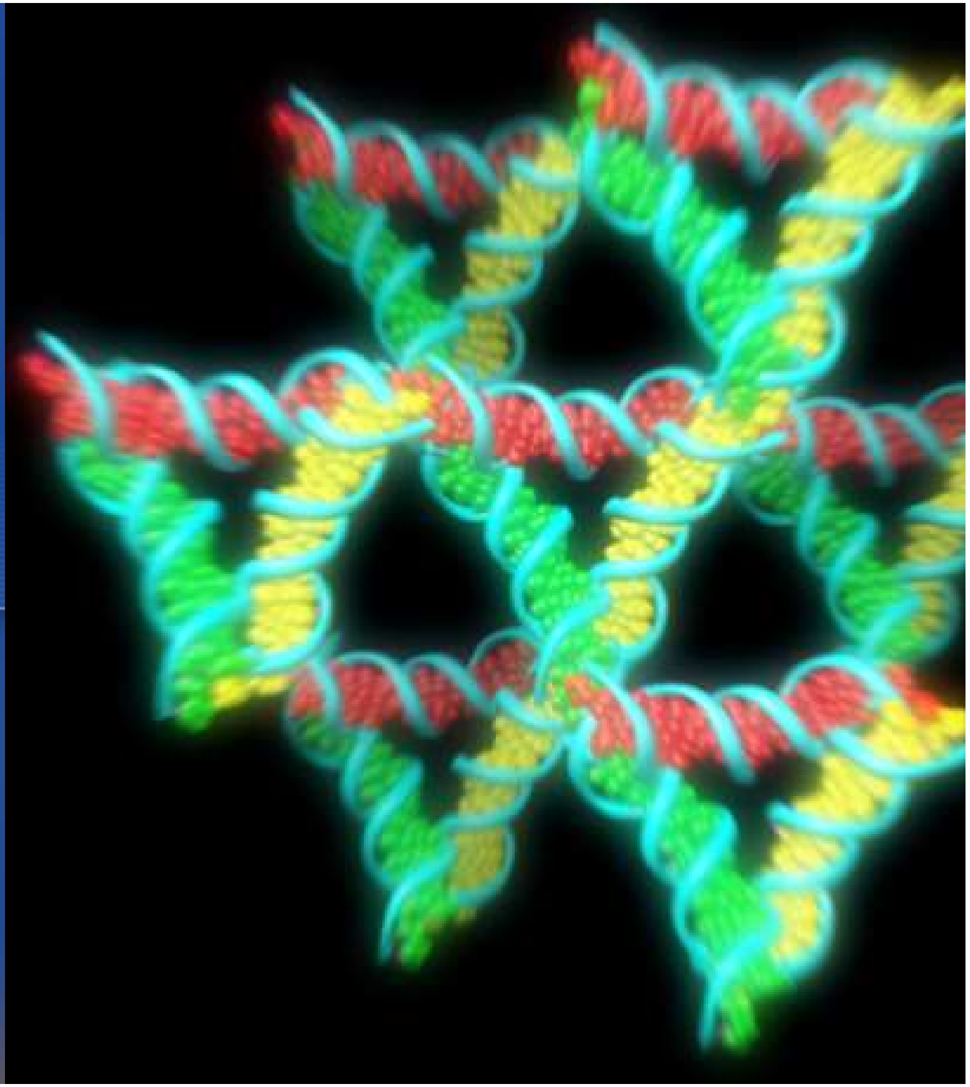
Molecular Programming

Luca Cardelli

University of Oxford

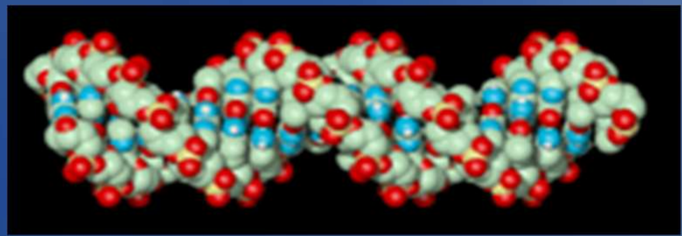
Future of Computing
2019-07-04, Porto



Objectives

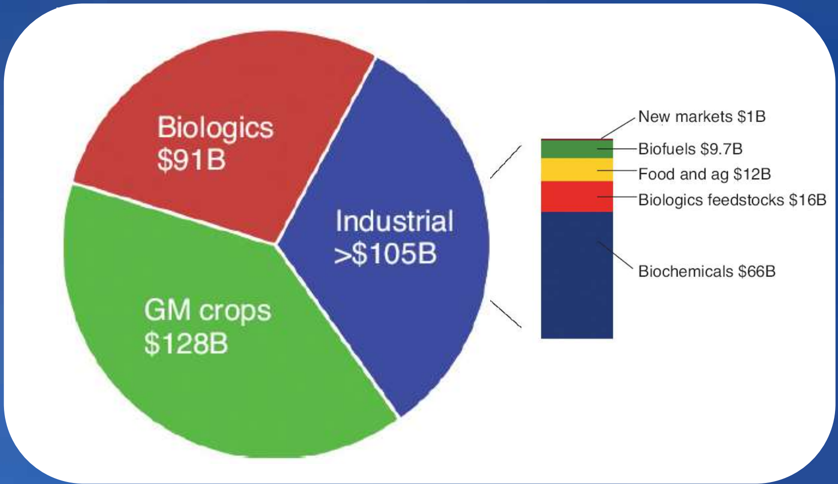
- The promises of Molecular Programming:
  - In Science & Medicine
  - In Engineering
  - In Computing

- The current practice of Molecular Programming
  - DNA technology
  - Molecular languages and tools
  - Molecular algorithms



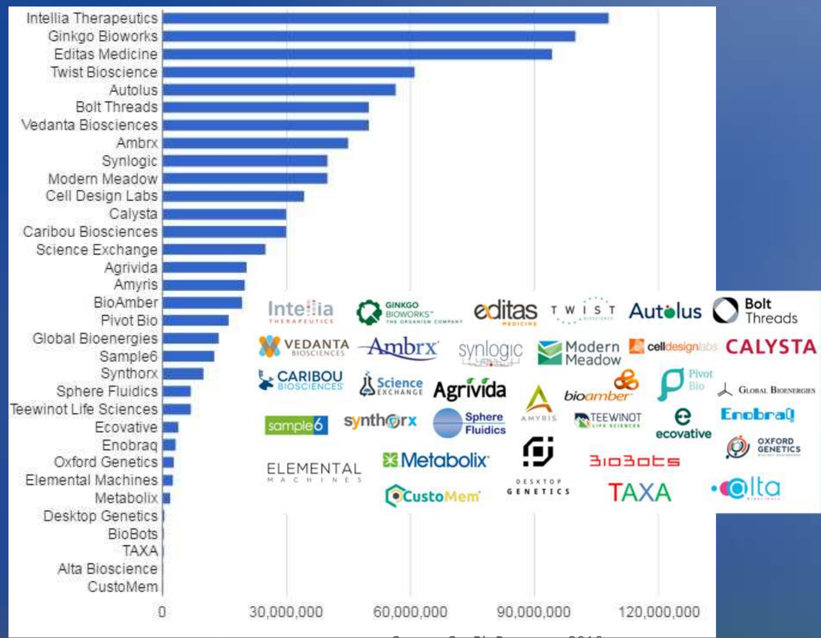
Synthetic Biology Market

Annual revenue from GMOs in the US exceeds $324Bn



Source: Rob Carlson, Nature Biotechnology, 2016

33 Programming Biology companies raised $900M in 2016



Source: SynBioBeta.com, 2016

Some (ongoing) successes stories

JUNO THERAPEUTICS

- ($4Bn) Reprogram a patient's own blood cells to recognise and destroy specific cancers.
- 90% remission in terminally ill leukemia patients

AMYRIS

- ($300M) Reprogra… …ise chemicals
- Antimalaria… …th Sanofi)
- Jet f… …flights (with Total)

Molecular "hacking"

GIN… BIO… THE O…ISM COMPANY

- Supply custom organisms for bio fabrication

Modern Meadow

- Grow meat, leather ($100Bn market) in the lab
- Proofs of concept already in production

Molecular Programming

A *technology* (and theory of computation) based on information-bearing molecules of historically biological origin (DNA/RNA) non necessarily involving living matter

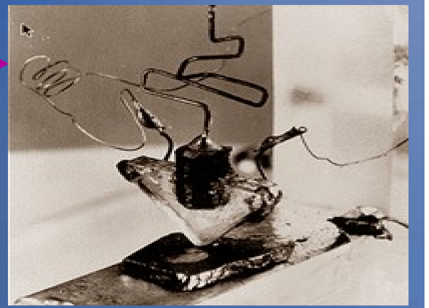
Molecular Programming:
The Hardware Aspect

Smaller and smaller things can be built

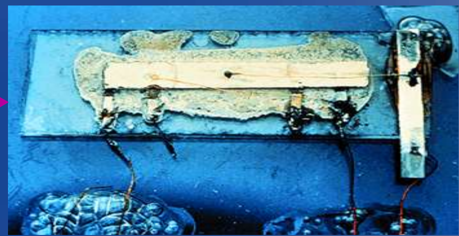
Smaller and Smaller

Very few Moore's cycles left!

First working transistor
John Bardeen and Walter Brattain , Dec. 23, 1947

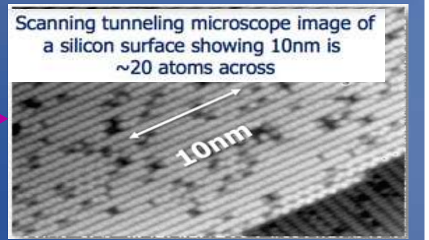


First integrated circuit
Jack Kilby, Sep. 1958.



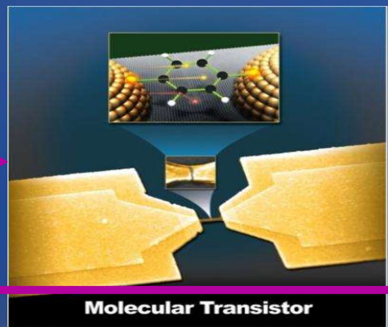
50+ years later

Jan 2010 25nm NAND flash
Intel&Micron. ~50atoms

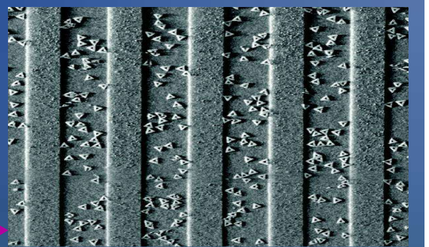


Jun 2018 7nm (54nm pitch)
TSMC, Intel, Samsung, GlobalFoundries - mass production

Single molecule transistor
Observation of molecular orbital gating
Nature, 2009; 462 (7276): 1039



Molecules on a chip



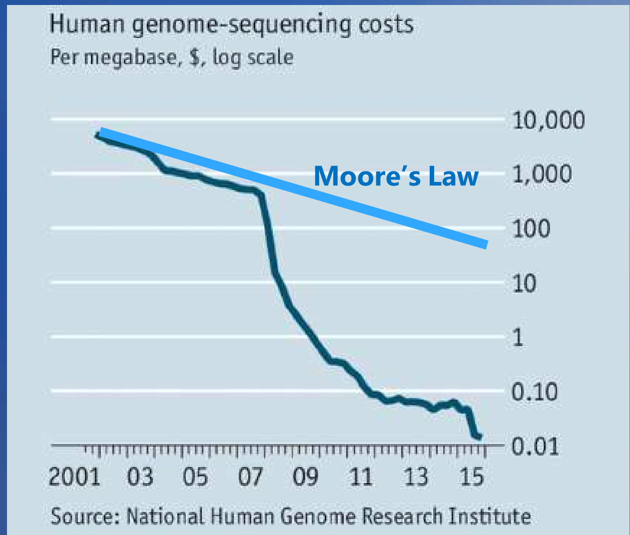
Placement and orientation of individual DNA shapes on lithographically patterned surfaces. Nature Nanotechnology 4, 557 - 561 (2009).

Race to the Bottom

Moore's Law is approaching the single-molecule limit

Carlson's Curve is the new exponential growth curve in technology

In both cases, we are now down to *molecules*



The Pace and Proliferation of Biological Technologies
March 4, 2004 by Rob Carlson

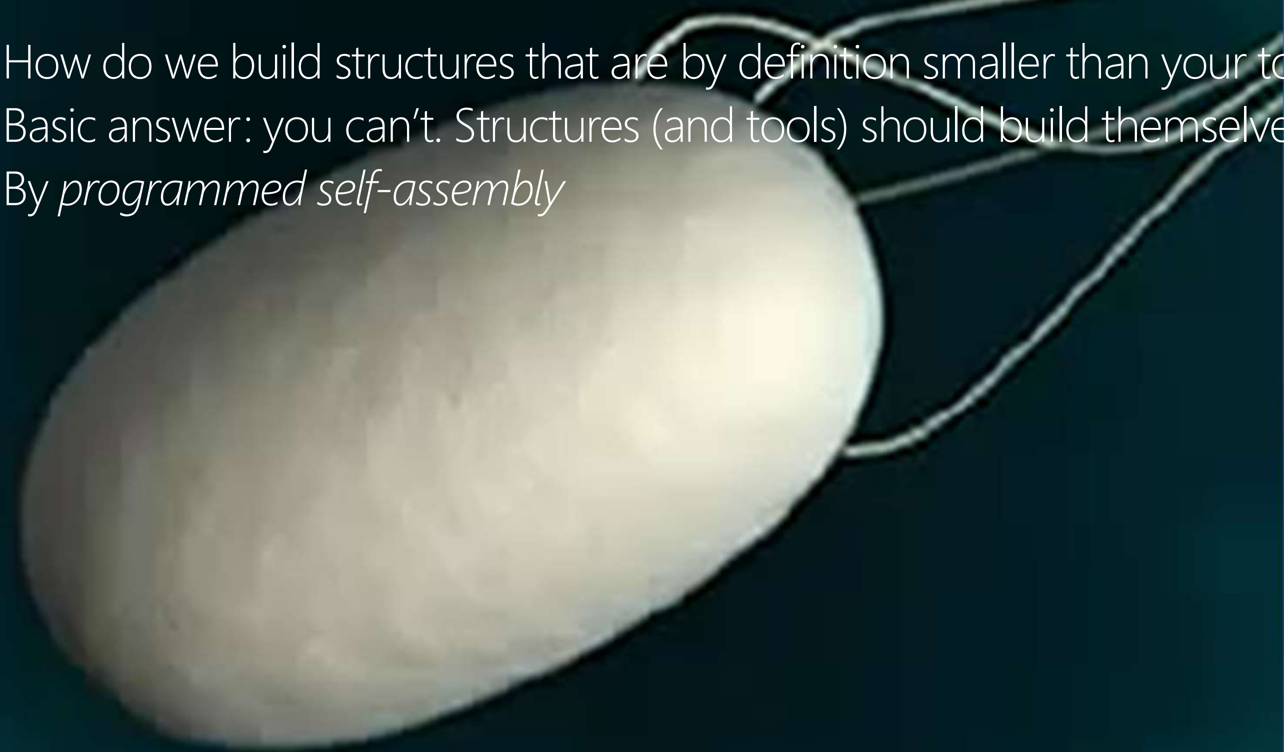
Waiter! There is fly DNA in my soup!

The SmidgION: A portable DNA sequencer that runs on an Iphone

Oxford Nanopore

Building the *Smallest* Things

- How do we build structures that are by definition smaller than your tools?
- Basic answer: you can't. Structures (and tools) should build themselves!
- By *programmed self-assembly*



www.youtube.com/watch?v=Ey7Emmddf7Y

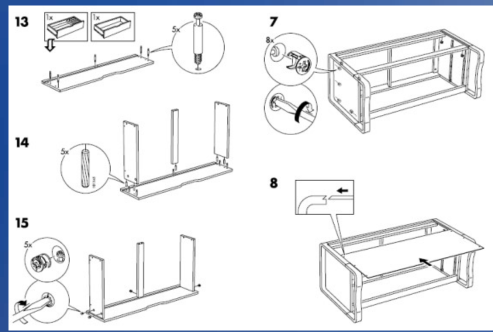
Molecular IKEA

- Nature can self-assemble.
  Can we?

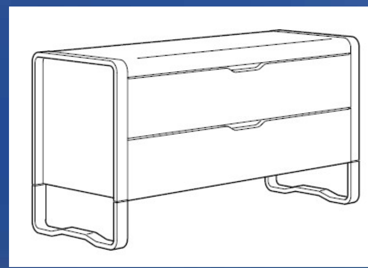
- *"Dear IKEA, please send me a chest of drawers that assembles itself."*

- We need a magical material where the pieces are pre-programmed to fit into to each other.

- At the molecular scale many such materials exist…



Add water

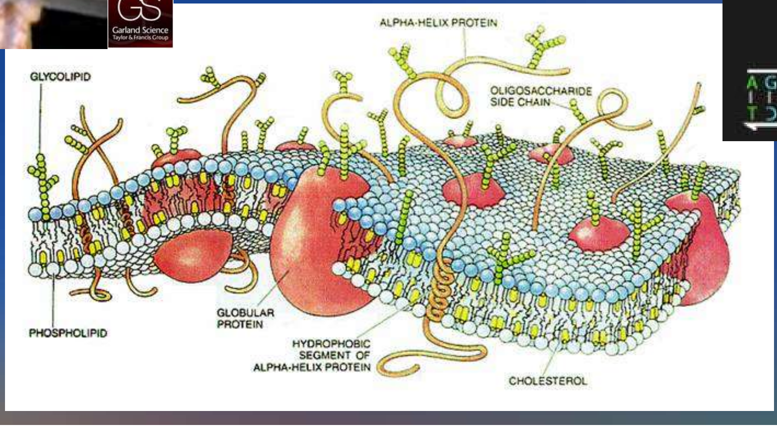


http://www.ikea.com/ms/en_US/customer_service/assembly_instructions.html

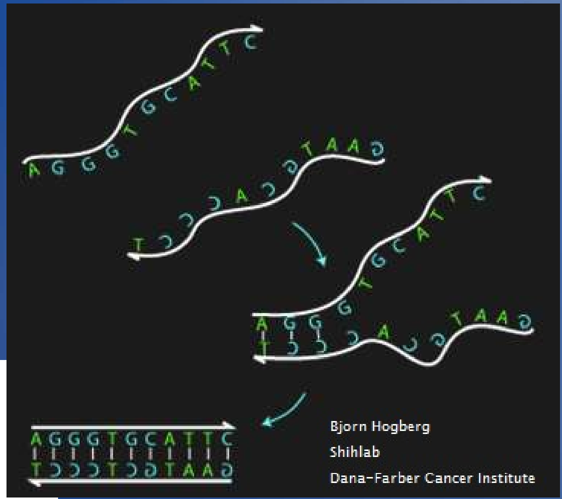
Programmed Self-Assembly

Proteins

Membranes



DNA/RNA



Molecular Programming:
The Software Aspect

Smaller and smaller things can be programmed

We can program...

