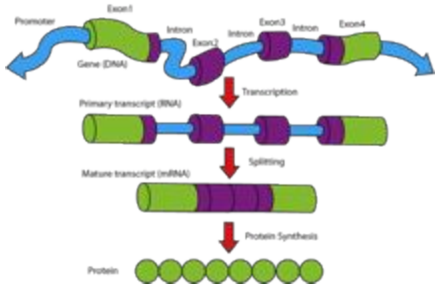
HOW TO GROW ALMOST ANYTHING

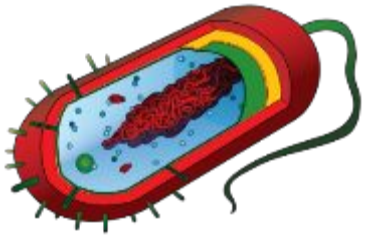
MIT MAS.885
2.10.26

Lecture 1: Gene Synthesis

Next Generation Gene Synthesis



Using Synthetic Genes…



to Program Cells to Become Factories to Make…

Devices, Sensors, Pharmaceuticals, Renewable Chemicals & Fuels and Food

HOW TO GROW ALMOST ANYTHING
MIT.MAS.885
February 10, 2026

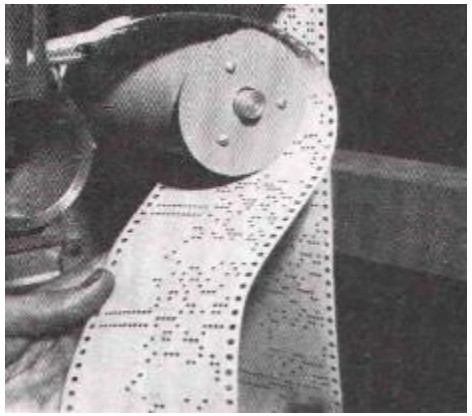
mit media lab

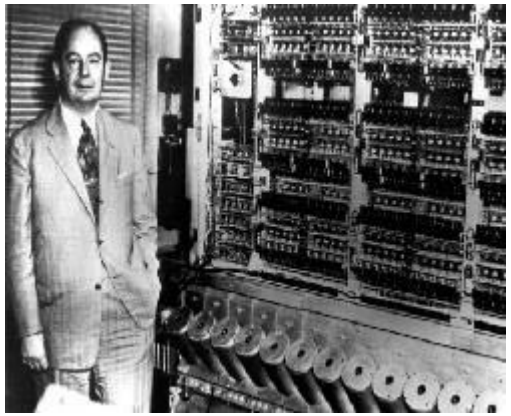
Massachusetts Institute of Technology

The Center for Bits and Atoms

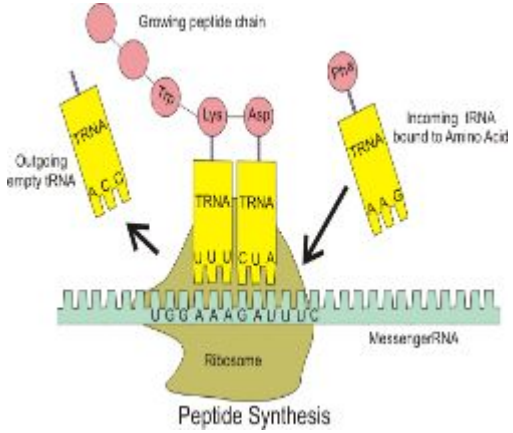
Synthetic Biology

Like this → only 10^6 times smaller

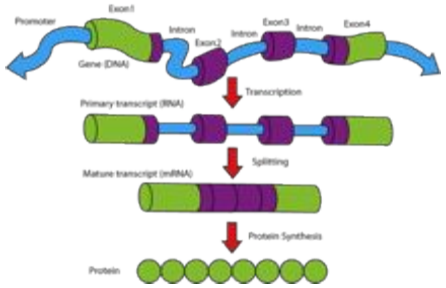




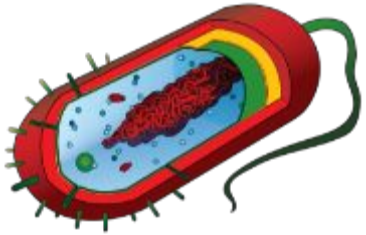
Tape Programming of the Mark I Computer



Ribosome



Using Synthetic Genes…



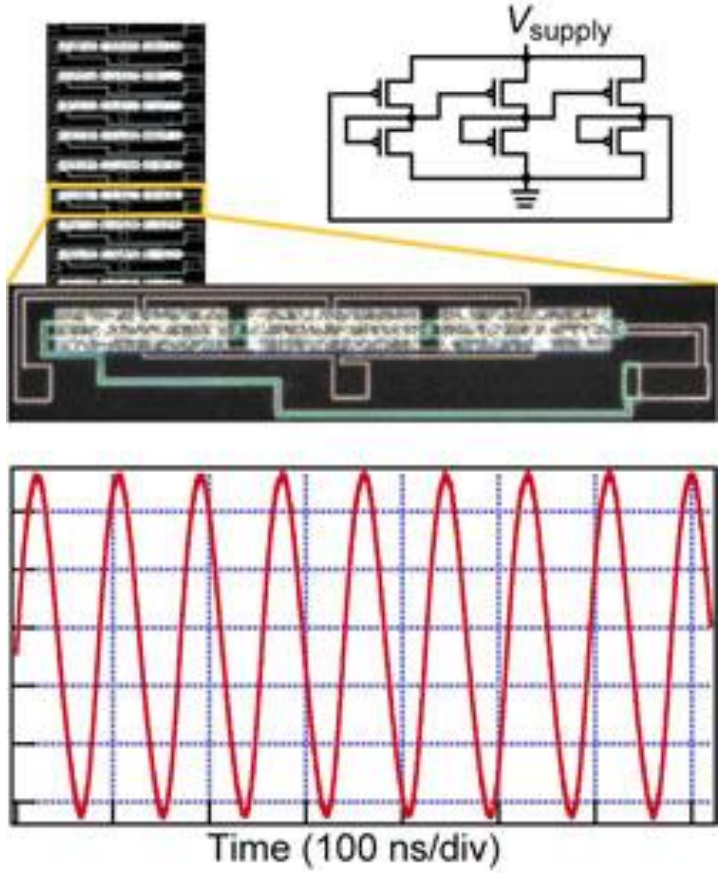
to Program Cells to Become Factories to Make…

Devices, Sensors, Pharmaceuticals, Renewable Chemicals & Fuels and Food

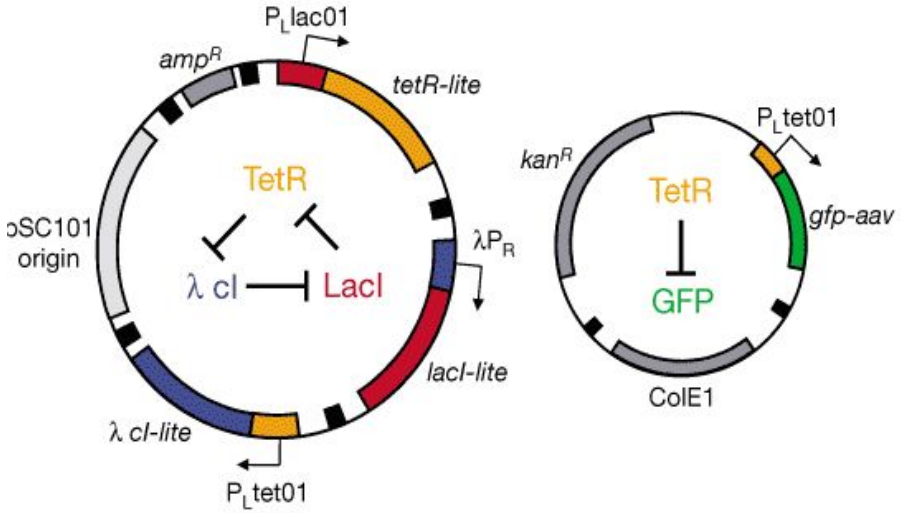
mit media lab

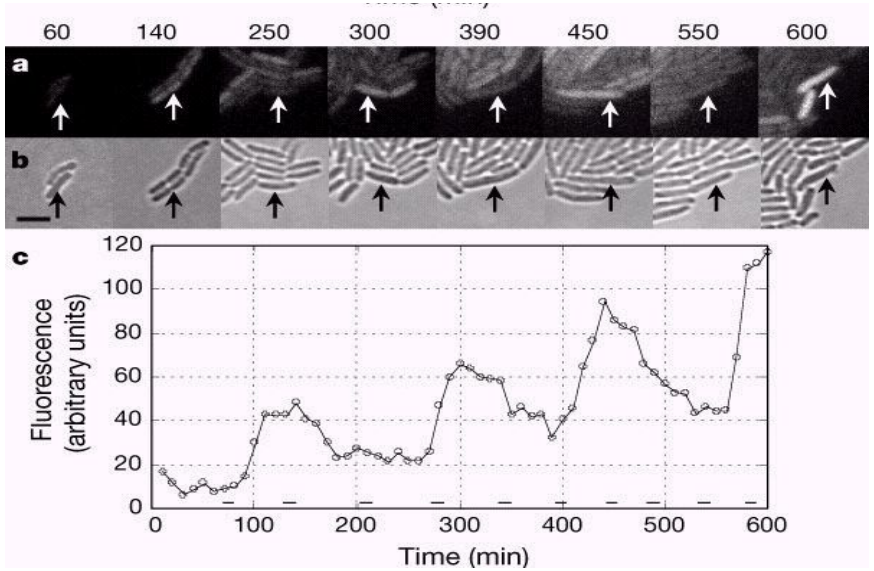
Cells as Computers and Logic

Ring Oscillators from Transcriptional-Translational Logic



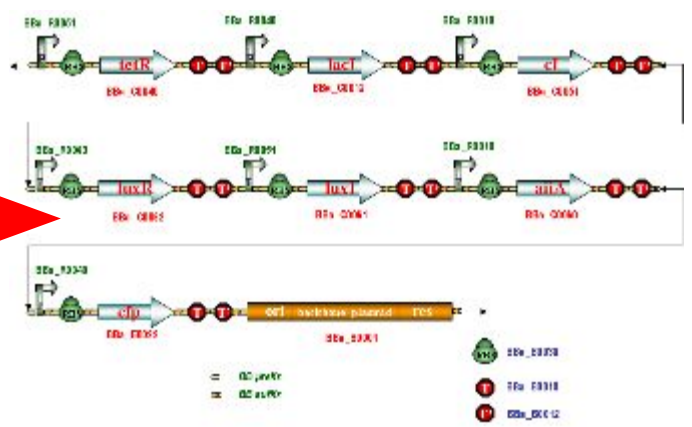
http://echinacea.harvard.edu/research/image-gallery/

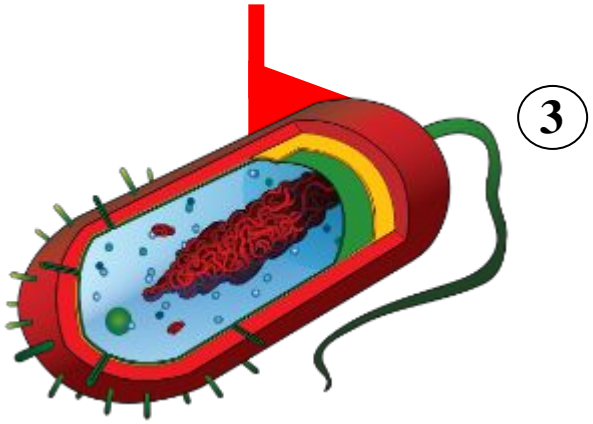


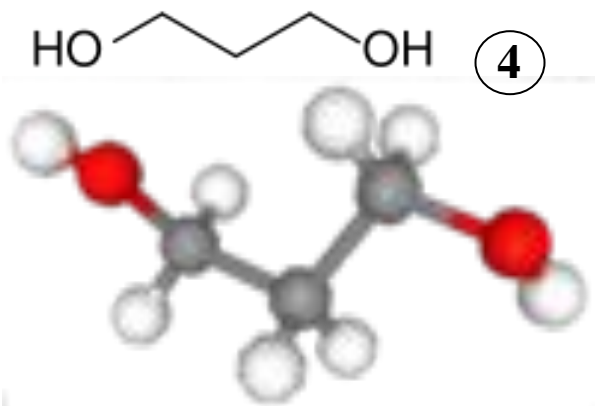


MB Elowitz MB and S Leibler (2000) A synthetic oscillatory network of transcriptional regulators.Nature.403 :335 - 338.

