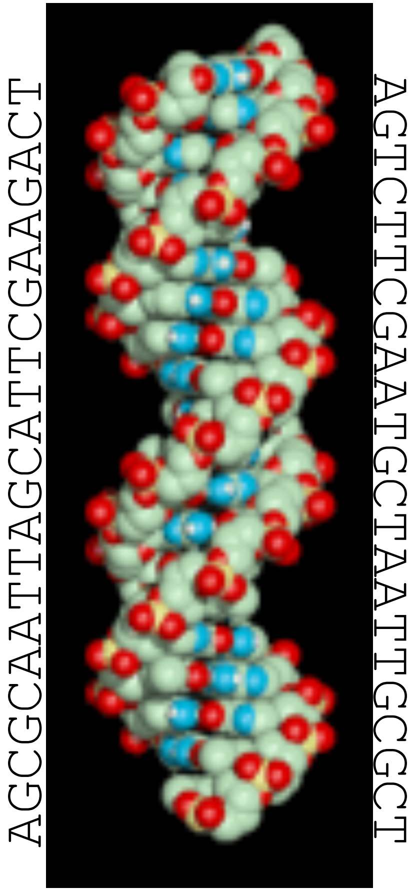
DNA as an Engineering Material

Ned Seeman
(New York University)

Wood.
Glue.
Wool.
Silk.
Leather.
DNA.



single-stranded



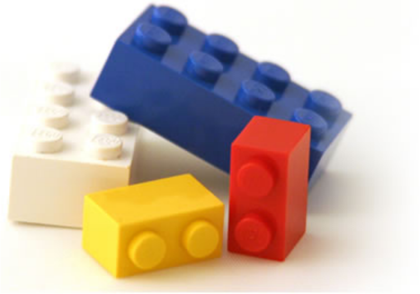
double-stranded

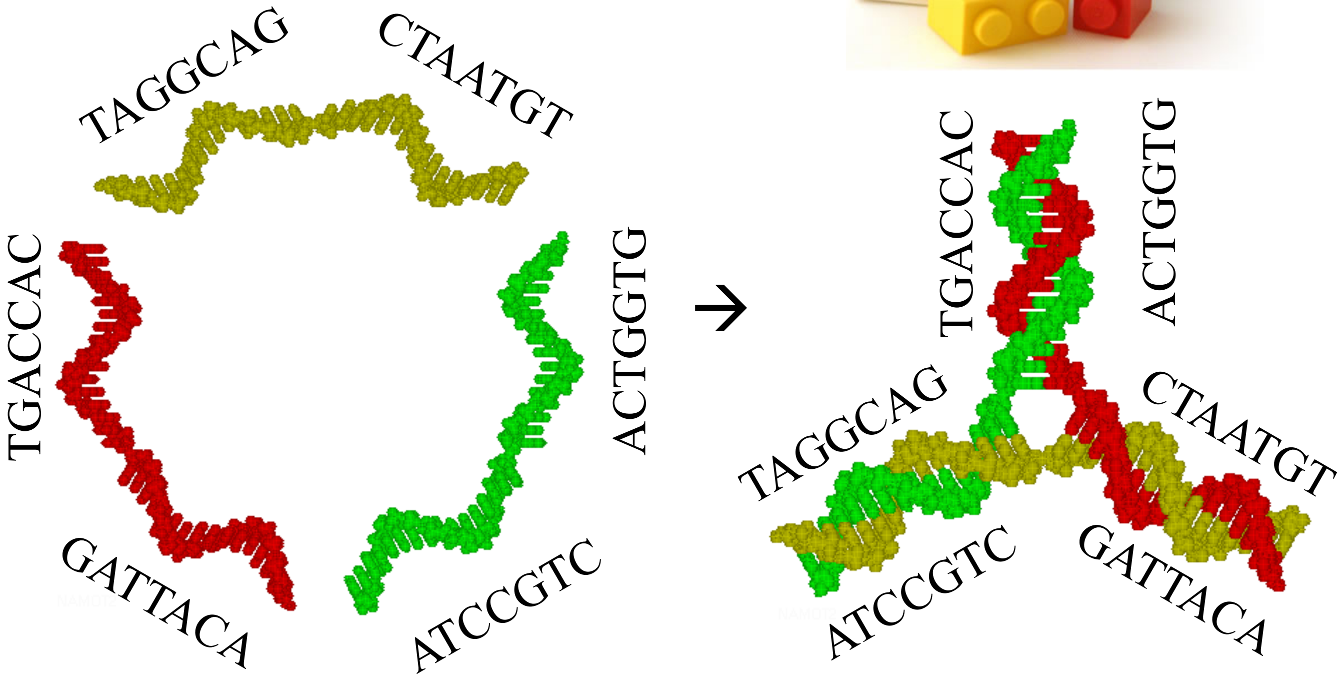
DNA as an Engineering Material

Ned Seeman

(New York University)

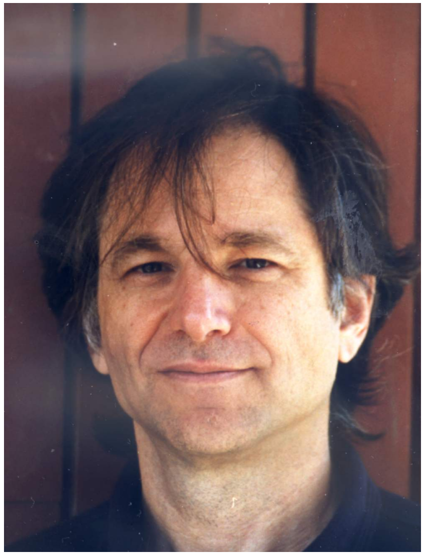
Wood.
Glue.
Wool.
Silk.
Leather.
DNA.





Seeman, *J. Theor. Biol.* 1982

DNA as a Computing Substrate



Len Adleman
(USC)

Gears.
Springs.
Fluids.
Vacuum tubes.
Transistors.
DNA.

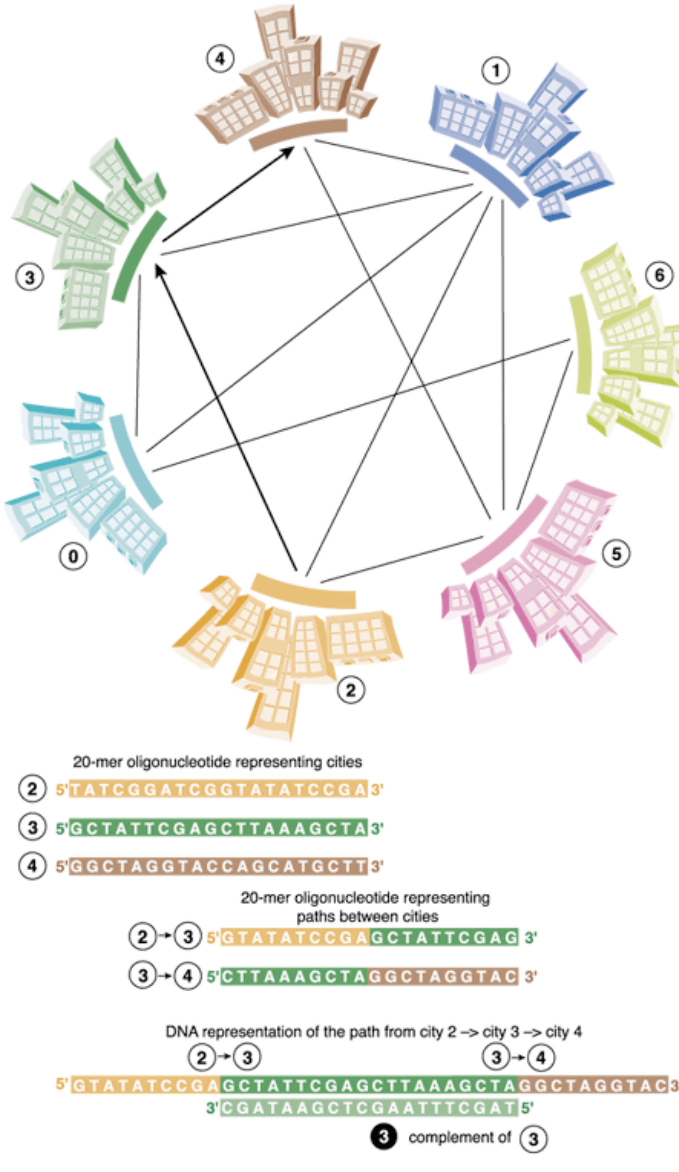
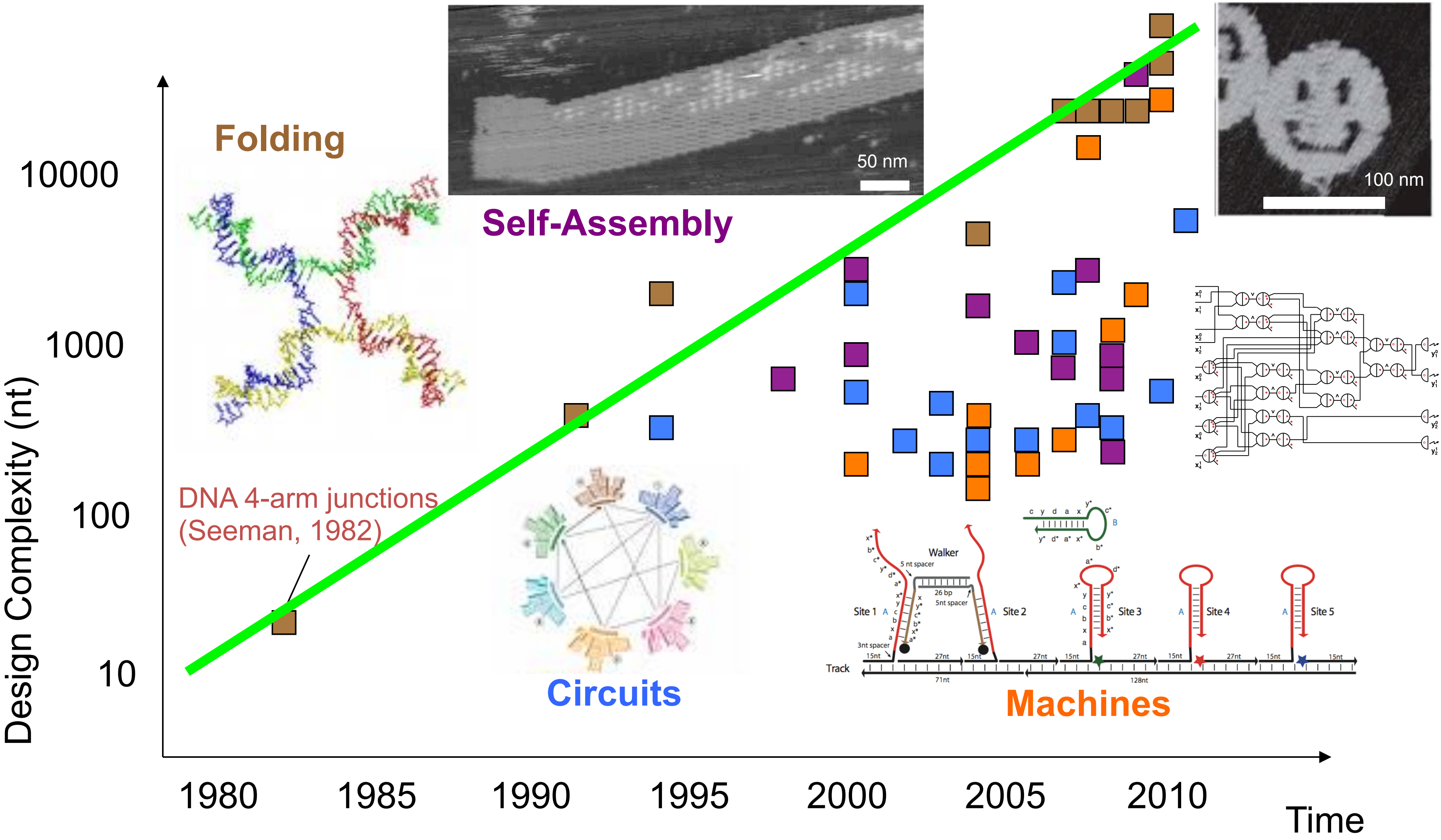

4
1
3
6
0
5
2
20-mer oligonucleotide representing cities
2 5'TATCGGATCGGTATATCCGA3'
3 5'GCTATTCGAGCTTAAAGCTA3'
4 5'GGCTAGGTACCAGCATGCTT3'
20-mer oligonucleotide representing paths between cities
2→3 5'GTATATCCGAGCTATTCGAG 3'
3→4 5'CTTAAAGCTAGGCTAGGTAC 3'
DNA representation of the path from city 2 –> city 3 –> city 4
2→3
3→4
5'GTATATCCGAGCTATTCGAGCTTAAAGCTAGGCTAGGTAC3'
3'CGATAAGCTCGAATTTCGAT5'
3 complement of 3

Figure by Jack Parker

based on Adleman, 1994, *Science*

Leonard M. Adleman, *Science*, 266: 1021-1024 (1994)

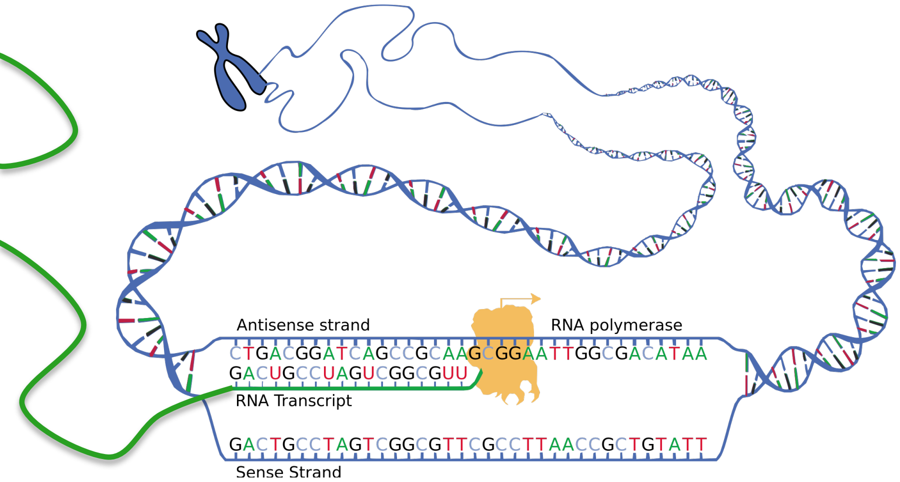
Growth of design complexity in DNA nanotechnology and DNA computing

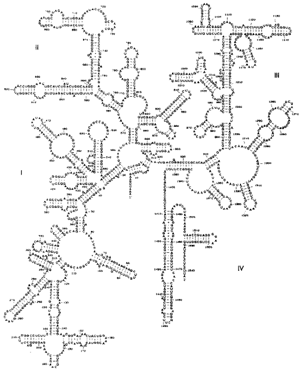


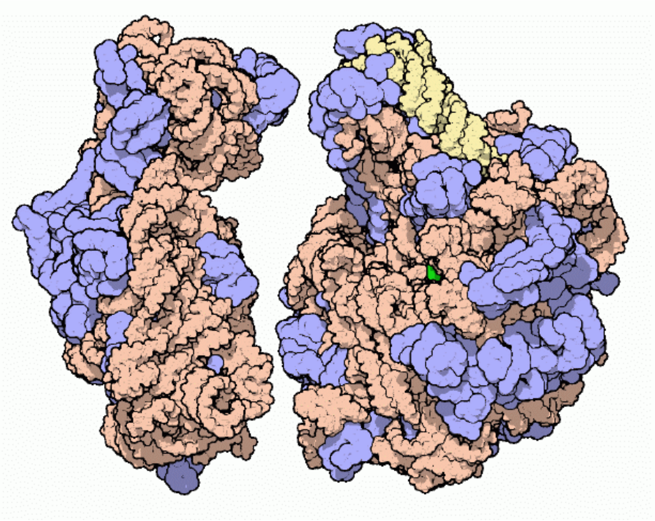
doubles every 3 years

Programmable information-based chemistry

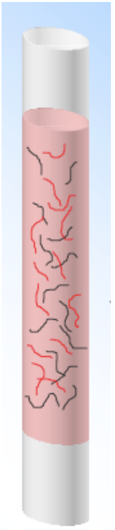
RNA biology

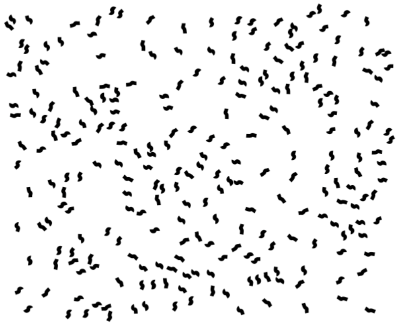


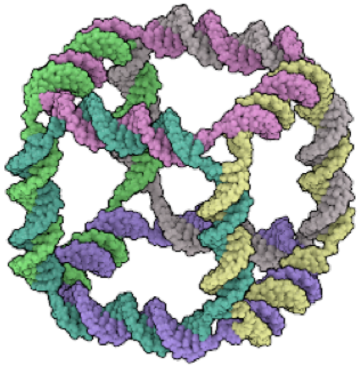


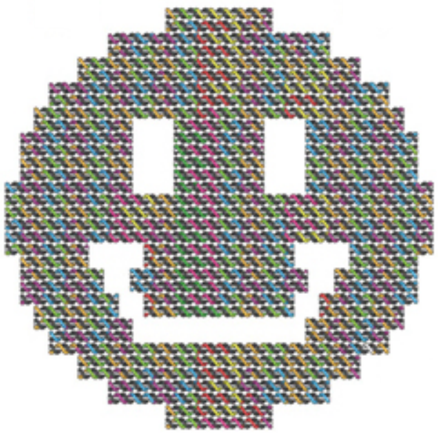


DNA nanotechnology

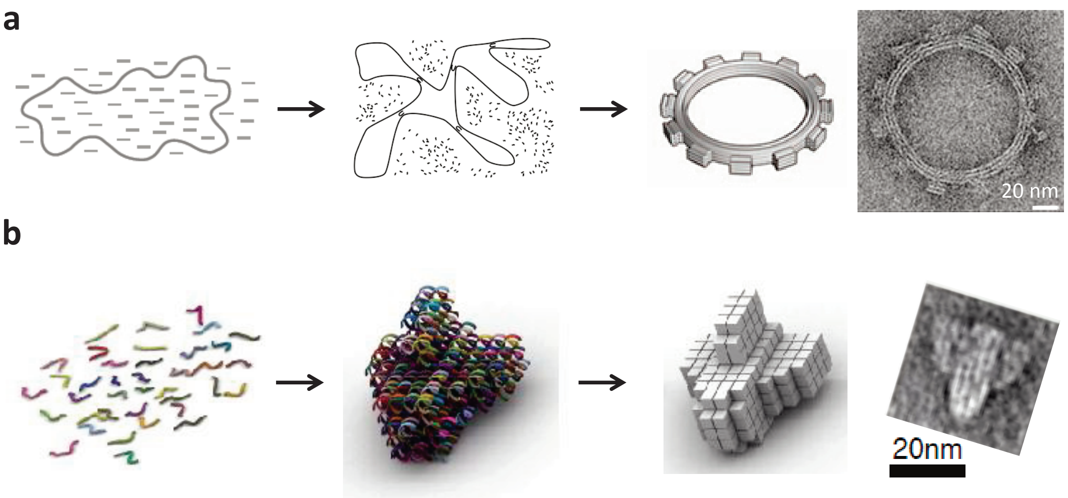






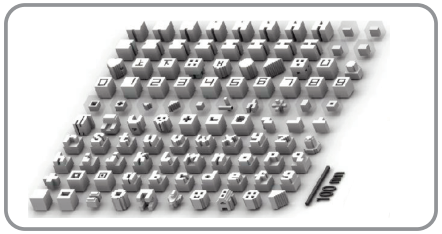


Architectures for structural molecular programs



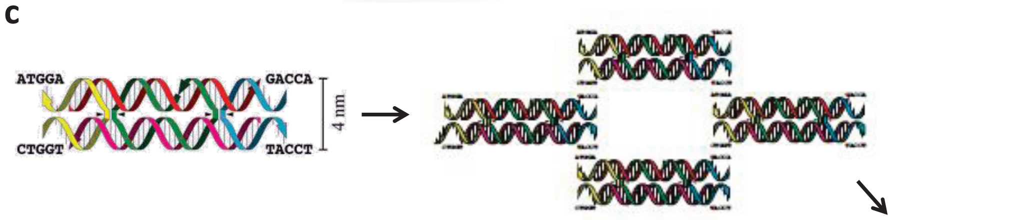
DNA origami

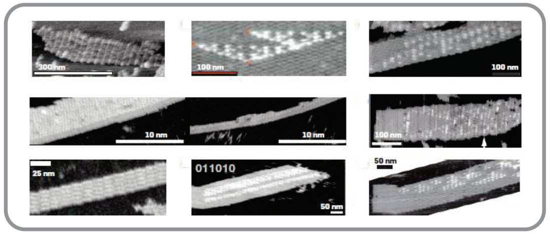
Rothemund, 2006
Shih & Lin, 2010



single-strand tiles (SST)