- Information
  - Completely!

Information

↓

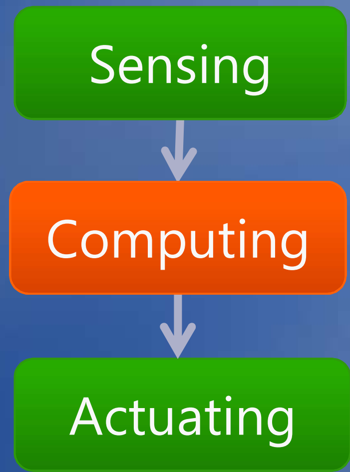
Computing

↓

Information

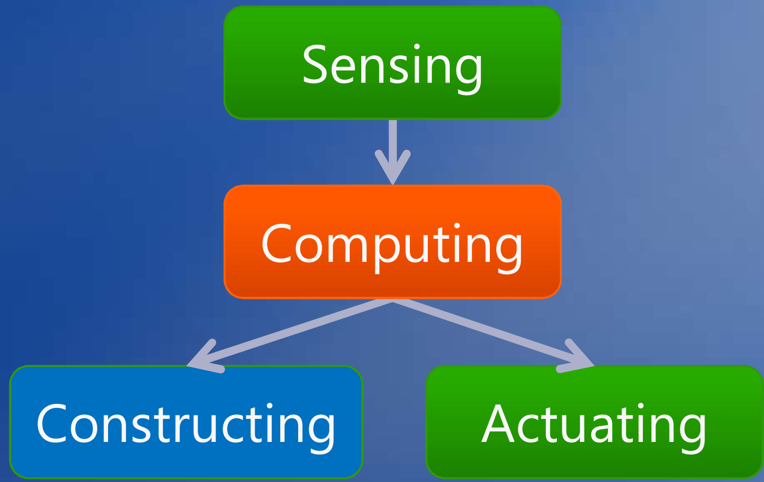
We can program...

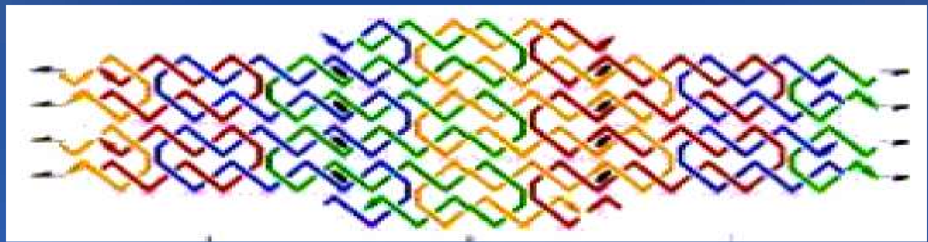
- Forces
  - Completely!
    (Modulo sensors/actuators)



We can program...

- Matter
  - Completely and directly! By self-assembly.
  - Currently: only DNA/RNA.

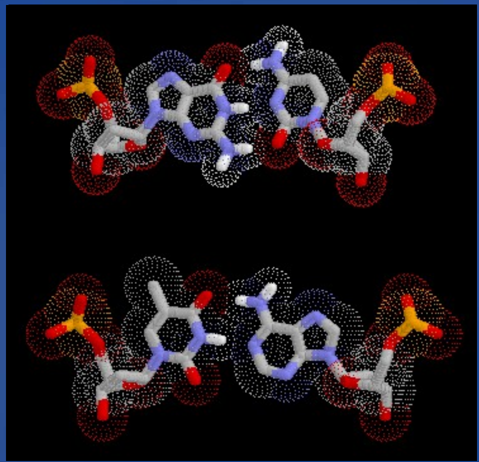




It's like a 3D printer without the printer!
[Andrew Hellington]

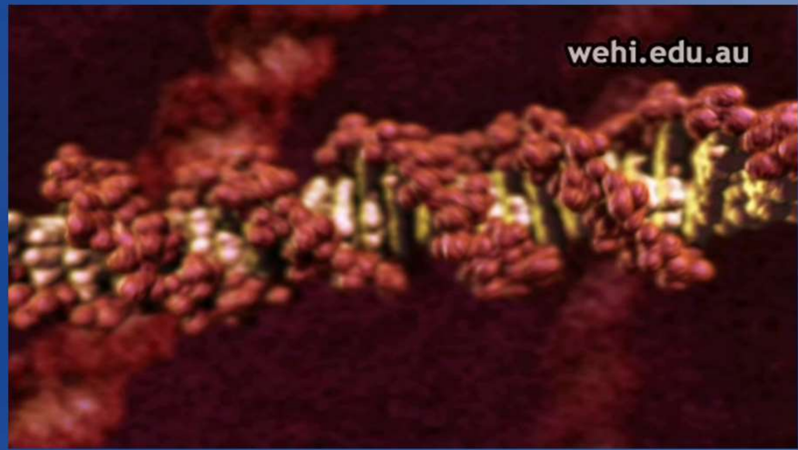
  - But DNA is an amazing *material*

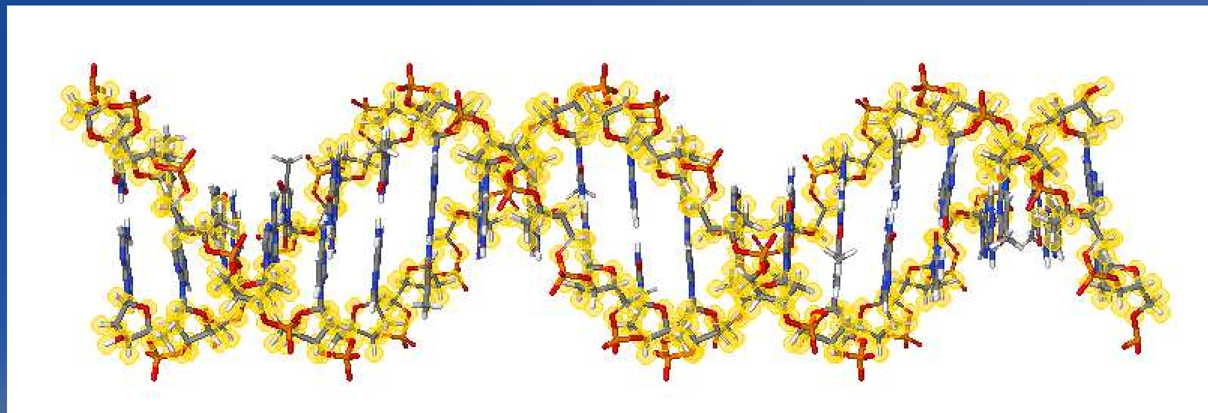
DNA



G-C Base Pair
Guanine-Cytosine

T-A Base Pair
Thymine-Adenine





Sequence of Base Pairs (GACT alphabet)

[Interactive DNA Tutorial](http://www.biosciences.bham.ac.uk/labs/minchin/tutorials/dna.html)

(http://www.biosciences.bham.ac.uk/labs/minchin/tutorials/dna.html)

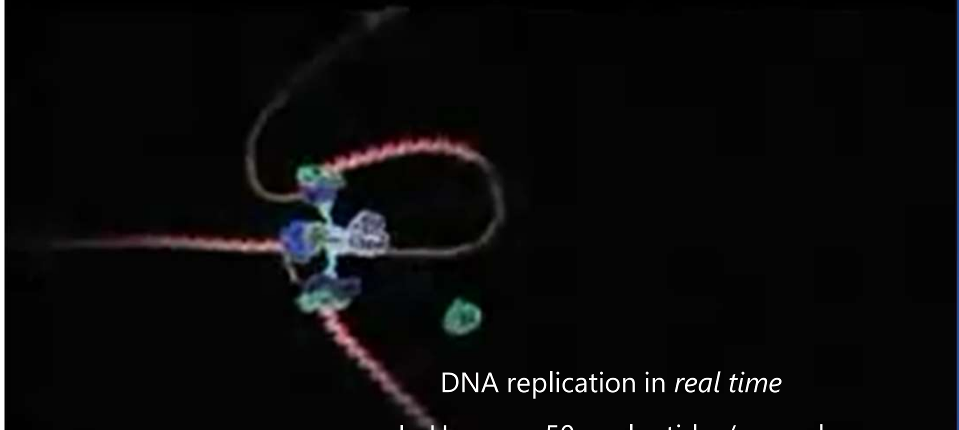
DNA Specs

- DNA in each human cell
  - 3 billion base pairs
  - 2nm thick = 4 silicon atoms!
  - 0.34nm per basepair = 2/3 silicon atom!
  - 2 meters long
    copied in parallel at each cell division!
  - 750 megabytes
    80% functional, but only 1.5% protein coding
  - folded into a 6μm spherical nucleus
    = 140 exabytes (million terabytes)/mm^3
    => *all* the data on the internet fits in a shoebox!
- DNA in each human body
  - 10 trillion cells
  - 133 Astronomical Units long
  - 7.5 octabytes (replicated)
- DNA in human population
  - 20 million light years long


DNA wrapping into chromosomes


Andromeda Galaxy
2.5 million light years away

DNA Benchmarks



DNA replication in *real time*

In Humans: 50 nucleotides/second
Whole genome in a few hours (with parallel processing)

In Bacteria: 1000 nucleotides/second
(higher error rate)

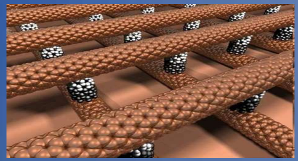
DNA transcription in *real time*

RNA polymerase II: 15-30 base/second

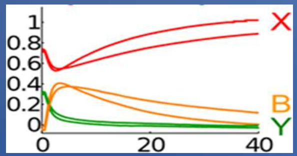
Drew Berry
http://www.wehi.edu.au/wehi-tv

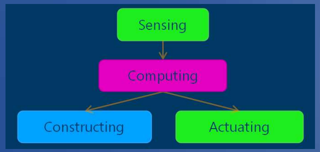
One molecule to rule them all

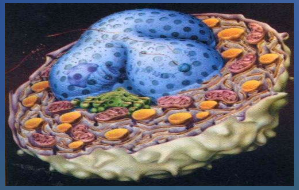
- There are many, many nanofabrication techniques and materials



- But only DNA (and RNA) can:
  - Organize ANY other matter [caveats apply]
  - Execute ANY kinetics [caveats: up to time scaling]
  - Assemble Nano-Control Devices
  - Interface to Biology







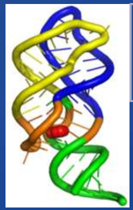
H.Lodish & al. Molecular Cell Biology 4th ed.

"Modern" DNA Computing

- Non-goals
  - Not to solve NP-complete problems with large vats of DNA
  - Not to replace silicon
- Bootstrapping a carbon-based technology
  - To precisely control the organization and dynamics of matter and information at the molecular level
  - DNA is our engineering material
    - Its biological origin is “accidental” (but convenient)
    - It is an information-bearing programmable material
    - Other such materials will be (are being) developed

Building Nano-Control Devices

All the components of nanocontrollers can already be built entirerly and solely with DNA, and interfaced to the environment

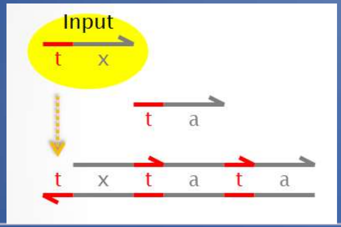


DNA Aptamers

Sensing

↓

Computing

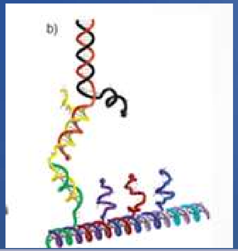


DNA Logical Gates

Constructing

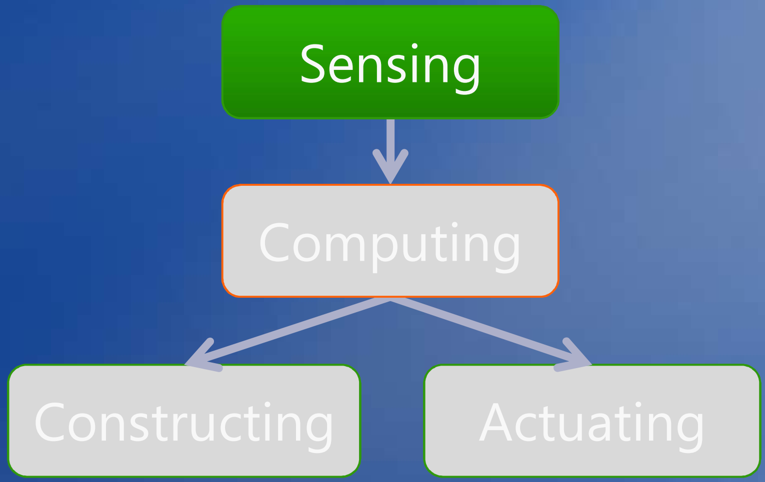
Actuating

Self-assembling DNA Tiles



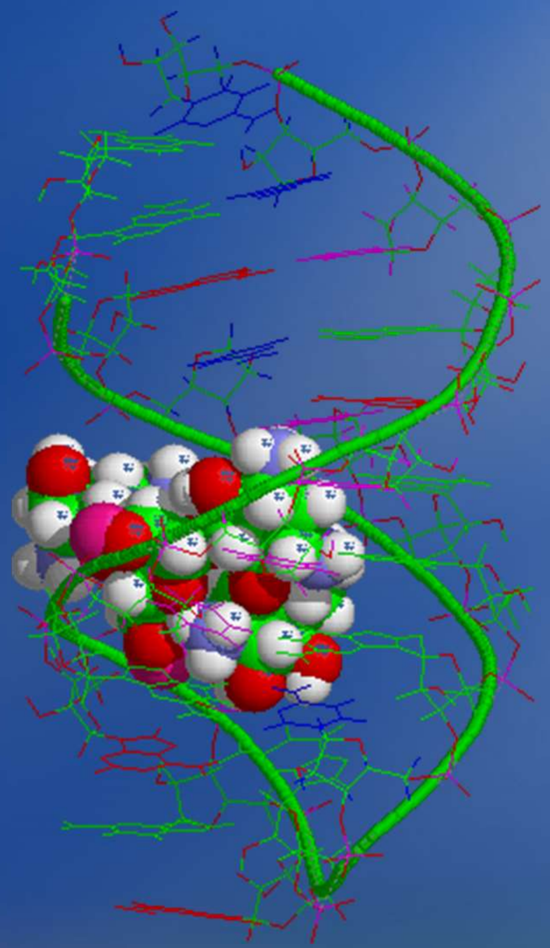
DNA Walkers & Cages

Sensing



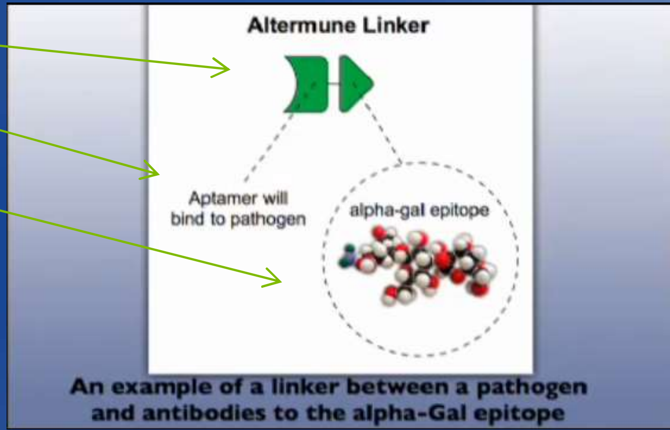
Aptamers

Artificially evolved DNA molecules that stick to (almost) anything you like highly selectively

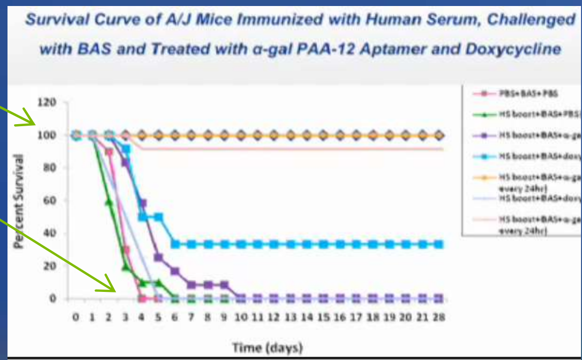


Pathogen Spotlights

- DNA aptamer binds to:
  - A) a pathogen
  - B) a molecule our immune system (when allergic) hates and immediately removes (eats) along with anything attached to it!