Metabolic Circuit



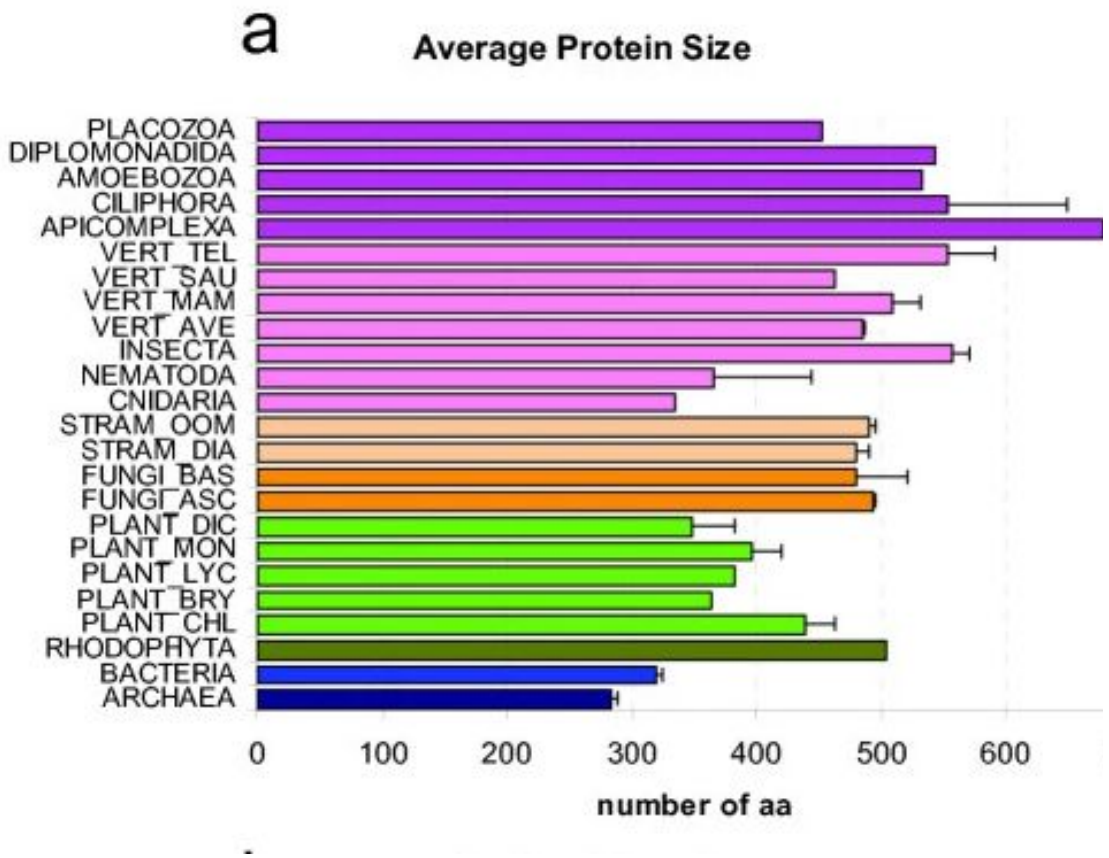


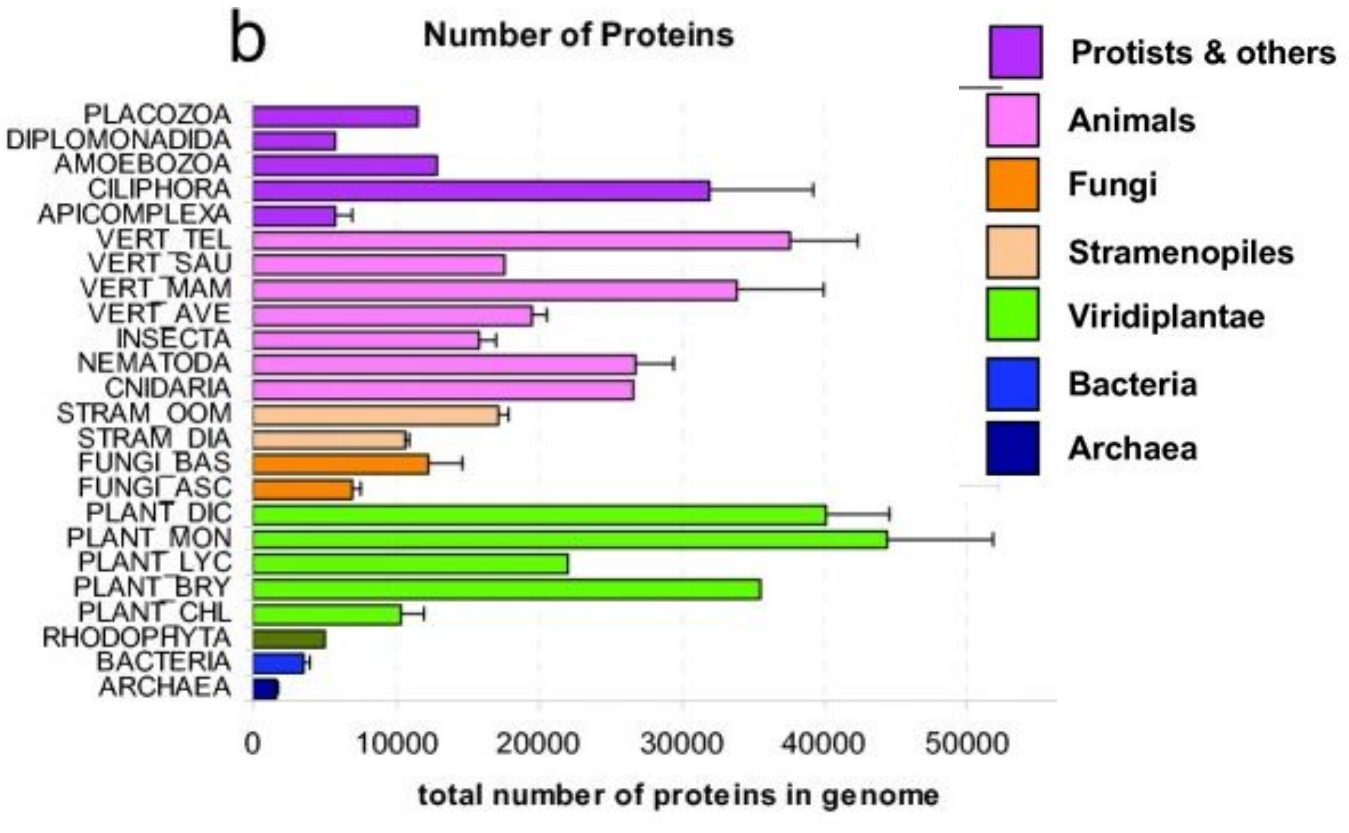


to Make High Value Products (Biofuels, Chemicals, Pharmaceuticals)

to Program Cells to Become

How Many Base Pairs Do We Need To Synthesize?



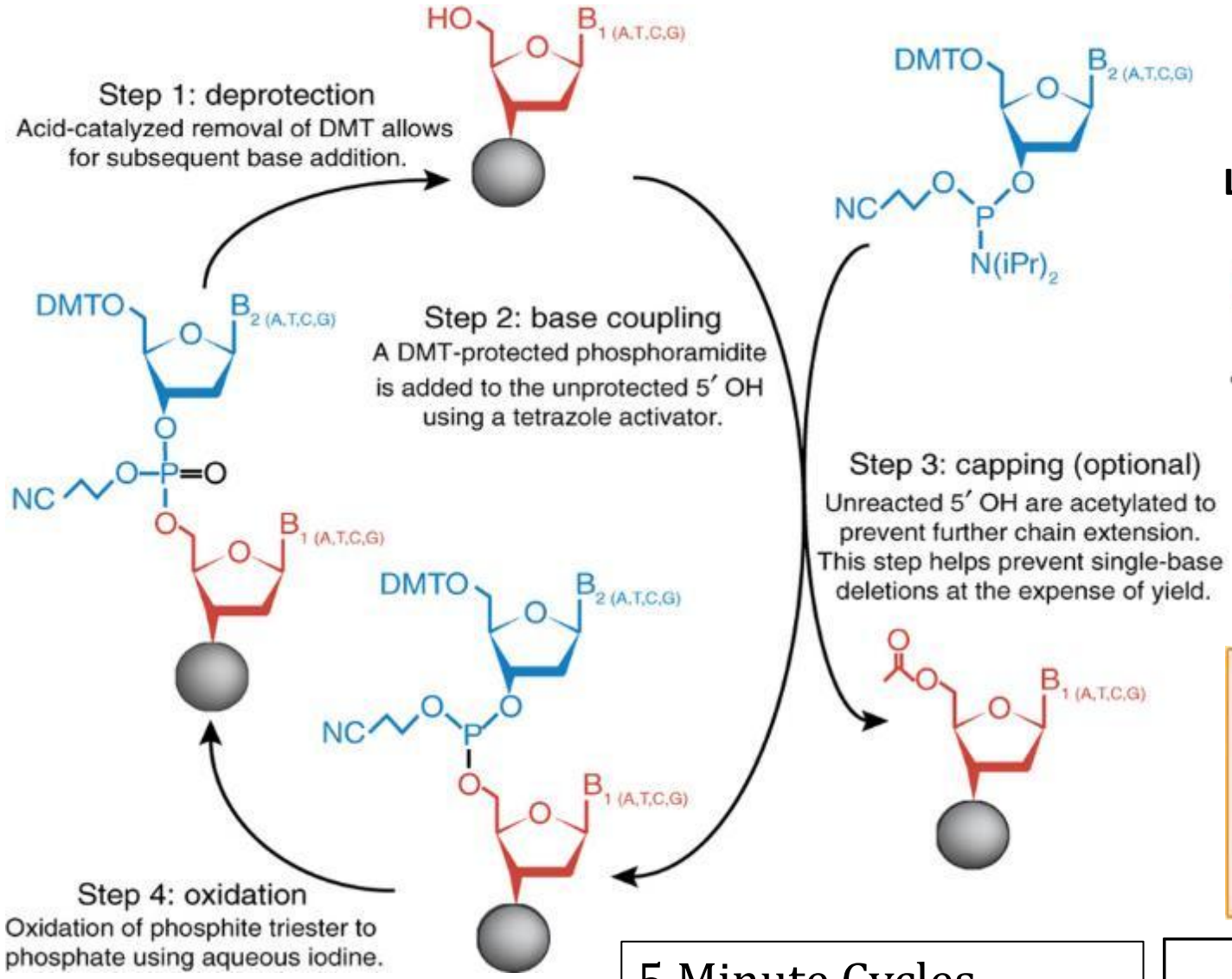


- **Average Bacterial Protein: 960bp**
- **Average Human Protein: 1036 bp**
- **Longest Human Proteins (PKS): >100kbp**
- **Metabolic Pathway: ~3-10kbp**

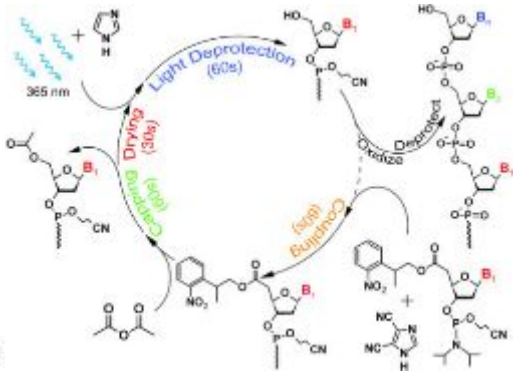
https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-5-85/

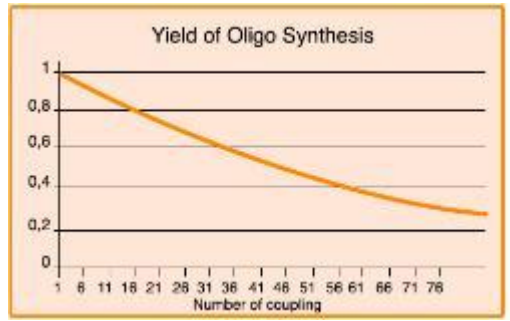
Phosphoramidite DNA Synthesis Cycle

Acid Based Deprotection



Light Based Deprotection





5 Minute Cycles
Machine ~ .5 Mbp / yr

$$(1-\frac{1}{N})^N \approx 37\%$$

http://www.nature.com/nmeth/journal/v11/n5/images/nmeth.2918-F2.jpg

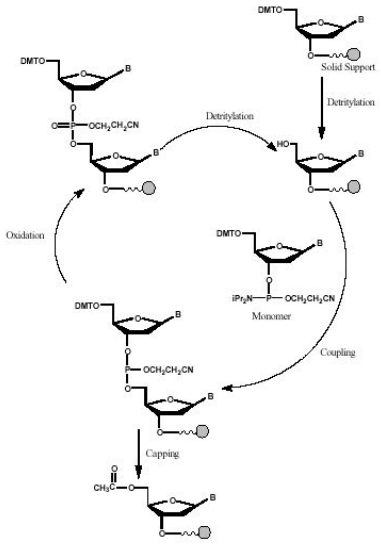
DNA Synthesis

Chemical Synthesis (Open Loop Protection Group)

Error Rate: $1{:}10^2$

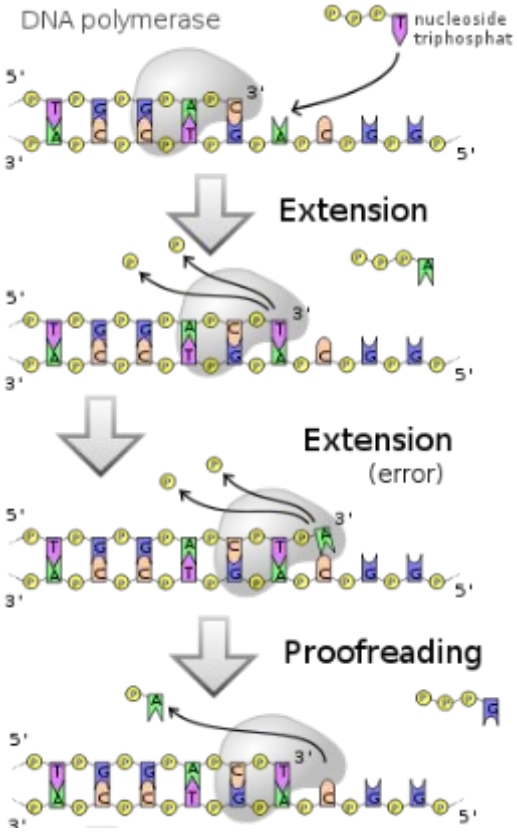
Throughput: 300 S per Base Addition

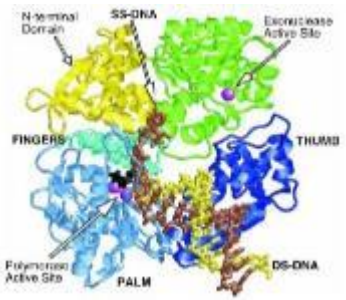
http://www.med.upenn.edu/naf/services/catalog99.pdf



DNA Synthesis Cycle

Biological Synthesis (Error Correcting Polymerase)





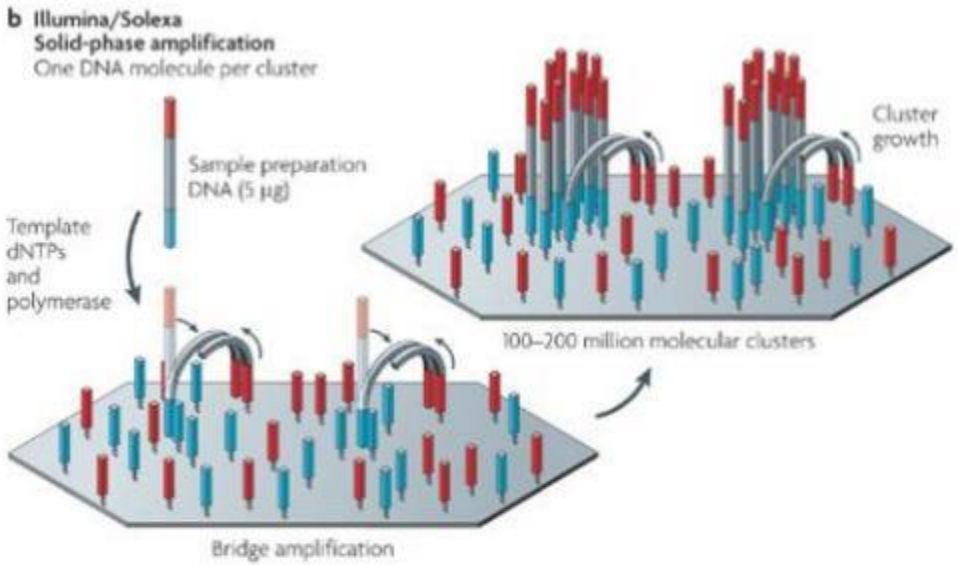
Error Rate: $1{:}10^6$

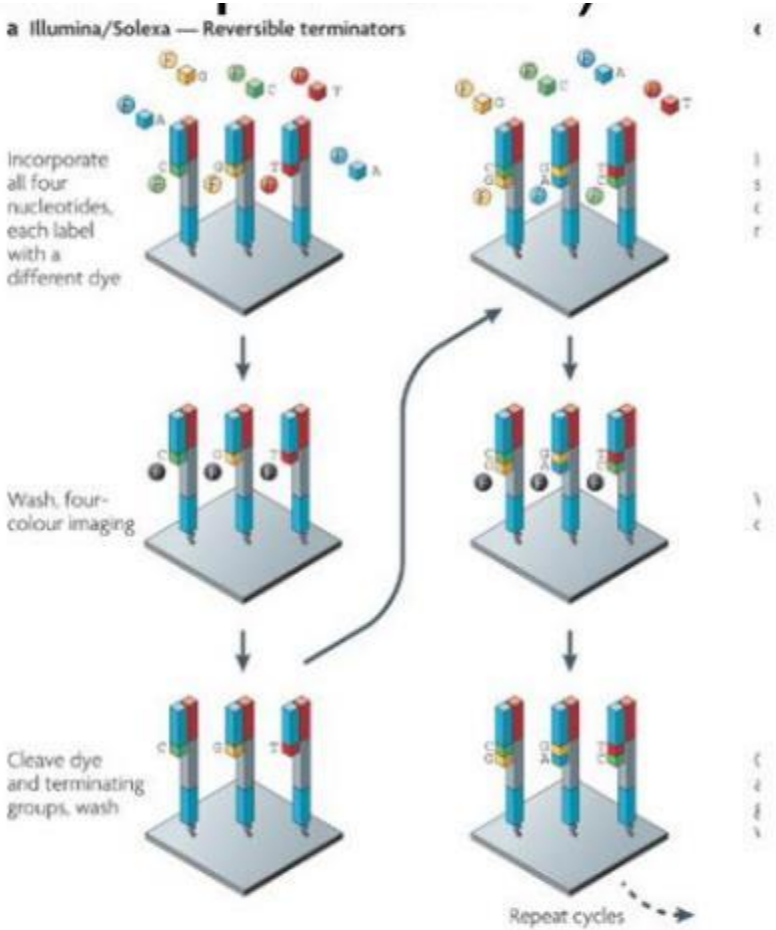
Throughput: 10 mS per Base Addition

Beese *et al.* (1993), *Science*, **260**, 352-355.
http://www.biochem.ucl.ac.uk/bsm/xtal/teach/repl/klenow.html