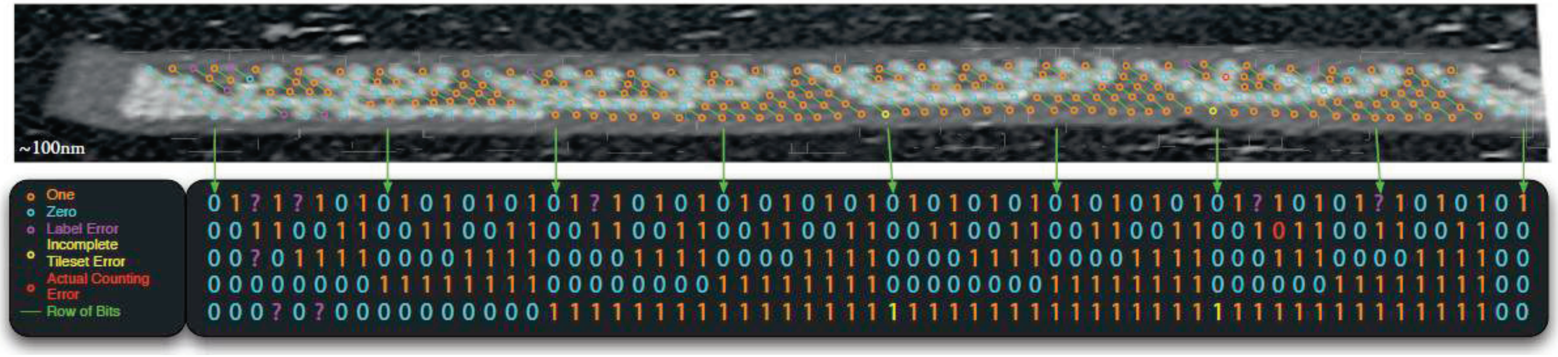
Wei et al, 2012
Ke et al, 2012





algorithmic self-assembly

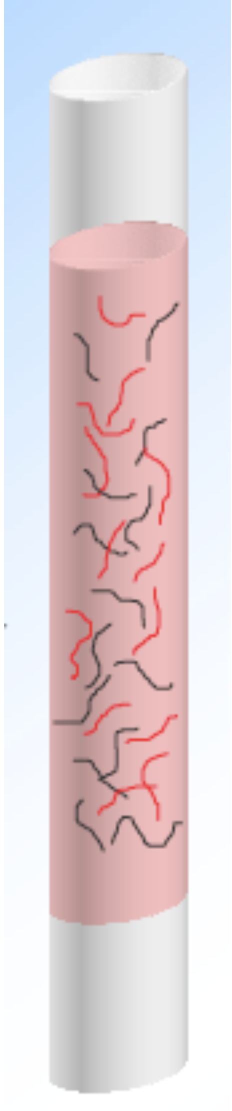
Barish et al, 2010



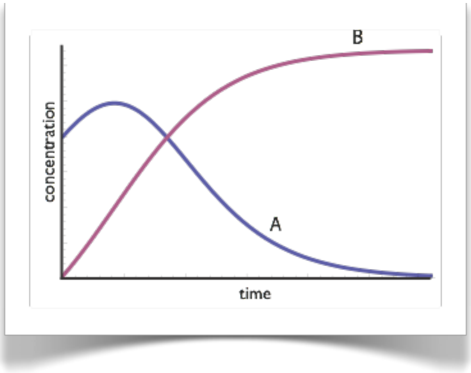
```
0 1 ? 1 ? 1 0 1 0 1 0 1 0 1 0 1 0 1 ? 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 ? 1 0 1 0 1 ? 1 0 1 0 1 0 1
0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 0 1 1 0 0 1 1 0 0 1 1 0 0
0 0 ? 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 1 1 1 0 0 0 0 1 1 1 1 0 0
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0
0 0 0 ? 0 ? 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0
```

Constantine Evans, 2014

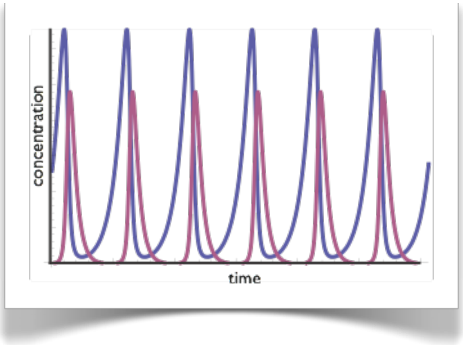
What kinds of dynamical behaviors are nucleic acid systems capable of?



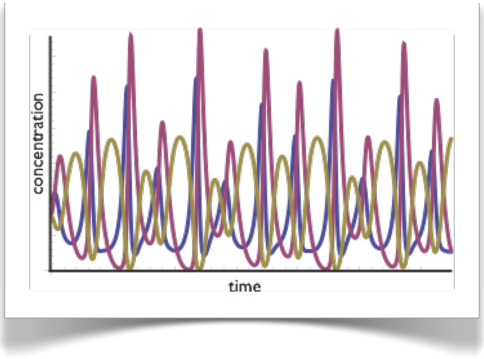
goes to completion…?



oscillates…?



does something complex…?



DNA strand displacement circuits

Bernie
Yurke

Andrew
Turberfield

Dave
Zhang

Georg
Seelig

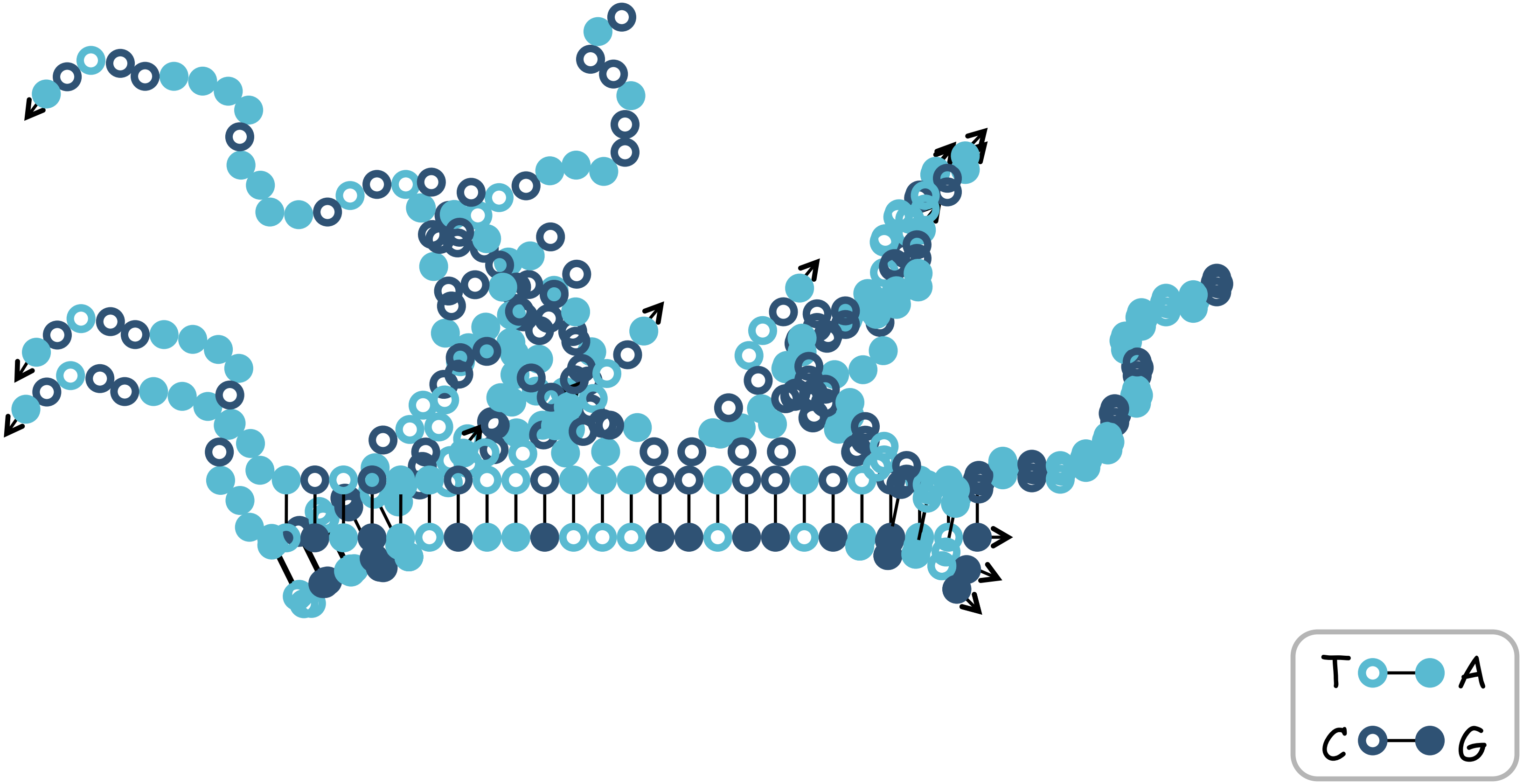
David
Soloveichik

Lulu
Qian

Shuki
Bruck

B. Yurke, A. Turberfield, A. Mills, F. Simmel, J. Neumann, *Nature*, 2000
A. Turberfield, J Mitchell, B. Yurke, A. Mills, M. Blakey, F. Simmel, *Phys Rev Lett*, 2003
Georg Seelig, David Soloveichik, Dave Zhang, Erik Winfree, *Science*, 2006
Dave Zhang, Andrew Turberfield, Bernie Yurke, Erik Winfree, *Science*, 2007
Lulu Qian & Erik Winfree, *Science*, 2011
Lulu Qian, Erik Winfree, Shuki Bruck, *Nature*, 2011

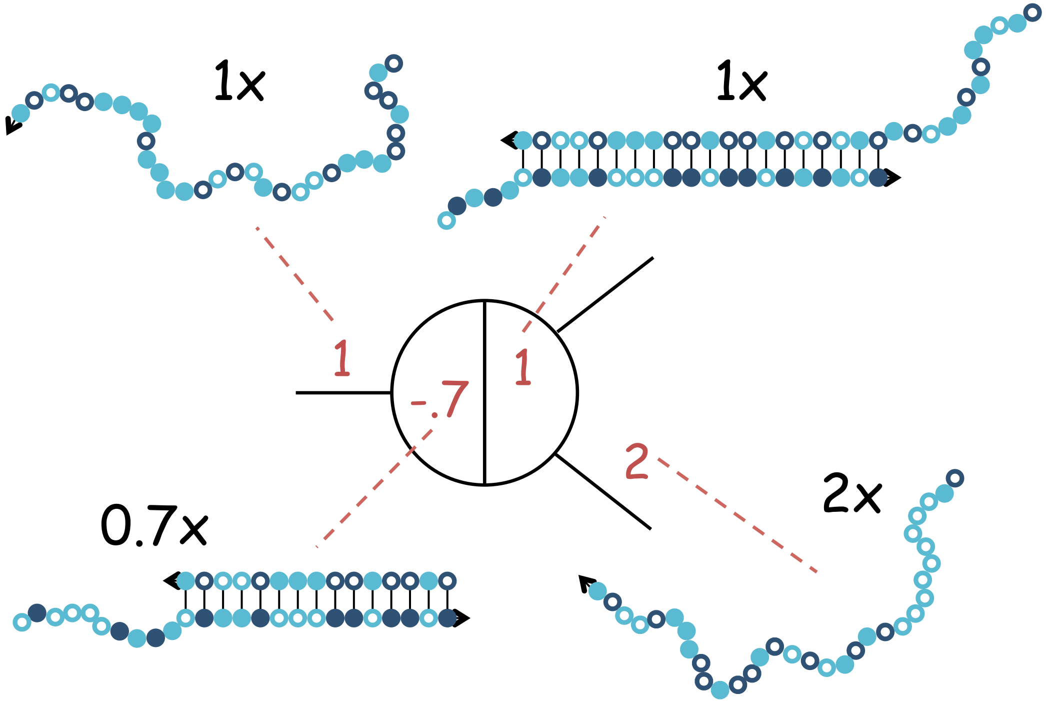
DNA strand displacement circuits



Lulu Qian, Erik Winfree, *Science*, 2011

DNA strand displacement circuits

node-wire abstraction

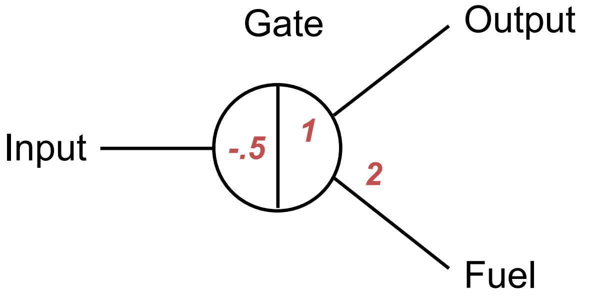


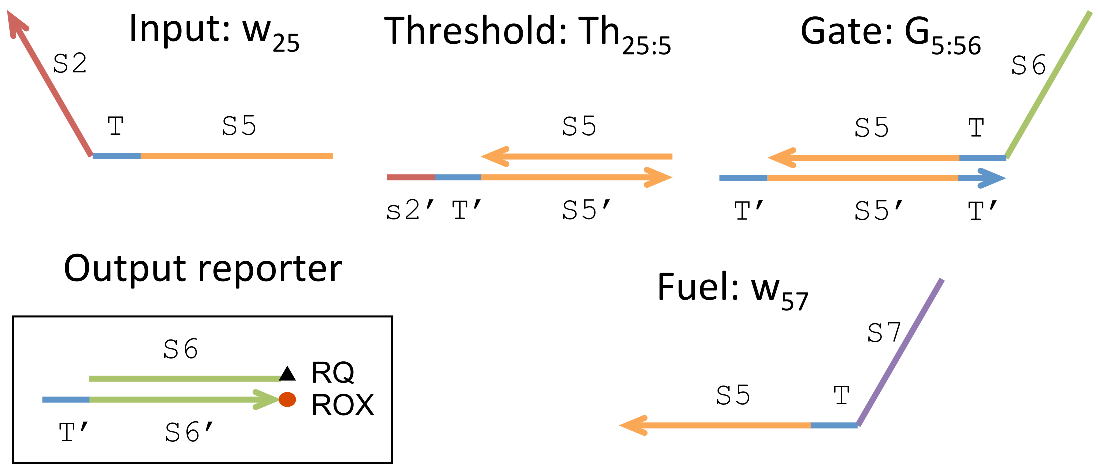
1x = 50 nM = $3\cdot10^{13}$ copies per milliliter

T ○—● A
C ○—● G

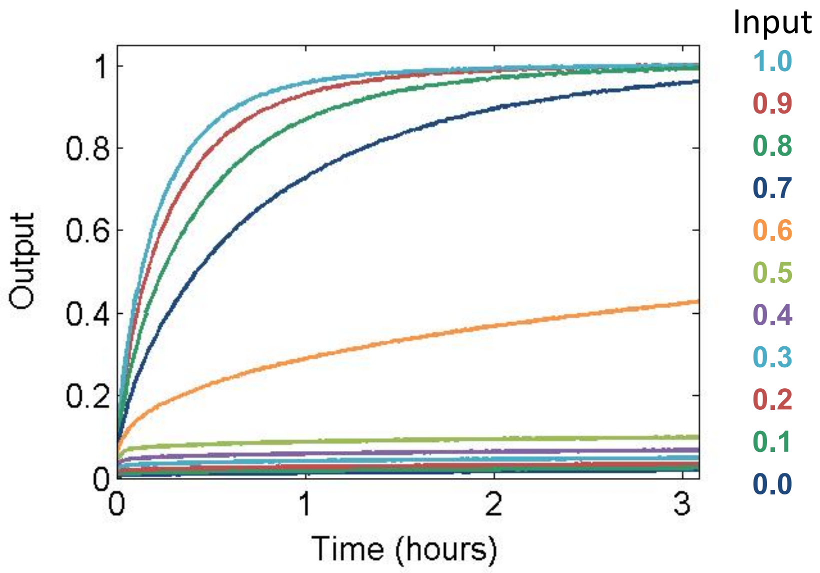
Lulu Qian, Erik Winfree, *Science*, 2011