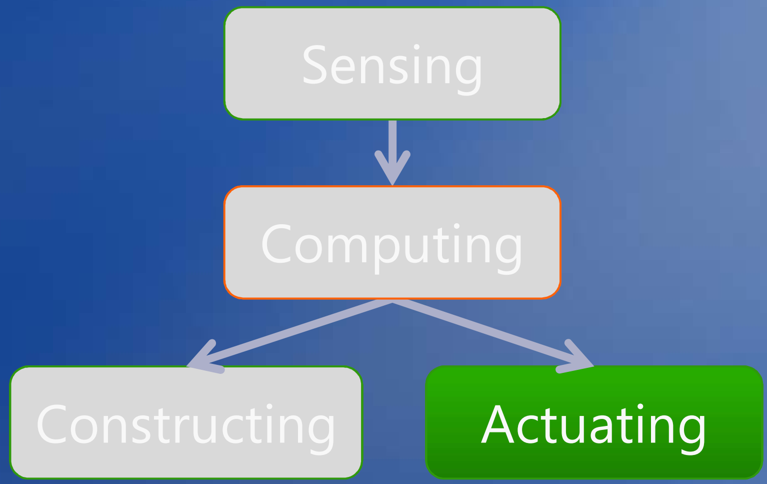


- Result: instant immunity
  - Mice poisoned with Anthrax plus aptamer (100% survival)
  - Mice poinsoned with Anthrax (not so good)



Kary Mullis (incidentally, also Nobel prize for inventing the Polymerase Chain Reaction)

Actuating



DNA Walkers

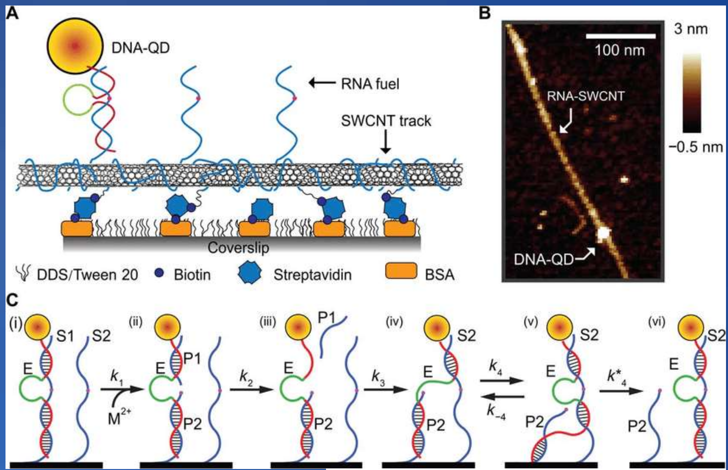
J|A|C|S
COMMUNICATIONS

Published on Web 05/17/2004

A Synthetic DNA Walker for Molecular Transport

Jong-Shik Shin† and Niles A. Pierce*,†,‡

Departments of Bioengineering and Applied & Computational Mathematics, California Institute of Technology, Pasadena, California 91125



Visible/near-infrared subdiffraction imaging reveals the stochastic nature of DNA walkers

January 2017 · Science Advances 3(1):e1601600

DOI: 10.1126/sciadv.1601600

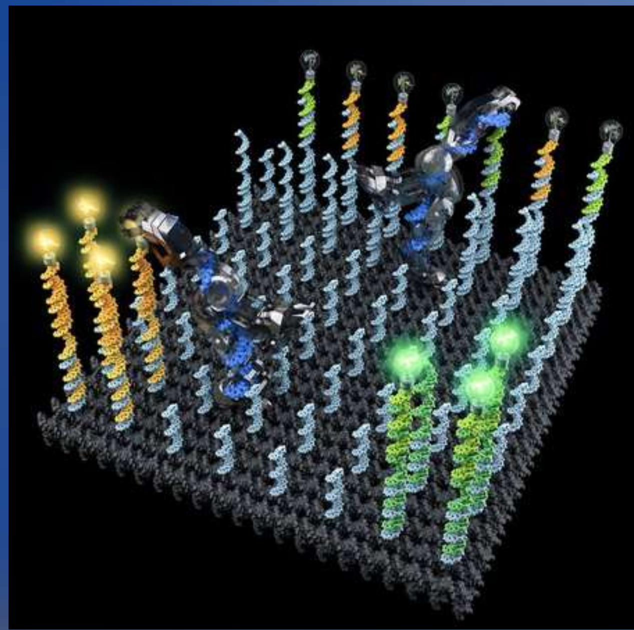
License · CC BY-NC 4.0

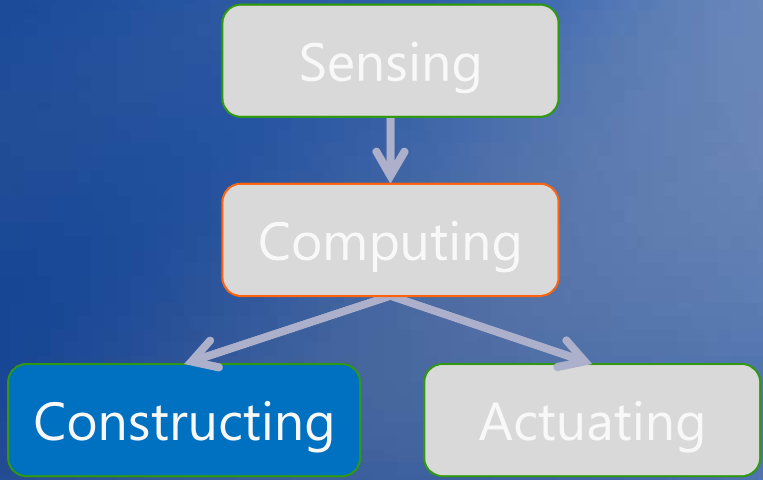
Jing Pan · Tae-Gon Cha · Feiran Li · Show all 6 authors · Jong Hyun Choi

DNA Robotics

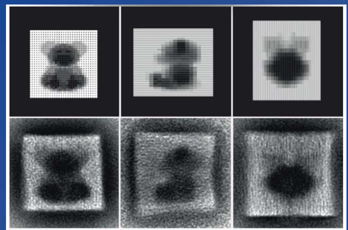
A cargo-sorting DNA robot

Thubagere, Anupama J. and Li, Wei and Johnson, Robert F. and Chen, Zibo and Doroudi, Shayan and Lee, Yae Lim and Izatt, Gregory and Wittman, Sarah and Srinivas, Niranjan and Woods, Damien and Winfree, Erik and Qian, Lulu (2017) *A cargo-sorting DNA robot*. Science, 357 (6356). Art. No. eaan6558. ISSN 0036-8075.



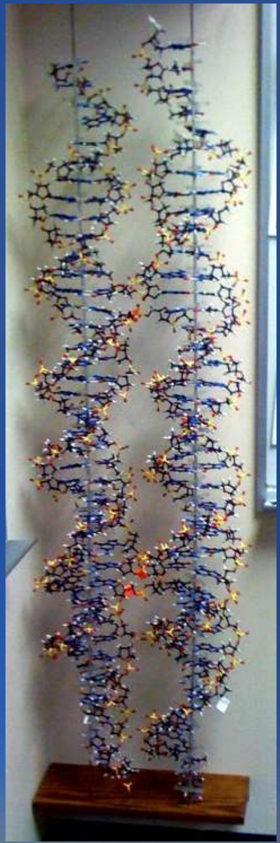


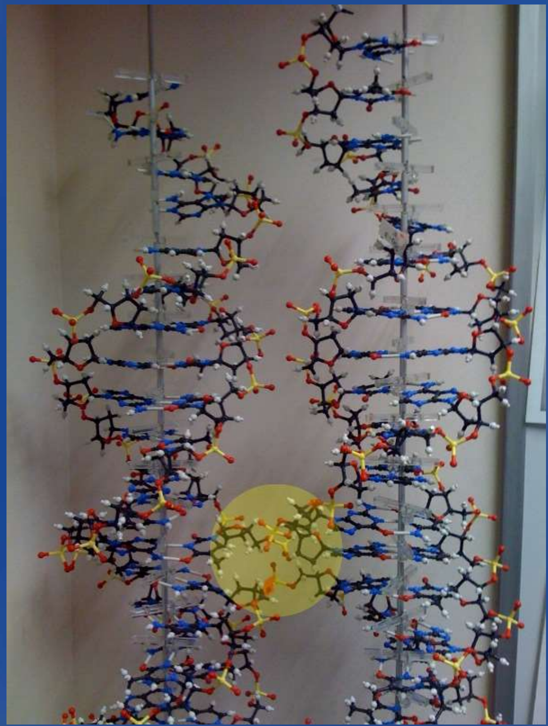
Constructing



The 3D model of the computer-designed bear shape shown on top was fabricated into the nanostructures visualized with transmission electron microscopy (below). Credit: Wyss Institute at Harvard University

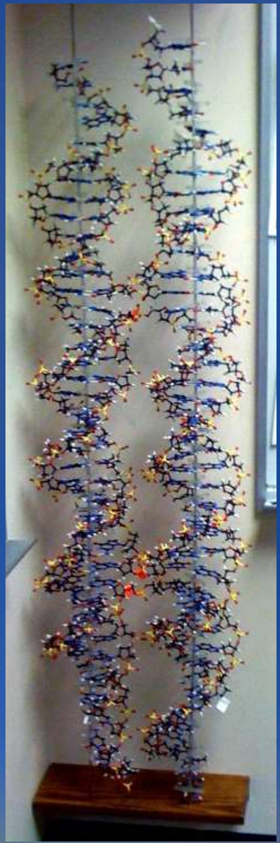
Crosslinking

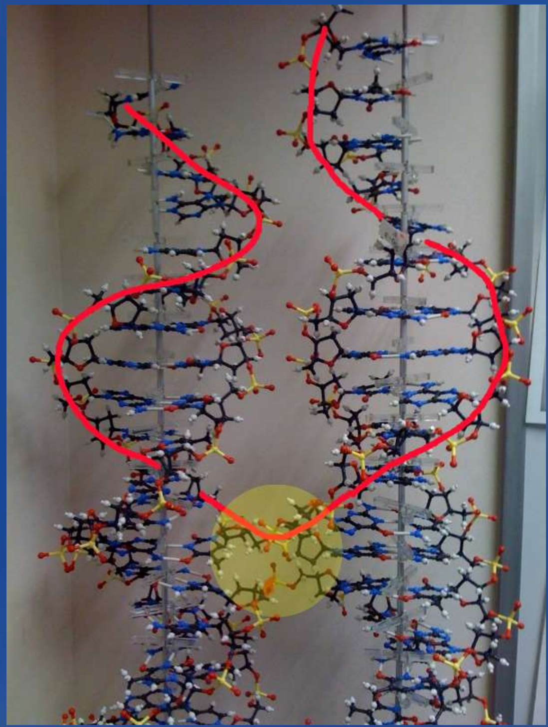




the dawn of structural DNA nanotechnology

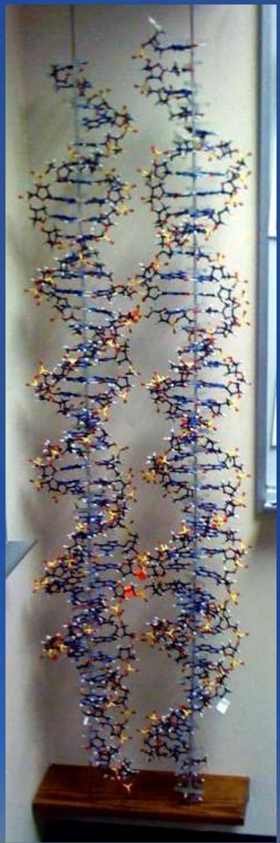
Crosslinking

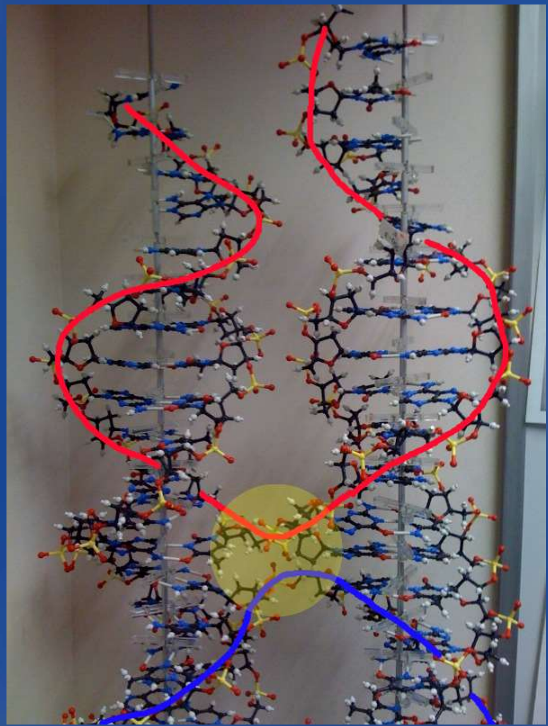




the dawn of
structural DNA
nanotechnology

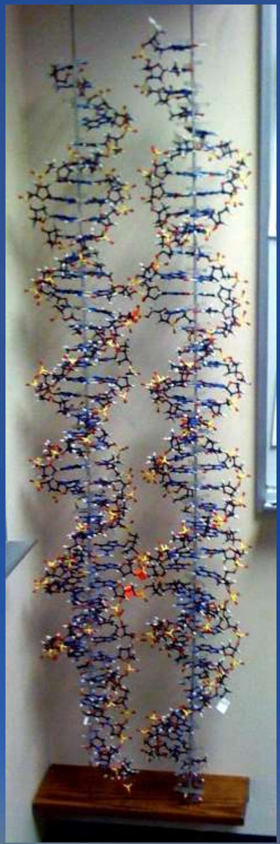
Crosslinking

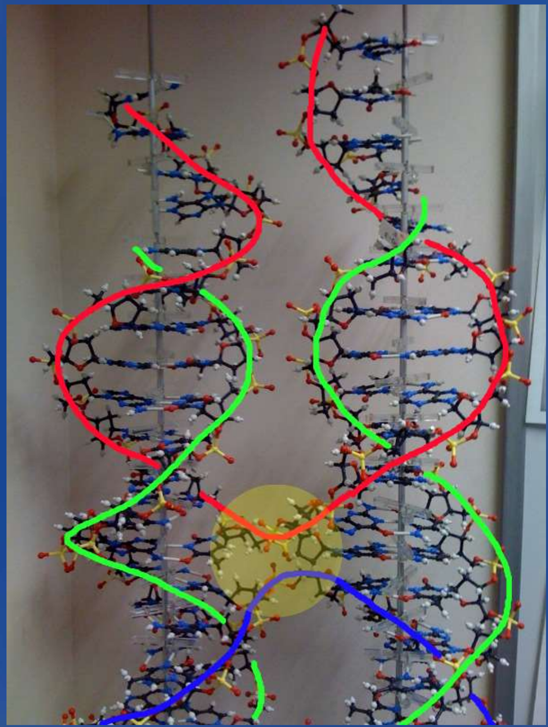




the dawn of
structural DNA
nanotechnology

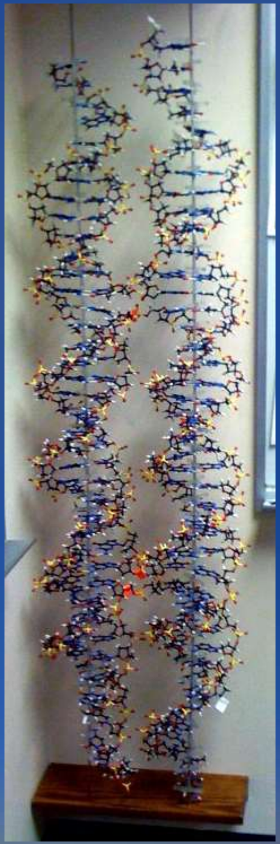
Crosslinking

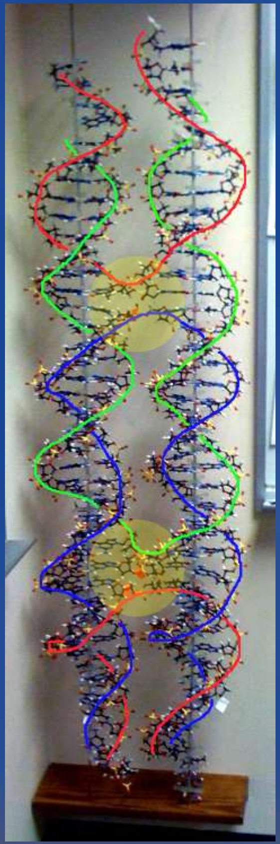




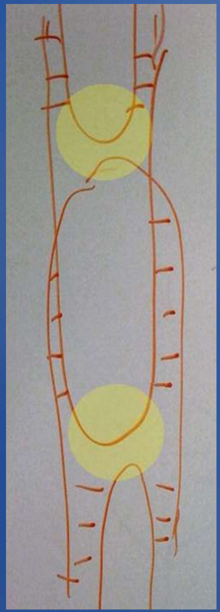
the dawn of
structural DNA
nanotechnology

Crosslinking





In nature, crosslinking is deadly (blocks DNA replication).

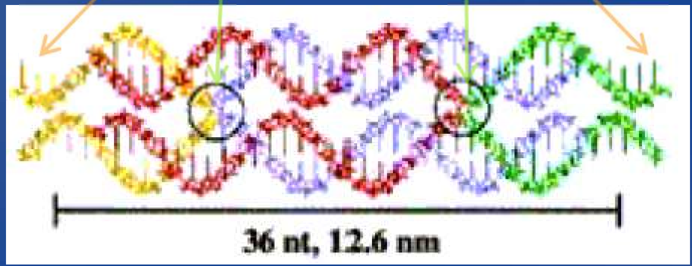


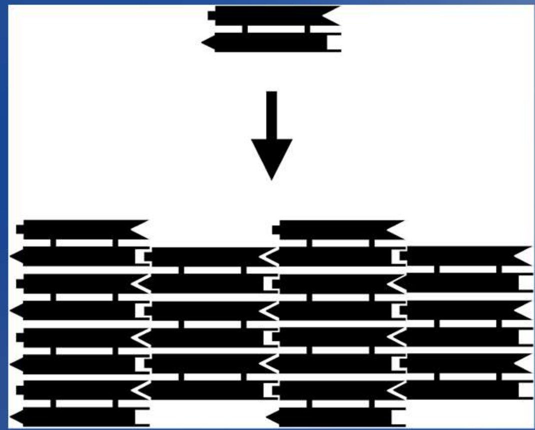
In engineering, crosslinking is the key to using DNA as a construction material.