Throughput Error Rate Product Differential: ~10^8

Next Generation Sequencing by Synthesis

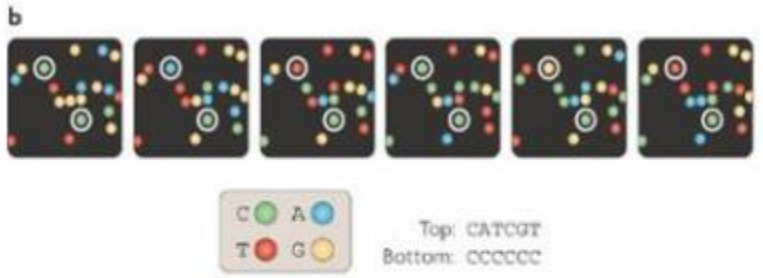




Each nucleotide has a dye with a different color

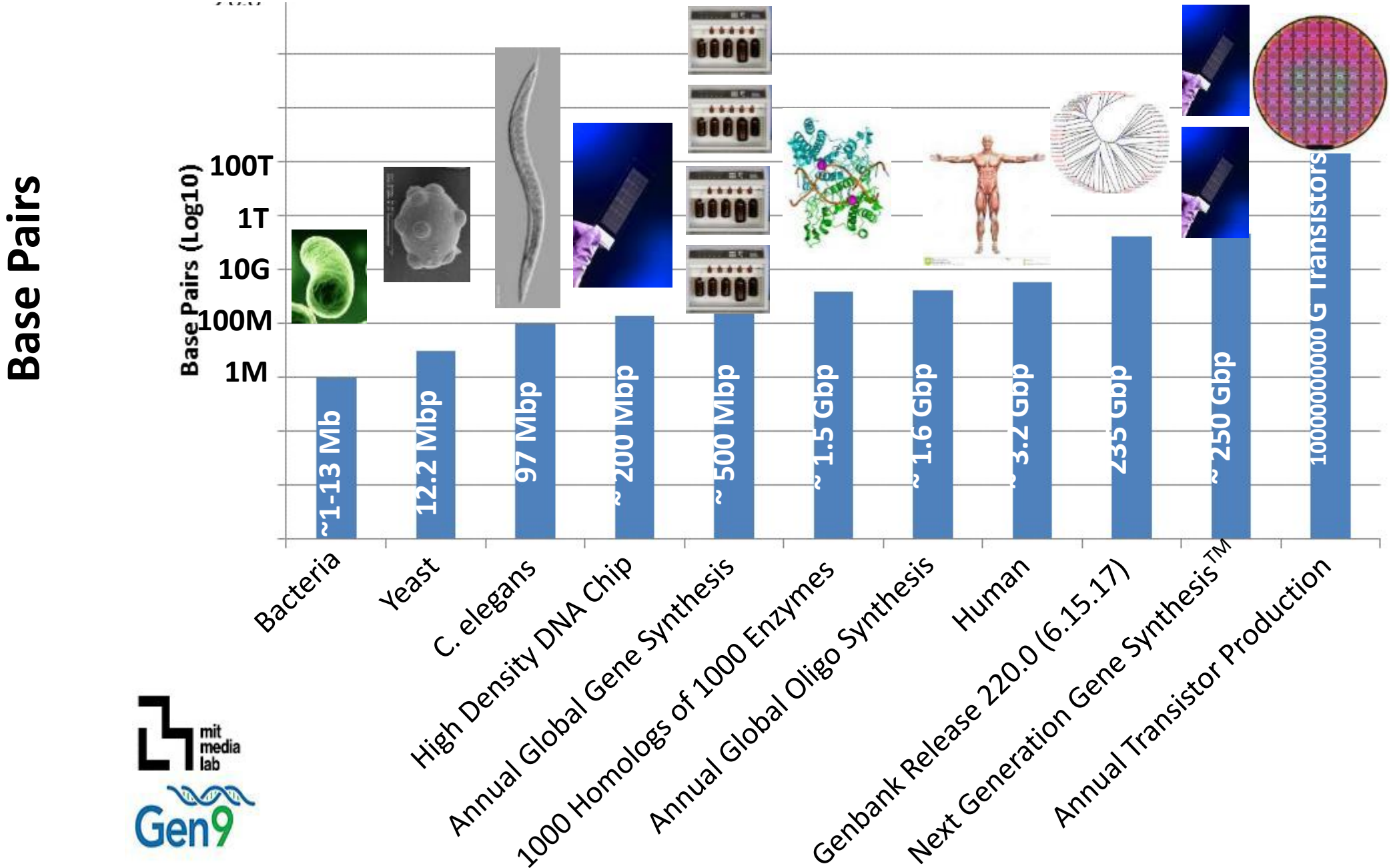
4-color fluorescent image of chip gathered after each chemical flows through

Register each image and follow color change of each colony to determine sequence



https://raw.githubusercontent.com/lmmx/shots/master/2015/May/Illinois-BioE416-Illumina-polonies.jpg

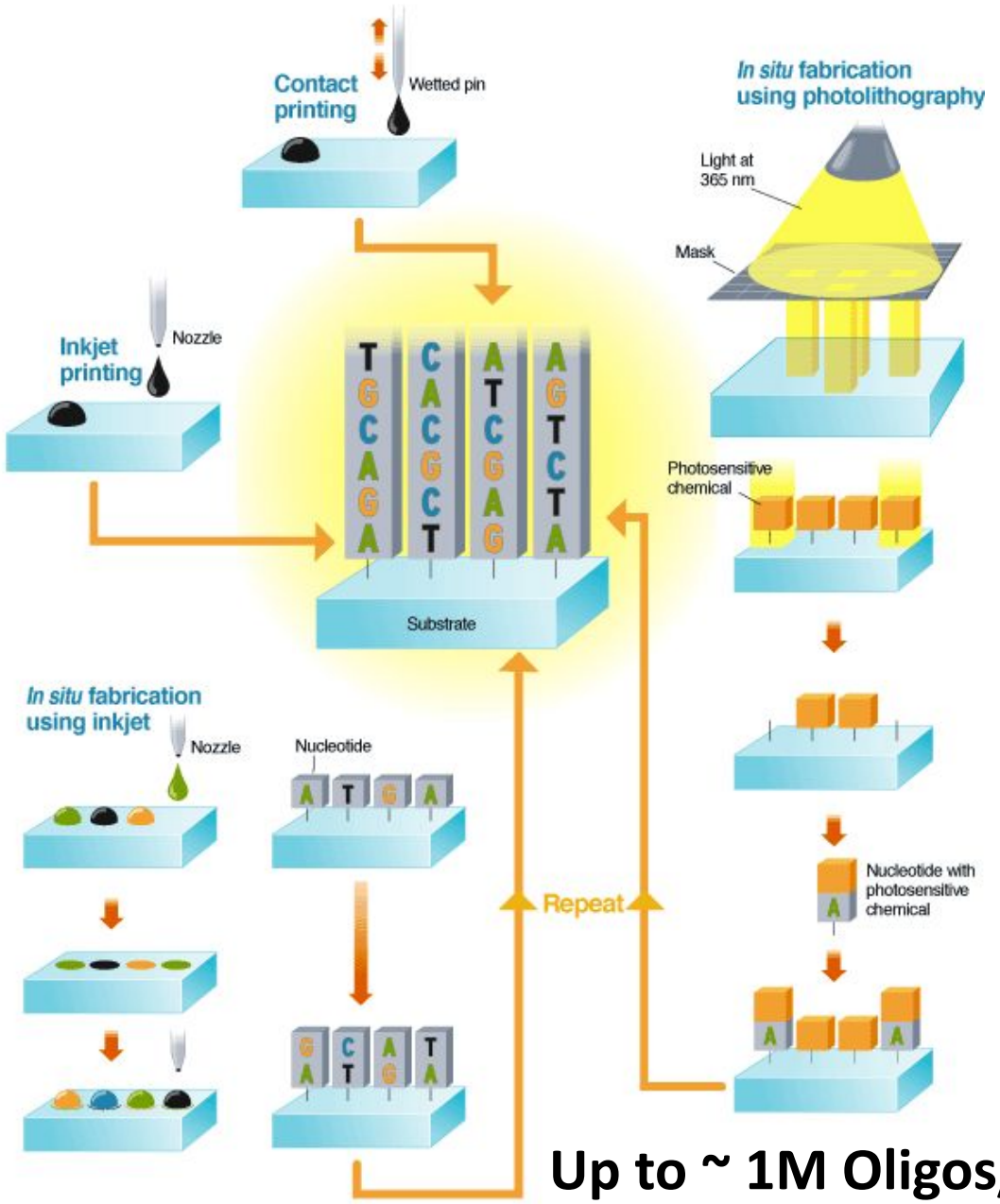
Next Generation (Chip Based) DNA Synthesis

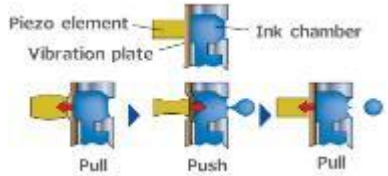


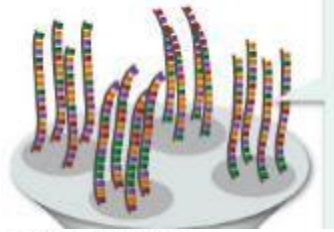
mit media lab

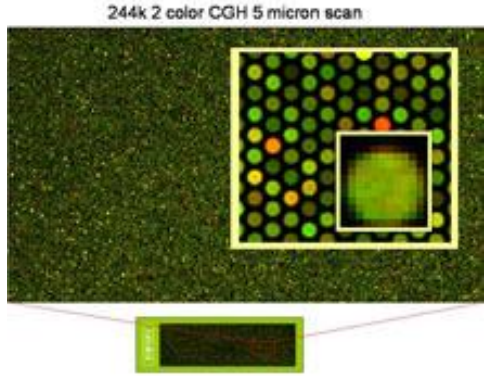
Gen9

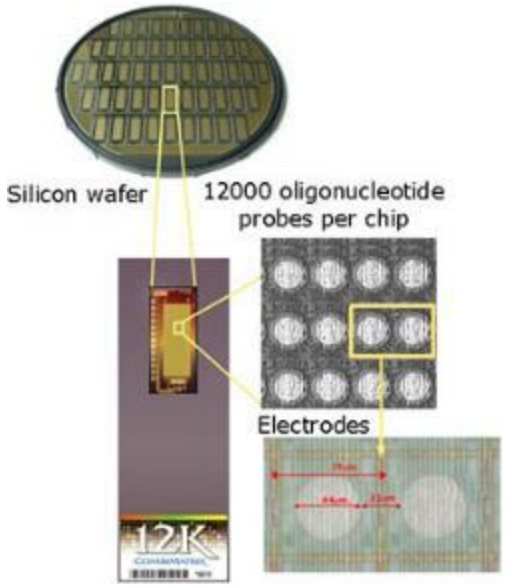
Next Generation (Chip Based) DNA Synthesis







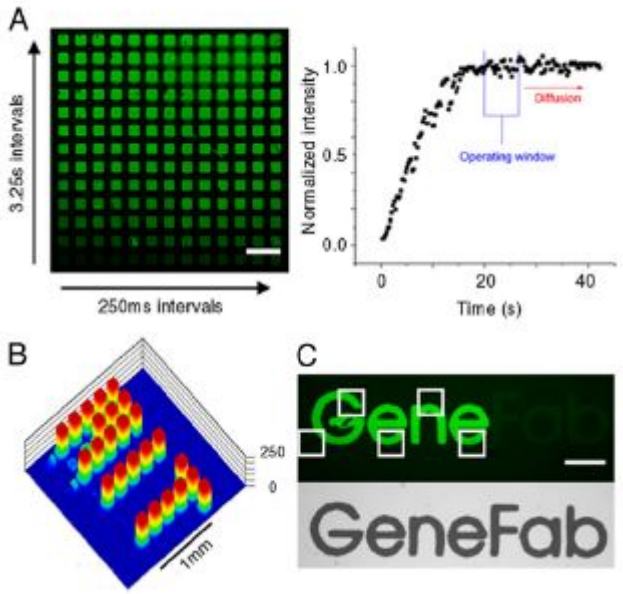


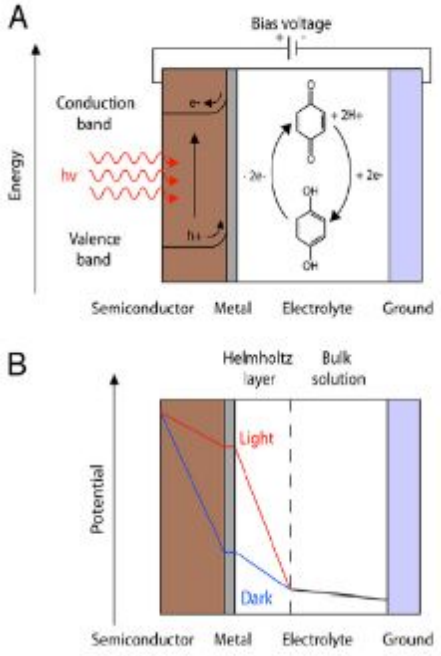


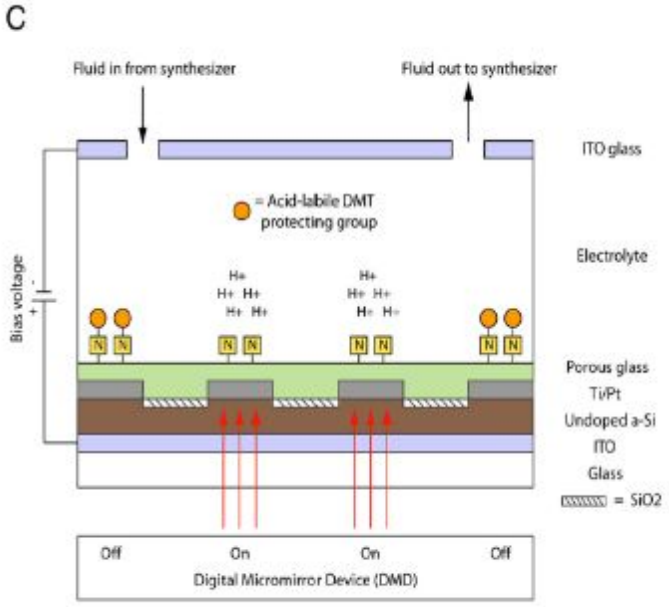
Up to ~ 1M Oligos/Chip

Next Generation (Chip Based) DNA Synthesis

~1000x Lower Oligonucleotide Cost







Chow, Brian Y., Christopher J. Emig, and Joseph M. Jacobson. "Photoelectrochemical synthesis of DNA microarrays." ***Proceedings of the National Academy of Sciences*** **106.36 (2009): 15219-15224.**