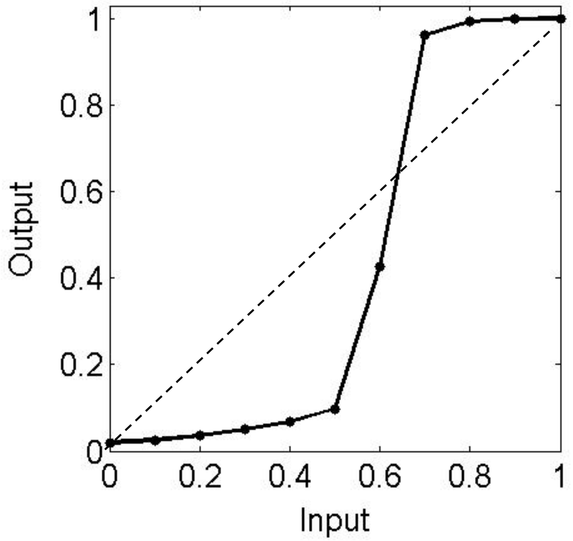
A seesaw catalyst with threshold



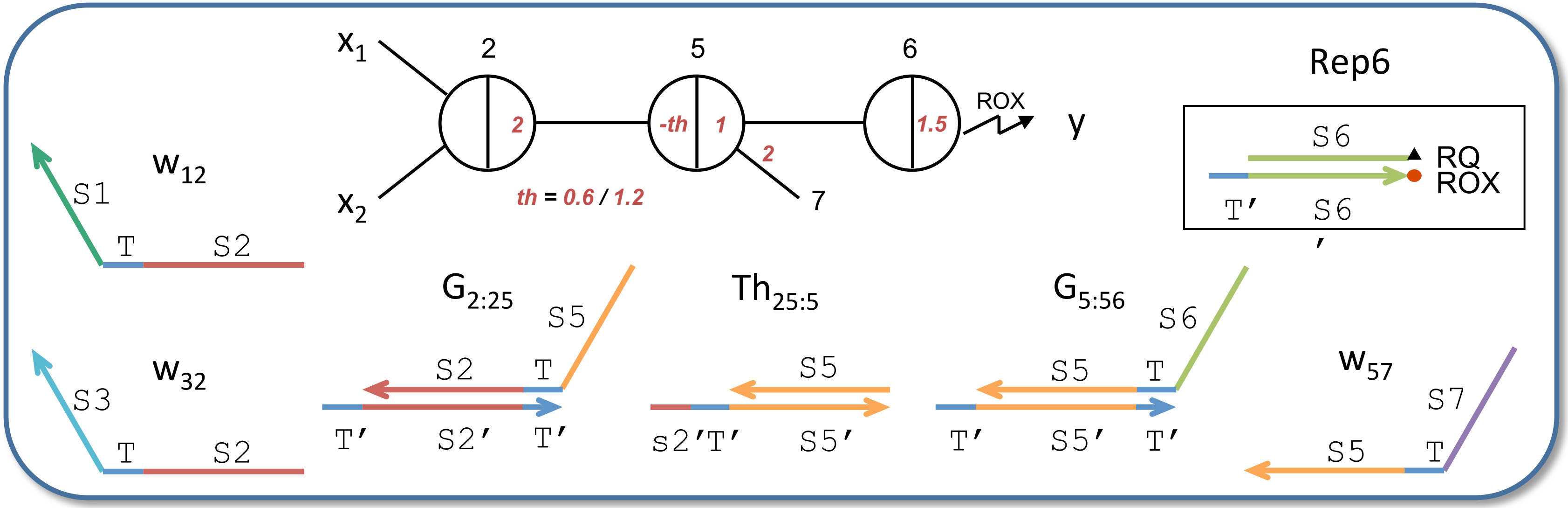


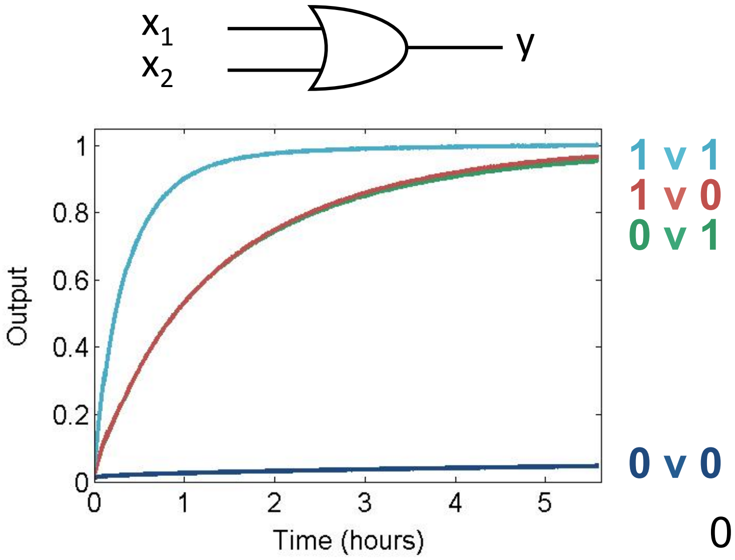
|T|=5 |Si|=15



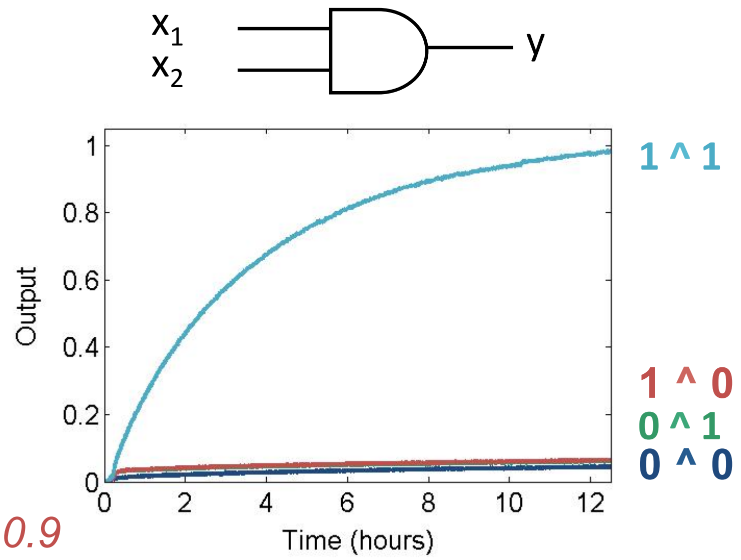


OR / AND gates

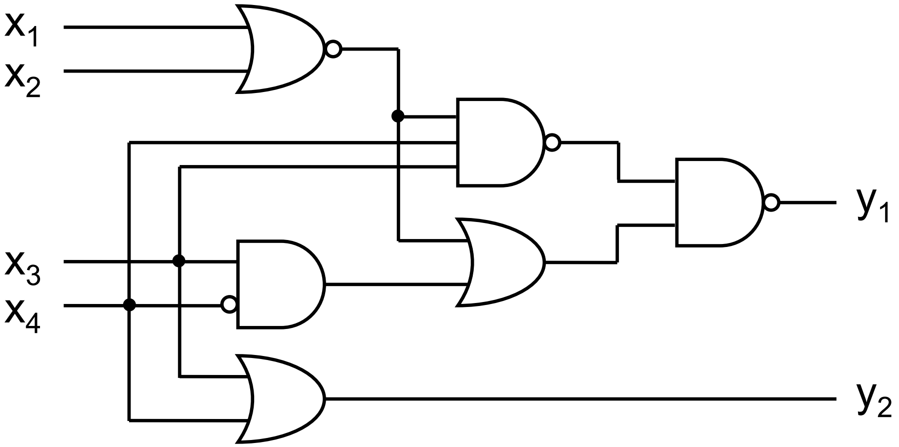




0: *0.1*   1: *0.9*



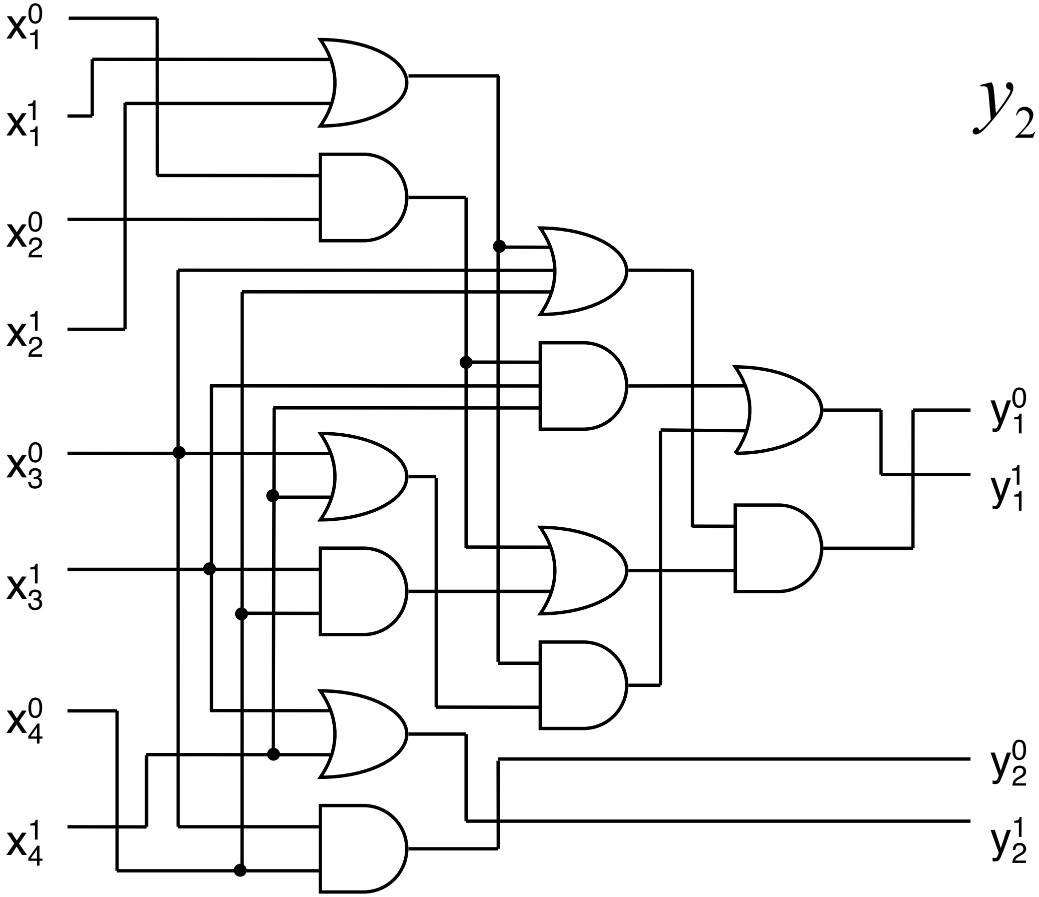
A square root circuit



$$y_2y_1 = \left\lfloor \sqrt{x_4x_3x_2x_1} \right\rfloor$$

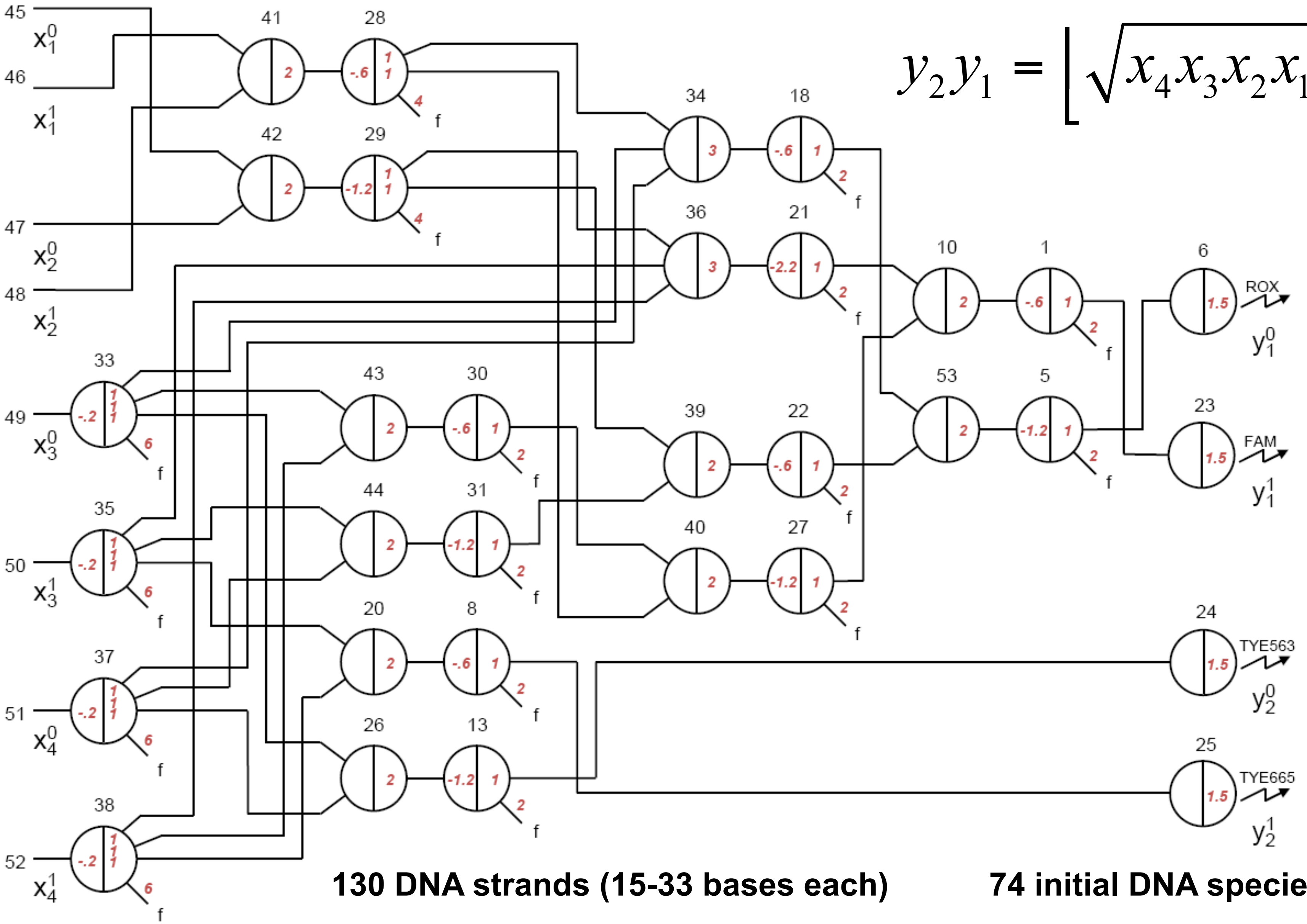
A square root circuit

$$y_2 y_1 = \left\lfloor \sqrt{x_4 x_3 x_2 x_1} \right\rfloor$$



A square root circuit

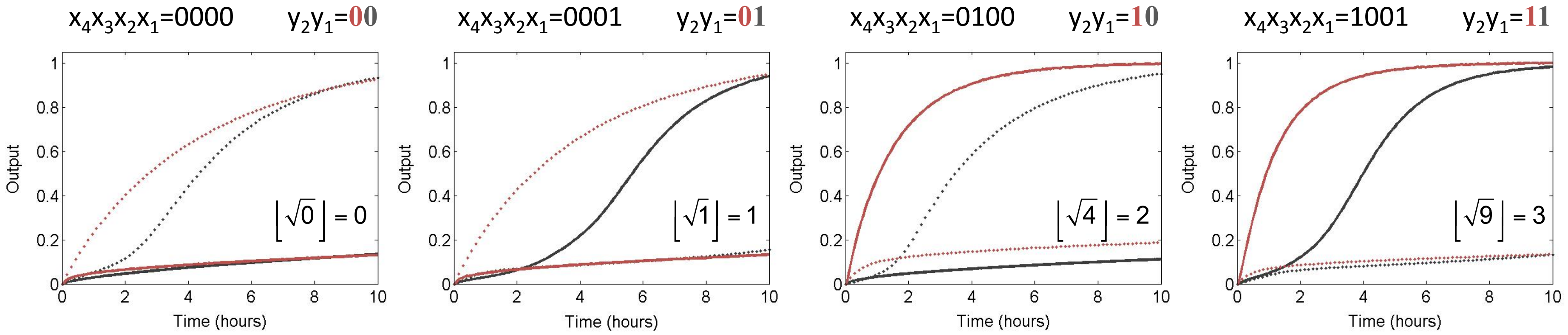
$$y_2 y_1 = \left\lfloor \sqrt{x_4 x_3 x_2 x_1} \right\rfloor$$