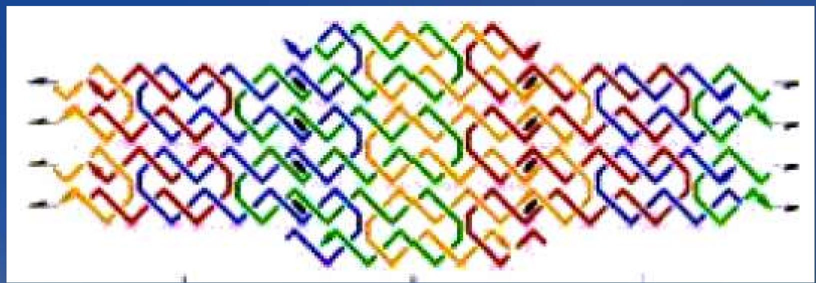
DNA Tiling

4 sticky ends

crosslinking







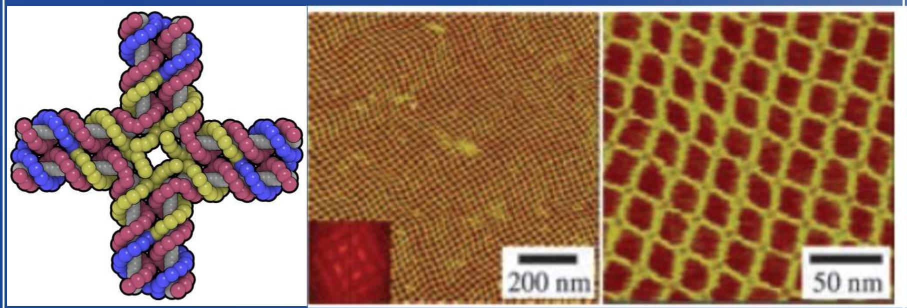
Construction and manipulation of DNA tiles in free space

Pankhudi

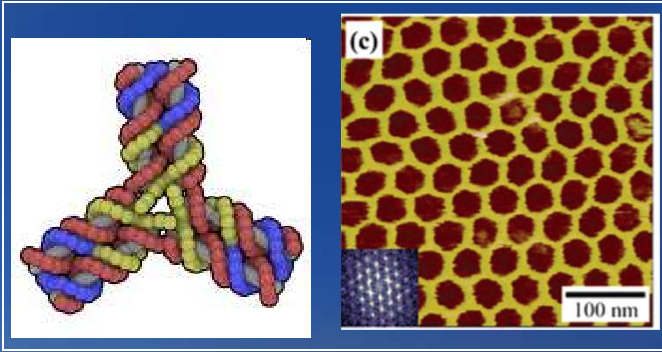
2D DNA Lattices

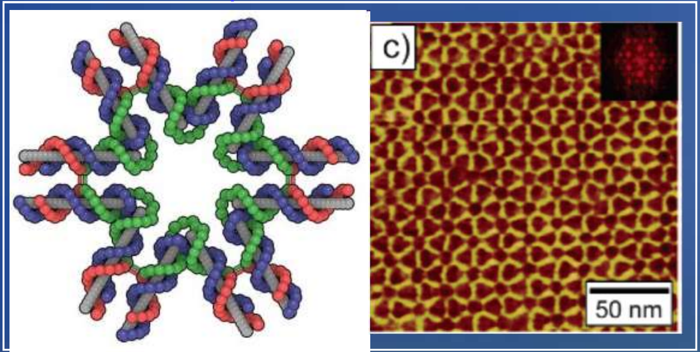
Chengde Mao

Purdue University, USA



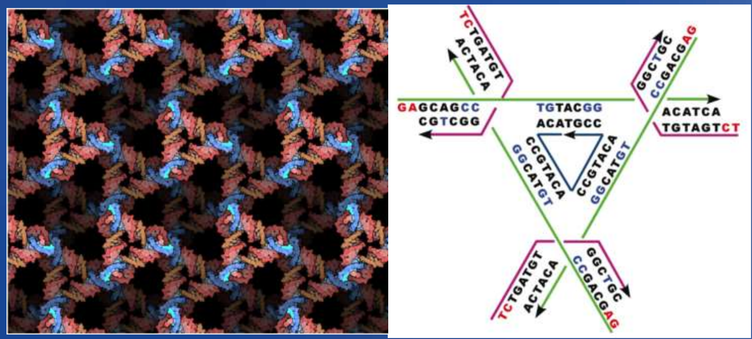
N-point Stars





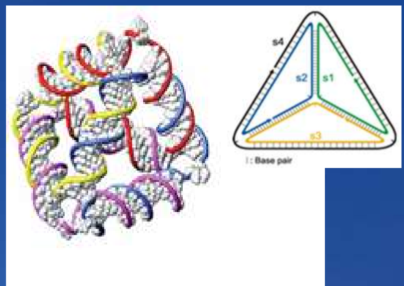
3D DNA Structures

Ned Seeman
NYU



3D Cyrstal

Andrew Tuberfield
Oxford

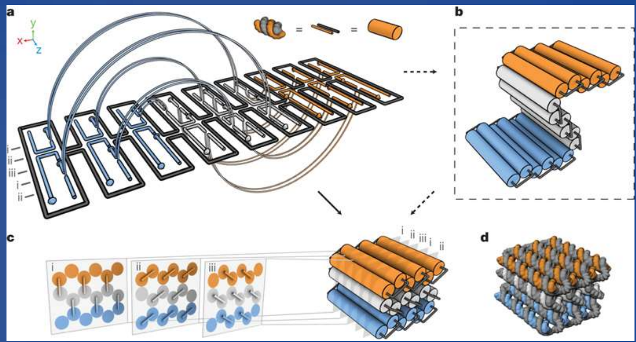


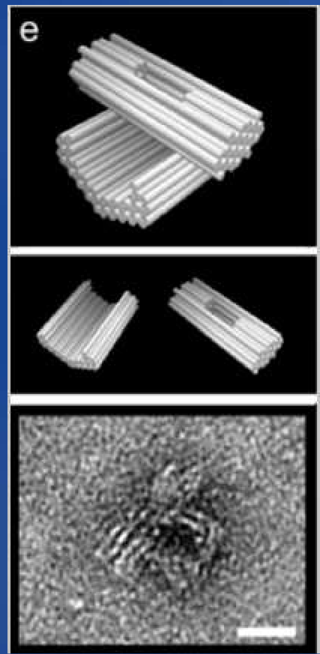
Tetrahedron

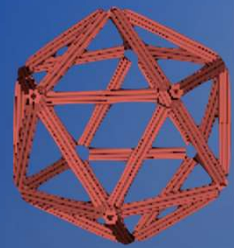
Friedrich Simmel
Munich

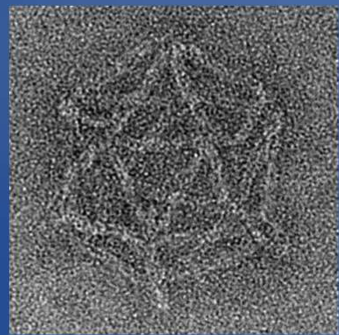
Robotic Arm

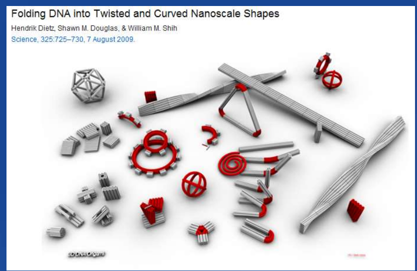
CADnano











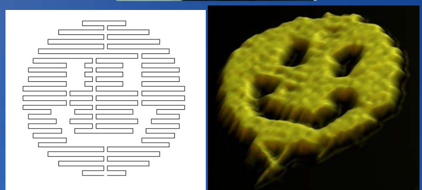
William Shih
Harvard

https://www.youtube.com/watch?v=Ek-FDPymyyg

S.M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf and W. M. Shih
Self-assembly of DNA into nanoscale three-dimensional shapes, Nature (2009)

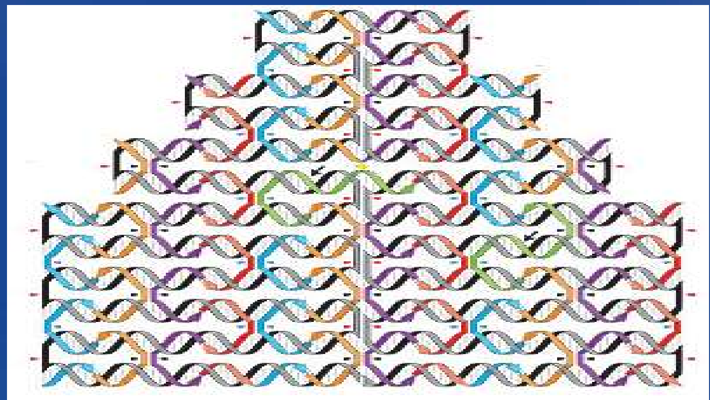
DNA Origami

Folding a long (6407bp) naturally occurring circular ssDNA (from bacteriophage M13) via lots of short 'staple' strands that constrain its shape

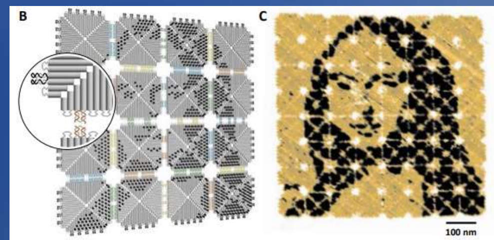


AFM image

Paul Rothemund's
"Disc with three holes" (2006)
Nature 440, 297, 2006



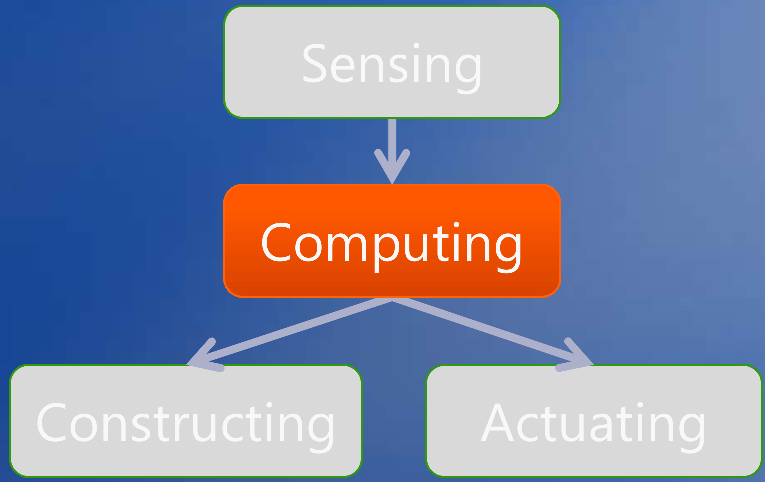
Black/gray: 1 long viral strand (natural DNA)
Color: many short staple strands (synthetic DNA)



AFM image

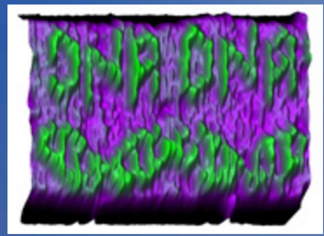
Lulu Qian's
Hierarchical assembly (2017)
Nature, 552(7683):67–71, 2017

Computing



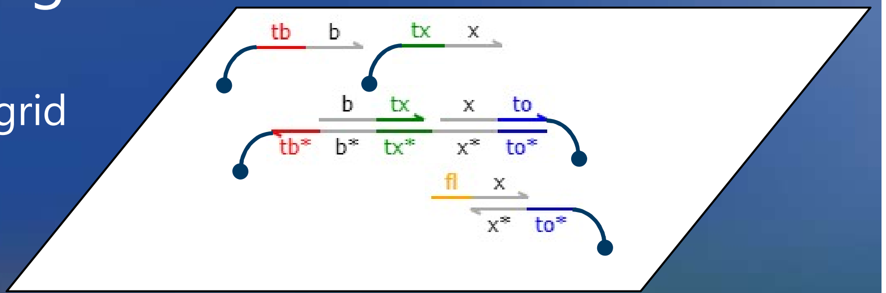
DNA Circuit Boards

- DNA origami are arrays of uniquely-addressable locations
  - Each staple is different and binds to a unique location on the origami
  - It can be extended with a unique sequence so that something else will attach uniquely to it.



Some staples are attached to “green blobs” (as part of their synthesis) Other staples aren’t

- More generally, we can bind “DNA gates” to specific locations
  - And so connect them into “DNA circuits” on a grid
  - Only neighboring gates will interact



Dalchau, Chandran, Gopalkrishnan, Reif, Phillips. 2014

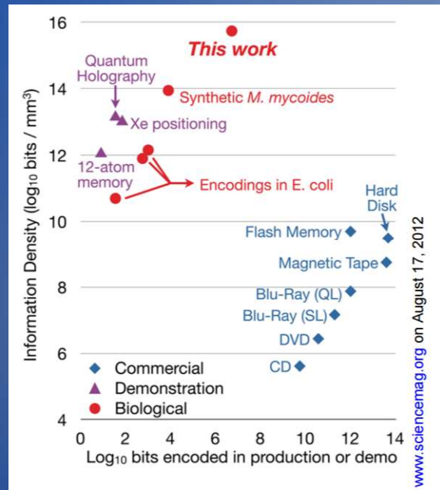
DNA Storage (Read/Write)

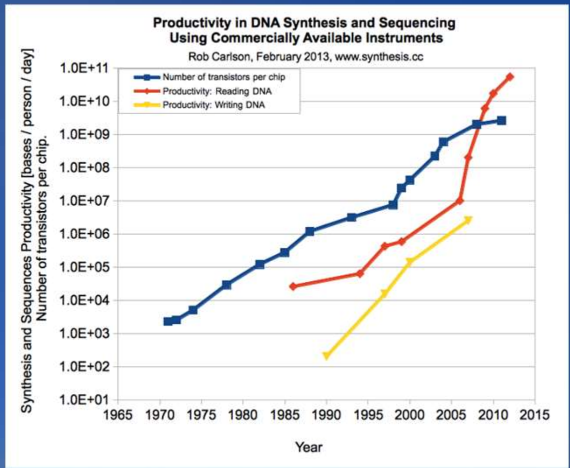
Information-rich physical structures can be used for storage.

DNA has a data density of 140 exabytes (1.4×10^{20} bytes) per mm^3 compared to state-of the art storage media that reaches ~500 megabytes (5×10^8 bytes) per mm^3
DNA has been shown to be stable for millions of years

Next-Generation Digital Information Storage in DNA

George M. Church,[1,2] Yuan Gao,[3] Sriram Kosuri[1,2*]



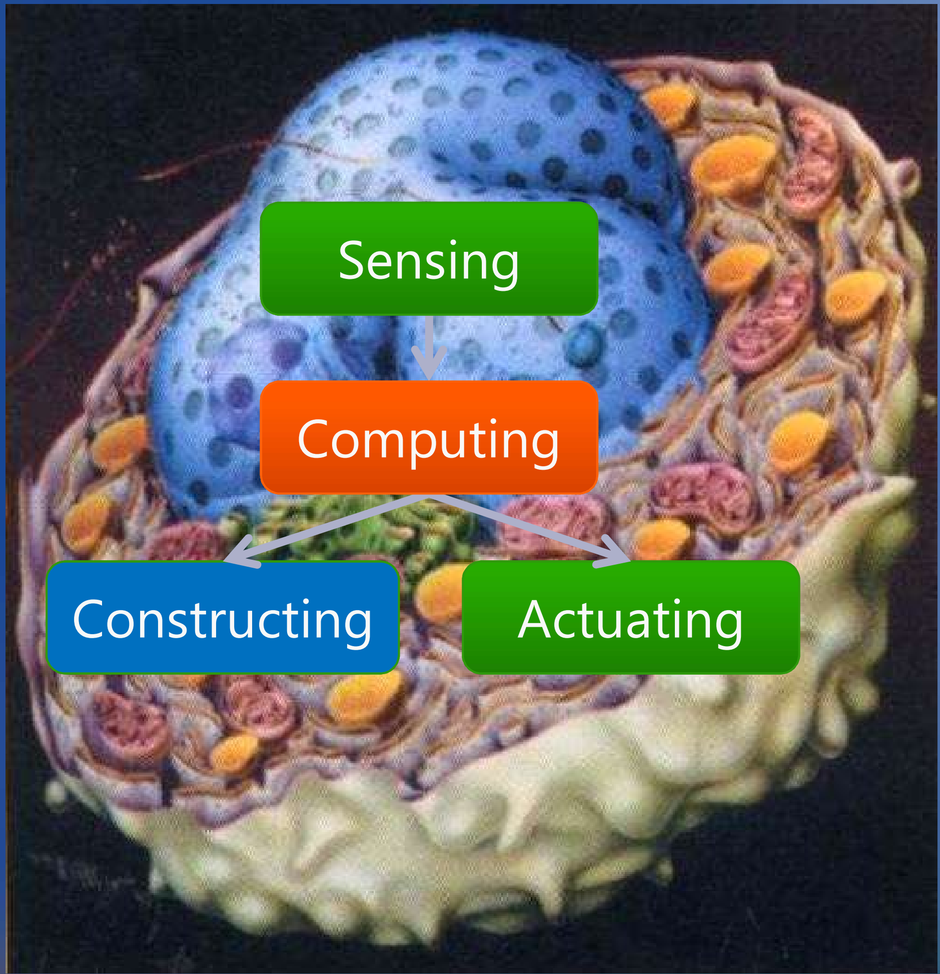


The Pace and Proliferation of Biological Technologies
March 4, 2004 by Rob Carlson

We have machines that can read (sequence) and write (synthesize) DNA. The Carslon Curve of "productivity" is growing much faster than Moore's Law.

Cost of sequencing is decreasing rapidly ($1000 whole human genome), while cost of synthesis is decreasing very slowly.
[Rob Carlson, www.synthesis.cc]

Curing



H.Lodish & al. Molecular Cell Biology 4th ed.

Interfacing to Biology

- A doctor in each cell

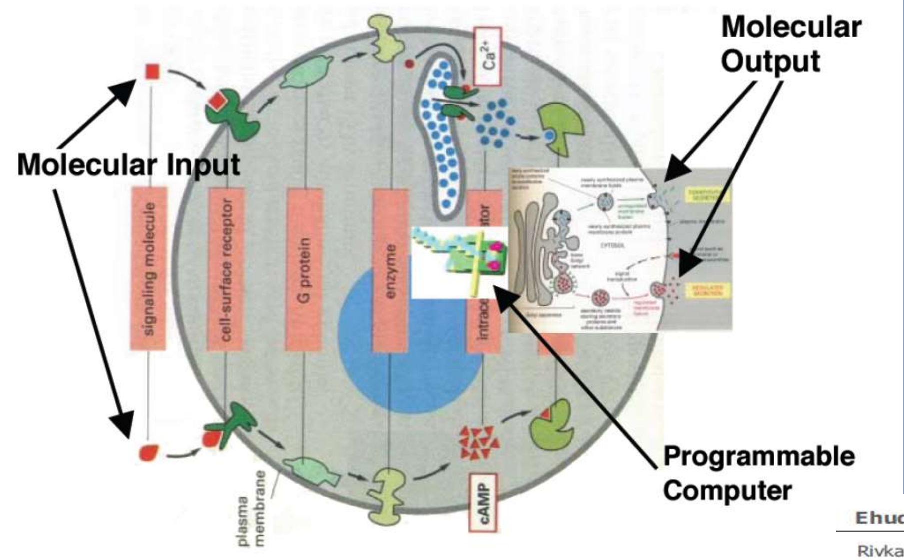


Fig. 1 Medicine in 2050: "Doctor in a Cell"

Ehud Shapiro
Rivka Adar
Kobi Benenson
Gregory Linshitz
Aviv Regev
William Silverman

Molecules and computation

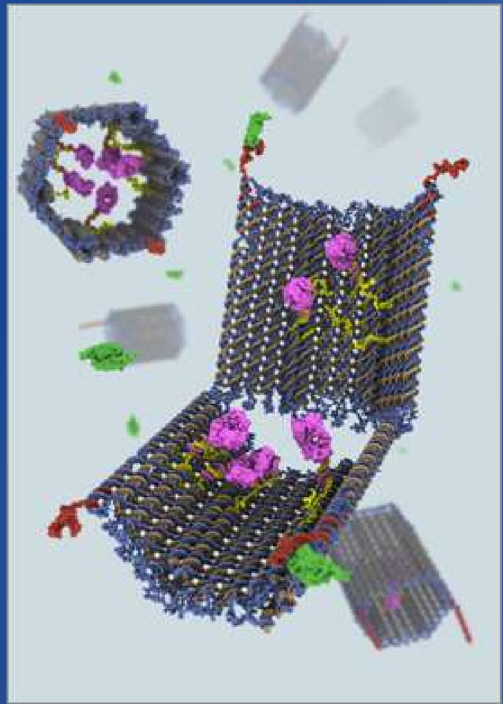
~2002

Programmed Drug Delivery

A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads

Shawn M. Douglas*, Ido Bachelet*, George M. Church†
+ See all authors and affiliations

Science 17 Feb 2012:
Vol. 335, Issue 6070, pp. 831-834
DOI: 10.1126/science.1214081