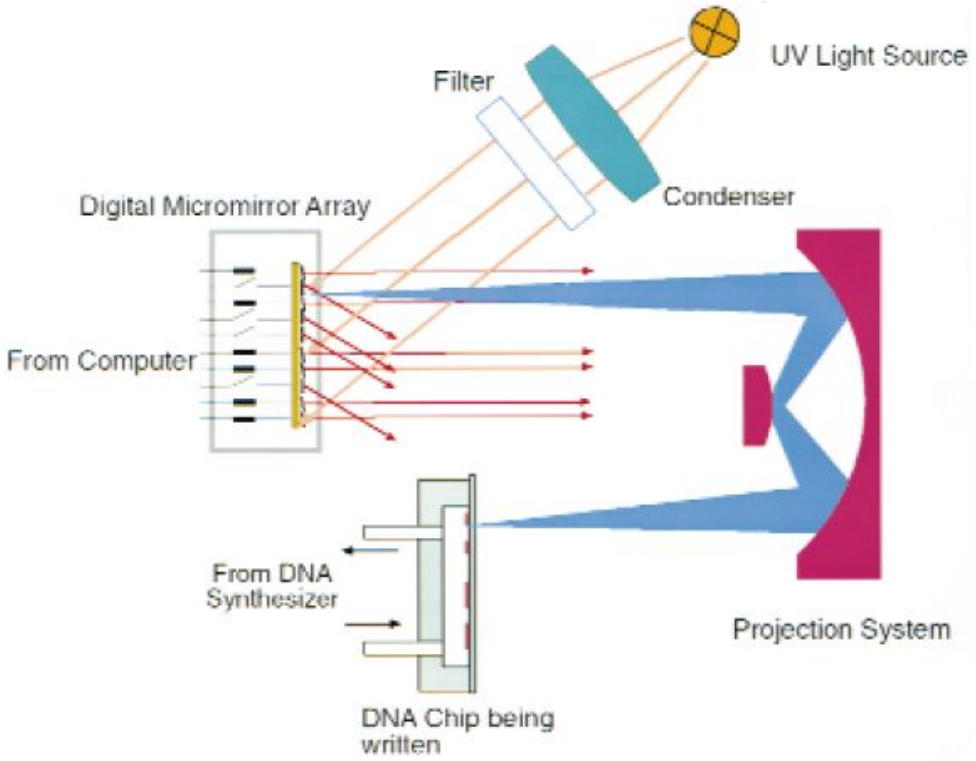
http://www.technologyreview.com/biomedicine/20035/

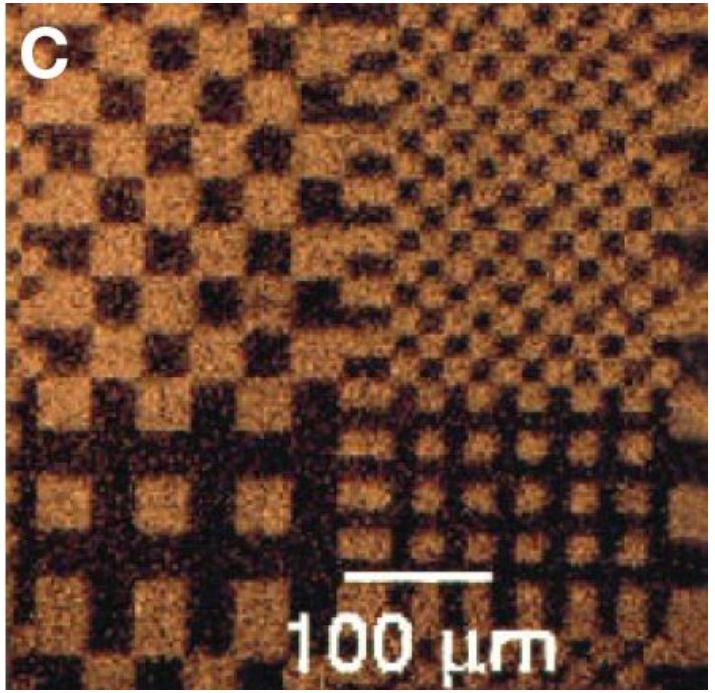
http://learn.genetics.utah.edu/content/labs/microarray/analysis

Accurate multiplex gene synthesis from programmable DNA microchips
Jingdong Tian, Hui Gong, Nijing Sheng, Xiaochuan Zhou, Erdogan Gulari, Xiaolian Gao and George Church
Nature **432**, 1050-1054(23 December 2004)
doi:10.1038/nature03151

Chip Based Oligo Nucleotide Synthesis

~1000x Lower Oligonucleotide Cost



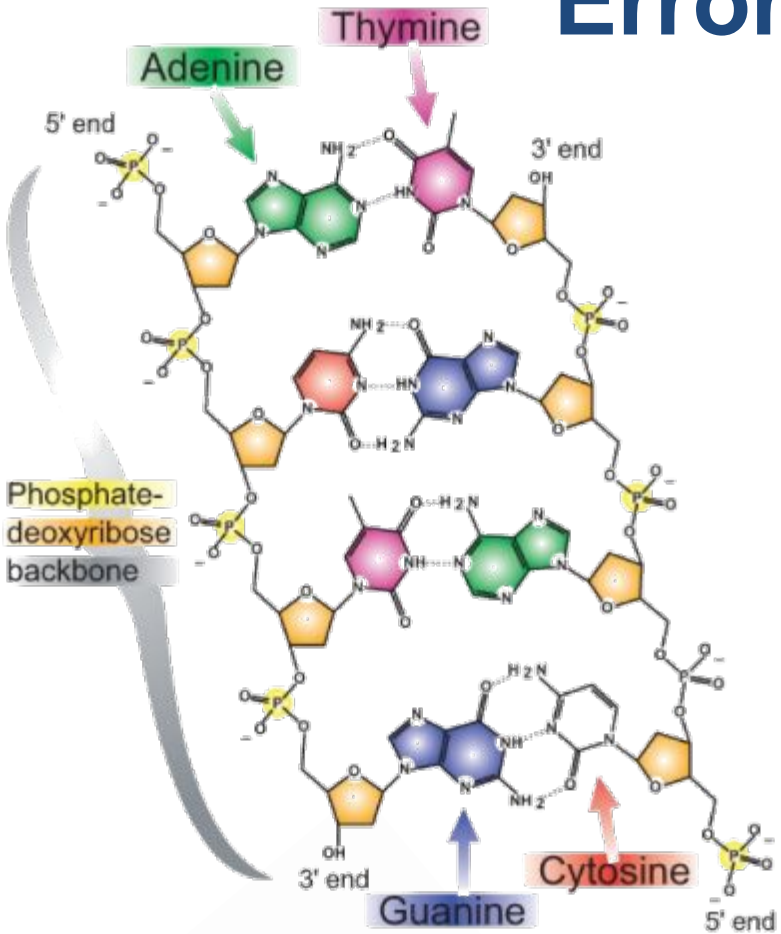


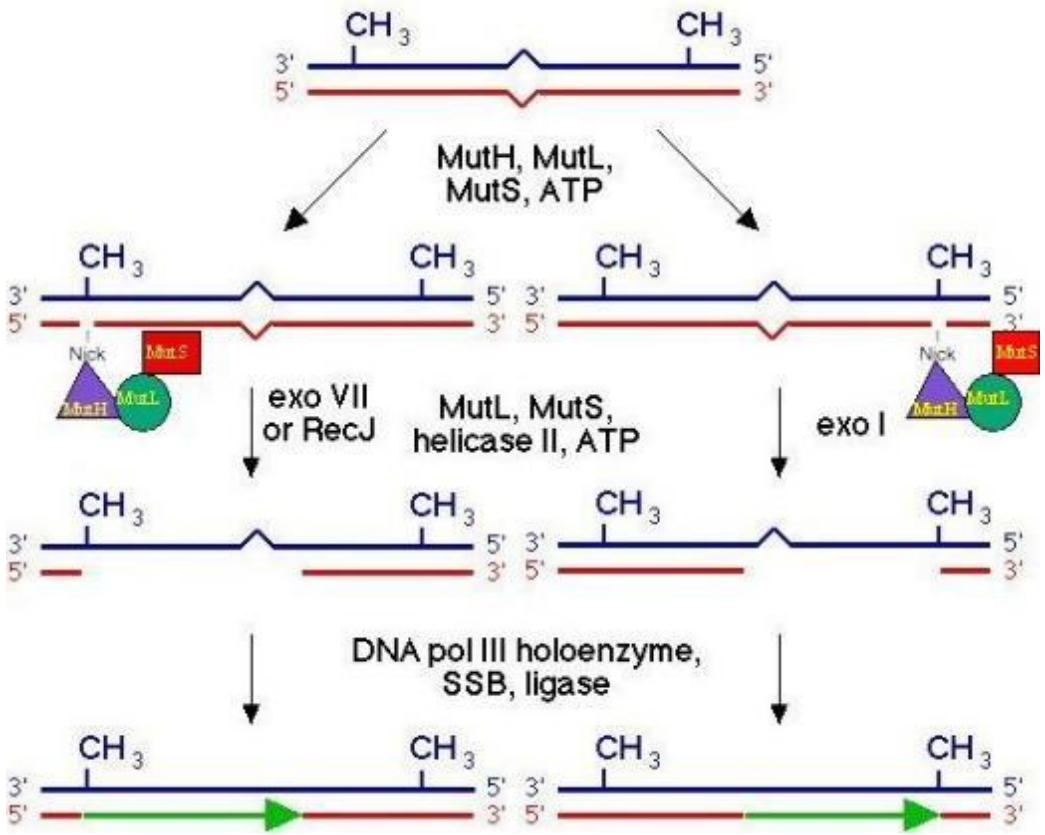
Up to 1M Oligos/Chip

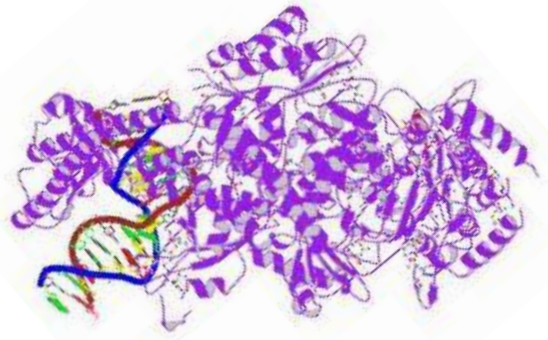
50 Mbp for ~ $1K (instead of ~$1M)

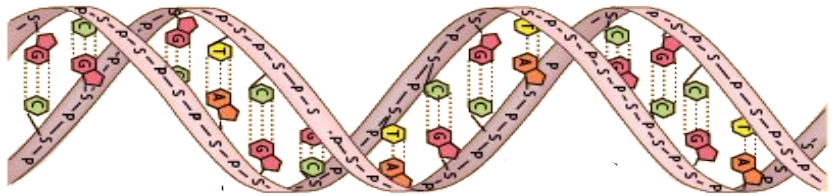
F. Cerrina et. al Nature Biotech 1999 p. 974

Error Correction





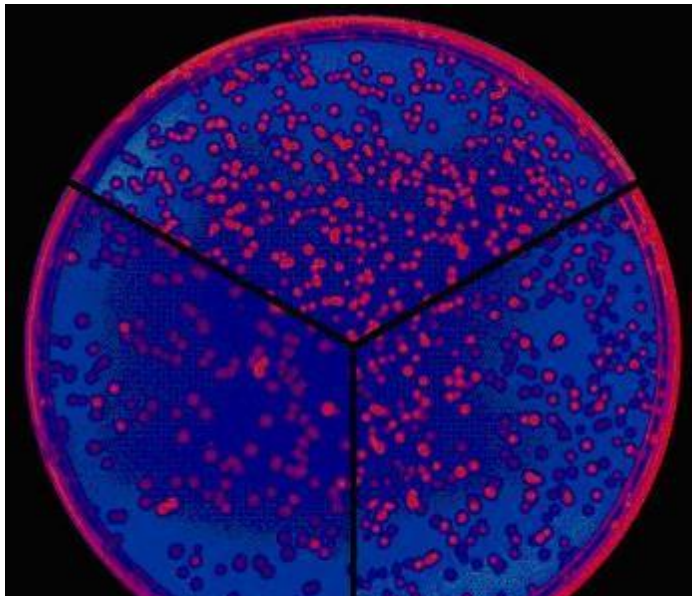


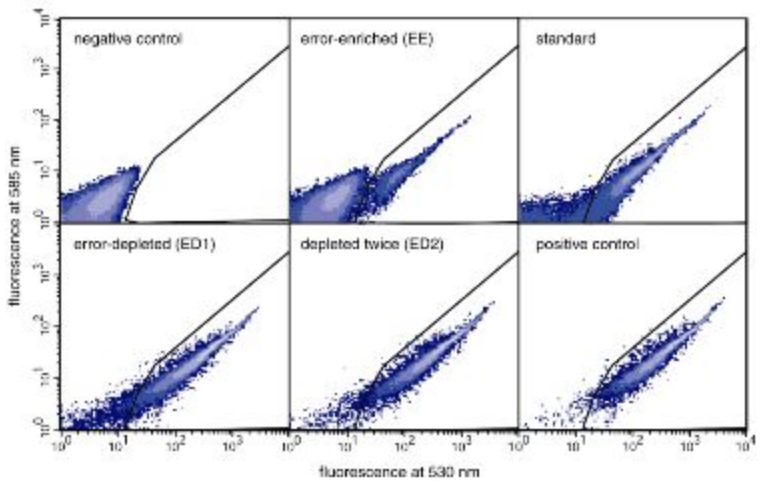


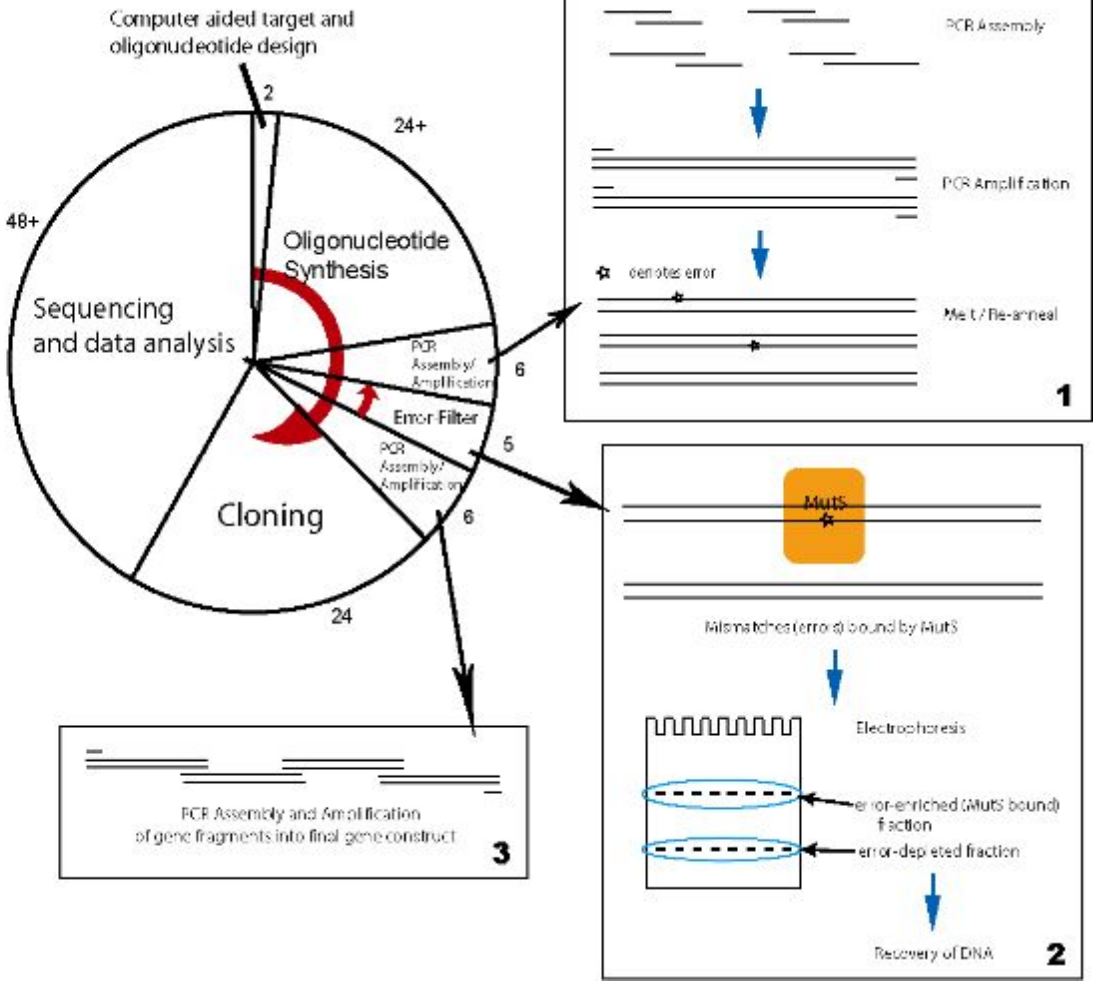
MutS Repair System

Lamers et al. Nature 407:711 (2000)