130 DNA strands (15-33 bases each)

74 initial DNA species

A square root circuit



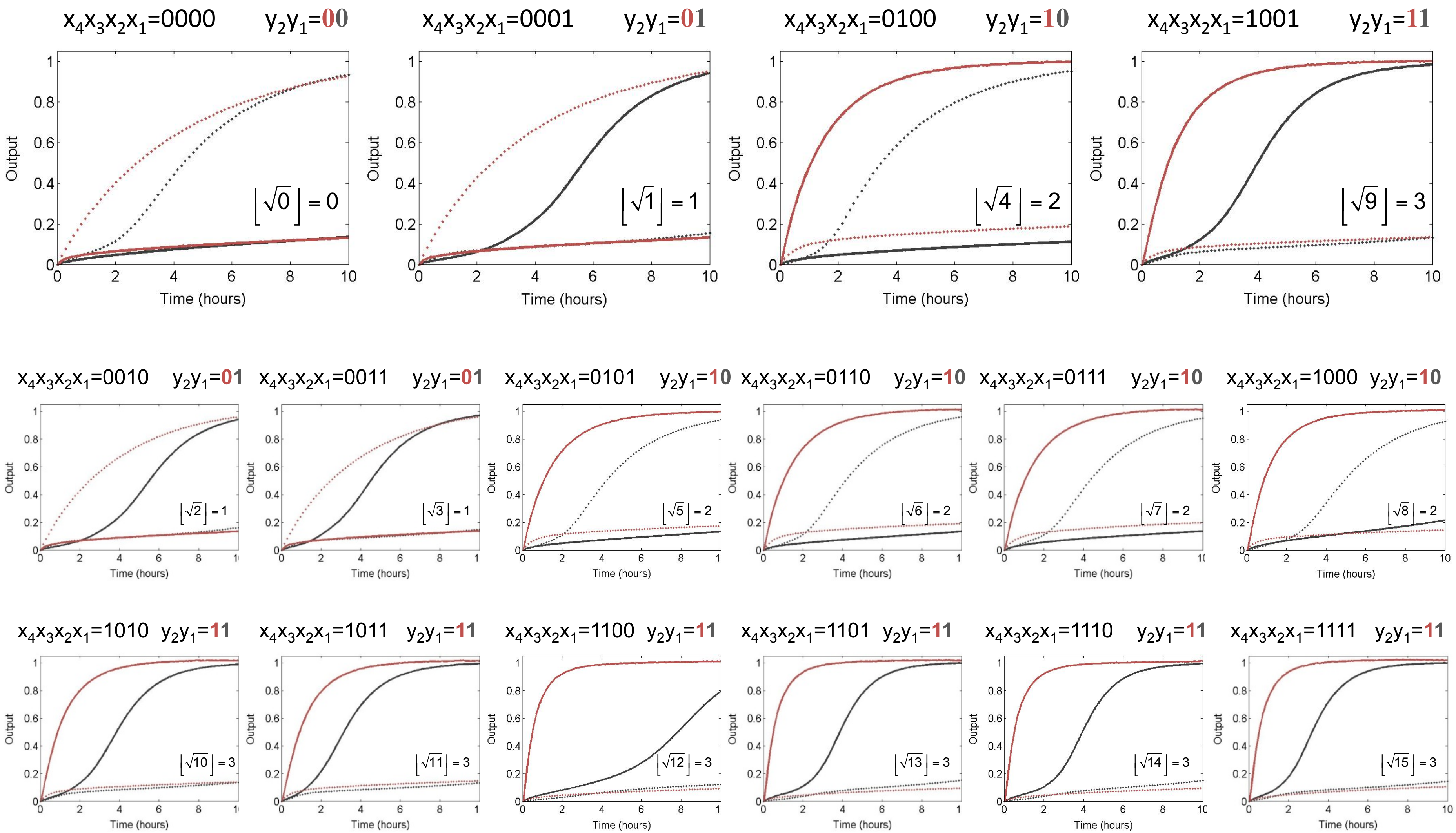
$x_4x_3x_2x_1$=0000 y_2y_1=00

$x_4x_3x_2x_1$=0001 y_2y_1=01

$x_4x_3x_2x_1$=0100 y_2y_1=10

$x_4x_3x_2x_1$=1001 y_2y_1=11

A square root circuit



$x_4x_3x_2x_1$=0000 y_2y_1=**00**

$x_4x_3x_2x_1$=0001 y_2y_1=**01**

$x_4x_3x_2x_1$=0100 y_2y_1=**10**

$x_4x_3x_2x_1$=1001 y_2y_1=**11**

$x_4x_3x_2x_1$=0010 y_2y_1=**01**

$x_4x_3x_2x_1$=0011 y_2y_1=**01**

$x_4x_3x_2x_1$=0101 y_2y_1=**10**

$x_4x_3x_2x_1$=0110 y_2y_1=**10**

$x_4x_3x_2x_1$=0111 y_2y_1=**10**

$x_4x_3x_2x_1$=1000 y_2y_1=**10**

$x_4x_3x_2x_1$=1010 y_2y_1=**11**

$x_4x_3x_2x_1$=1011 y_2y_1=**11**

$x_4x_3x_2x_1$=1100 y_2y_1=**11**

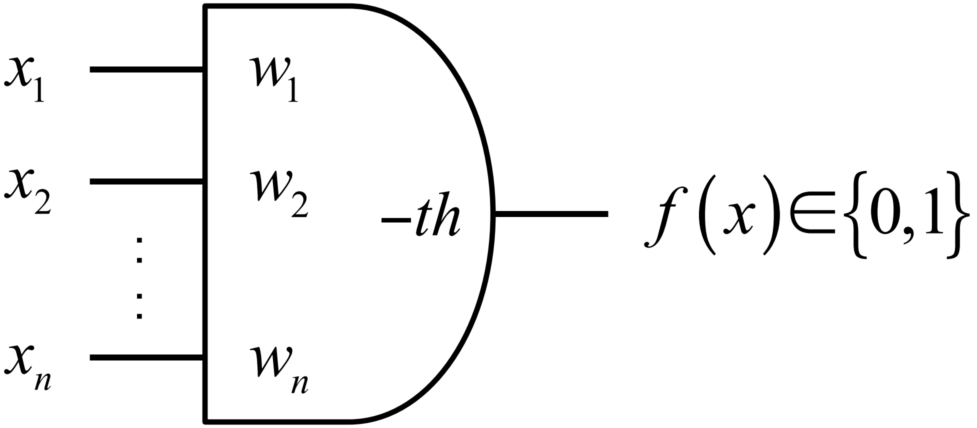
$x_4x_3x_2x_1$=1101 y_2y_1=**11**

$x_4x_3x_2x_1$=1110 y_2y_1=**11**

$x_4x_3x_2x_1$=1111 y_2y_1=**11**

Linear Threshold (LT) gate

$x_i \in \{0,1\}$

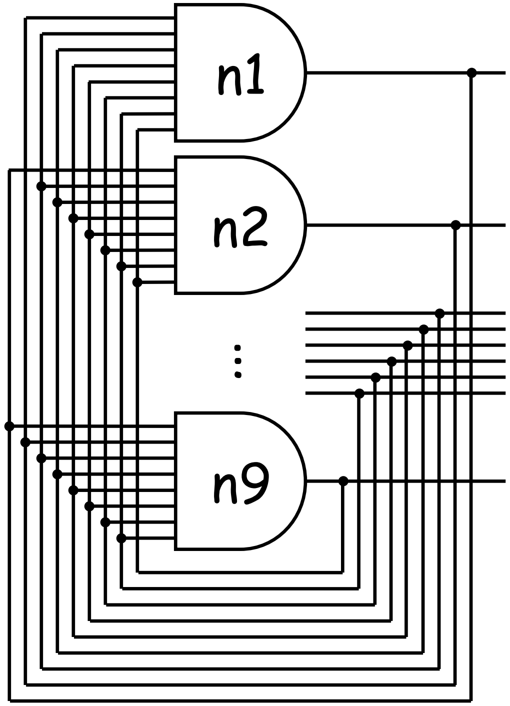


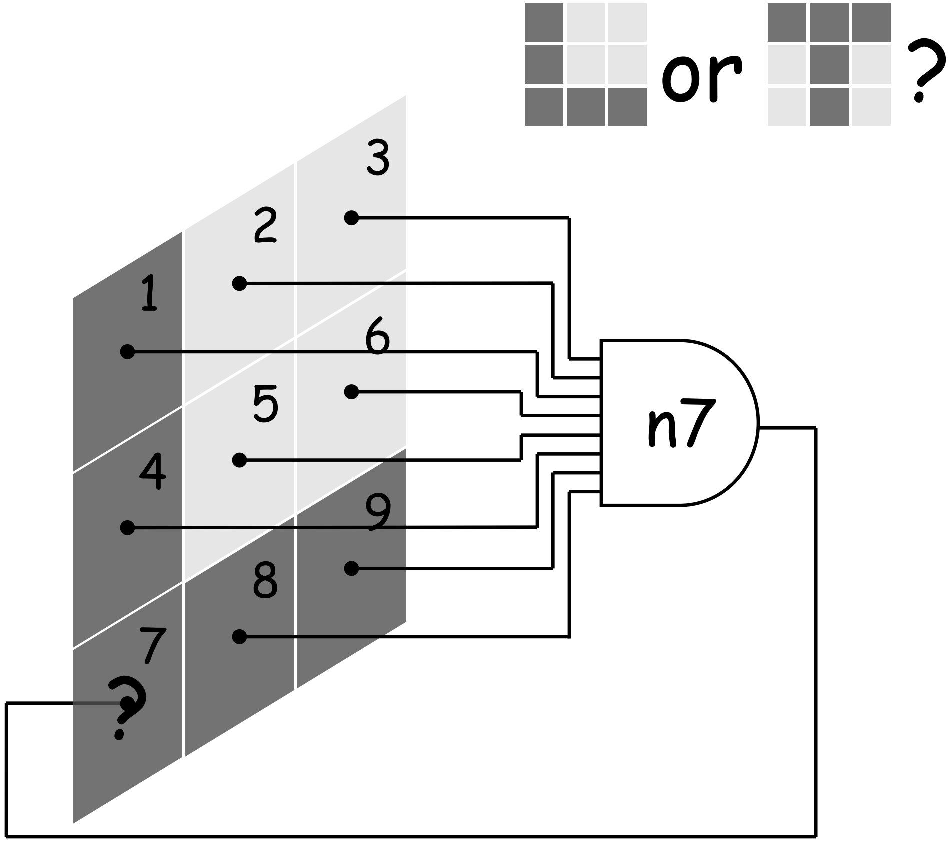
$$F(x) = -th + \sum_{i=1}^{n} w_i x_i$$

$$f(x) = \begin{cases} 1 & \text{if } F(x) \geq 0 \\ 0 & \text{if } F(x) < 0 \end{cases}$$

Artificial neural network

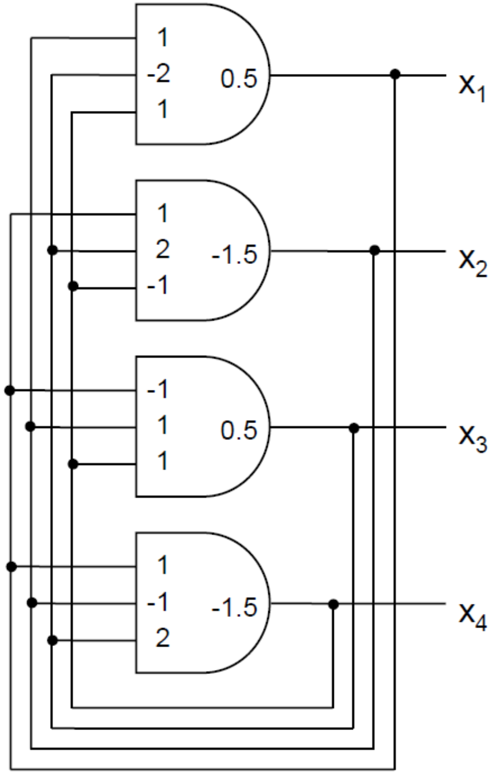
Hopfield associative memory





J.J. Hopfield, **Neural networks and physical systems with emergent collective computational abilities**, *PNAS*, 1982.

Hopfield associative memory



A “read your mind” game

Q1: Did the scientist study neural networks?

Q2: Was the scientist British?

Q3: Was the scientist born in the 20th century?

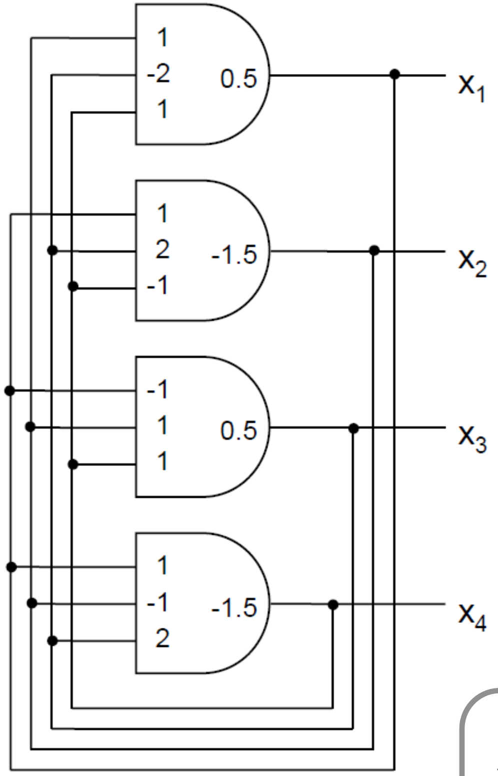
Q4: Was the scientist a mathematician?

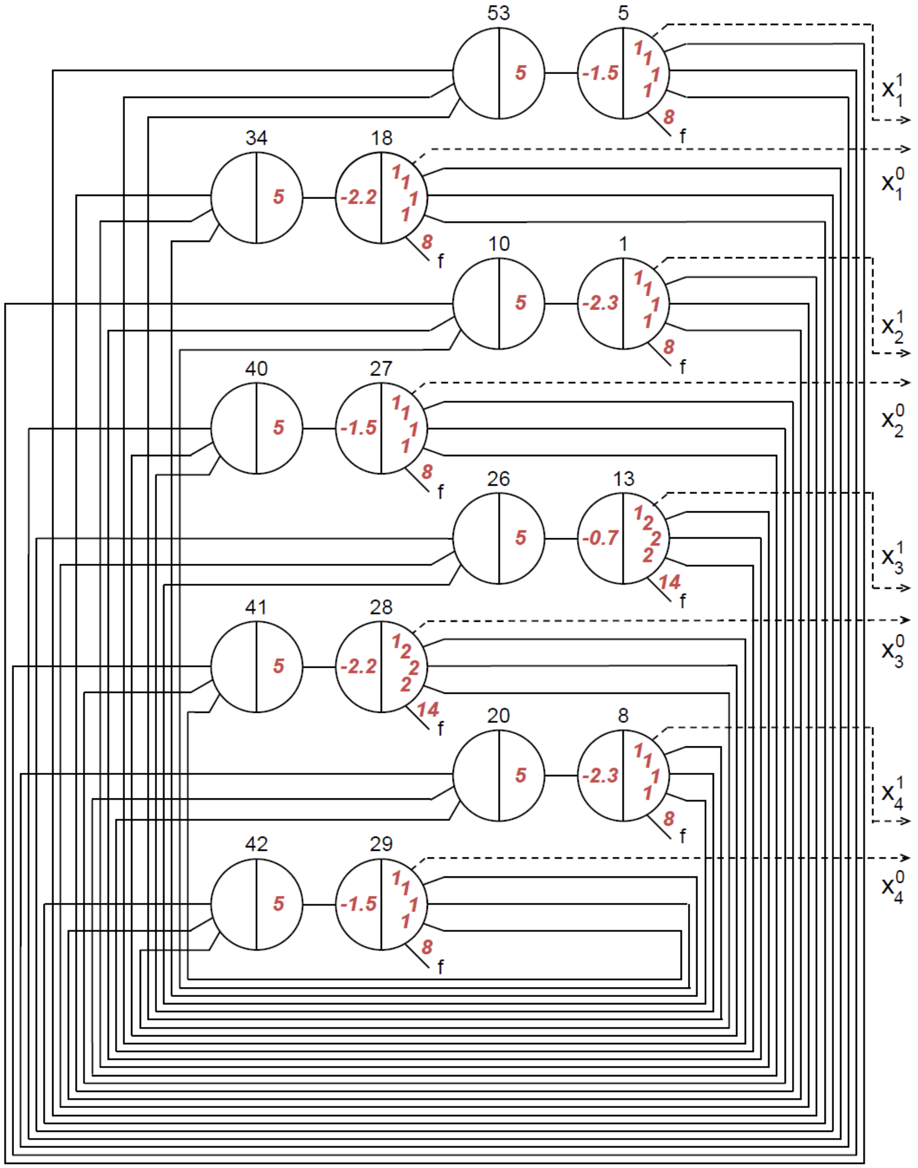
Answers: Yes (1), No (0) or I don’t know (?)

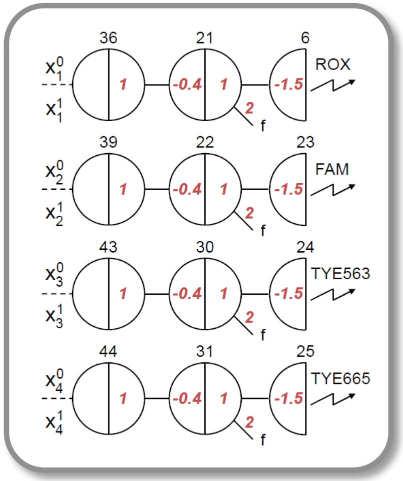
0 1 1 0	**Rosalind Franklin**
1 1 1 1	**Alan Turing**
0 0 1 1	**Claude Shannon**
1 0 0 0	**Santiago Ramon y Cajal**

Lulu Qian, Erik Winfree, Jehoshua Bruck, *Nature*, 2011

Hopfield associative memory



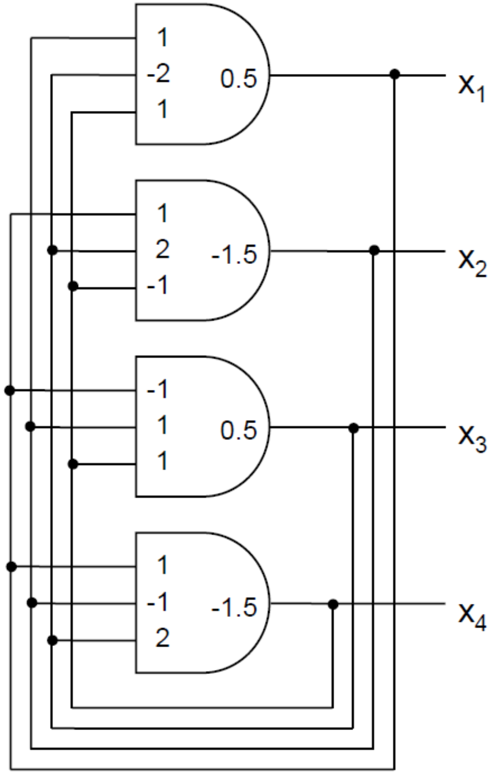


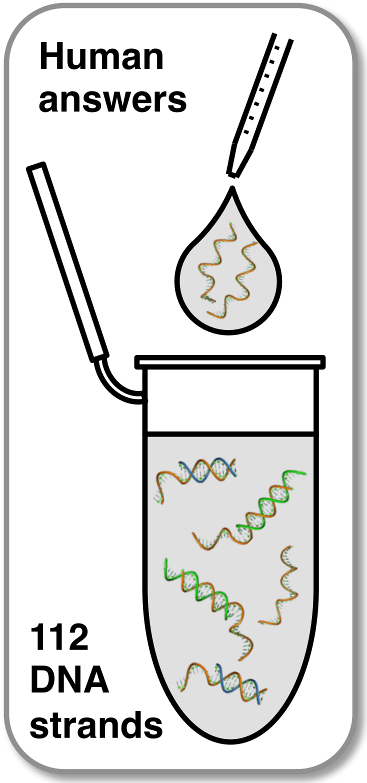


112 DNA strands (15-33 bases each)

72 initial DNA species

Hopfield associative memory





A "read your mind" game

Q1: Did the scientist study neural networks?

Q2: Was the scientist British?

Q3: Was the scientist born in the 20th century?

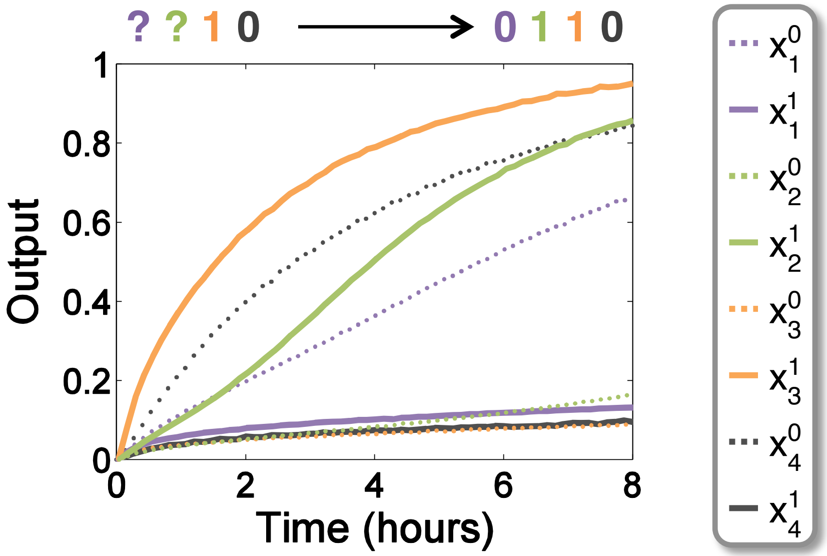
Q4: Was the scientist a mathematician?

Answers: Yes (1), No (0) or I don't know (?)

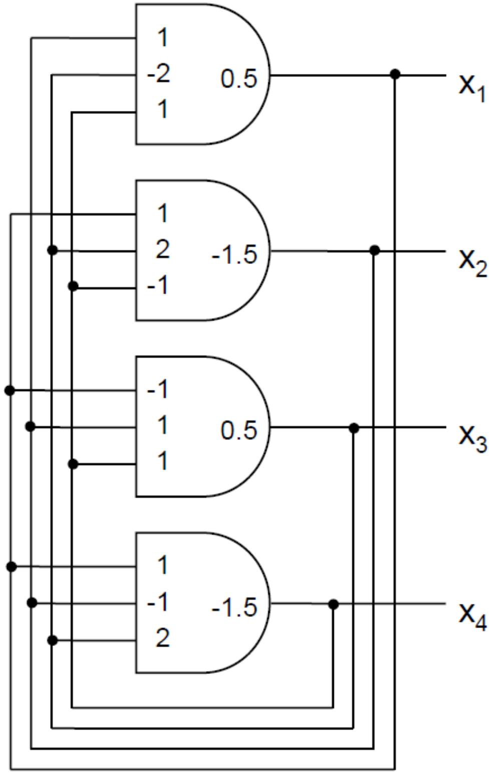
0 1 1 0	Rosalind Franklin
1 1 1 1	Alan Turing
0 0 1 1	Claude Shannon
1 0 0 0	Santiago Ramon y Cajal

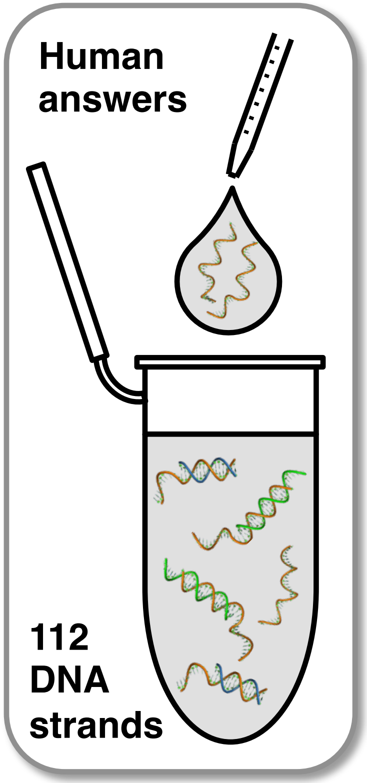
Human: The scientist I am thinking of was born in the 20th century (1) but was not a mathematician (0).

DNA associative memory: The scientist you are thinking of didn't study neural networks (0) and was British (1), so I guess she is Rosalind Franklin (0 1 1 0).



Hopfield associative memory





Human answers

112 DNA strands

A “read your mind” game

Q1: Did the scientist study neural networks?

Q2: Was the scientist British?

Q3: Was the scientist born in the 20th century?

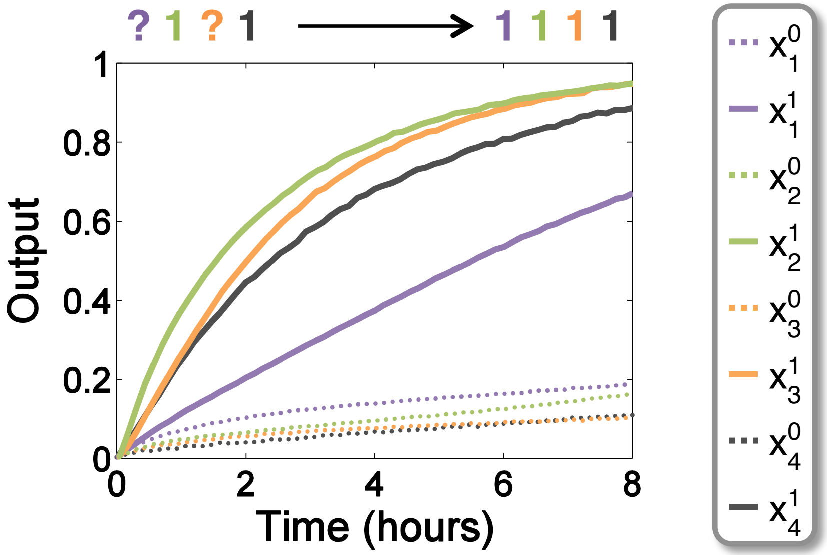
Q4: Was the scientist a mathematician?

Answers: Yes (1), No (0) or I don’t know (?)