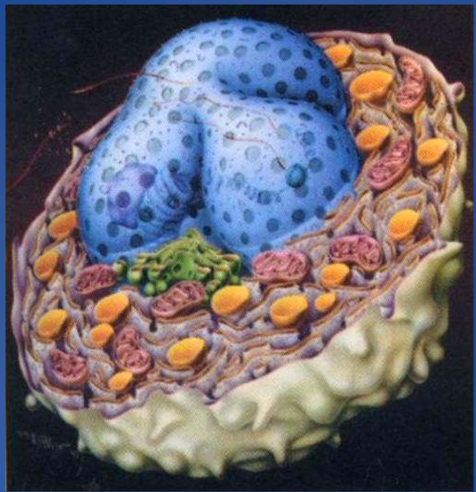


~2002

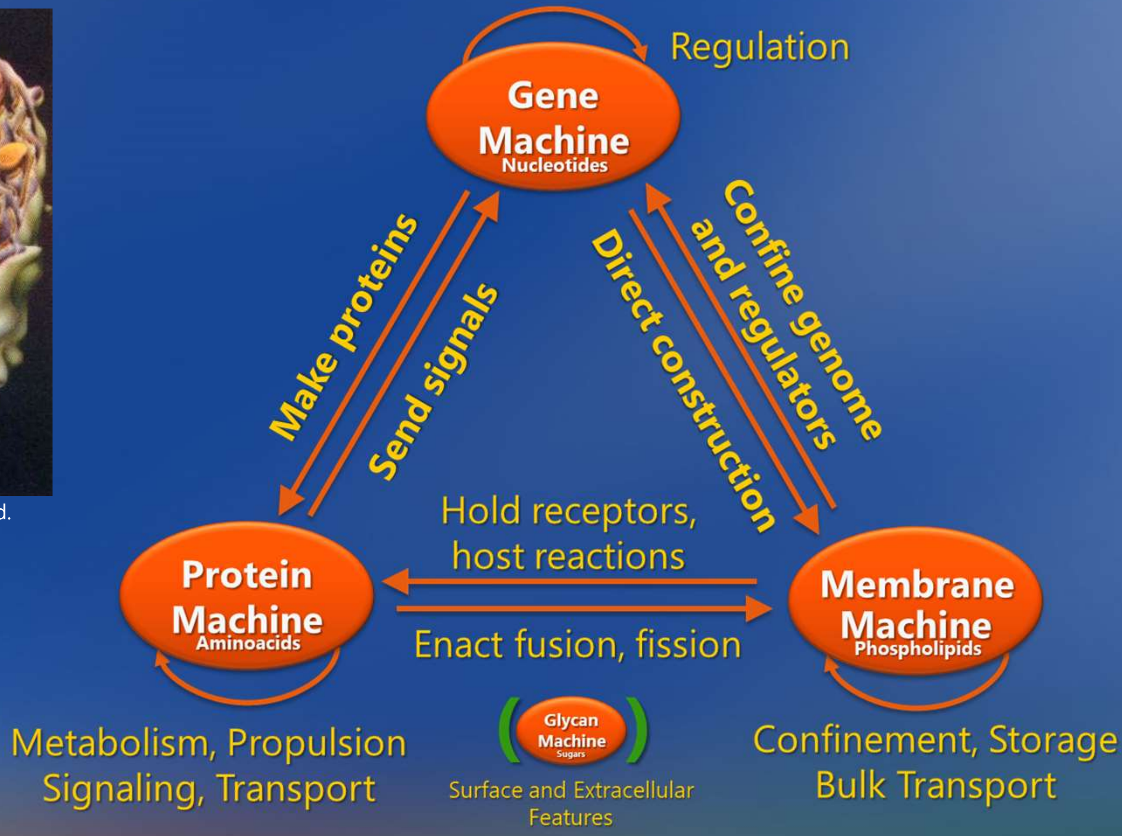
Molecular Programming:
The Biological Aspect

Biological systems are already
'molecularly programmed'

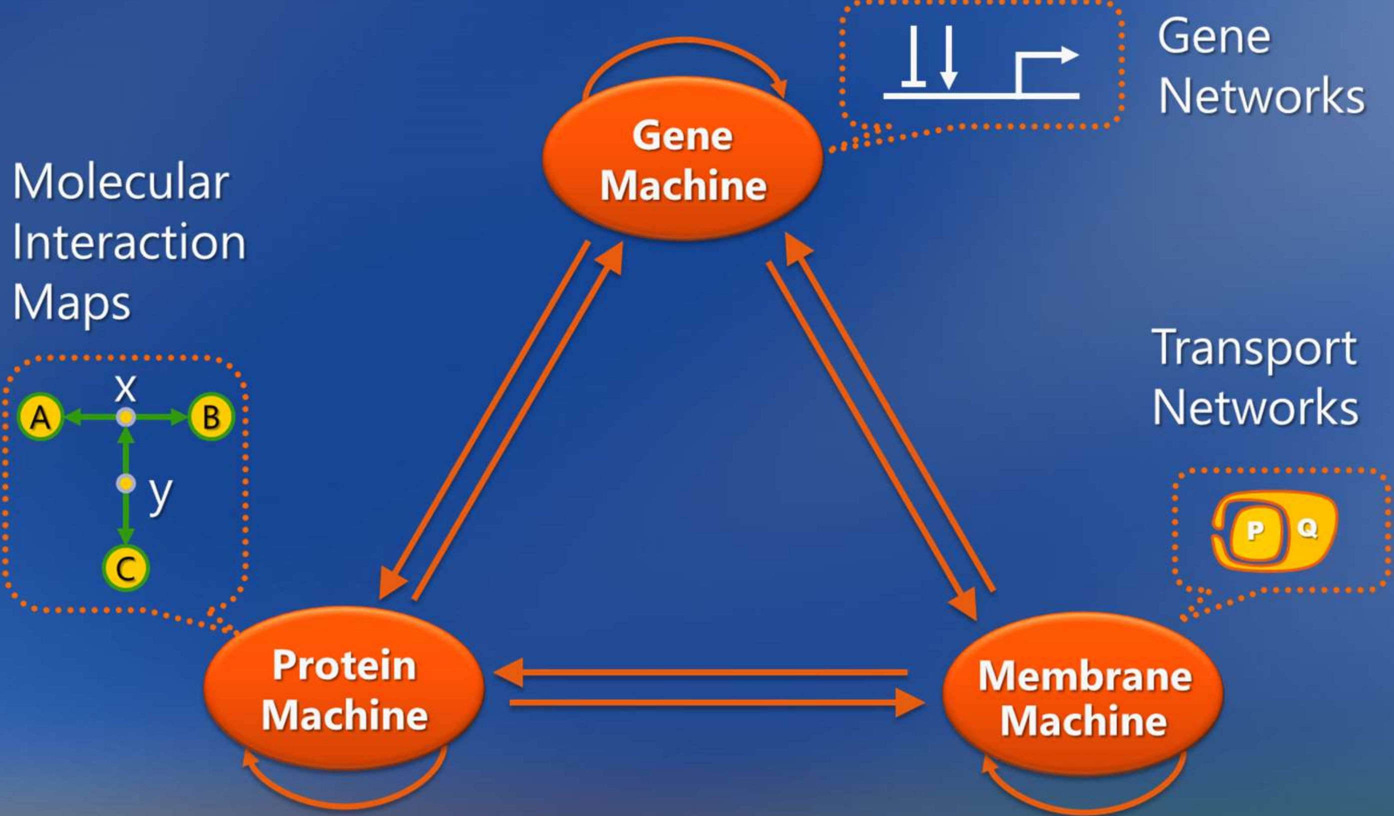
Abstract Machines of Biology



H.Lodish & al. Molecular Cell Biology 4th ed.



Biological Languages



But ...

- Biology is programmable, but (mostly) not by us!
- Still work in progress:
  - Gene networks are being programmed in synthetic biology, but using existing 'parts'
  - Protein networks are a good candidate, but we cannot yet effectively design proteins
  - Transport networks are being investigated for programming microfluidic devices that manipulate vesicles

Molecular Programming:
The Execution Aspect

How do we "run" a molecular program?

Programming Language: Chemistry

- A Lingua Franca between Biology, Dynamical Systems, and Concurrent Languages

- Chemical Reaction Networks
  - $A + B \rightarrow_r C + D$ (the program)

- Ordinary Differential Equations
  - $d[A]/dt = -r[A][B]$ ... (the behavior)

- Rich analytical techniques based on Calculus and more recently on stochastic models

Chemical Programming Examples

specification | *program*

Y := min(X1, X2)

X1 + X2 -> Y

Y := max(X1, X2)

X1 -> L1 + Y
X2 -> L2 + Y
L1 + L2 -> K
Y + K -> 0

max(X1,X2)=
(X1+X2)-min(X1,X2)

(but is not computed "sequentially": it is a form of concurrent computation)

chemical reaction network

Chemical Reaction Networks

- Finite list of chemical reactions over a finite set of species
  - N.B.: "abstract" species, not specific atoms/molecules that physically exist
- Computationally Powerful
  - Turing-complete up to an arbitrarily small error
- Full Turing Completeness
  - When including complexation (polymerization), which DNA enables
    (complexation encodes an actual infinity of chemical reactions by finite means)

51

How do we “run” Chemistry?

- Chemistry is not easily executable
  - “Please Mr Chemist, execute me this bunch of reactions that I just made up”
- Most molecular languages are not executable
  - They are descriptive (modeling) languages
- How can we execute molecular languages?
  - With real molecules?
  - That we can design ourselves?
  - And that we can buy on the web?

DNA Strand Displacement

An "unnatural" use of DNA for emulating *any* system of chemical reactions with real molecules

Domains

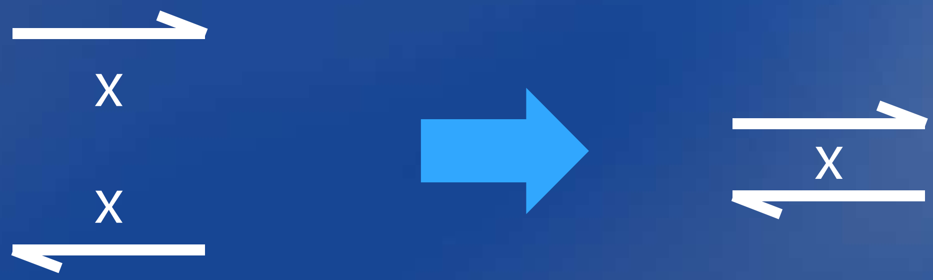
- Subsequences on a DNA strand are called domains
  - *provided* they are “independent” of each other

- Differently named domains must not hybridize
  - With each other, with each other’s complement, with subsequences of each other, with concatenations of other domains (or their complements), etc.

Short Domains

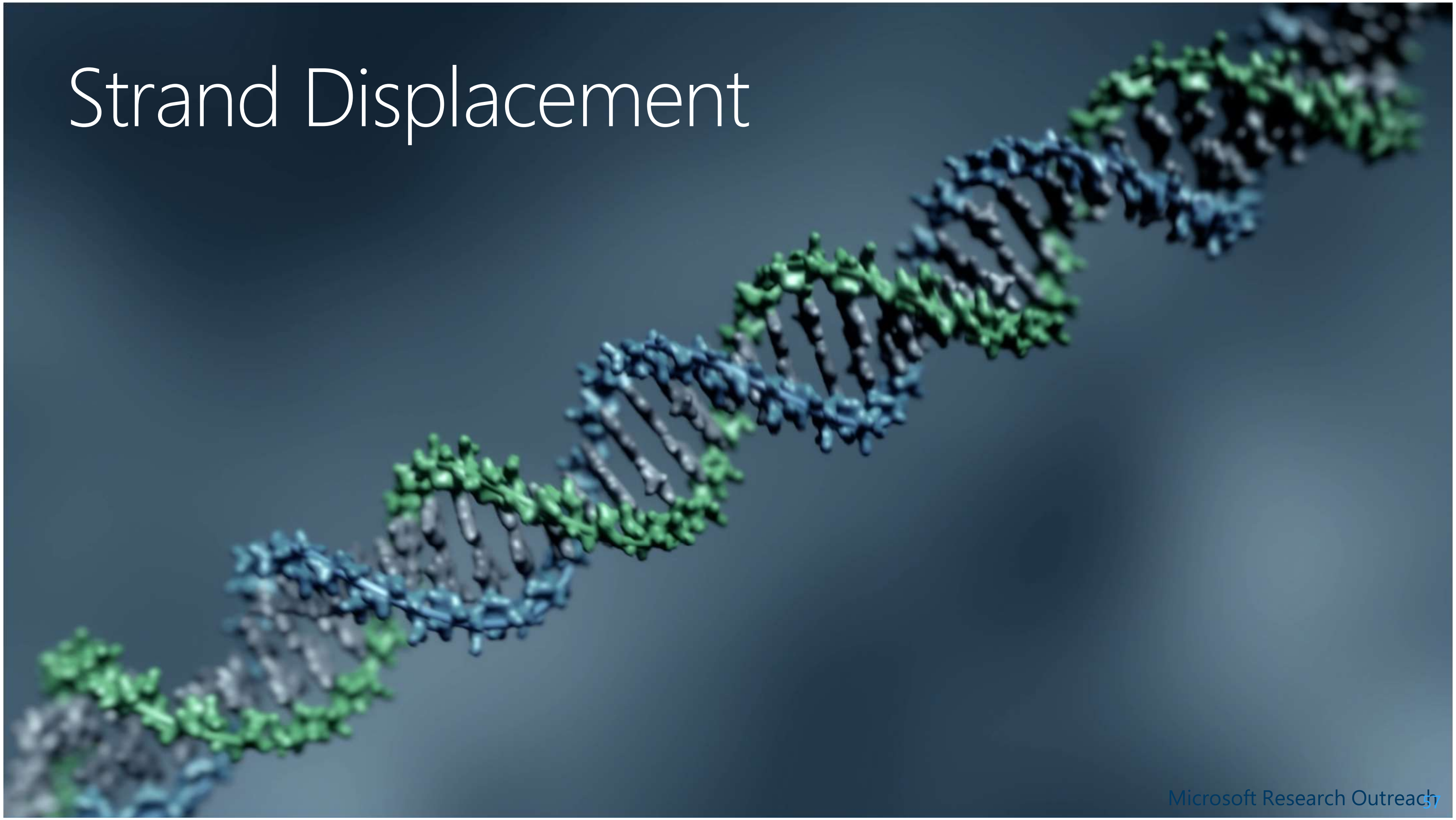
Reversible Hybridization

Long Domains

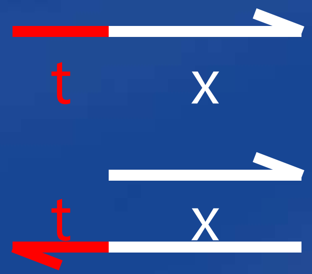


Irreversible Hybridization

Strand Displacement

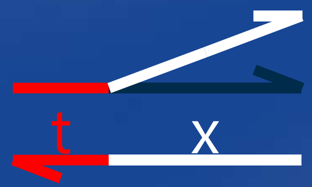


Strand Displacement



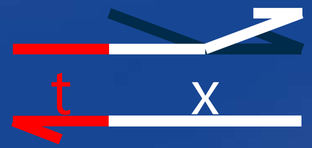
"Toehold Mediated"