Error Correcting Gene Synthesis



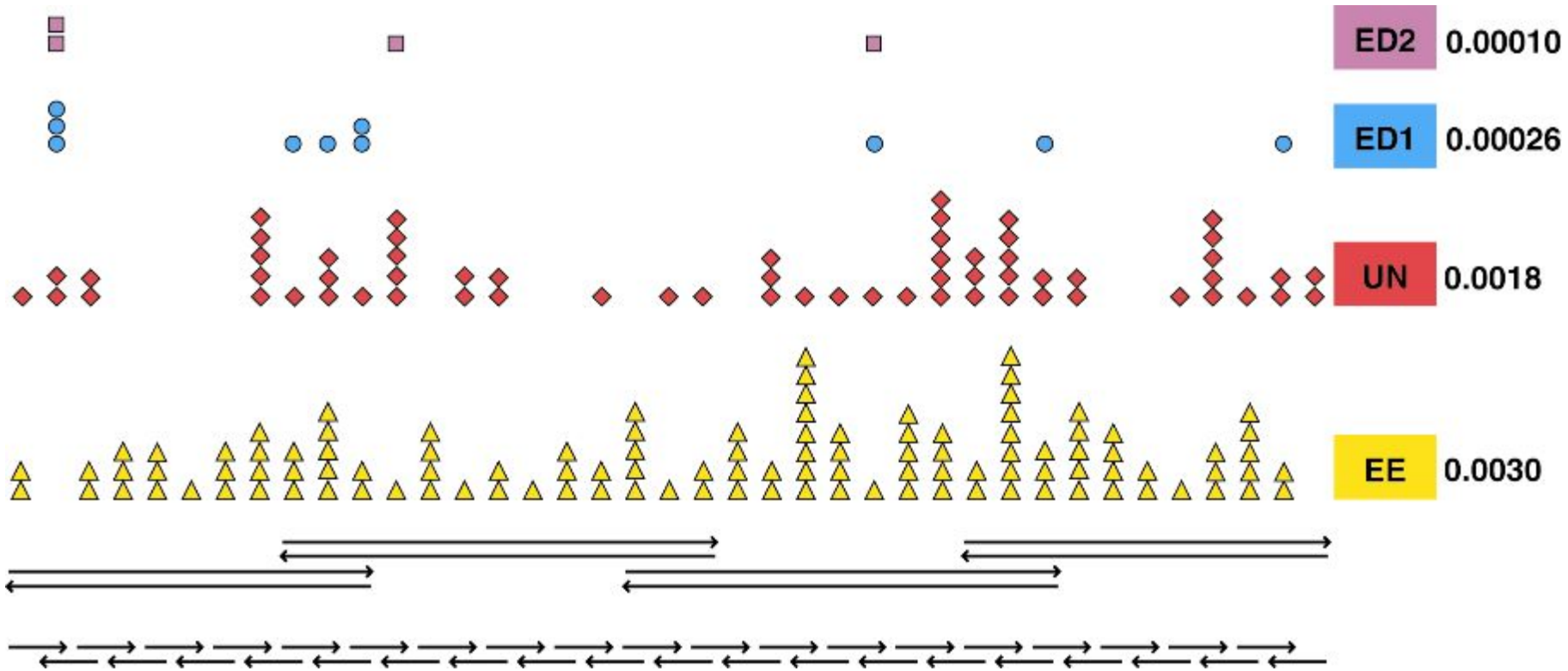




Error Rate 1:10^4

Nucleic Acids Research 2004
32(20):e162

Error Reduction: GFP Gene synthesis



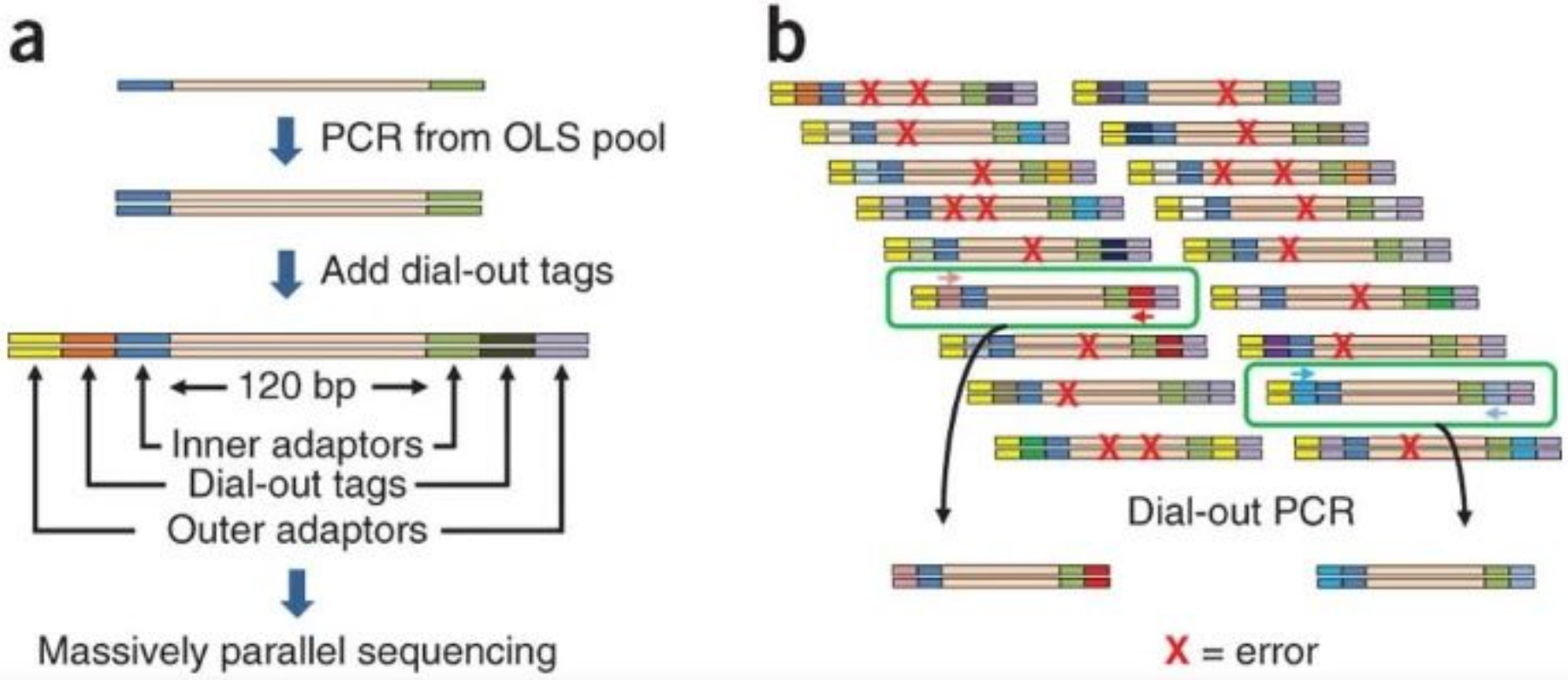
Brief Communication | Published: 12 August 2012

Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules

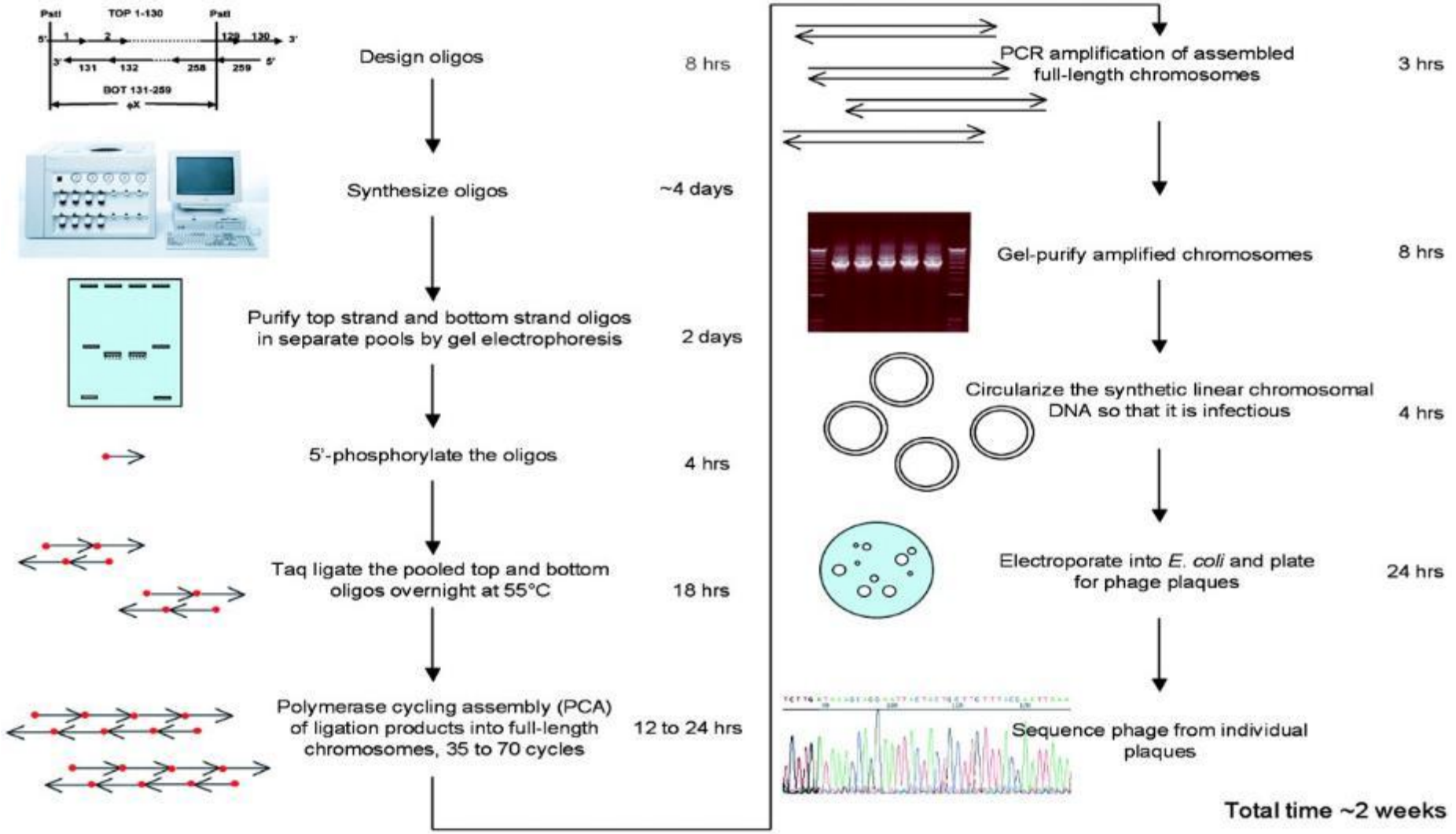
Jerrod J Schwartz, Choli Lee & Jay Shendure

Nature Methods **9**, 913–915 (2012) | Cite this article

Figure 1: Dial-out PCR for retrieving accurate sequences from a nonuniform, error-rich library of synthetic DNA molecules.



Generating a synthetic genome by whole genome assembly: φX174 bacteriophage from synthetic oligonucleotides



AGCTAGCTAGTTTTTTTTTT GGGGGGGGGGAAGGCC

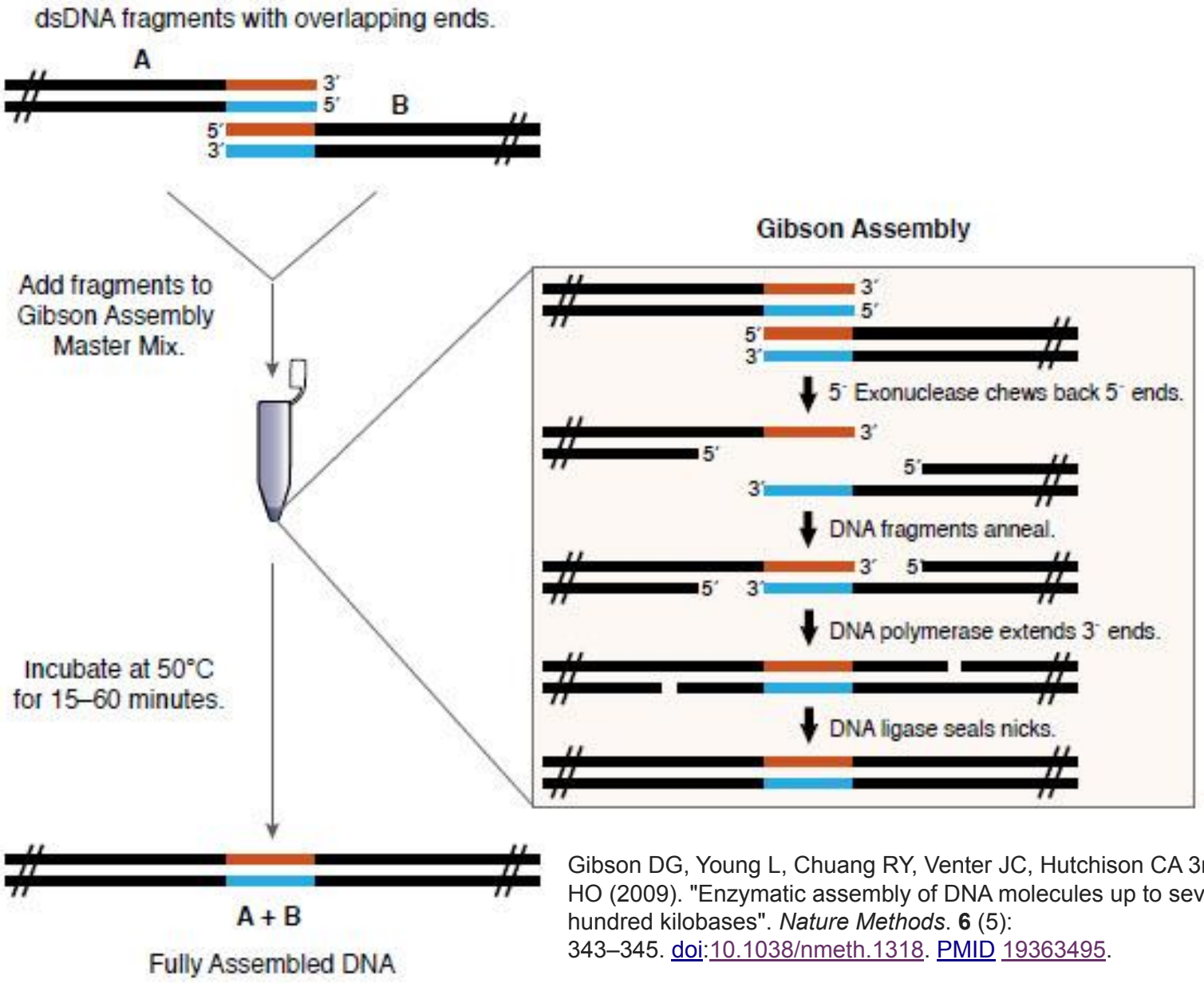
TTAA

AAAAAAAAAACCCCCCCCCC

2003

http://www.pnas.org/content/100/26/15440/F1.large.jpg

Gibson Assembly (2009)



Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases". *Nature Methods*. **6** (5): 343–345. doi:10.1038/nmeth.1318. PMID 19363495.

https://www.neb.com/~/media/Catalog/All-Products/0AA961B294E444AFBEDD5C4A904C76E6/Long%20Description/E2611a.jpg

Multiplex Synthesis (2010)

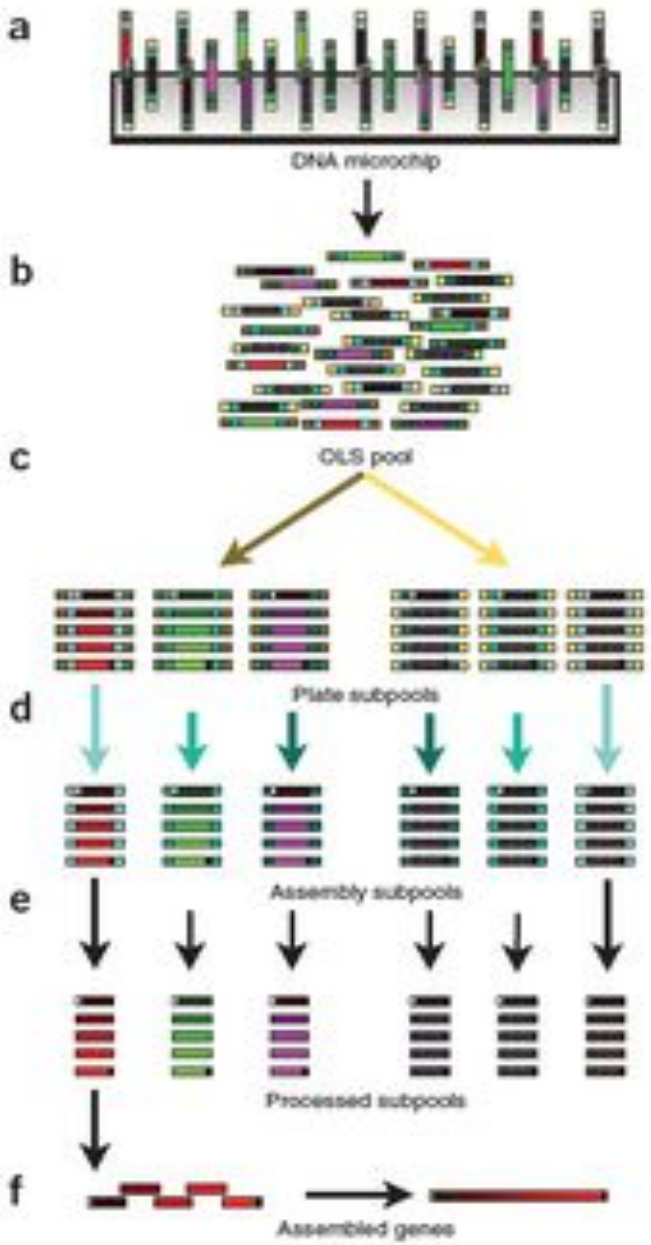


Figure 1 Schematic for scalable gene synthesis from OLS pool 2. (**a**,**b**) Pre-designed oligonucleotides (no distinction is made between dsDNA and ssDNA in the figure) are synthesized on a DNA microchip (**a**) and then cleaved to make a pool of oligonucleotides (**b**). (**c**) Plate-specific primer sequences (yellow or brown) are used to amplify separate plate subpools (only two are shown), which contain DNA to assemble different genes (only three are shown for each plate subpool). (**d**) Assembly-specific sequences (shades of blue) are used to amplify assembly subpools that contain only the DNA required to make a single gene. (**e**) The primer sequences are cleaved using either type IIS restriction enzymes (resulting in dsDNA) or by DpnII/USER/λ exonuclease processing (producing ssDNA). (**f**) Construction primers (shown as white and black sites flanking the full assembly) are then used in an assembly PCR reaction to build a gene from each assembly subpool. Depending on the downstream application, the assembled products are then cloned either before or after an enzymatic error correction step.

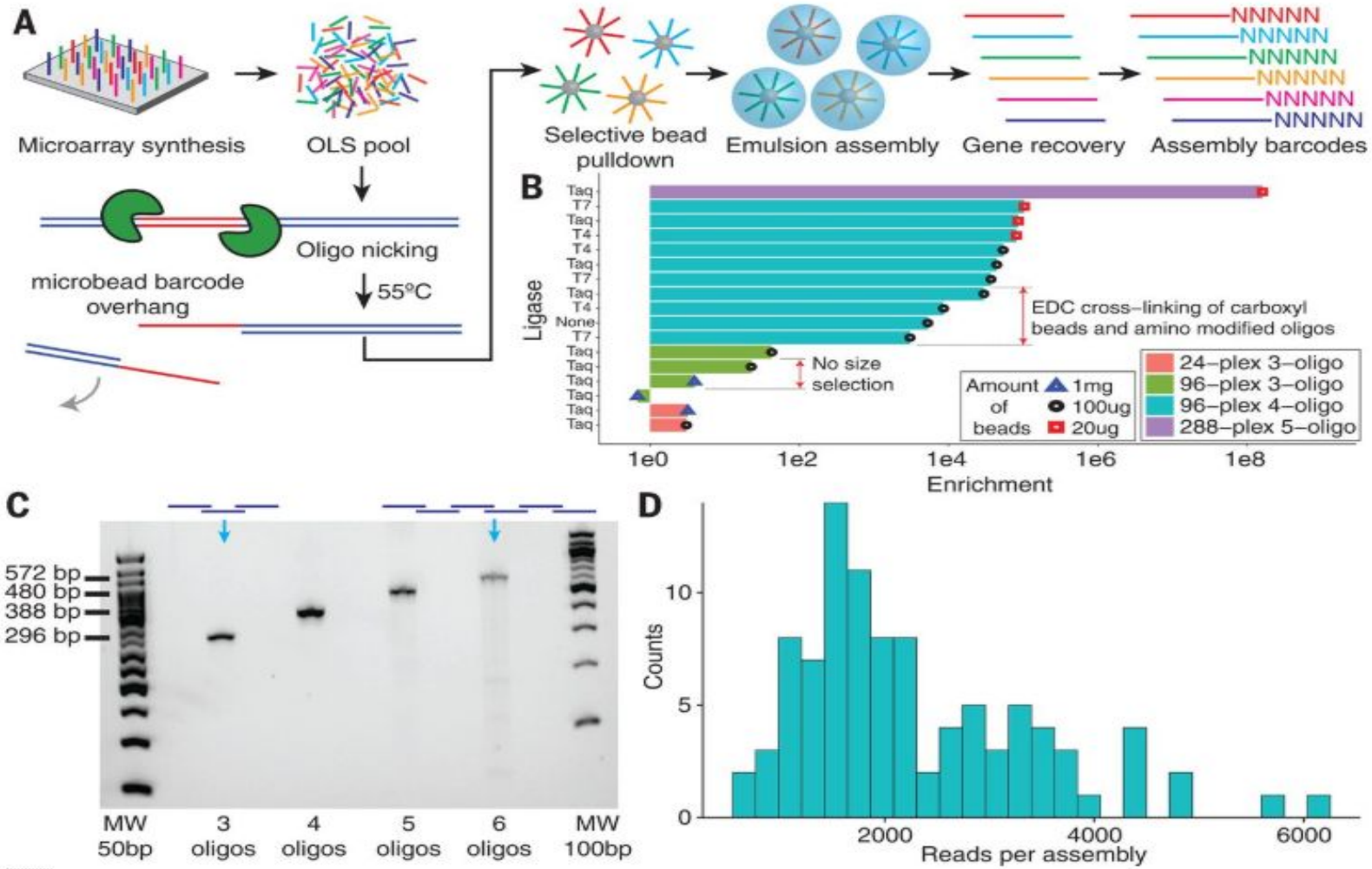
Kosuri, Sriram, et al. "Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips." *Nature biotechnology* 28.12 (2010): 1295-1299.

REPORT

Multiplexed gene synthesis in emulsions for exploring protein functional landscapes

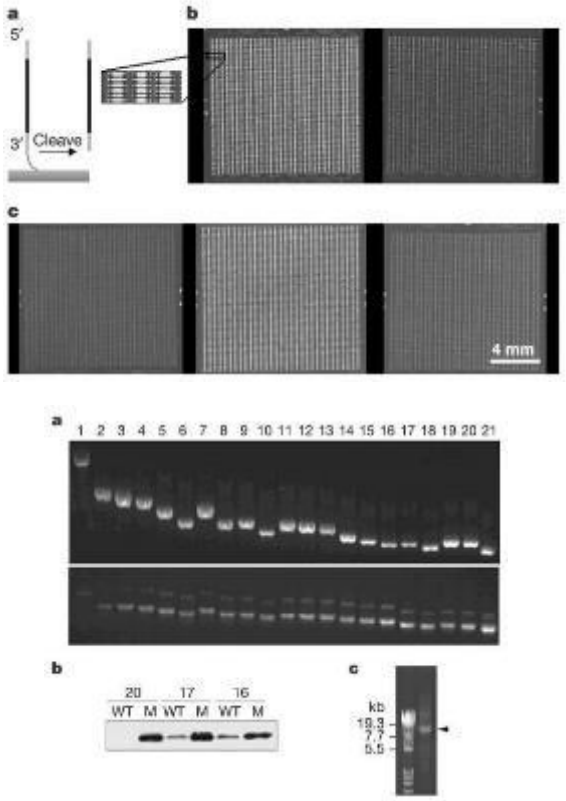
Calin Plesa[1,*], Angus M. Sidore[2,*], Nathan B. Lubock[1], Di Zhang[3], Sriram Kosuri[1,4,†]
+ See all authors and affiliations

Science 19 Jan 2018:
Vol. 359, Issue 6373, pp. 343-347
DOI: 10.1126/science.aao5167



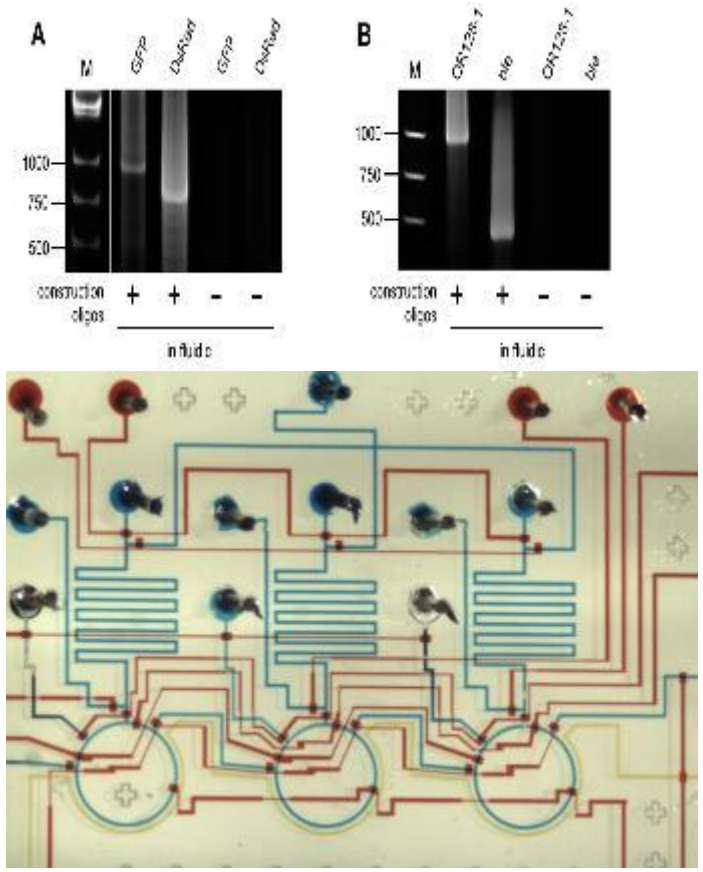
Chip Based Gene Synthesis

1

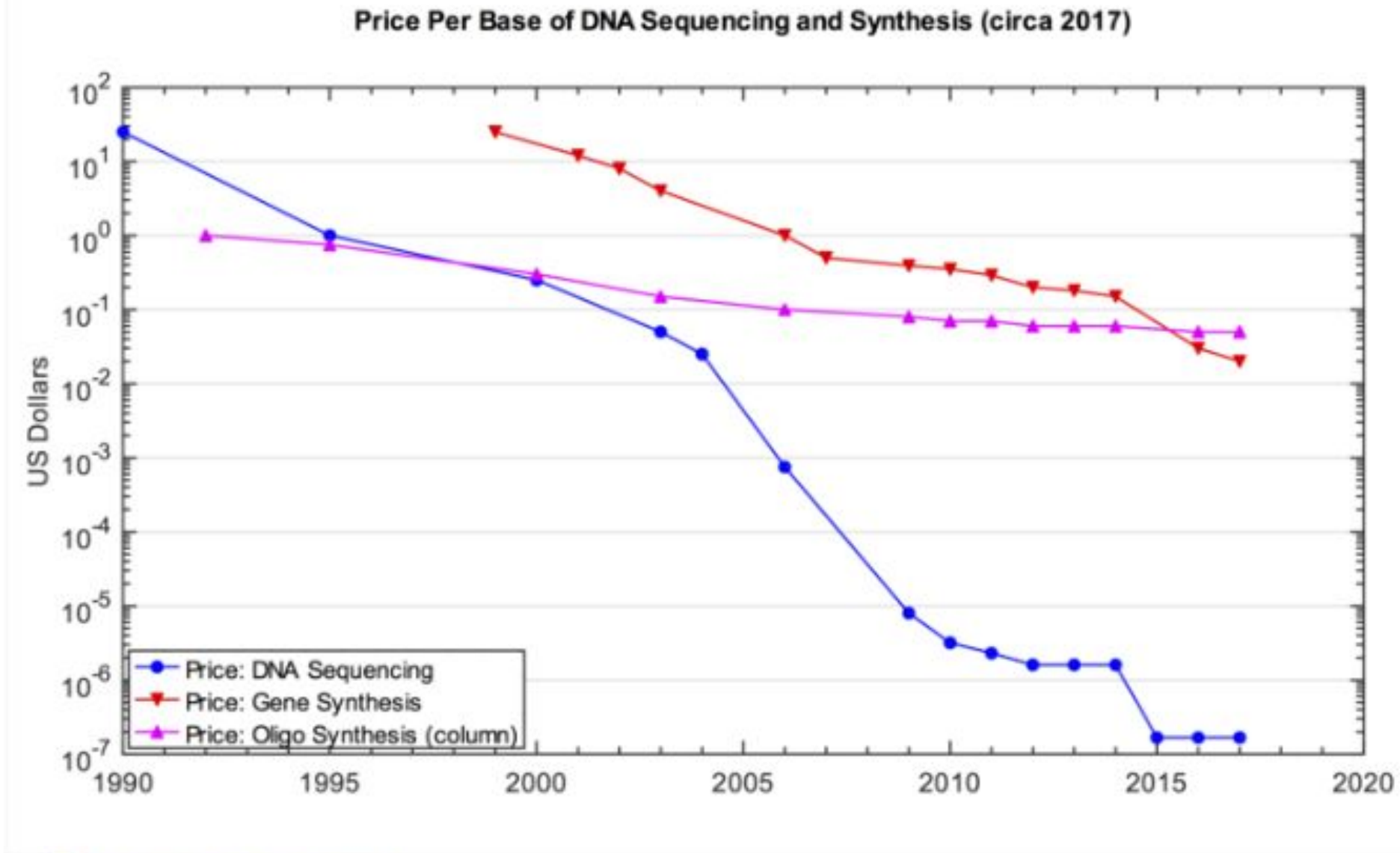


Accurate multiplex gene synthesis from programmable DNA microchips
Jingdong Tian, Hui Gong, Nijing Sheng, Xiaochuan Zhou, Erdogan Gulari, Xiaolian Gao and George Church
Nature **432**, 1050-1054(23 December 2004)
doi:10.1038/nature03151