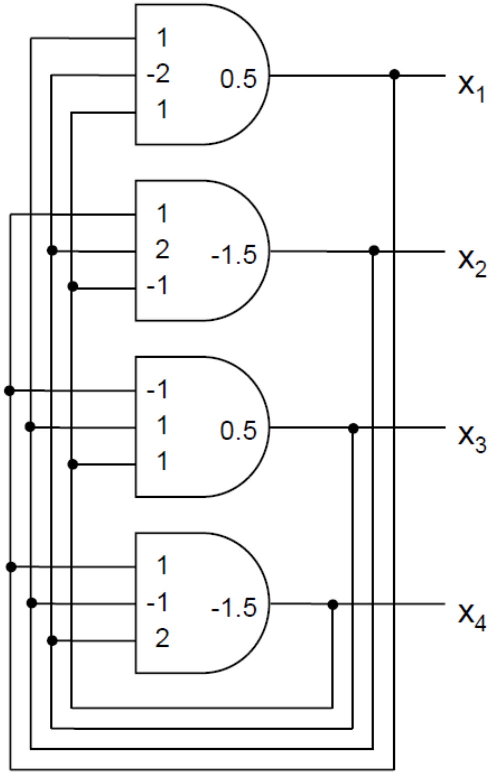
0	1	1	0	Rosalind Franklin
1	1	1	1	Alan Turing
0	0	1	1	Claude Shannon
1	0	0	0	Santiago Ramon y Cajal

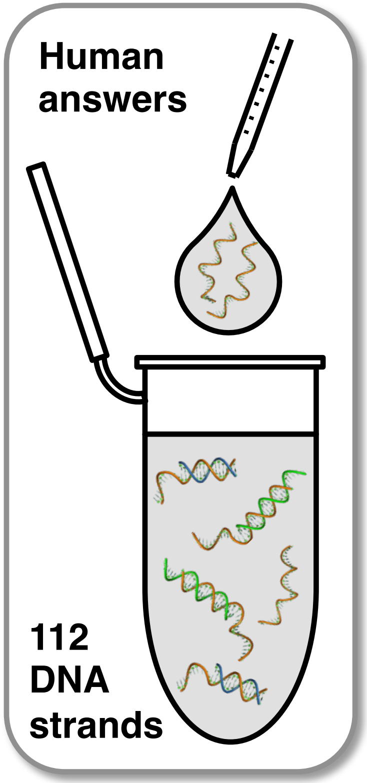
Human: The scientist I am thinking of was British (1) and a mathematician (1).

DNA associative memory: The scientist you are thinking of studied neural networks (1) and was born in the 20th century (1), so I guess he is Alan Turing (1 1 1 1).



Hopfield associative memory





Human answers

112 DNA strands

A “read your mind” game

Q1: Did the scientist study neural networks?

Q2: Was the scientist British?

Q3: Was the scientist born in the 20th century?

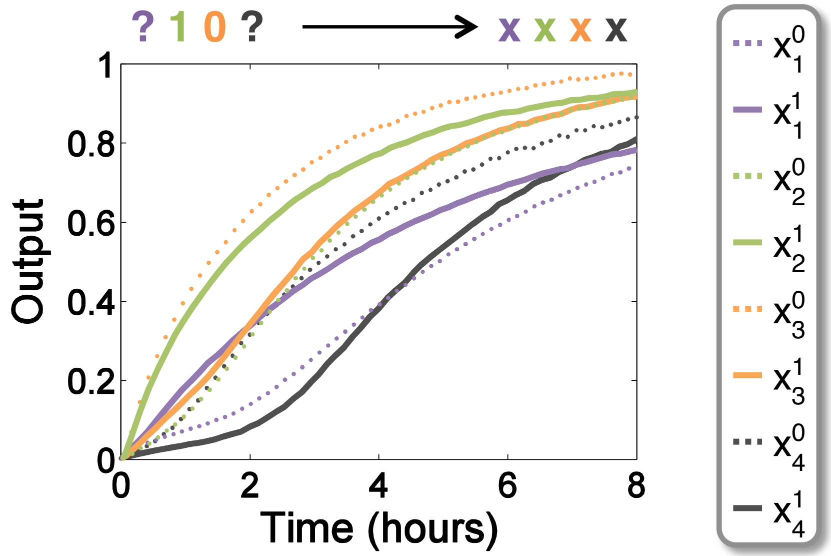
Q4: Was the scientist a mathematician?

Answers: Yes (1), No (0) or I don’t know (?)

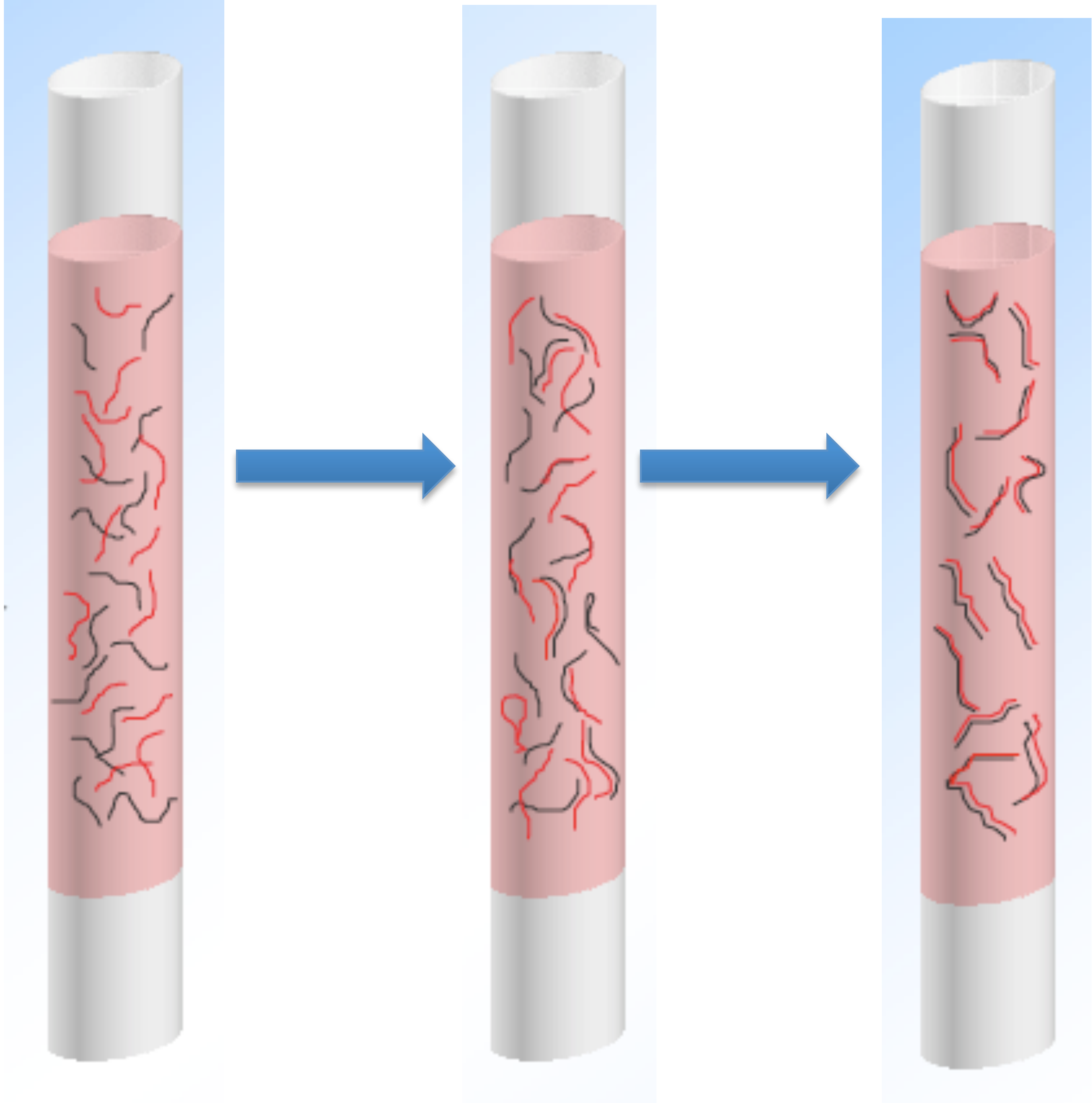
0 1 1 0	Rosalind Franklin
1 1 1 1	Alan Turing
0 0 1 1	Claude Shannon
1 0 0 0	Santiago Ramon y Cajal

Human: The scientist I am thinking of was British (1) but was not born in the 20th century (0).

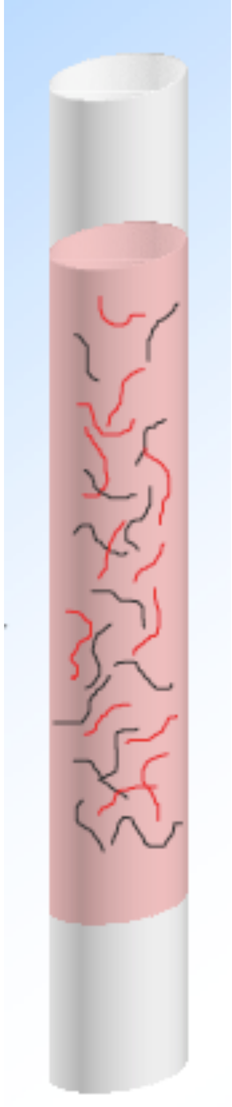
DNA associative memory: There’s wrong information that you provided, I cannot recognize this scientist (x x x x).



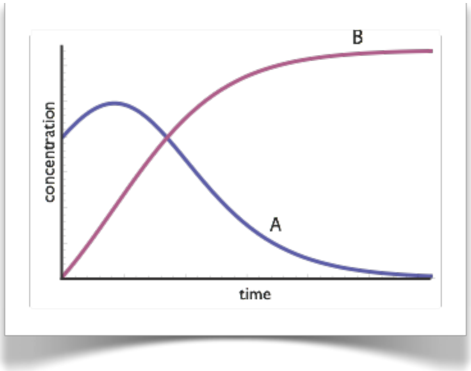
Toss a mixture of DNA in a test tube… What can it do? What can’t it do?



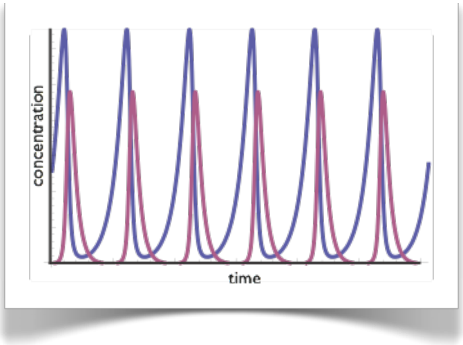
What kinds of dynamical behaviors are nucleic acid systems capable of?



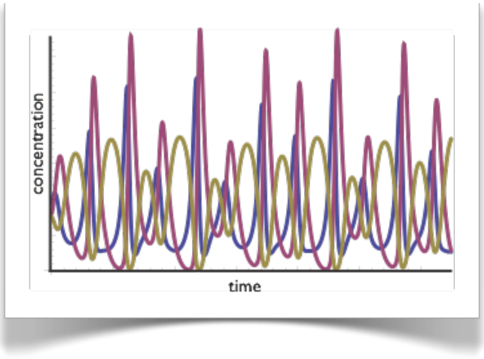
goes to completion…?



oscillates…?

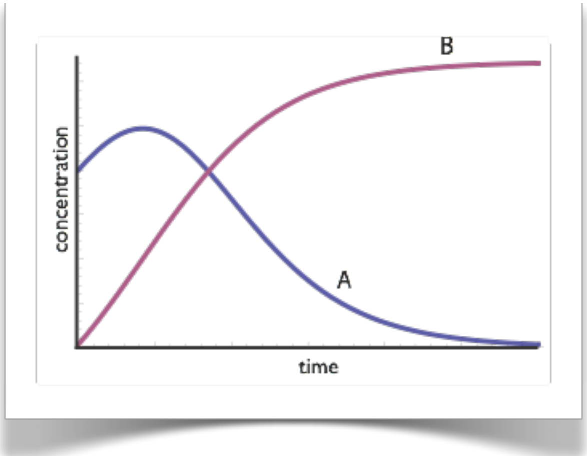


does something complex…?



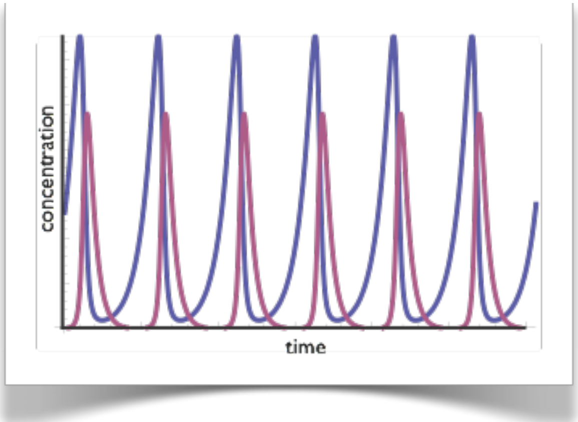
Some behaviors of simple chemical systems (mass action, well-mixed)

stabilization



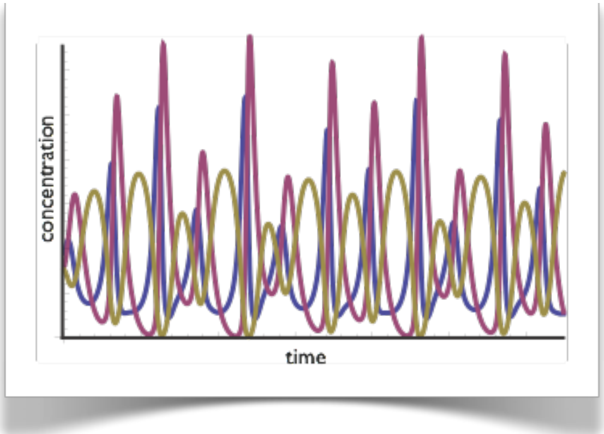
$$\begin{array}{rcl} A & \longrightarrow & B \\ A+B & \longrightarrow & B \\ A & \longrightarrow & 2A \end{array}$$

oscillation



$$\begin{array}{rcl} A & \longrightarrow & 2A \\ A+B & \longrightarrow & 2B \\ B & \longrightarrow & \end{array}$$

chaos



$$\begin{array}{rcl} A & \xrightarrow{30} & 2A \\ 2A & \xrightarrow{0.5} & A \\ B+A & \xrightarrow{1} & 2B \\ B & \xrightarrow{10} & \\ A+C & \xrightarrow{1} & \\ C & \xrightarrow{16.5} & 2C \\ 2C & \xrightarrow{0.5} & C \end{array}$$

open systems: no conservation of mass/energy

Some behaviors of simple chemical systems (mass action, well-mixed)

Chemical reaction networks (CRNs) as a programming language…
…use the formalism prescriptively, not descriptively…

- Mathematically equivalent to Petri Nets, Vector Addition Systems, etc.

  e.g. Petri (1966); Karp (1969); Goss & Peccaud (1998)

- Can implement arbitrary sequential digital logic circuits

  e.g. Magnasco (1997)

- Can implement (nearly) arbitrary dynamical system behaviors

  e.g. Korzuhin (1967); Klonowski (1983)

- Can can simulate Turing machine computations with arbitrarily small error

  e.g. Soloveichik, Cook, Winfree, Bruck (2008); Cummings, Doty, Soloviechik (2014)

- Etc. etc…

But does every formal chemical reaction network exist?

Compiling CRNs into DNA

CRN program:

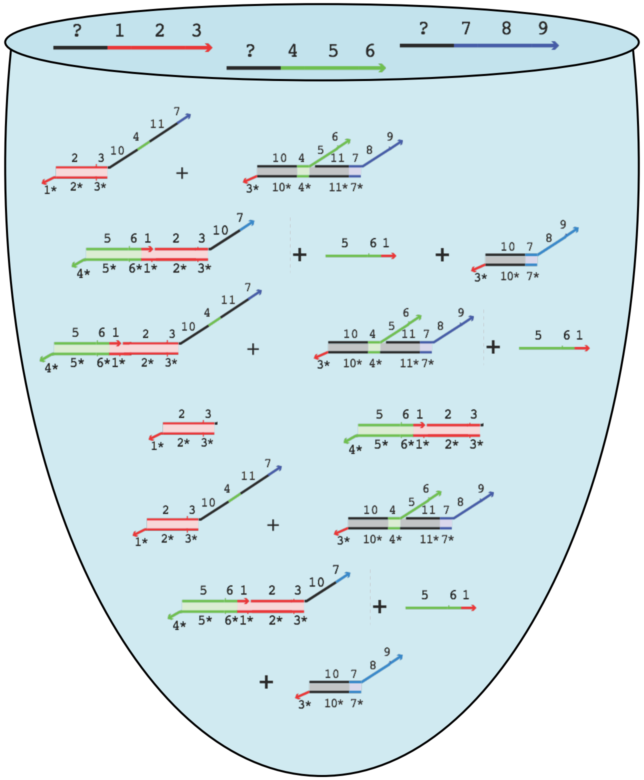
$$A \xrightarrow{30} 2A$$
$$2A \xrightarrow{0.5} A$$
$$B+A \xrightarrow{1} 2B$$
$$B \xrightarrow{10}$$
$$A+C \xrightarrow{1}$$
$$C \xrightarrow{16.5} 2C$$
$$2C \xrightarrow{0.5} C$$

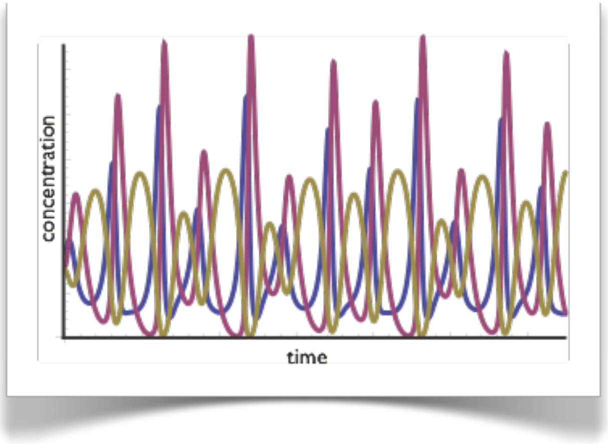
Initial conditions:

5 A's

3 B's

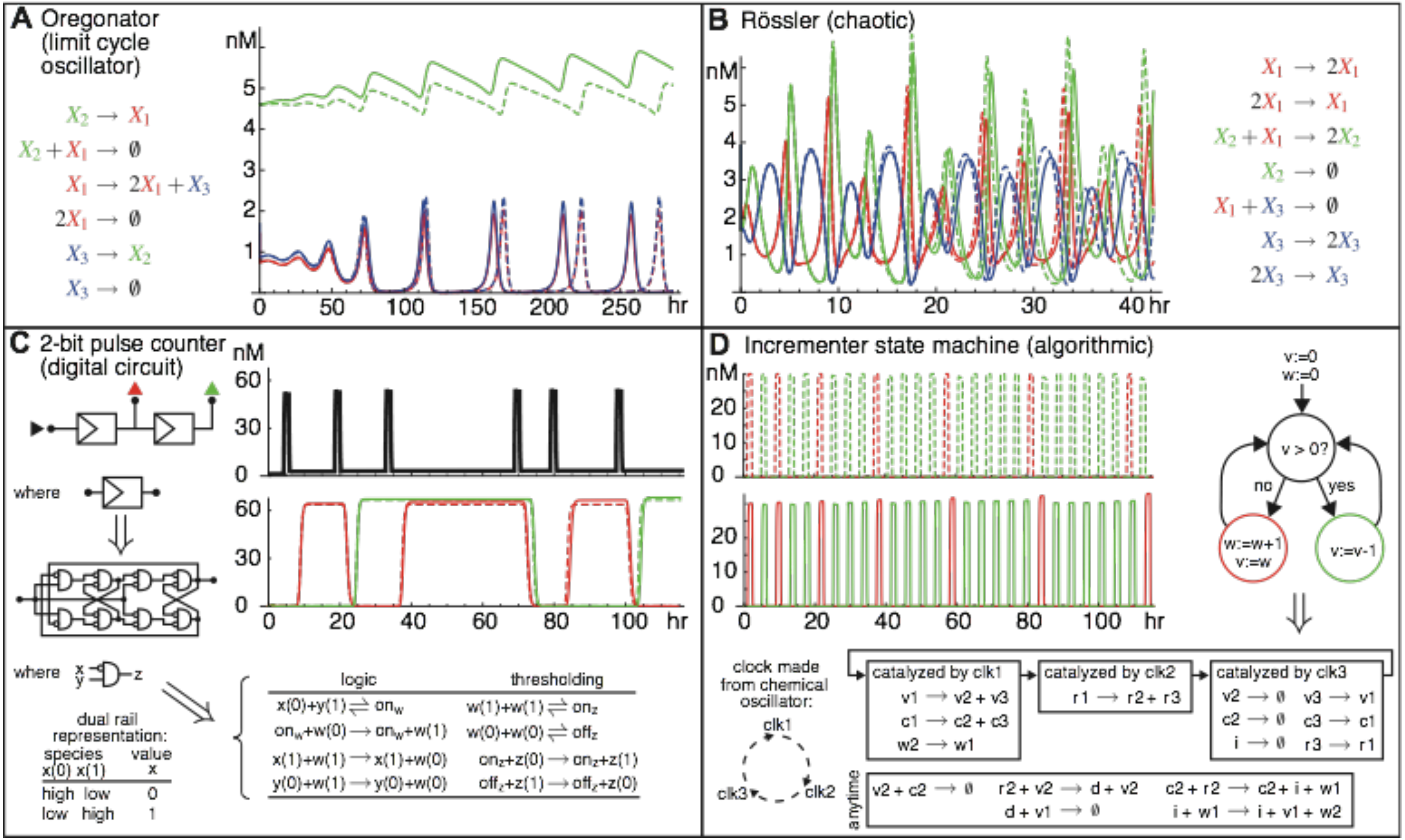
10 C's



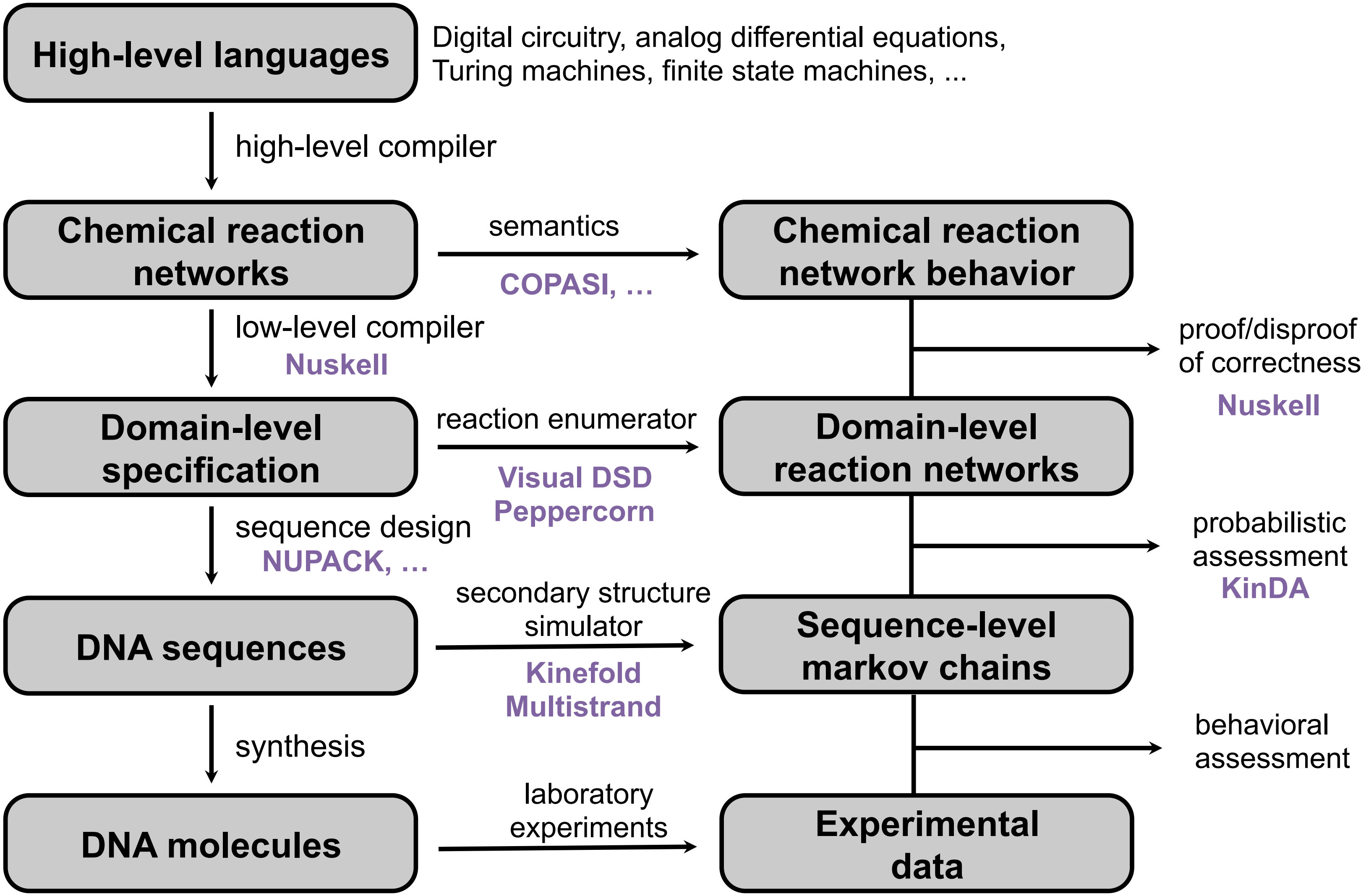


Soloveichik, Seelig, Winfree, PNAS, 2010

Some behaviors of simple chemical systems (mass action)



A compiler & verification hierarchy



High-level languages

Digital circuitry, analog differential equations, Turing machines, finite state machines, ...

high-level compiler

Chemical reaction networks

semantics

COPASI, …

Chemical reaction network behavior

low-level compiler

Nuskell

proof/disproof of correctness

Nuskell

Domain-level specification

reaction enumerator

Visual DSD

Peppercorn

Domain-level reaction networks

sequence design

NUPACK, …

probabilistic assessment

KinDA

DNA sequences

secondary structure simulator

Kinefold

Multistrand

Sequence-level markov chains

synthesis

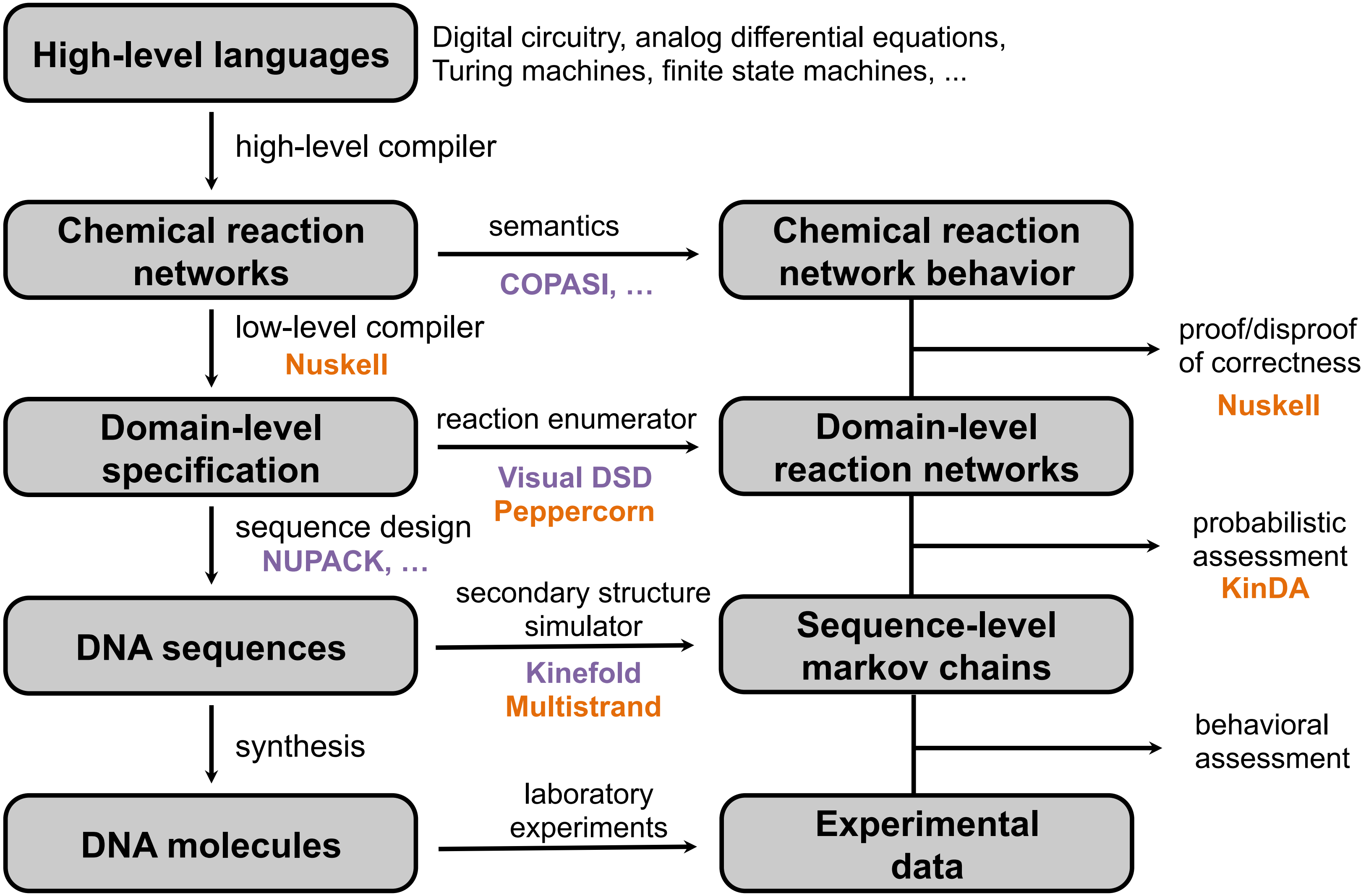
behavioral assessment

DNA molecules

laboratory experiments

Experimental data

A compiler & verification hierarchy



High-level languages

Digital circuitry, analog differential equations, Turing machines, finite state machines, ...

high-level compiler

Chemical reaction networks

semantics

COPASI, …

Chemical reaction network behavior

low-level compiler

Nuskell

proof/disproof of correctness

Nuskell

Domain-level specification

reaction enumerator

Visual DSD

Peppercorn

Domain-level reaction networks

sequence design

NUPACK, …

probabilistic assessment

KinDA

DNA sequences

secondary structure simulator

Kinefold

Multistrand

Sequence-level markov chains

synthesis

behavioral assessment

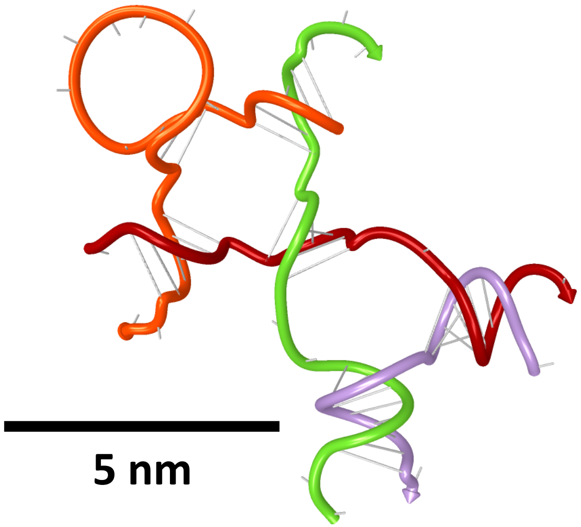
DNA molecules

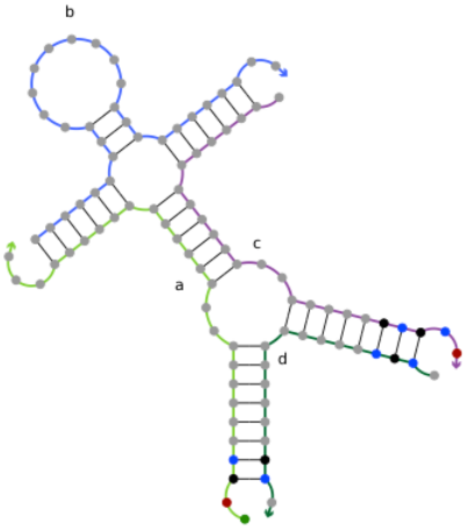
laboratory experiments

Experimental data

Abstractions for molecular programs

NUPACK: rigorous thermodynamic and kinetic sequence design and analysis





```
material = rna
temperature[C] = 23.0
sodium[M] = 1.0

structure stickfigure =
    U2D8(U2D6(D6(U3+
       )D3U9D6(U2+U1))U2D8(U2+U1))U1

domain a = AUGC N23
domain b = N29
domain c = N20 GCGCU
domain d = N18

stickfigure.seq = a b c d
```

sequence structure

$$\Delta G(\phi, s) = \sum_{\text{loop}\in s} \Delta G(\text{loop})$$

Equilibrium probability

$$p(\phi, s) = \frac{1}{Q(\phi)} e^{-\Delta G(\phi,s)/k_B T}$$

Partition function

$$Q(\phi) = \sum_{s\in\Gamma} e^{-\Delta G(\phi,s)/k_B T}$$

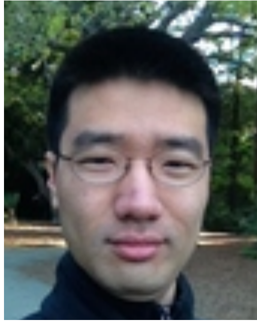
Average number of incorrectly paired nucleotides, given target s

Dynamic programming: $\Theta(N^3)$

$$n(\phi, s) \quad = \quad N - \sum_{1\leq i\leq N} P_{i,j}(\phi) S_{i,j}(s)$$

Zadeh, Steenberg, Bois, Wolfe, Pierce, Khan, Dirks, Pierce, J Comp Chem, 2011

Wolfe, Pierce, ACS Synthetic Biology, 2015



Nuskell: A verifying compiler for domain-level DNA implementations of chemical reaction networks

"Compiling and verifying DNA-based chemical reaction network implementations"
Seung Woo Shin, Master's Thesis (2011)

"A bisimulation approach to verification of molecular implementations of formal chemical reaction networks"
Qing Dong, Master's Thesis (2012)

"Verifying Chemical Reaction Network Implementations: A Pathway Decomposition Approach"
Seung Woo Shin, Chris Thachuk, Erik Winfree, arXiv:1411.0782 (2014)

Verification pipeline

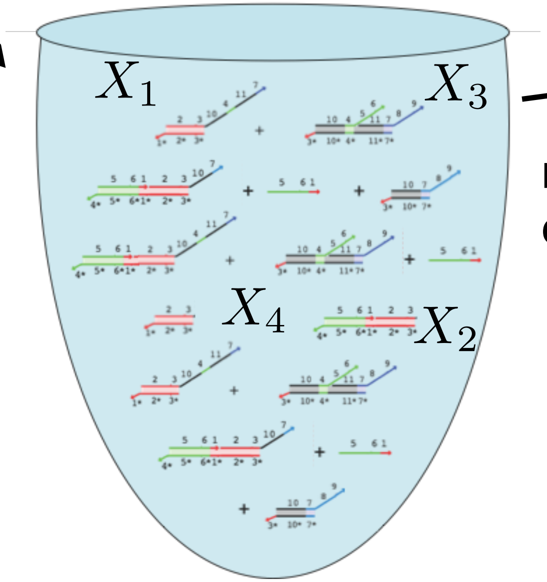
Target CRN

$$X3 + X4 \xrightarrow{k_1} X5$$
$$X5 \xrightarrow{k_2} X1$$
$$X1 + X2 \xrightarrow{k_3} X3$$

$$\overset{?}{=}$$

compile

Claimed implementation



reaction enumerator

Implementation CRN

$$X3 + g1 \xrightarrow{k_4} i + g2$$
$$i + g2 \xrightarrow{k_5} X3 + g1$$
$$i + x4 \xrightarrow{k_6} j + w1$$
$$j + g3 \xrightarrow{k_7} X5 + w2$$
$$X5 + g4 \xrightarrow{k_8} k + w3$$
$$k + g5 \xrightarrow{k_9} X1 + w4$$
$$X1 + g6 \xrightarrow{k_{10}} l + g7$$
$$l + g7 \xrightarrow{k_{11}} X1 + g6$$
$$l + X2 \xrightarrow{k_{12}} m + w5$$
$$m + g8 \xrightarrow{k_{13}} X3 + w6$$

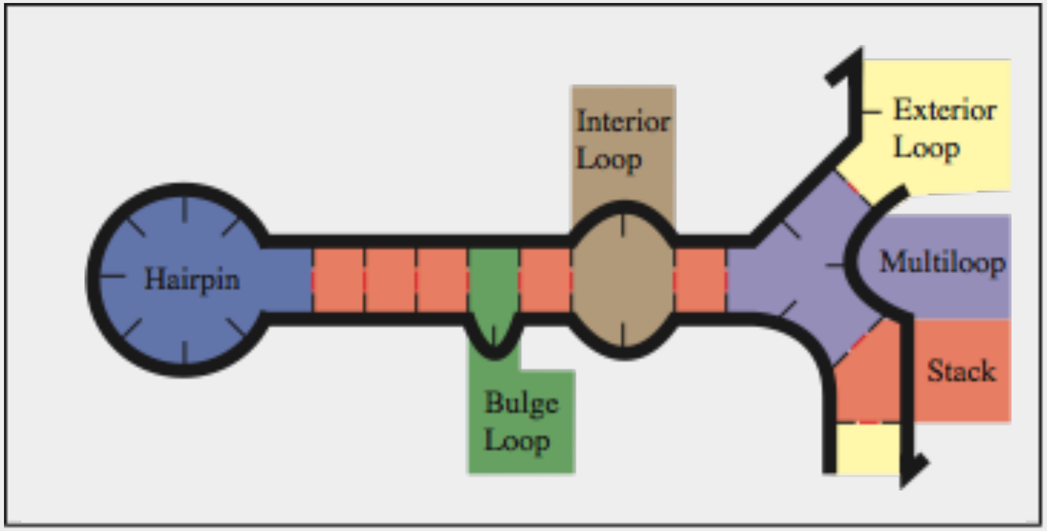
Multistrand: stochastic simulation of the kinetics of multiple interacting nucleic acid strands

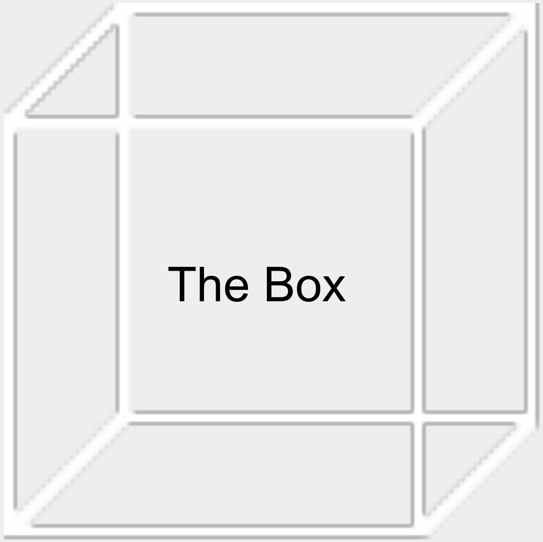
"Stochastic Simulation of the Kinetics of Multiple Interacting Nucleic Acid Strands"
Joseph M. Schaeffer, PhD Thesis (2013)

"Stochastic Simulation of the Kinetics of Multiple Interacting Nucleic Acid Strands"
Joseph M. Schaeffer, Chris Thachuk, and Erik Winfree, LNCS 9211 (2015)

Hands-on live demo: http://www.multistrand.org

Secondary Structure State Space





((.((.((.((((.......)))))).)).(.+).))