Strand Displacement



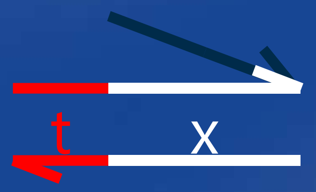
Toehold Binding

Strand Displacement



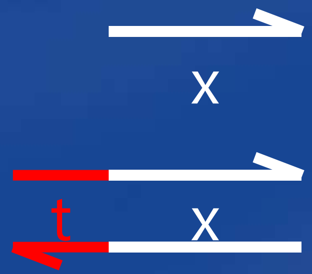
Branch Migration

Strand Displacement



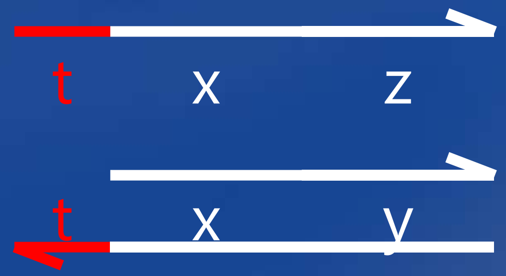
Displacement

Strand Displacement

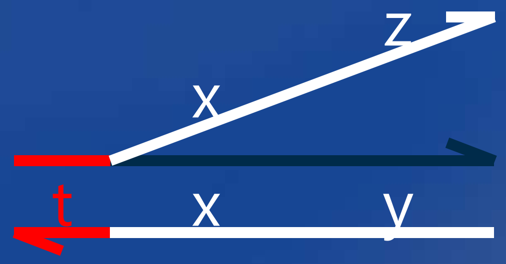


Irreversible release

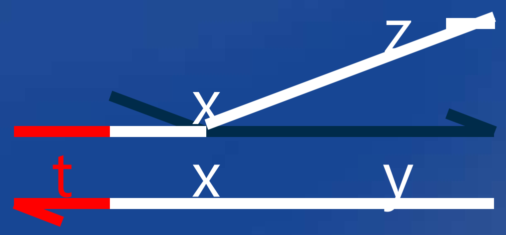
Bad Match



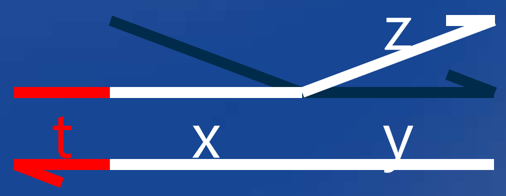
Bad Match



Bad Match



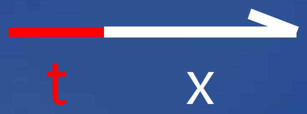
Bad Match



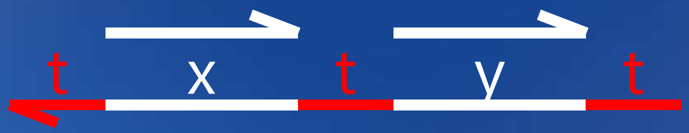
Cannot proceed
Hence will undo

Two-Domain Architecture

- Signals: 1 toehold + 1 recognition region



- Gates: “top-nicked double strands” with open toeholds



Garbage collection “built into” the gate operation

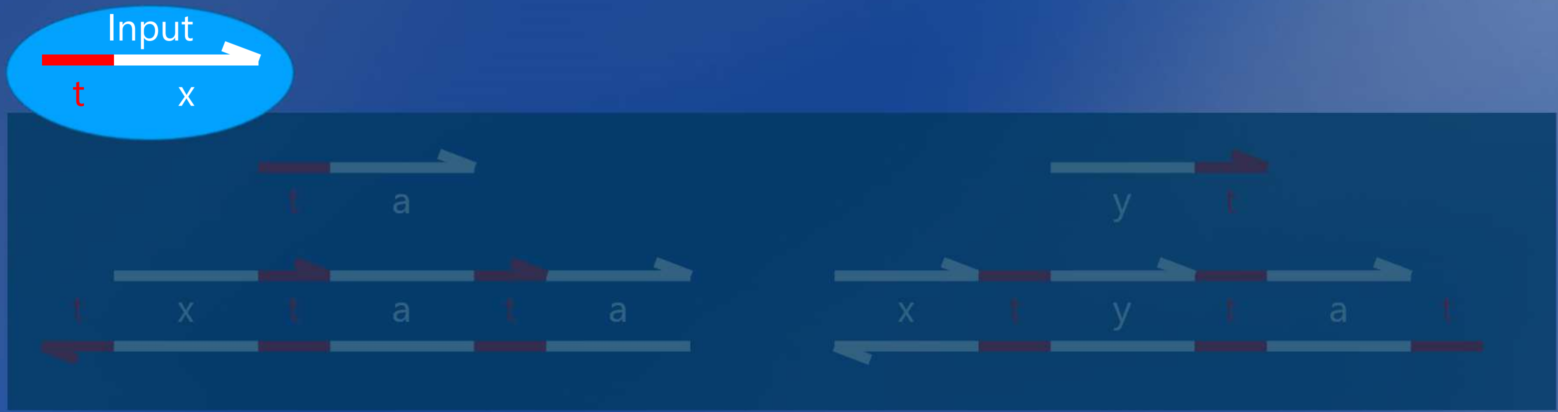
Two-Domain DNA Strand Displacement

Luca Cardelli

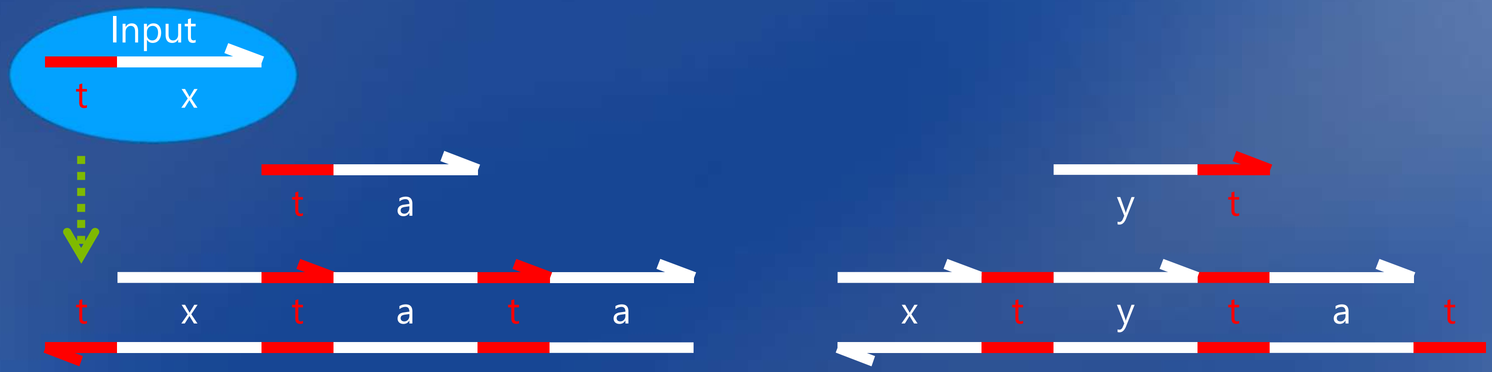
In S. B. Cooper, E. Kashefi, P. Panangaden (Eds.): Developments in Computational Models (DCM 2010). EPTCS 25, 2010, pp. 33-47. May 2010.

Transducer

Transducer x→y



Transducer x→y

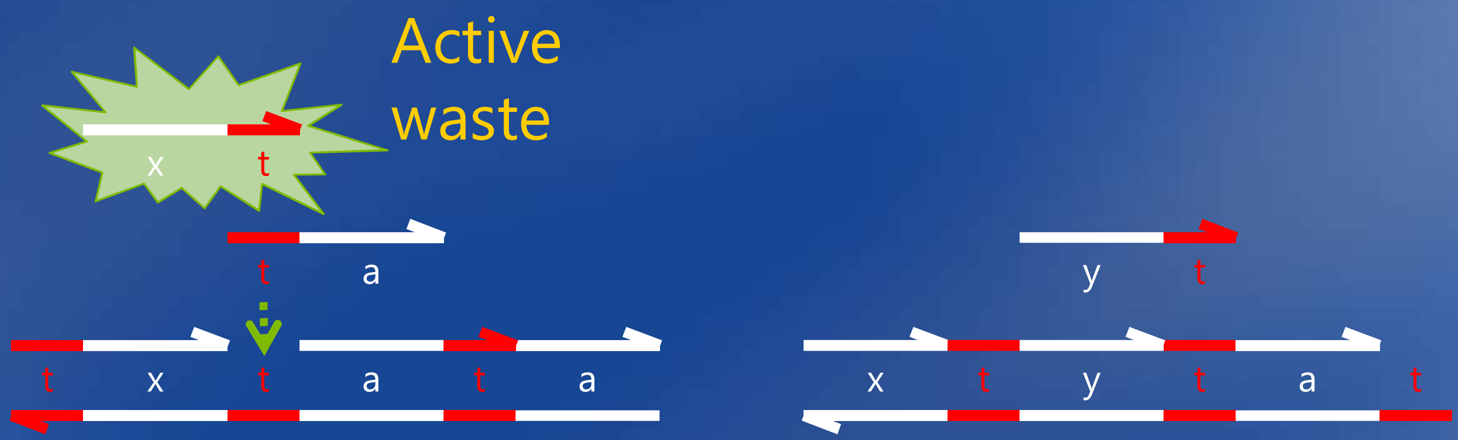


Built by self-assembly!

ta is a *private* signal (a different 'a' for each xy pair)

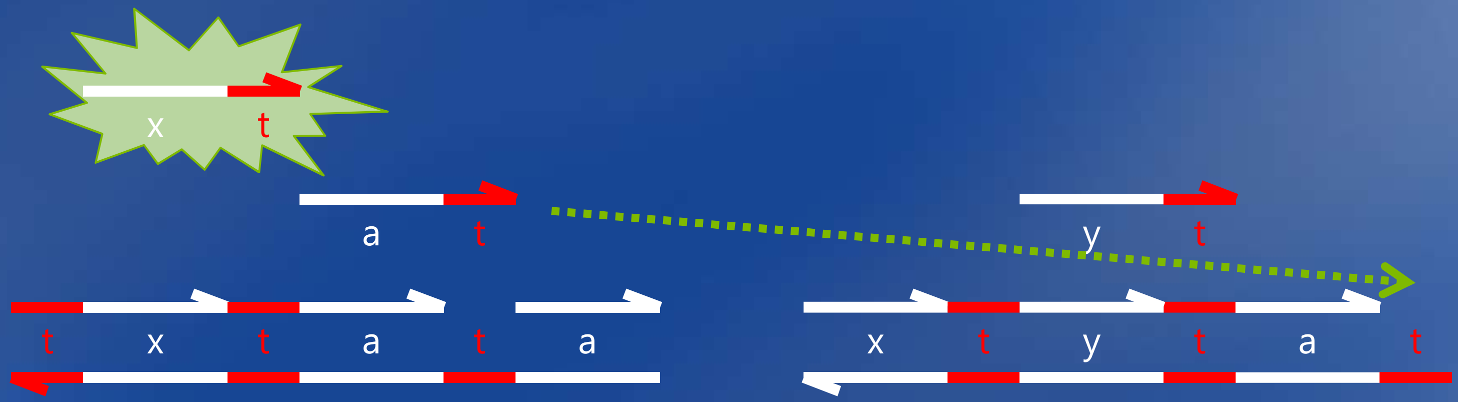
Transducer x→y

Transducer x→y



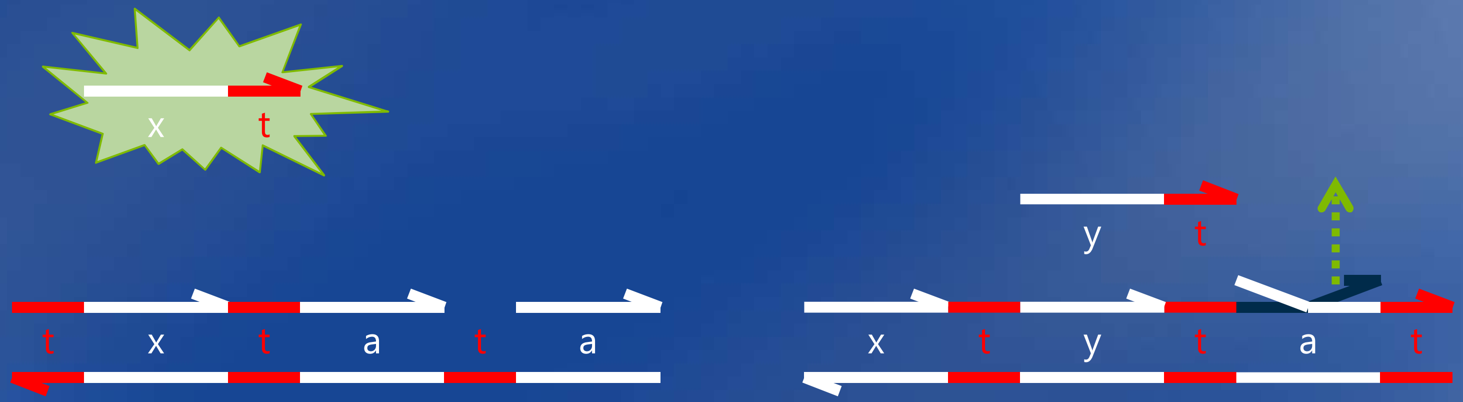
Transducer x→y

Transducer x→y

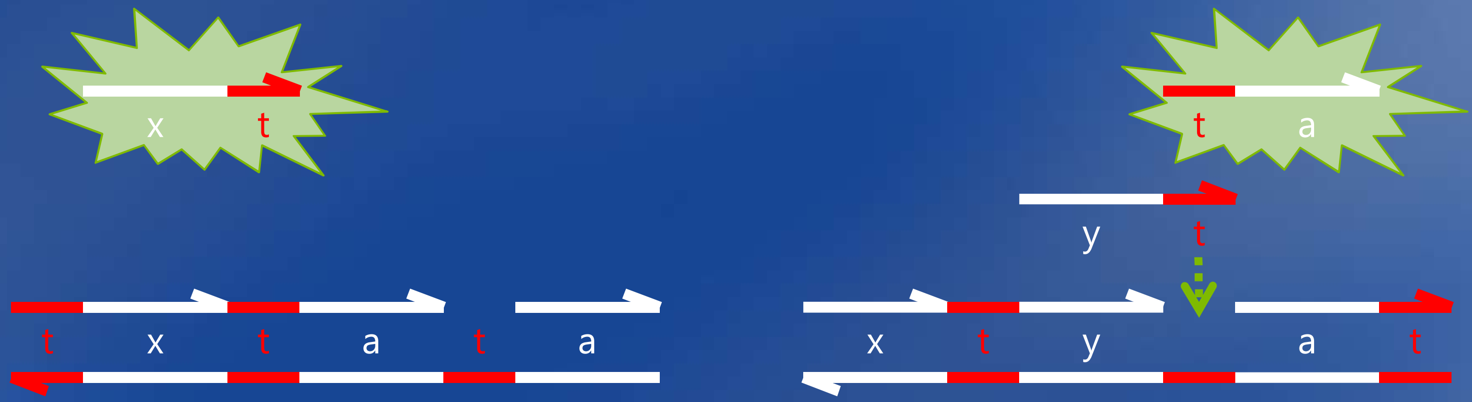


So far, a **tx** *signal* has produced an **at** *cosignal*.
But we want signals as output, not cosignals.

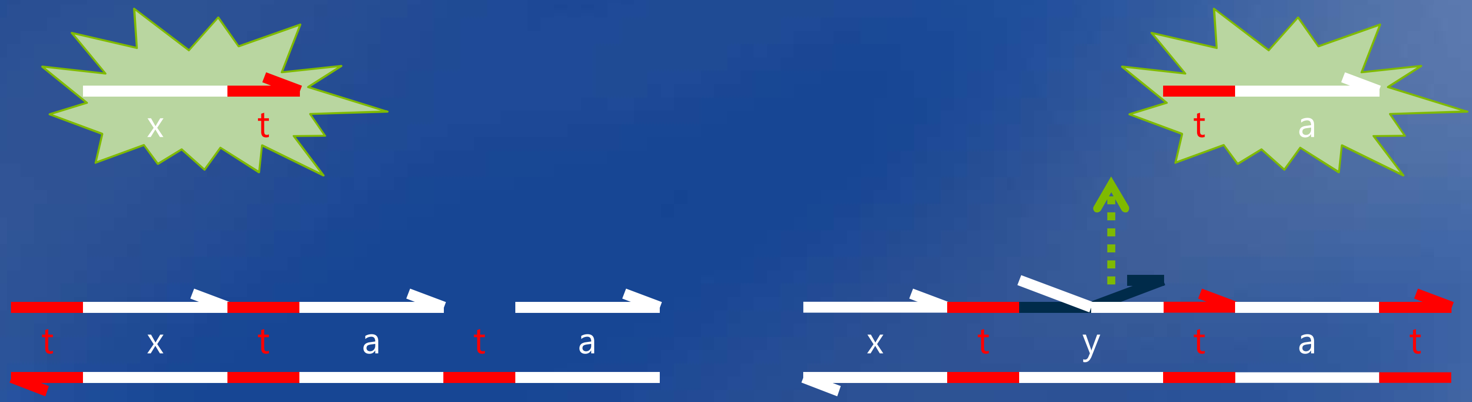
Transducer x→y



Transducer x→y



Transducer x→y



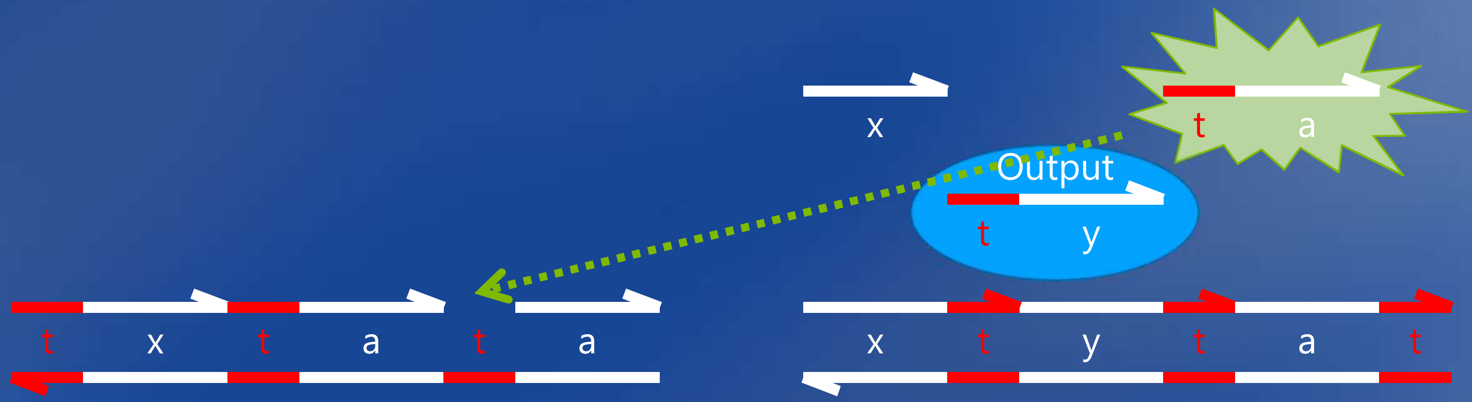
Transducer x→y

Here is our output **ty** *signal*.
But we are not done yet:
1) We need to make the output irreversible.
2) We need to remove the garbage.
We can use (2) to achieve (1).

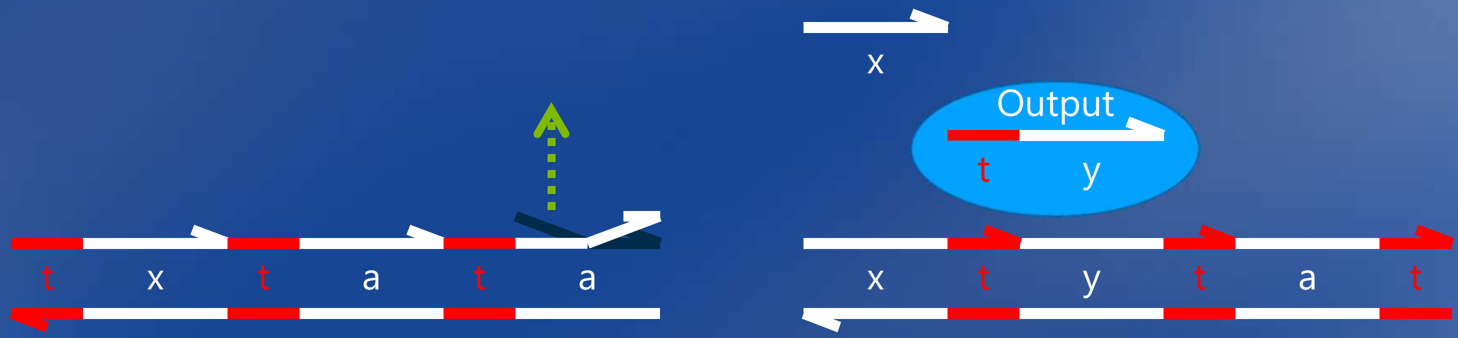
Transducer x→y

Transducer x→y

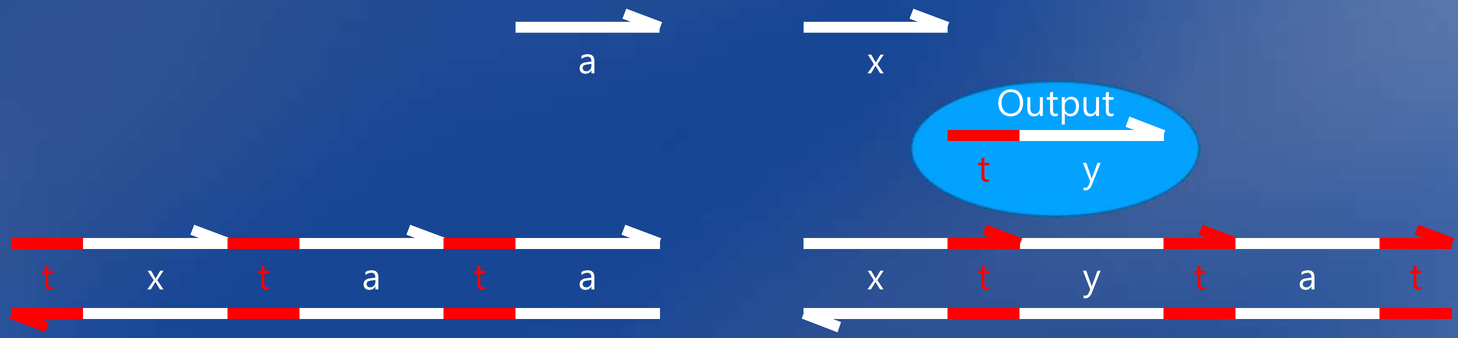
Transducer x→y



Transducer x→y



Transducer x→y



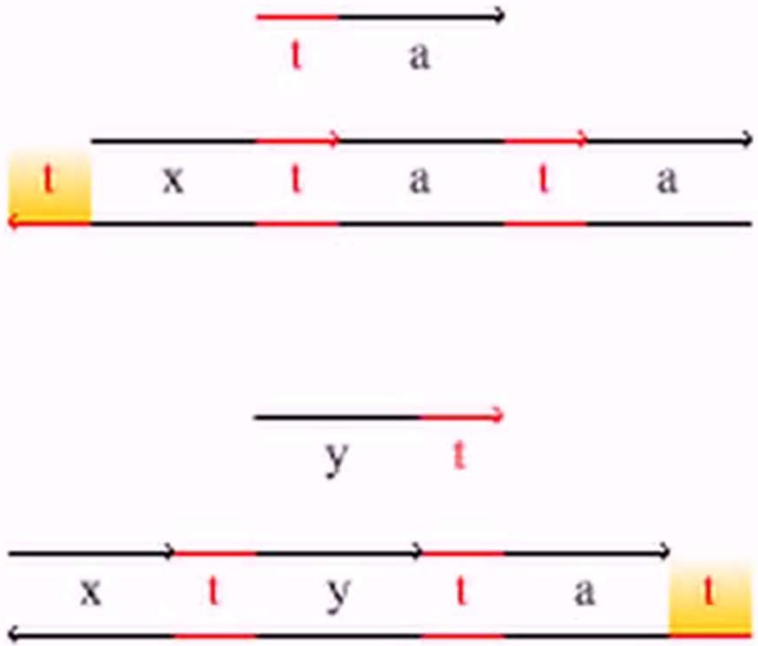
Transducer x→y

Done.