2



Parallel gene synthesis in a microfluidic device
D.S. Kong, P.A. Carr, L. Chen, S. Zhang, J.M. Jacobson Nucleic Acid Research , 2007, Vol. 35, No. 8 e61

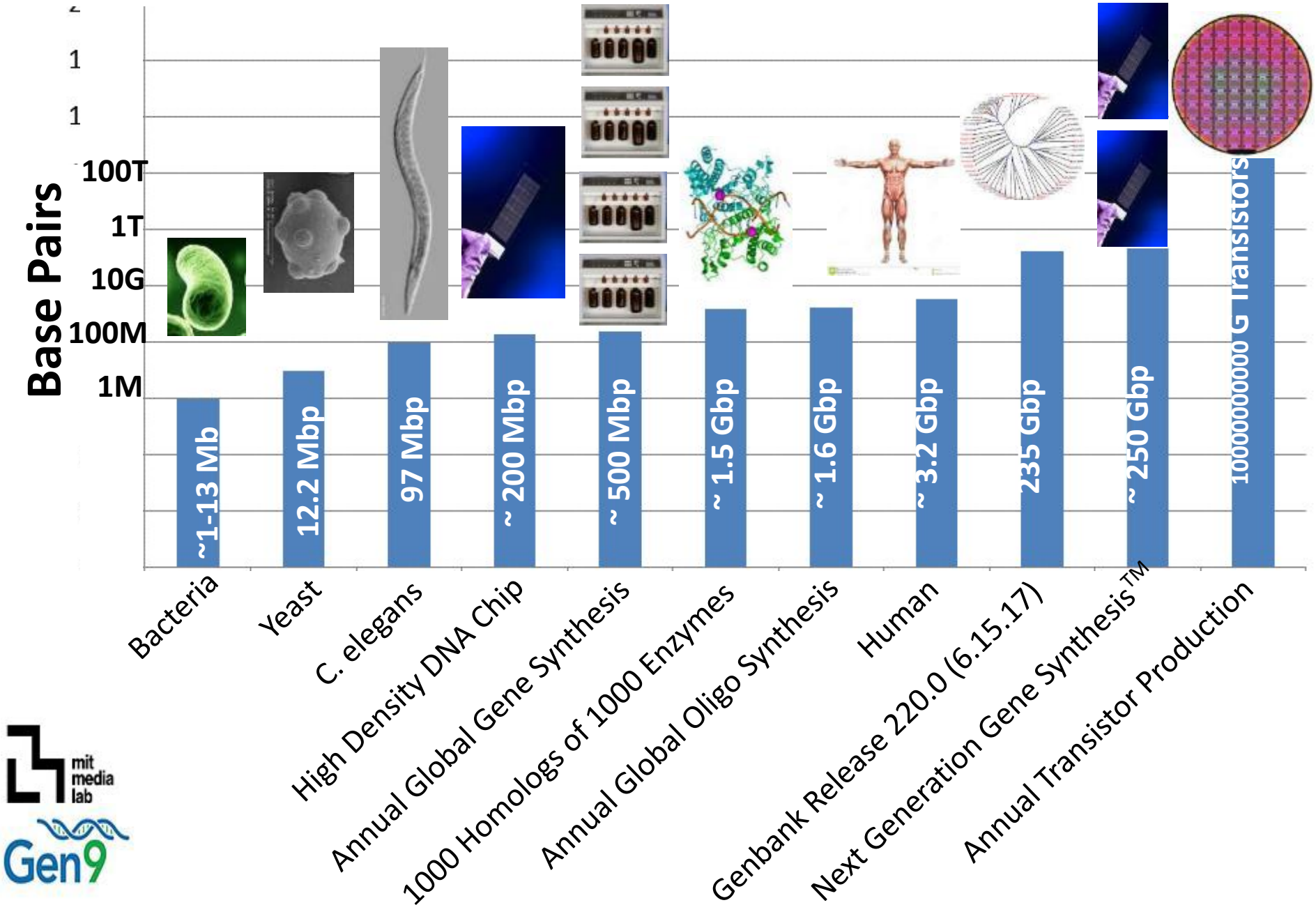


Bioeconomy CAPITAL

20 January, 2018

http://www.synthesis.cc/assets_c/2014/02/Carlson_Price_Seq_Synth_Feb2014.html

Next Generation (Chip Based) DNA Synthesis



mit media lab

Gen9

Letter | Published: 18 June 2018

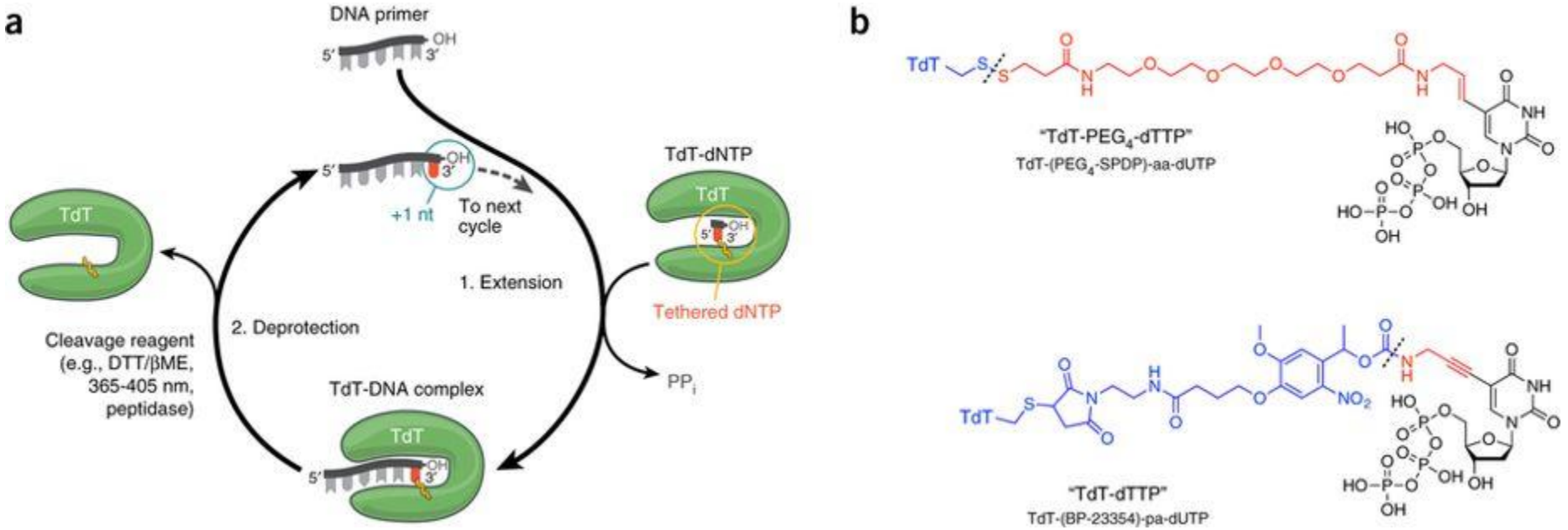
De novo DNA synthesis using polymerase-nucleotide conjugates

Sebastian Palluk, Daniel H Arlow, Tristan de Rond, Sebastian Barthel, Justine S Kang, Rathin Bector, Hratch M Baghdassarian, Alisa N Truong, Peter W Kim, Anup K Singh, Nathan J Hillson & Jay D Keasling

Nature Biotechnology **36**, 645–650 (2018) | Download Citation

Enzymatic Synthesis

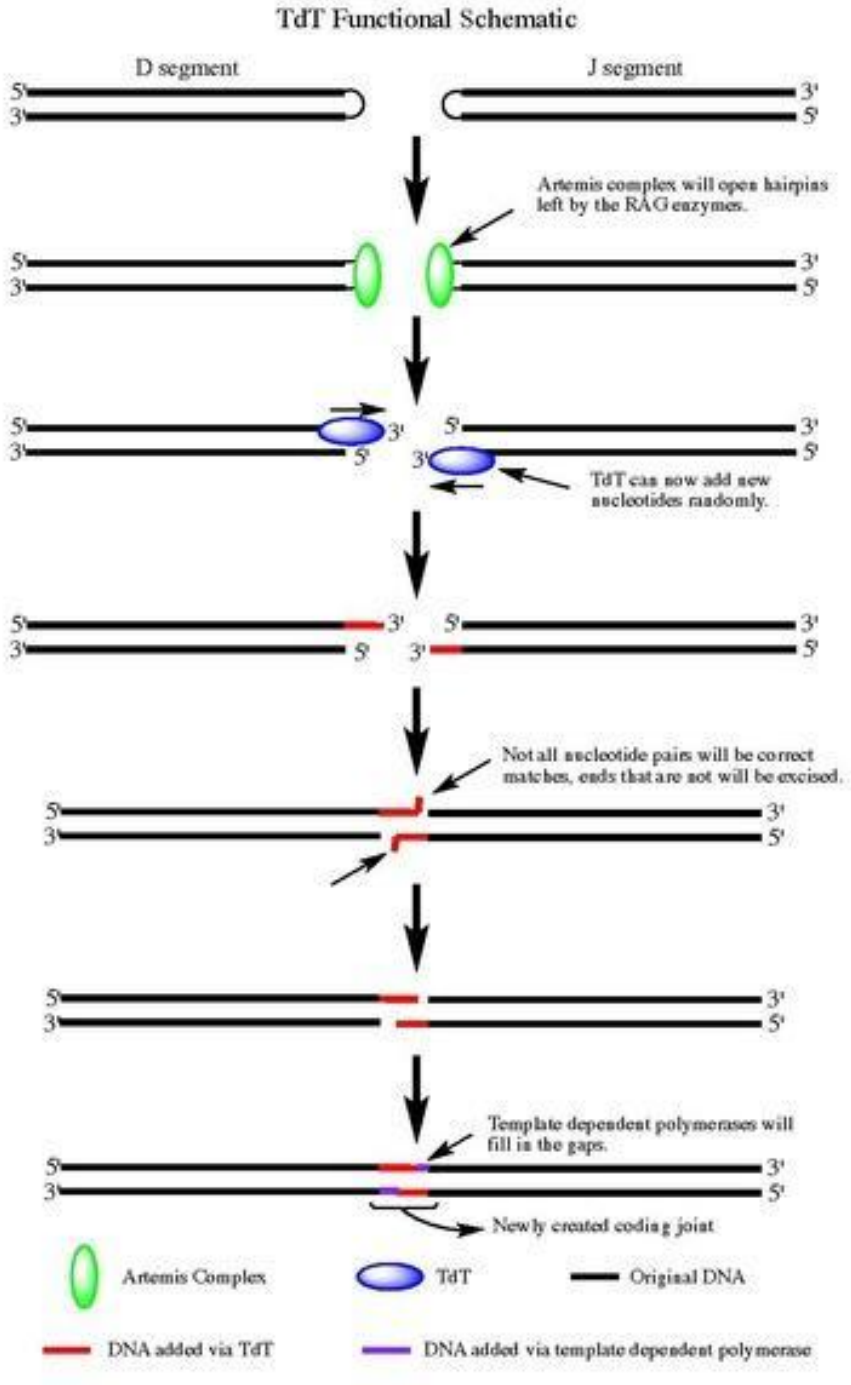
97.7% yield
20 S reaction time



(**a**) Scheme for two-step oligonucleotide extension using TdT–dNTP conjugates consisting of a TdT molecule site-specifically labeled with a dNTP via a cleavable linker. In the extension step, a DNA primer is exposed to an excess of TdT–dNTP conjugate. Upon incorporation of the tethered nucleotide into the 3′ end of the primer, the conjugate becomes covalently attached and prevents further extensions by other TdT–dNTP molecules. In the deprotection step, the remaining TdT–dNTP conjugates are inactivated (or removed) and the linkage between the incorporated nucleotide and TdT is cleaved by addition of the cleavage reagent (e.g., DTT, 365 nm light, peptidase), thereby releasing the primer for subsequent extension. The cycle can be iterated to extend a primer by a defined sequence. (**b**) Chemical structures of two types of TdT–linker-dNTP conjugates used in this study, based on the amine-to-thiol crosslinkers PEG_4-SPDP (upper, "TdT–PEG_4-dTTP") and BP-23354 (lower, "TdT–dTTP") and the dTTP analogs 5-aminoallyl dUTP (aa-dUTP) and 5-propargylamino dUTP (pa-dUTP), respectively. Upon cleavage of the linker, the atoms indicated in red remain attached to the nucleobase and are referred to as a scar. The

Terminal Deoxynucleotidyl Transferase(TdT)

Synthesizing DNA Enzymatically



Article | Open access | Published: 21 January 2026

Construction of complex and diverse DNA sequences using DNA three-way junctions

Noah Evan Robinson, Weilin Zhang (张炜林), Rajesh Ghosh, Bryan Gerber, Hanqiao Zhang (张汉翘), Charles Sanfiorenzo, Sixiang Wang (王思翔), Dino Di Carlo & Kaihang Wang (王开航)

Nature (2026) | Cite this article

Fig. 2: Sidewinder reliably assembles large multifragment assemblies with high accuracy.