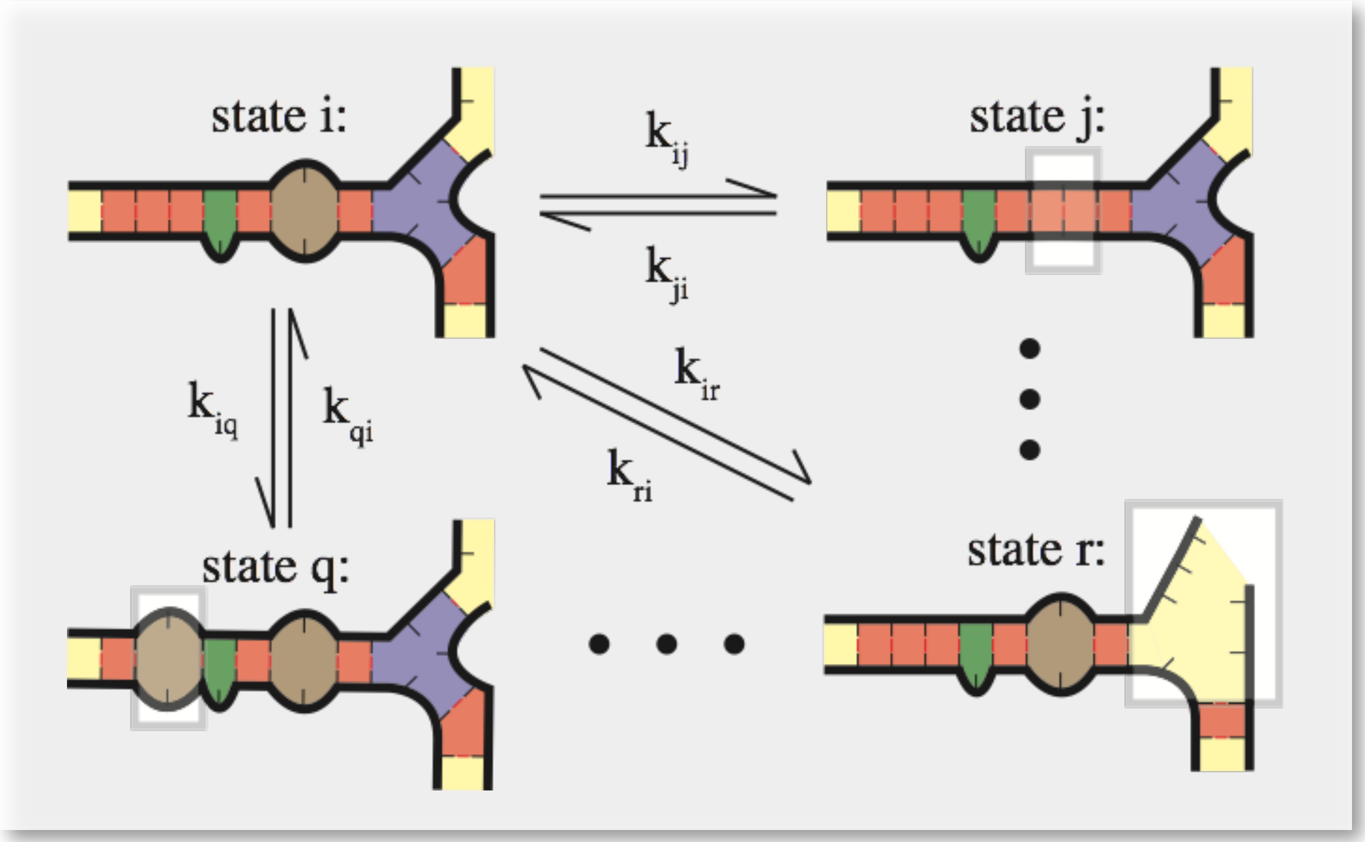
$$\Delta G_{box}(s) = \sum_{c\,\in s} \Delta G^{*}(c)$$

$$\Delta G^{*}(c) = \left(\sum_{loop\,\in c} \Delta G(loop) \right) + (L_c - 1) * (\Delta G_{assoc} + \Delta G_{volume})$$

* same energy model as NUPACK 3.0

Kinetics:

continuous-time markov chain representing a random walk on the energy landscape for The Box.



Elementary steps: single base pairs form and break.

Move selection:

$$Pr(m) = \frac{k_{im}}{\Sigma_j k_{ij}}$$

$$Pr(\Delta t) = \Sigma_j k_{ij} * e^{-\Sigma_j k_{ij} \Delta t}$$

Detailed balance:

$$\frac{k_{ij}}{k_{ji}} = e^{-\frac{\Delta G(j) - \Delta G(i)}{RT}}$$

$$Pr(i) = \frac{e^{-\Delta G_{box}(i)/RT}}{Q_{kin}}$$

$$Q_{kin} = \sum_i e^{-\Delta G_{box}(i)/RT}$$

Kinetics:

continuous-time markov chain representing a random walk on the energy landscape for The Box.

Unimolecular
(i to j is downhill)

(Metropolis)

$$k_{ij} = k_{uni}$$

$$k_{ji} = k_{uni} * e^{-\frac{\Delta G_{box}(i) - \Delta G_{box}(j)}{RT}}$$

Bimolecular
(i to j is joining)

$$k_{ij} = k_{bi} * e^{-\frac{\Delta G_{volume}}{RT}} = k_{bi} * u$$

$$k_{ji} = k_{bi} * e^{-\frac{\Delta G_{box}(i) - \Delta G_{box}(j) + \Delta G_{volume}}{RT}}$$

$$= k_{bi} * e^{-\frac{\Delta G_{loops}(i,j) - \Delta G_{assoc}}{RT}}$$

Elementary steps: single base pairs form and break.

Move selection:

$$Pr(m) = \frac{k_{im}}{\Sigma_j k_{ij}}$$

$$Pr(\Delta t) = \Sigma_j k_{ij} * e^{-\Sigma_j k_{ij} \Delta t}$$

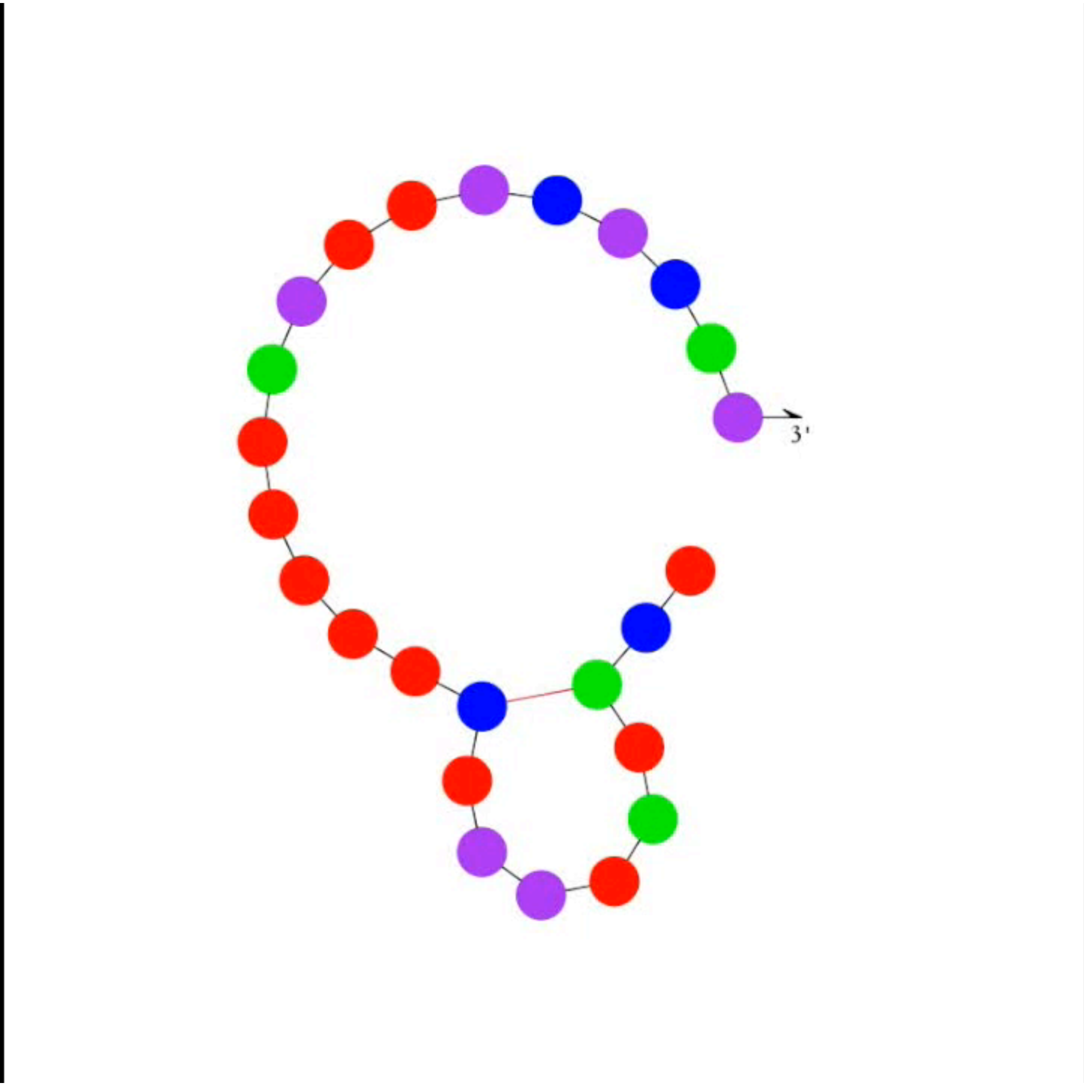
Detailed balance:

$$\frac{k_{ij}}{k_{ji}} = e^{-\frac{\Delta G(j) - \Delta G(i)}{RT}}$$

$$Pr(i) = \frac{e^{-\Delta G_{box}(i)/RT}}{Q_{kin}}$$

$$Q_{kin} = \sum_i e^{-\Delta G_{box}(i)/RT}$$

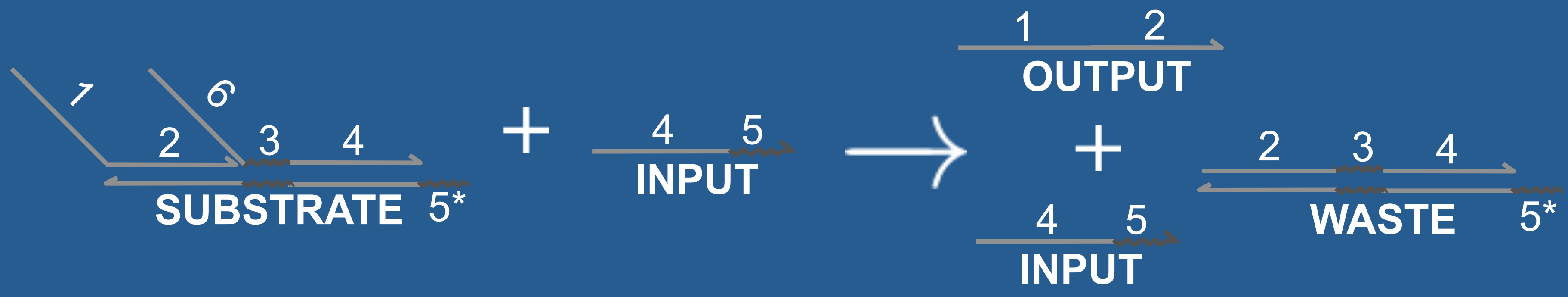
Simulating a single strand



KinDA: design and analysis of kinetics for domain-level DNA strand displacement systems

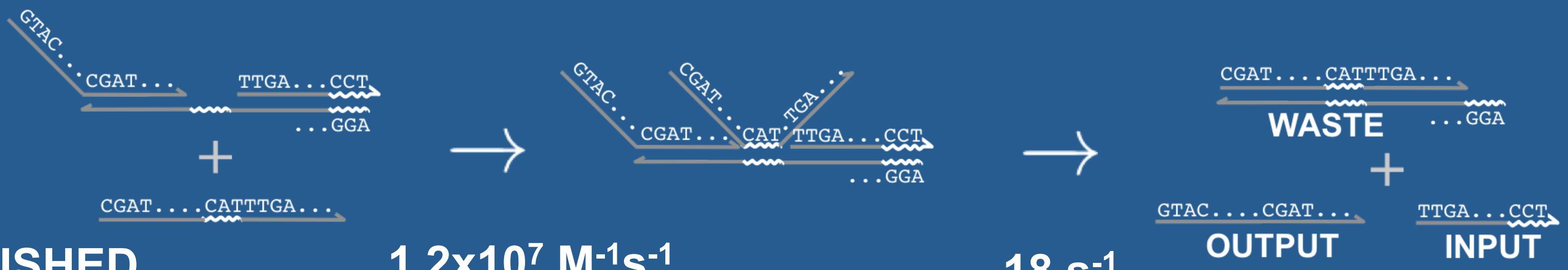
"Automated Sequence Analysis for Domain-level DNA Strand Displacement Systems"
Joseph Berleant, Chris Berlind, Joseph Schaeffer, Niranjan Srinivas, Chris Thachuk, Erik Winfree (in preparation)

Case Study: Entropy-driven catalyst

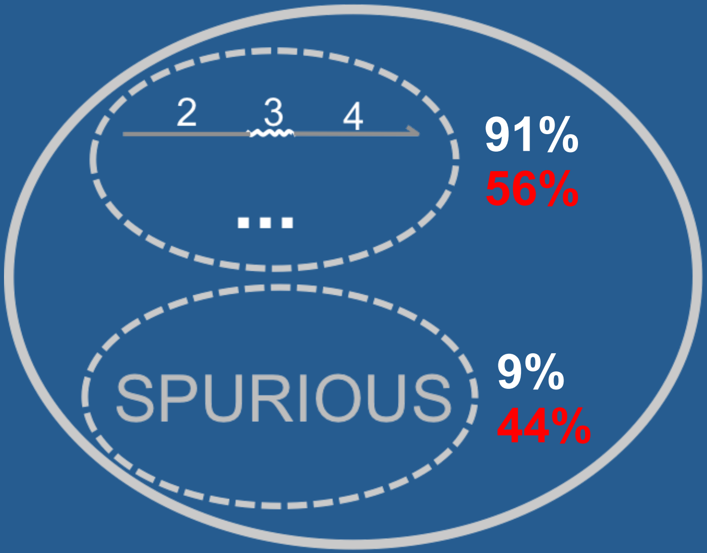


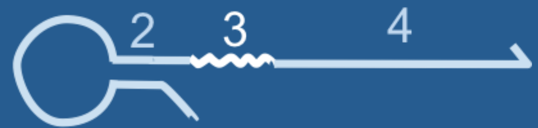
DOMAIN	PUBLISHED	RANDOM
1	CTTTCCTACA	AAACGCGAAA
2	CCTACGTCTCCAACTAACTTACGG	CTTTAATATGGATTTTGCACCAGT
3	CCCT	TTCC
4	CATTCAATACCCTACG	AGTCTTAATTGACCCA
5	TCTCCA	CAAAGA
6	CCACATACATCATATT	ACAAGAGCCTTGATTA

Case Study: Entropy-driven catalyst

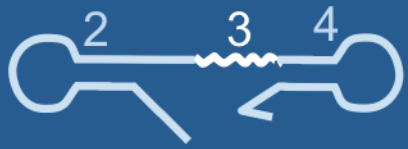


PUBLISHED	$1.2x10^7$ $M^{-1}s^{-1}$	18 s^{-1}
RANDOM	$4.4x10^6$ $M^{-1}s^{-1}$	73 s^{-1}





```
CCTACGTCTCCAACTAACTTACGGCCCTCATTCAATACCCTACG
....(((.............))).....................
```

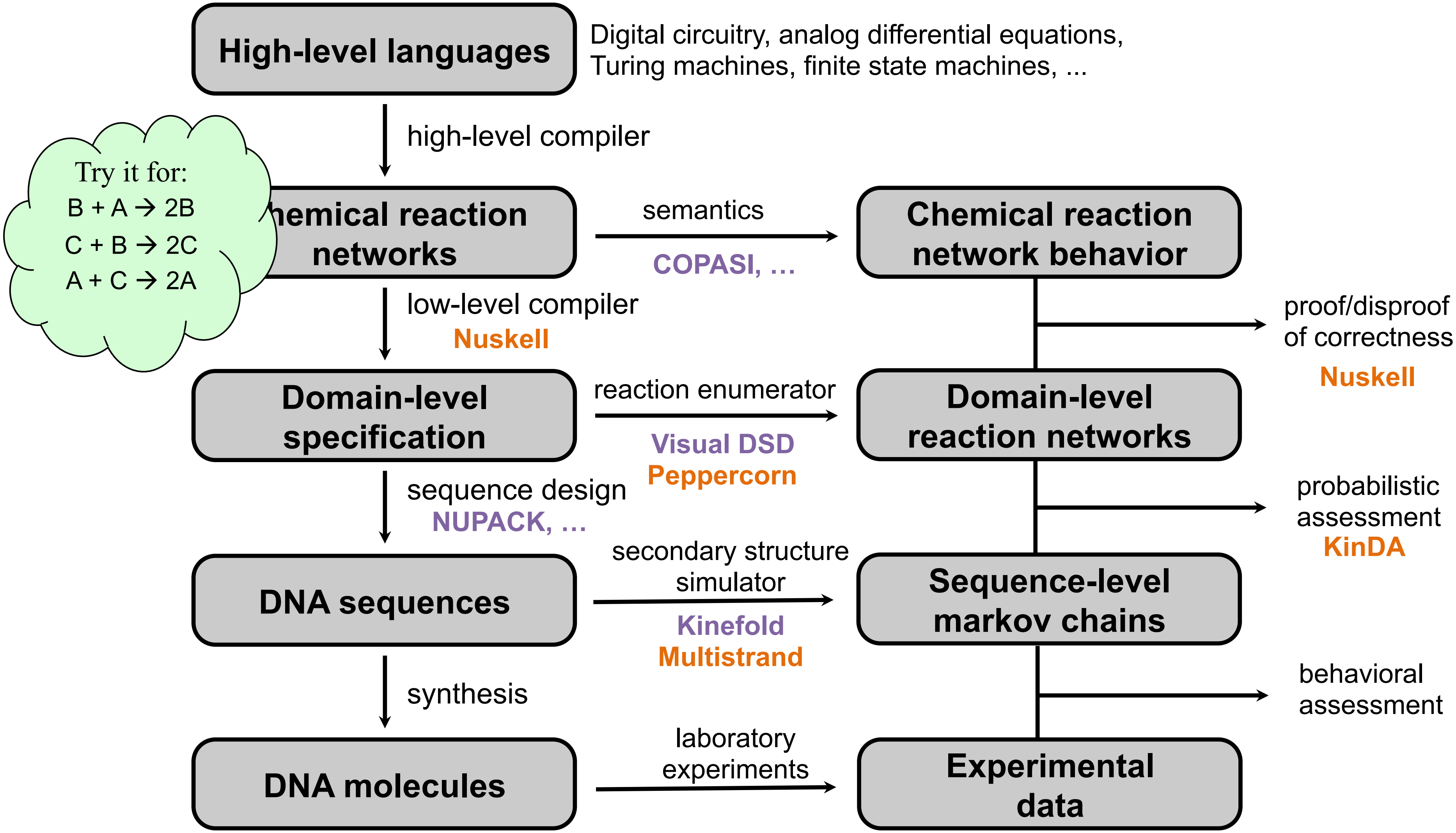


```
CTTTAATATGGATTTTGCACCAGTTTCCAGTCTTAATTGACCCA
........(((........))).......(((......)))...
```

Programming Dynamical Behaviors in Chemical Systems using DNA Strand Displacement Cascades

Niranjan Srinivas, James Parkin, Georg Seelig, Erik Winfree, David Soloveichik
(in preparation)

A compiler & verification hierarchy



High-level languages

Digital circuitry, analog differential equations, Turing machines, finite state machines, ...

high-level compiler

Try it for:

$B + A \rightarrow 2B$

$C + B \rightarrow 2C$

$A + C \rightarrow 2A$

hemical reaction networks

semantics

COPASI, …

Chemical reaction network behavior

low-level compiler

Nuskell

proof/disproof of correctness

Nuskell

Domain-level specification

reaction enumerator

Visual DSD

Peppercorn

Domain-level reaction networks

sequence design

NUPACK, …

probabilistic assessment

KinDA

DNA sequences

secondary structure simulator

Kinefold

Multistrand

Sequence-level markov chains

synthesis

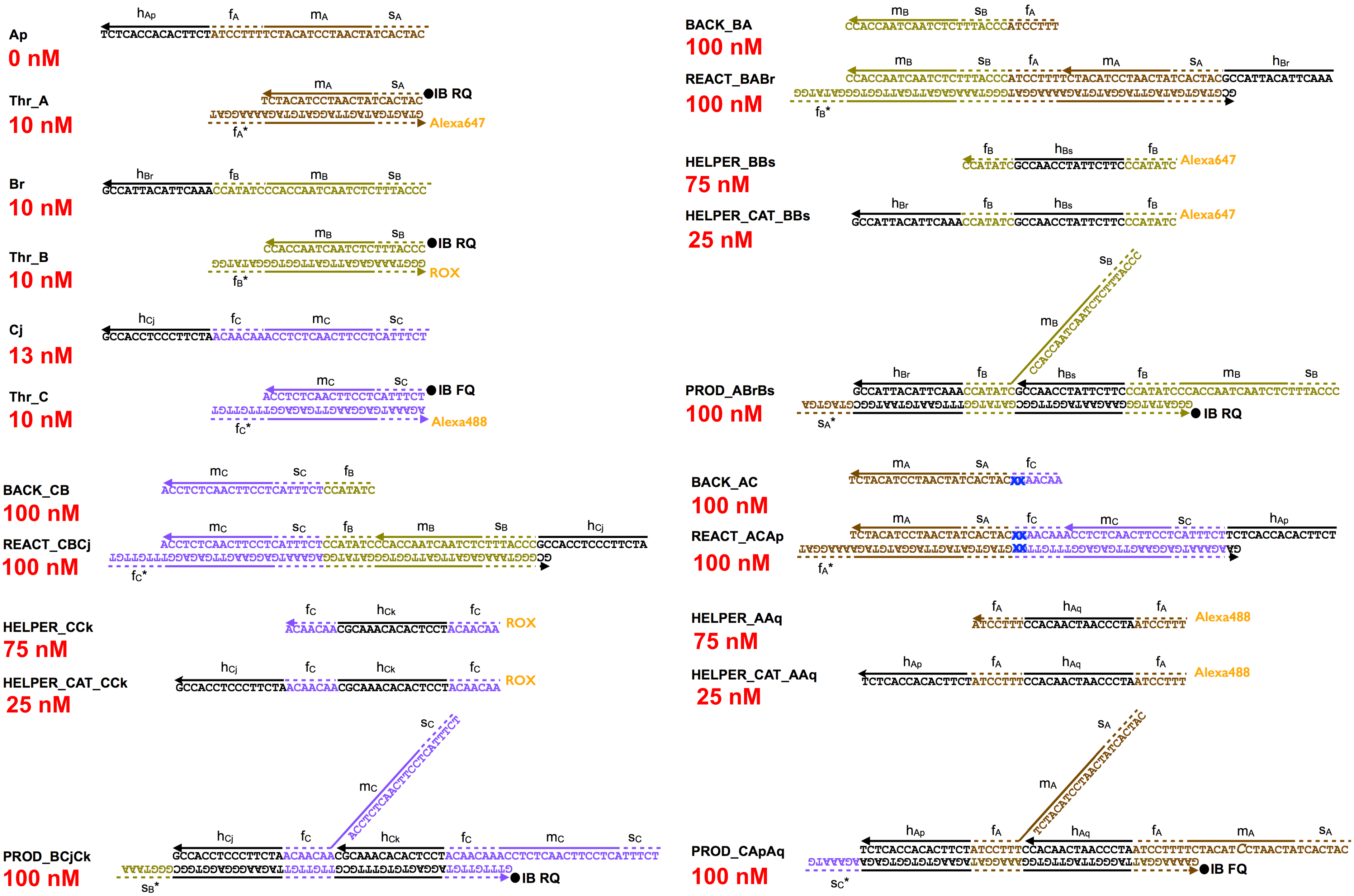
behavioral assessment

DNA molecules

laboratory experiments

Experimental data

A test tube of synthetic DNA molecules



Ap
0 nM

Thr_A
10 nM

Br
10 nM

Thr_B
10 nM

Cj
13 nM

Thr_C
10 nM

BACK_CB
100 nM

REACT_CBCj
100 nM

HELPER_CCk
75 nM

HELPER_CAT_CCk
25 nM

PROD_BCjCk
100 nM

BACK_BA
100 nM

REACT_BABr
100 nM

HELPER_BBs
75 nM

HELPER_CAT_BBs
25 nM

PROD_ABrBs
100 nM

BACK_AC
100 nM

REACT_ACAp
100 nM

HELPER_AAq
75 nM

HELPER_CAT_AAq
25 nM

PROD_CApAq
100 nM

hAp fA mA sA
TCTCACCACACTTCTATCCTTTTCTACATCCTAACTATCACTAC
mA sA
TCTACATCCTAACTATCACTAC
IB RQ
fA*
Alexa647
hBr fB mB sB
GCCATTACATTCAAACCATATCCCACCAATCAATCTCTTTACCC
mB sB
CCACCAATCAATCTCTTTACCC
IB RQ
fB*
ROX
hCj fC mC sC
GCCACCTCCCTTCTAACAACAAACCTCTCAACTTCCTCATTTCT
mC sC
ACCTCTCAACTTCCTCATTTCT
IB FQ
fC*
Alexa488
mC sC fB
ACCTCTCAACTTCCTCATTTCTCCATATC
mC sC fB mB sB hCj
ACCTCTCAACTTCCTCATTTCTCCATATCCCACCAATCAATCTCTTTACCCGCCACCTCCCTTCTA
fC*
fC hCk fC
ACAACAACGCAAACACACTCCTACAACAA
ROX
hCj fC hCk fC
GCCACCTCCCTTCTAACAACAACGCAAACACACTCCTACAACAA
ROX
sC
mC
ACCTCTCAACTTCCTCATTTCT
hCj fC hCk fC mC sC
GCCACCTCCCTTCTAACAACAACGCAAACACACTCCTACAACAAACCTCTCAACTTCCTCATTTCT
IB RQ
sB*
mB sB fA
CCACCAATCAATCTCTTTACCCATCCTTT
mB sB fA mA sA hBr
CCACCAATCAATCTCTTTACCCATCCTTTTCTACATCCTAACTATCACTACGCCATTACATTCAAA
fB*
fB hBs fB
CCATATCGCCAACCTATTCTTCCCATATC
Alexa647
hBr fB hBs fB
GCCATTACATTCAAACCATATCGCCAACCTATTCTTCCCATATC
Alexa647
sB
mB
CCACCAATCAATCTCTTTACCC
hBr fB hBs fB mB sB
GCCATTACATTCAAACCATATCGCCAACCTATTCTTCCCATATCCCACCAATCAATCTCTTTACCC
IB RQ
sA*
mA sA fC
TCTACATCCTAACTATCACTACXXAACAA
mA sA fC mC sC hAp
TCTACATCCTAACTATCACTACXXAACAAACCTCTCAACTTCCTCATTTCTTCTCACCACACTTCT
fA*
fA hAq fA
ATCCTTTCCACAACTAACCCTAATCCTTT
Alexa488
hAp fA hAq fA
TCTCACCACACTTCTATCCTTTCCACAACTAACCCTAATCCTTT
Alexa488
sA
mA
TCTACATCCTAACTATCACTAC
hAp fA hAq fA mA sA
TCTCACCACACTTCTATCCTTTCCACAACTAACCCTAATCCTTTTCTACATCCTAACTATCACTAC
IB FQ
sC*

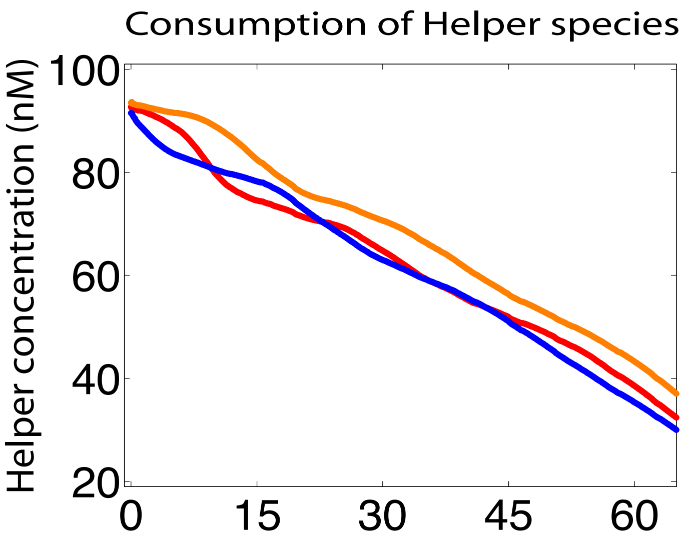
An enzyme-free oscillator: the Displacillator

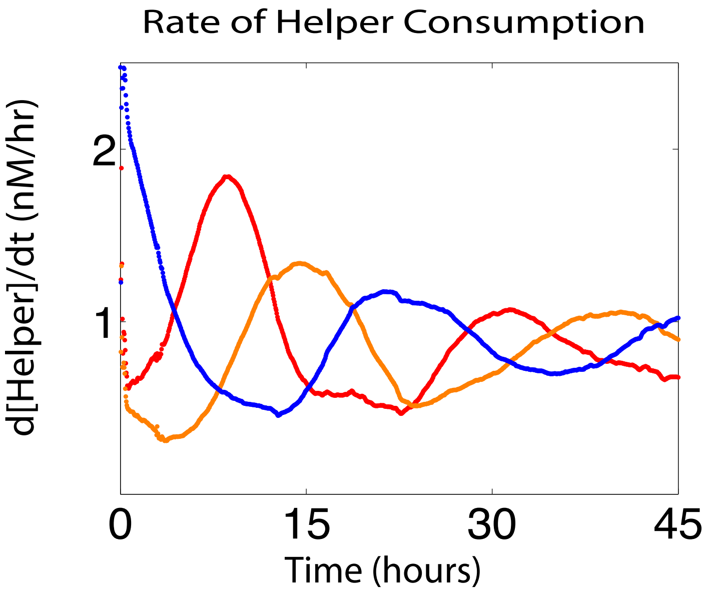
$$B + A \longrightarrow 2B$$

$$C + B \longrightarrow 2C$$

$$A + C \longrightarrow 2A$$

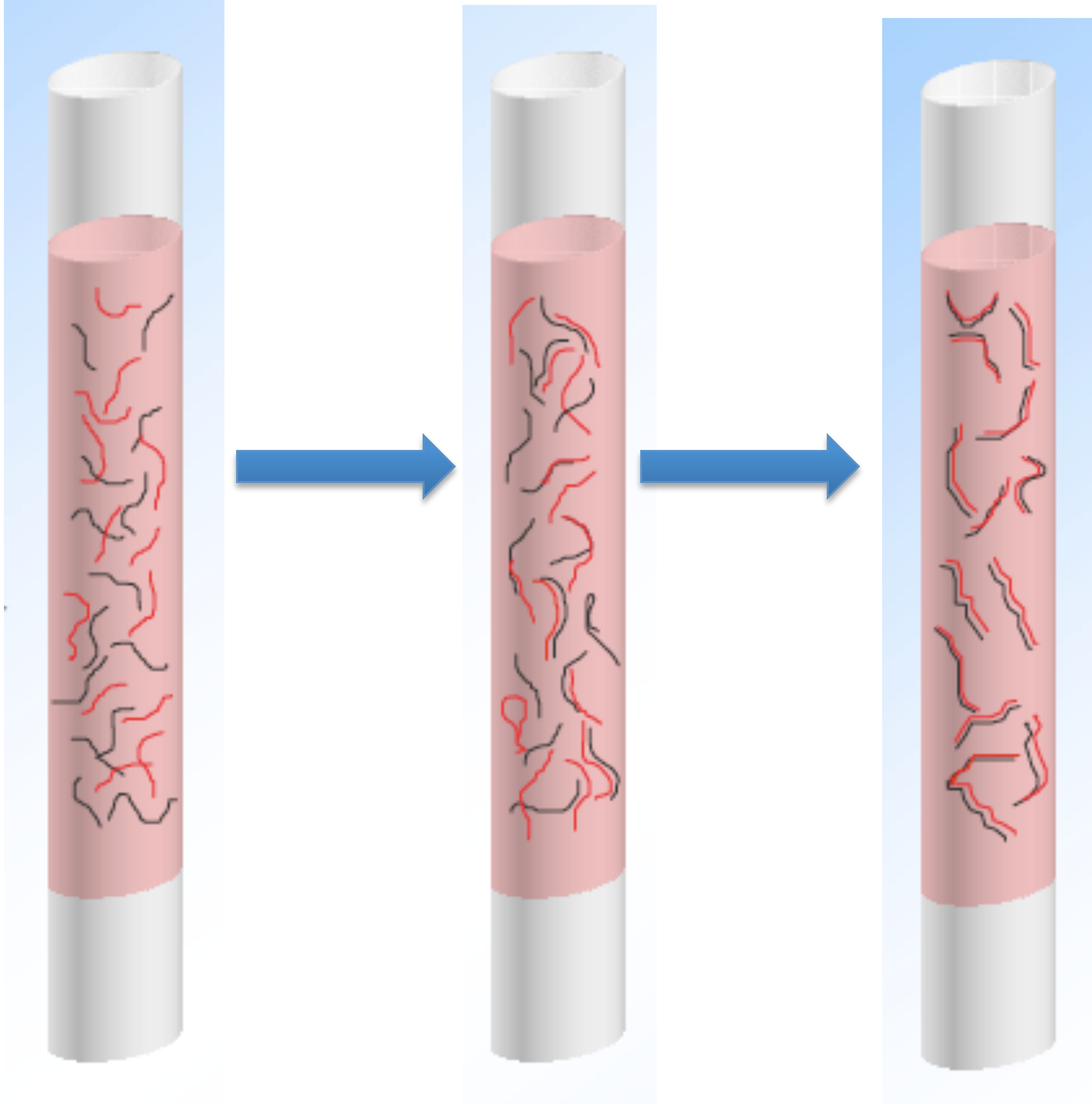
Rock-Paper-Scissors CRN

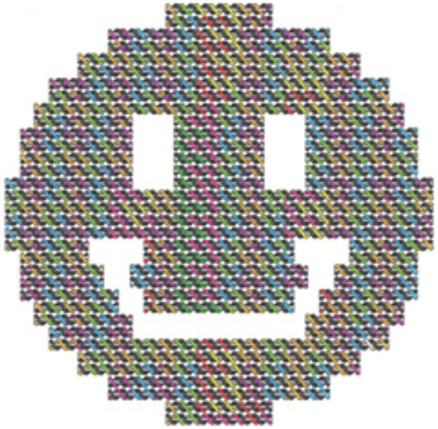


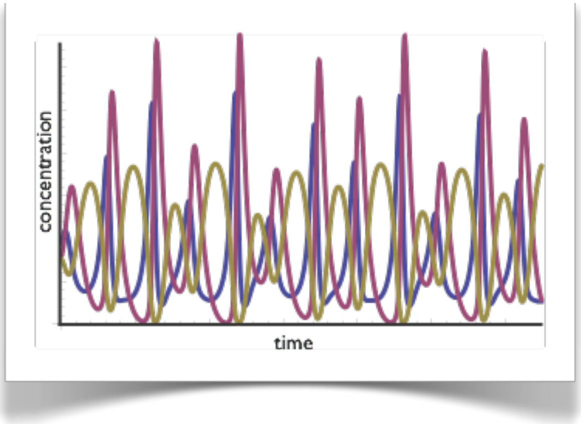


Niranjan Srinivas, James Parkin, Georg Seelig, Erik Winfree, David Soloveichik **(in preparation)**

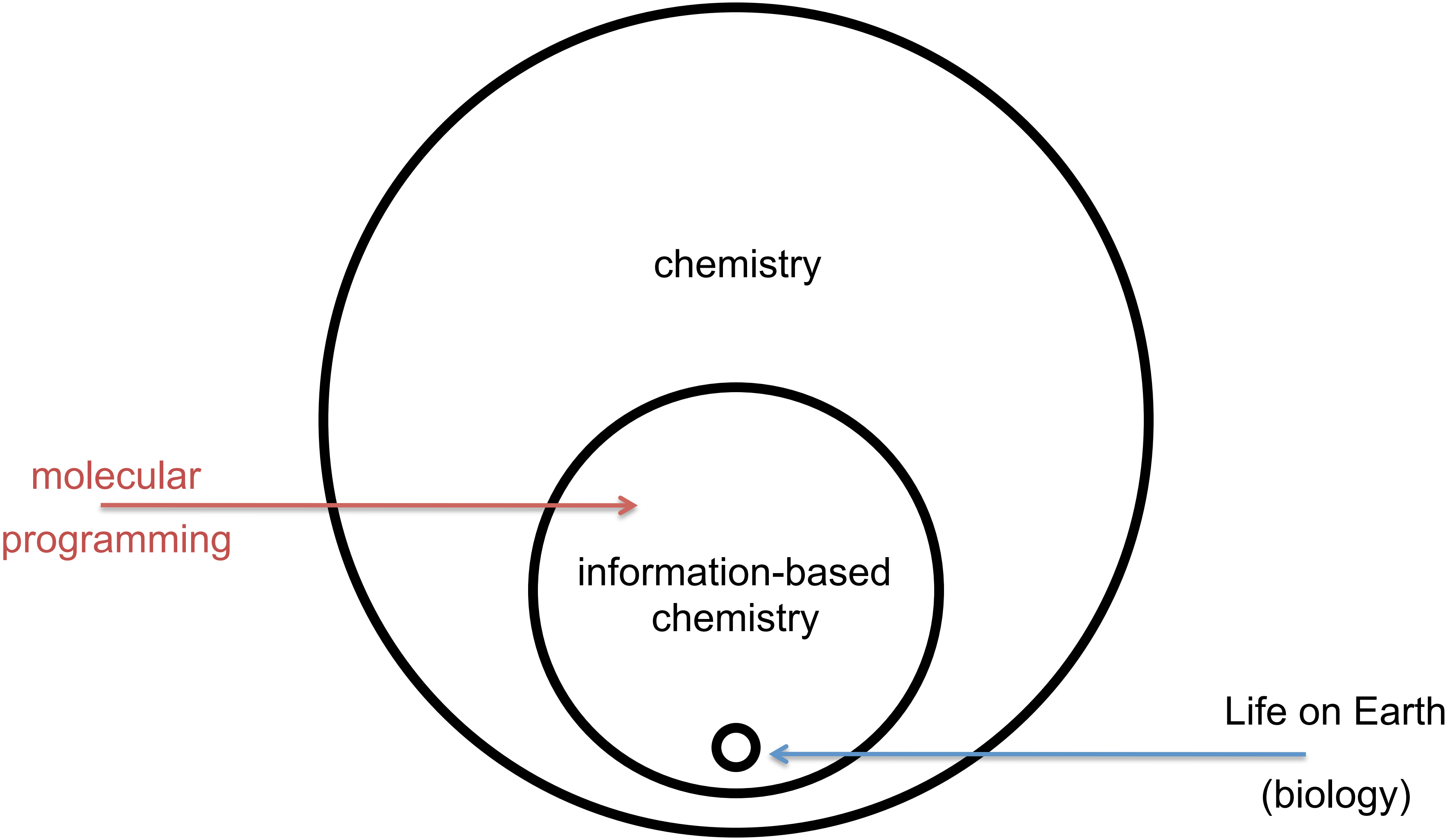
Toss a mixture of DNA in a test tube… What can it do? What can't it do?







What can information-based chemistry do?



chemistry

molecular programming

information-based chemistry

Life on Earth

(biology)

Why Molecular Programming?

Chemistry will be the new information technology of the 21st C

- Information encoded in synthetic molecules can direct processes such as folding, self-assembly, circuitry, and machinery, thereby providing programmable control of a wide range of chemical systems.

It will transform industry, much as electronics did in the 20th C

- Potential applications synthesizing programmable materials, devices, diagnostics, therapeutics… anything that chemistry can do… that will be empowered by pervasive embedded information processing and programmable behaviors.