N.B. the gate is consumed: it is the energy source

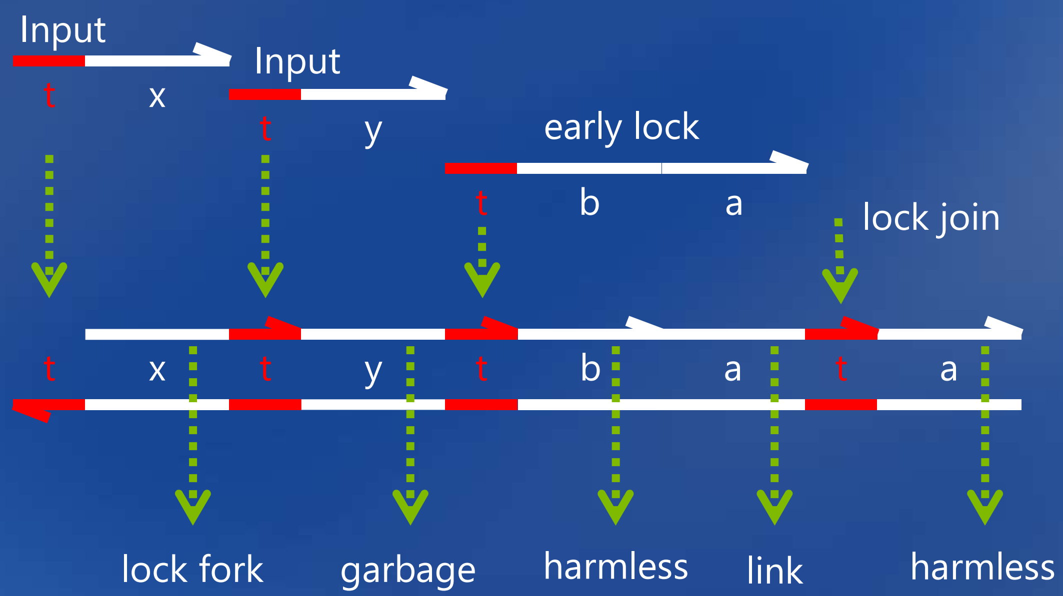
(no proteins, no enzymes, no heat-cycling, etc.; just DNA in salty water)

Transducer $x \rightarrow y$



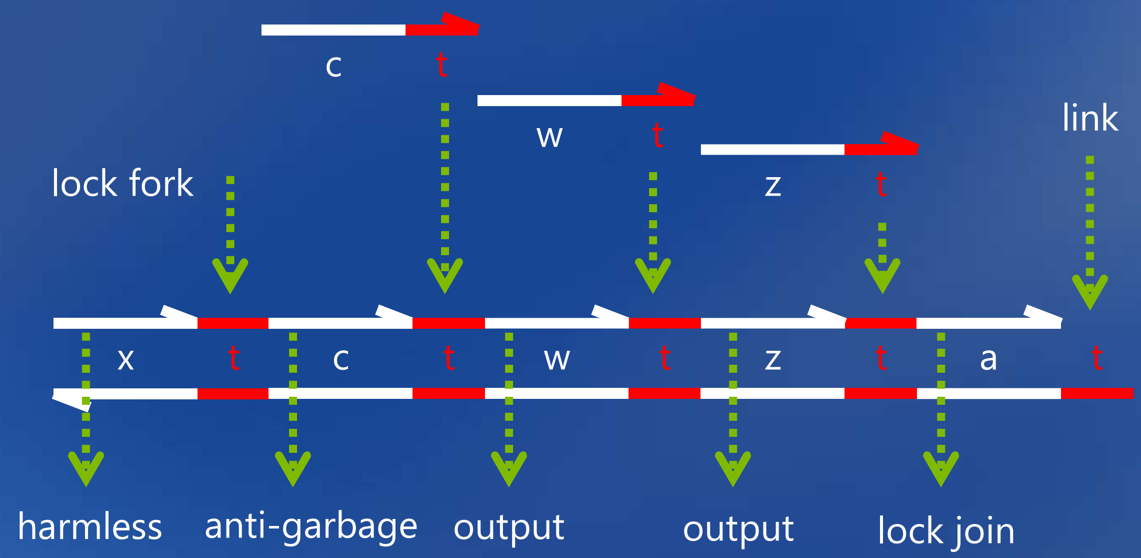
Reaction $x + y \rightarrow z + w$

join
half



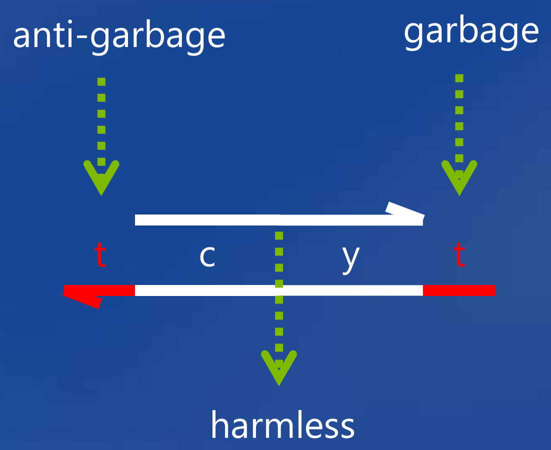
Reaction $x + y \rightarrow z + w$

fork half

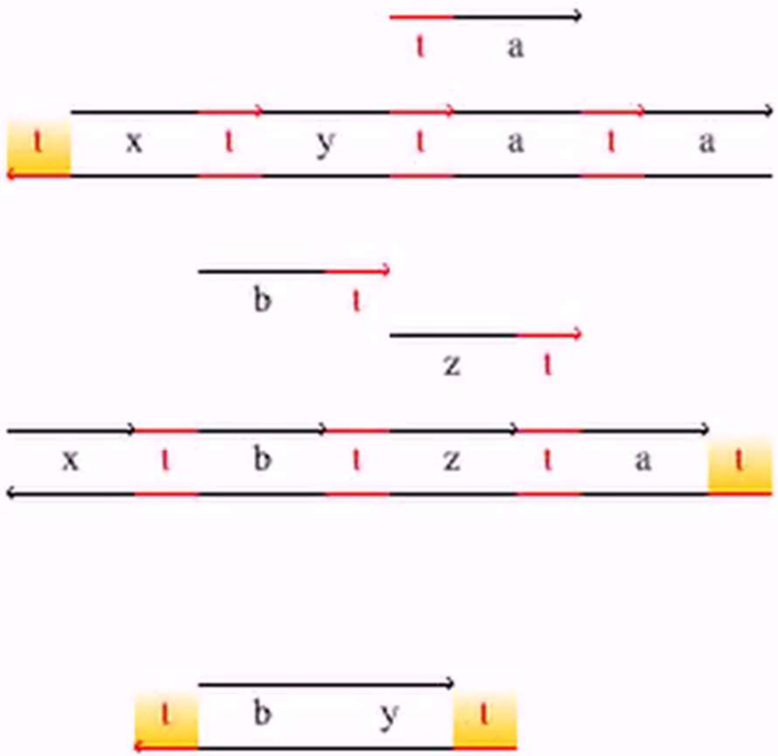


Reaction $x + y \rightarrow z + w$

garbage collection



Join x+y → z

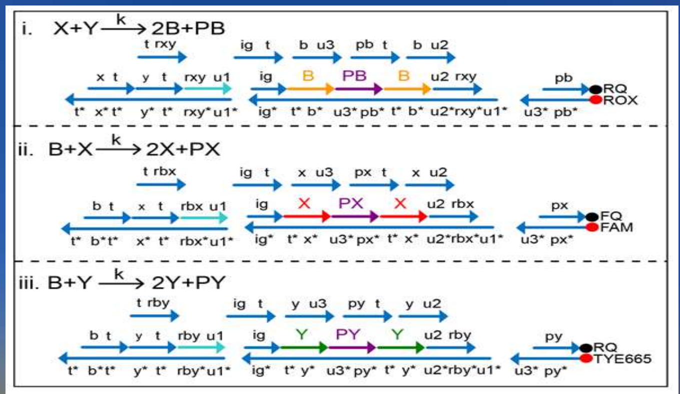


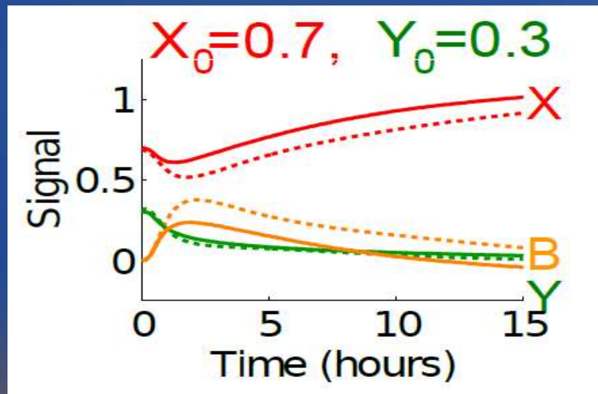
DNA Implementation of the Approximate Majority algorithm

nature
nanotechnology

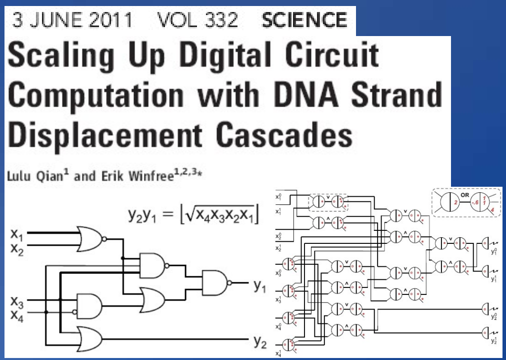
Programmable chemical controllers made from DNA

Yuan-Jyue Chen, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik & Georg Seelig

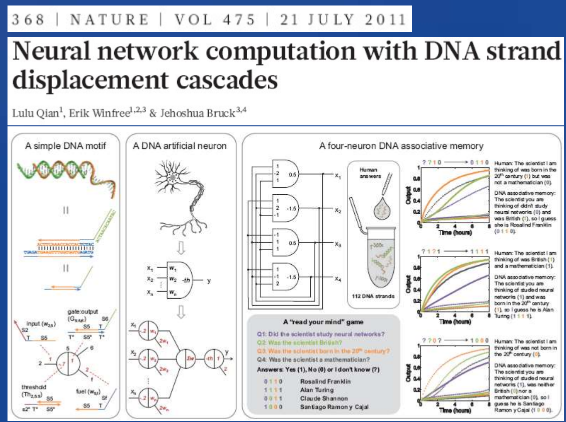




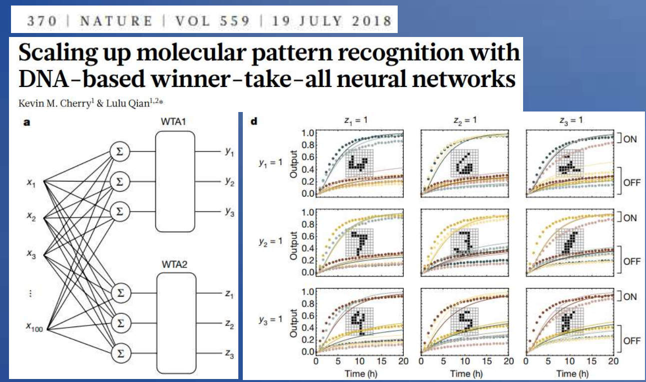
Some Large-scale Circuits (so far…)



Computing the square root of a 4-bit number



Classifying 4 distinct 4-bit patterns via 4 neurons



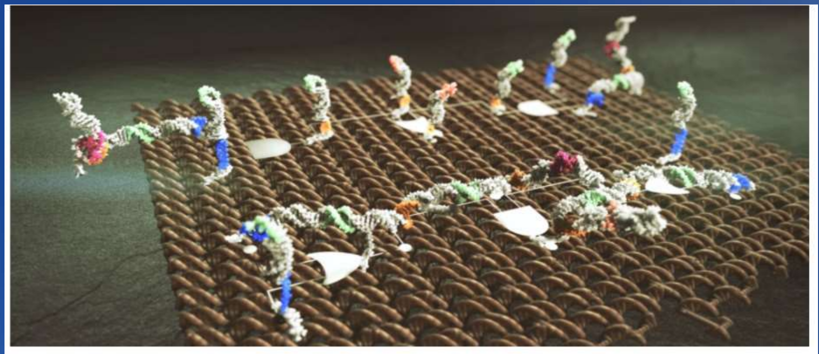
Classifying 9 distinct 100-bit patterns via WTA networks

Scaling up: DNA Circuit Boards

ARTICLES
PUBLISHED ONLINE: 24 JULY 2017 | DOI: 10.1038/NNANO.2017.127
nature nanotechnology

A spatially localized architecture for fast and modular DNA computing

Gourab Chatterjee[1], Neil Dalchau[2], Richard A. Muscat[3], Andrew Phillips[2]* and Georg Seelig[3,4]*



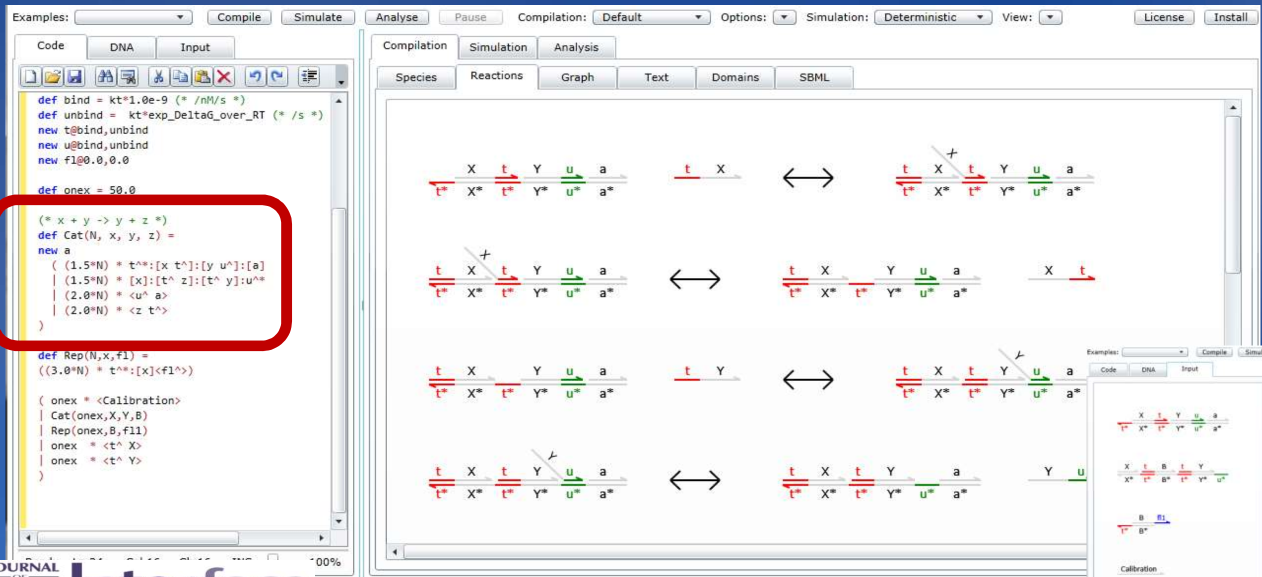
The first computational circuit boards made of DNA
https://www.microsoft.com/en-us/research/blog/researchers-build-nanoscale-computational-circuit-boards-dna

Physical Execution

A wetlab pipeline for Molecular Programming

Computer Aided Design

MSRC Biological Computation Group



Visual DSD

A Development Environment for DNA Strand Displacement

Journal of the Royal Society **Interface**

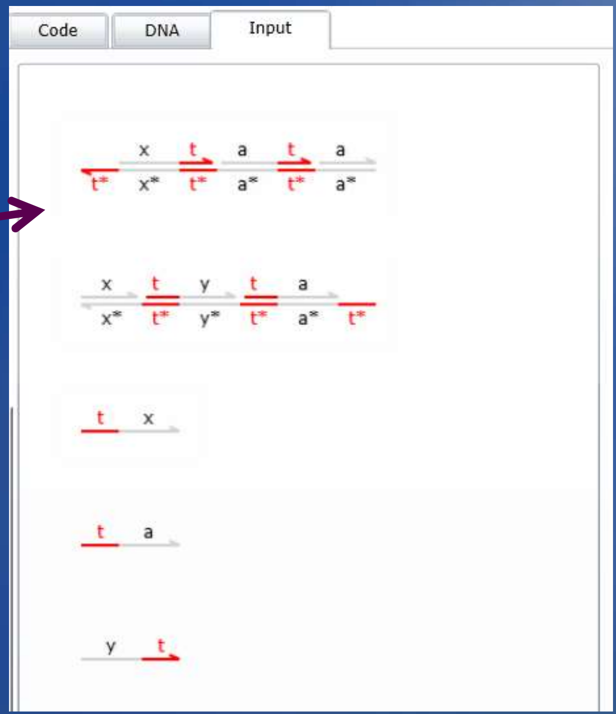
A programming language for composable DNA circuits

Andrew Phillips and Luca Cardelli

Output of Design Process

- Domain structures
  - (DNA sequences to be determined)

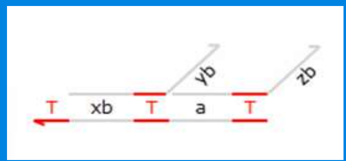
"Ok, how do I
run this for real"

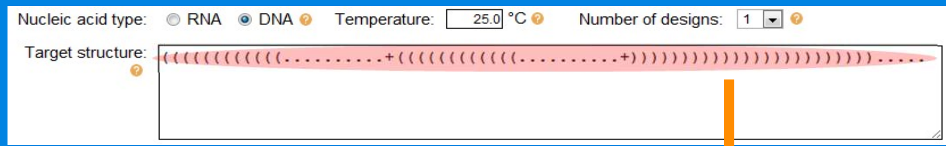


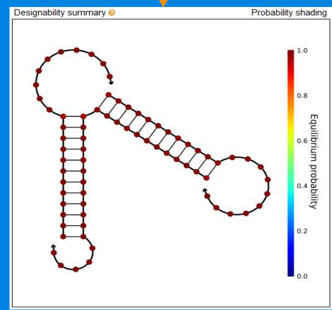
From Structures to Sequences

www.nupack.org

DSD Structure → "Dot-Paren" representation

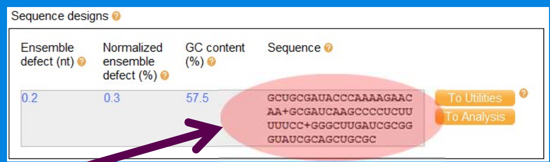






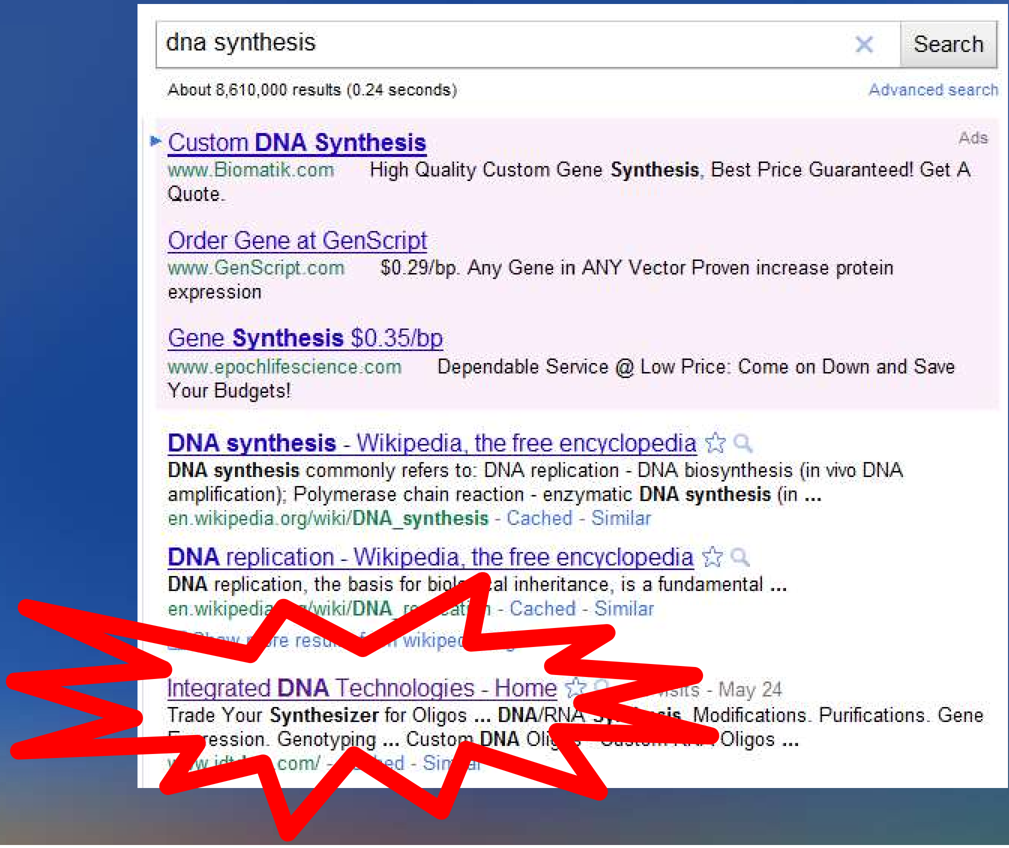
Thermodynamic Synthesis

Output Sequences



"Ok, where do I buy these?"

“DNA Synthesis”



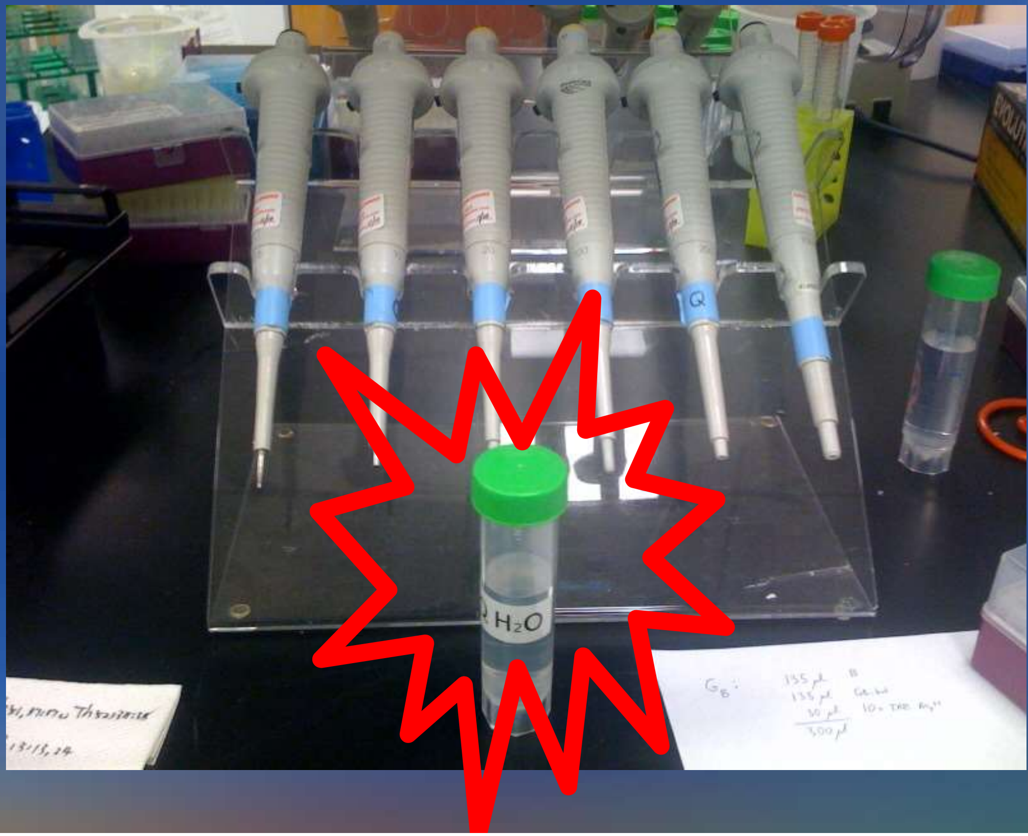
From Sequences to Molecules

- Copy&Paste from nupack

Molecules by FedEx

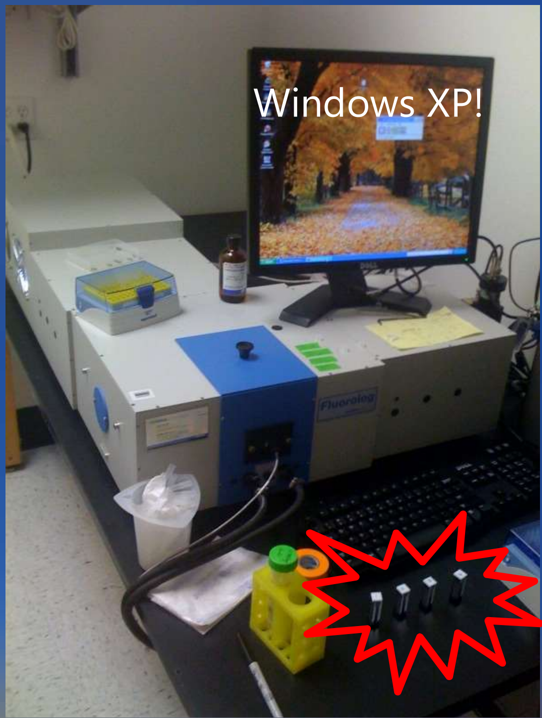
"Ok, how do I run these?"

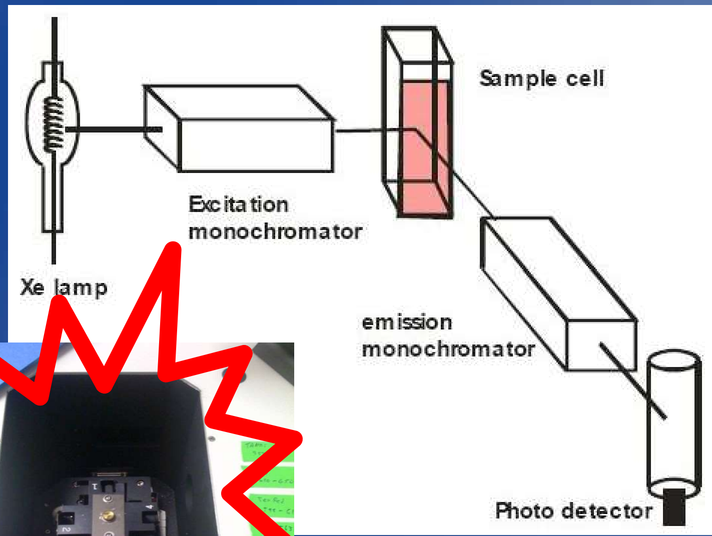
Add Water



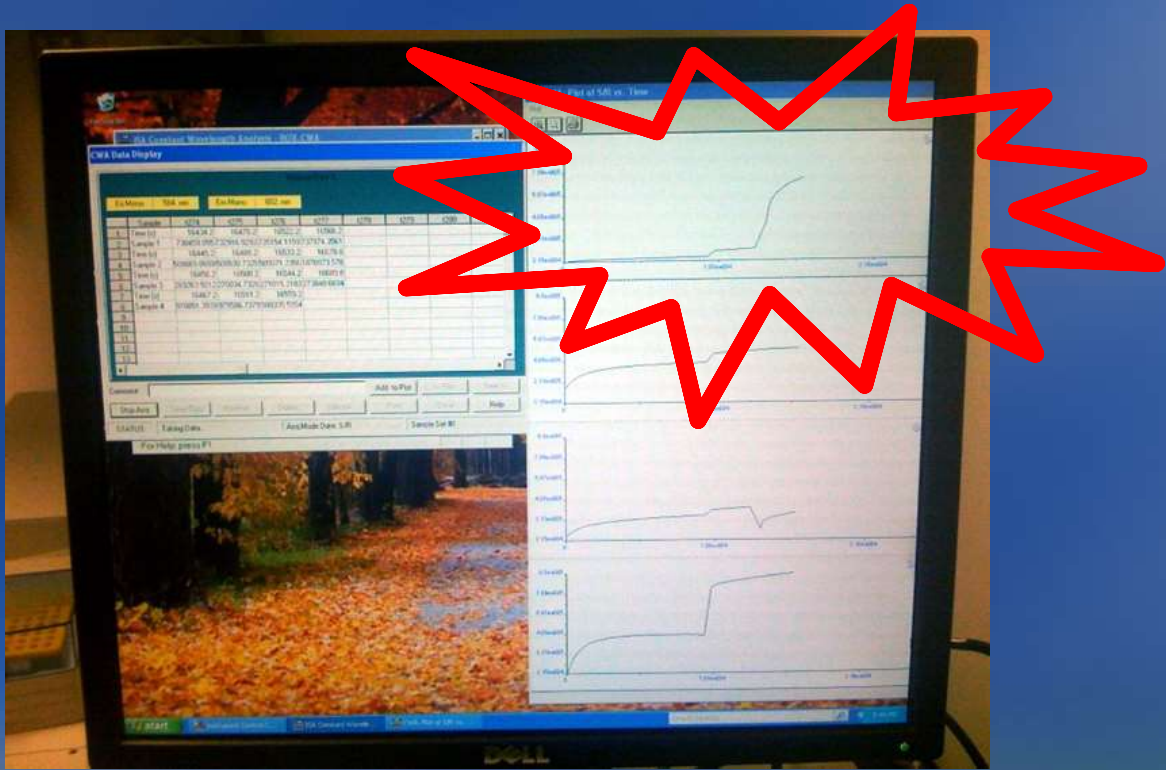
Execute (finally!)

- Fluorescence is your one-bit 'print' statement





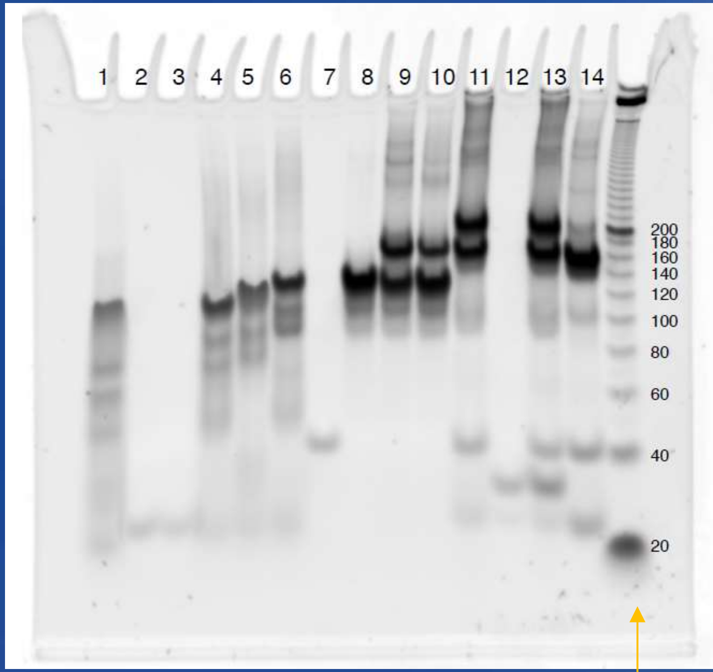
Output



Debugging

- A core dump

polyacrylamide gel electrophoresis



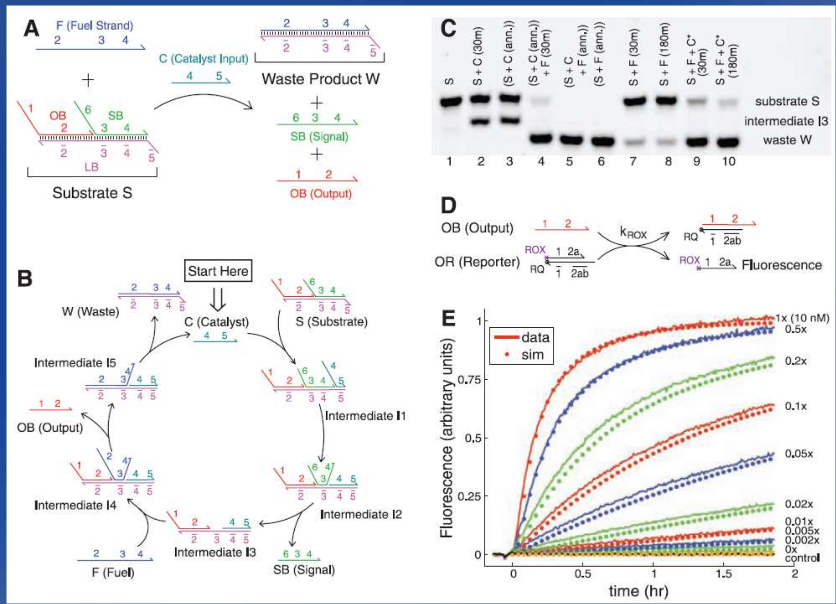
DNA strand length

Various processing stages

Calibration scale

Delivery!

Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA
David Yu Zhang, *et al.*
Science **318**, 1121 (2007);
DOI: 10.1126/science.1148532



Final Remarks

State of the art

- Building a full software/hardware pipeline for a new fundamental technology
  - Mathematical Foundations [~ concurrency theory in the 80's]
  - Programming Languages [~ software engineering in the 70's]
  - Analytical Methods and Tools [~ formal methods in the 90's]
  - Device Architecture and Manufacturing [~ electronics in the 60's]

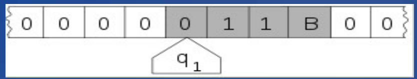
- To realize the potential of Molecular Programming

- "With *no alien technology*" [David Soloveichik]

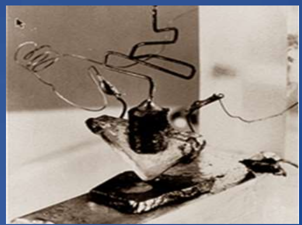
- We have some good strategies. Device design is now largely a 'software problem' but with a significant 'engineering scaleup and integration' problem

A Brief History of DNA

Turing Machine, 1936

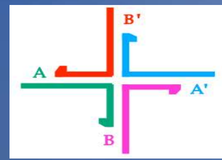


Transistor, 1947

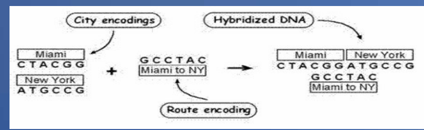


DNA, -3,800,000,000

Structural DNA Nonotech, 1982



DNA Algorithm, 1994



Computer programming

Systematic manipulation of information

Systematic manipulation of matter

Molecular programming

20th century

21th century

Resources

- DNA Computing and Molecular Programming Conference - incarnations since 1995
  http://www.dna-computing.org/
- Molecular Programming Project (Caltech - U.W. - Harvard - UCSF)
  http://molecular-programming.org/ (2008-2018 NSF Expeditions in Computing)
- Georg Seelig's DNA Nanotech Lab at U.W. CS&E
  http://homes.cs.washington.edu/~seelig/
- Biological Computation Group at Microsoft
  https://www.microsoft.com/en-us/research/group/biological-computation/