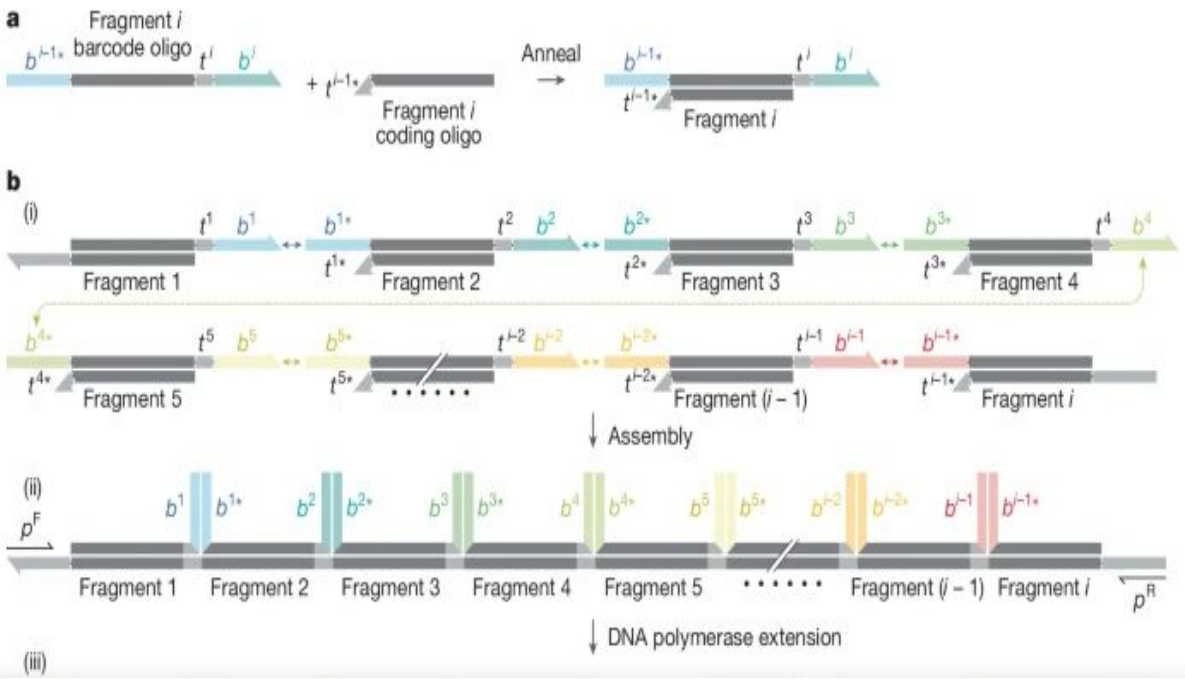
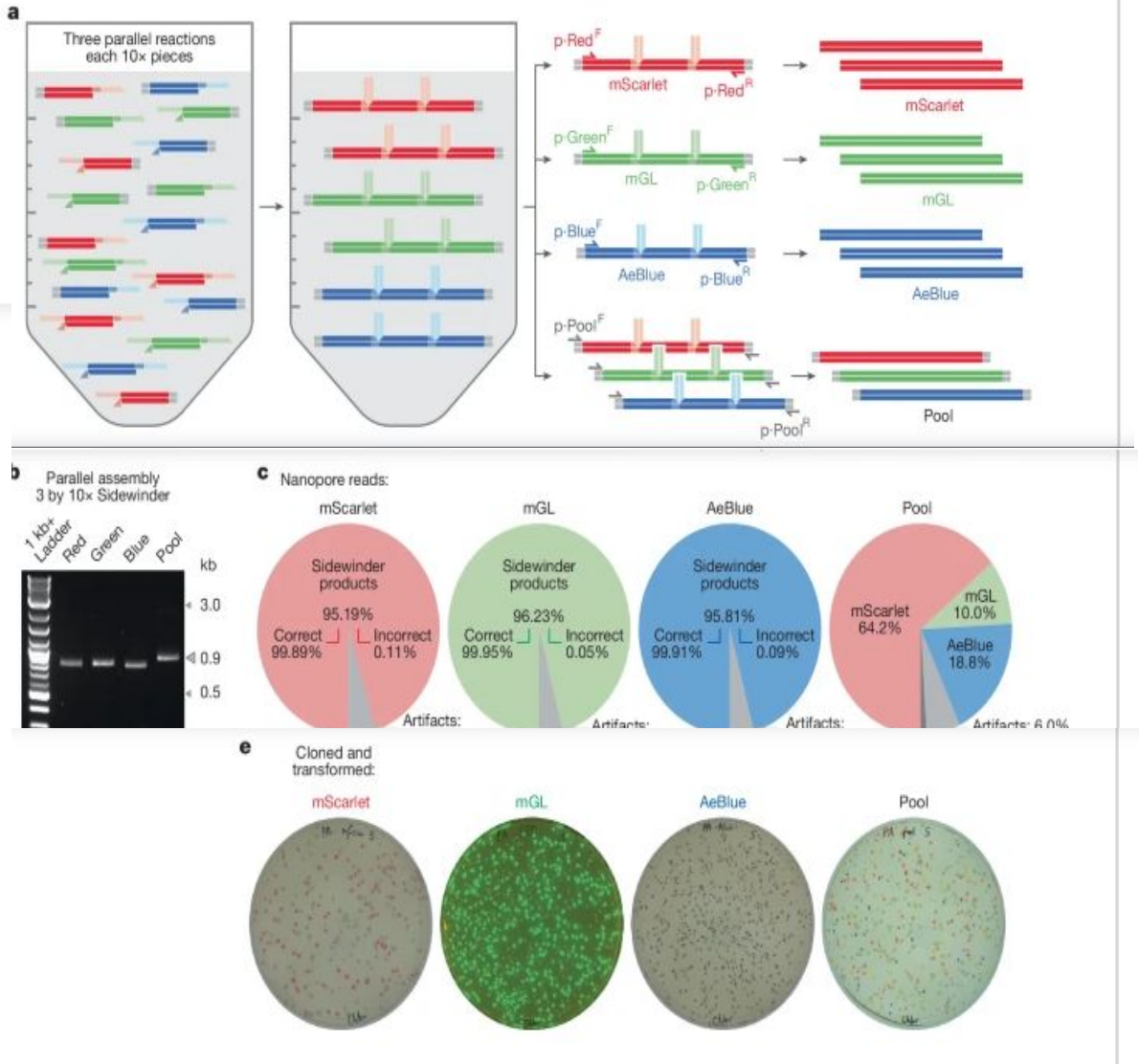


Fig. 4: Sidewinder independently assembles multiple distinct constructs in one pot with high fidelity.



DNASCRIPT

INTRODUCING THE SYNTAX SYSTEM:

The first DNA printer powered by Enzymatic DNA Synthesis (EDS)

REQUEST QUOTE

SYSTEM CAPABILITIES

- Fully automated DNA printer
- Plug-and-play operation

- Parallel synthesis in 96-well plates
- Up to 96 oligos per run

- 15-min setup time per run
- Same-day synthesis of 15- to 30-mers
- Longer oligos available for next-day use

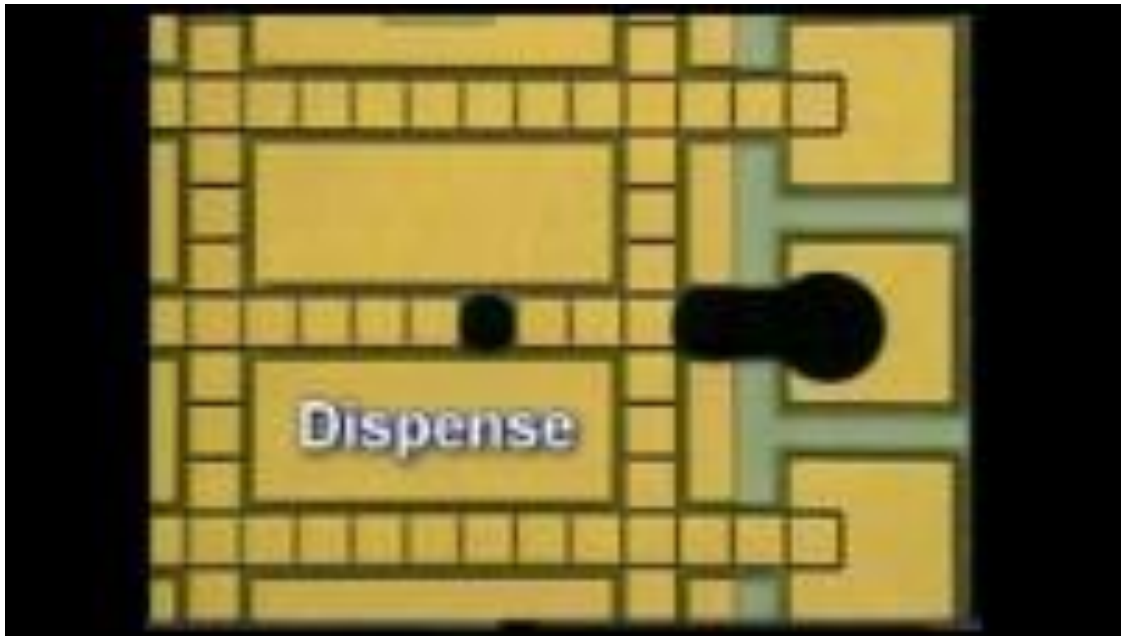
OLIGO SPECIFICATIONS

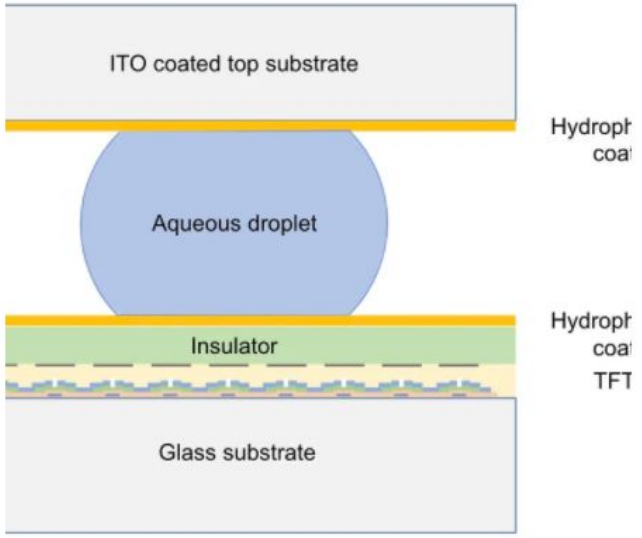
- 15–60 nt *de novo* synthesis (A, T, G, and C)
- Custom iDNA 15–40 nt

- Molecular biology-ready DNA
- Default 5'-phosphate

- 200–300 pmol per well
- Normalized to 4–7 µM

nuclera


Dispense


ITO coated top substrate
Hydroph
coa
Aqueous droplet
Hydroph
coa
Insulator
TFT
Glass substrate

Applications

Cells as Chemical Factories

1,3 Propanediol

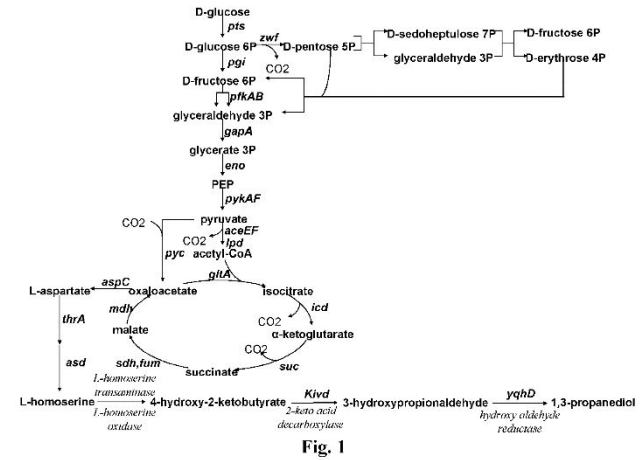
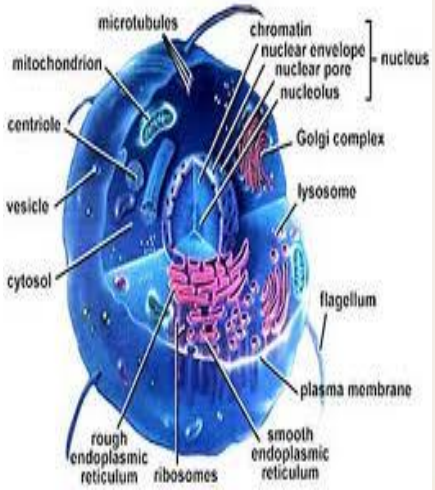


Fig. 1



E. coli fermentation

HO OH

1,3-Propanediol

Polypropylene terephthalate

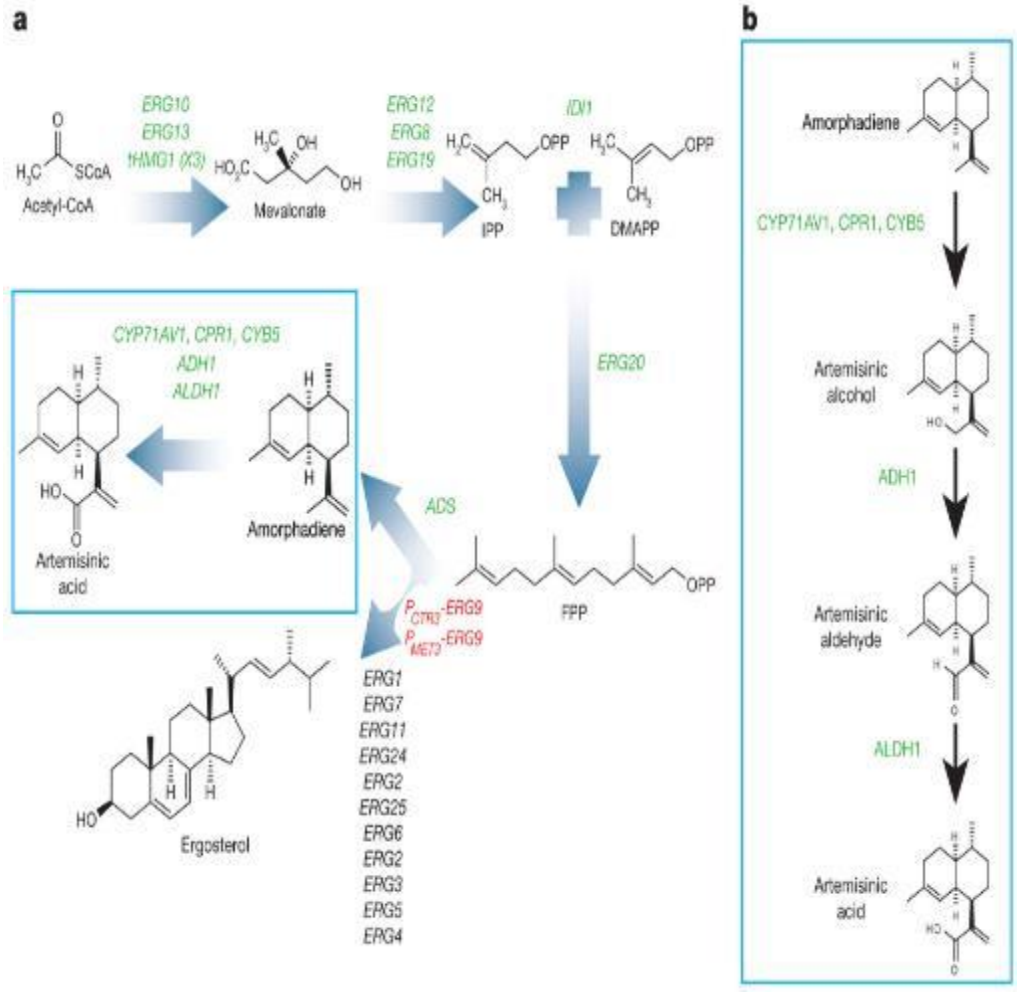
http://www.latonkorea.com/Plant.html

http://3rdpartylogistics.blogspot.com/2011/10/genetic-bacteria-genetic-modification.html

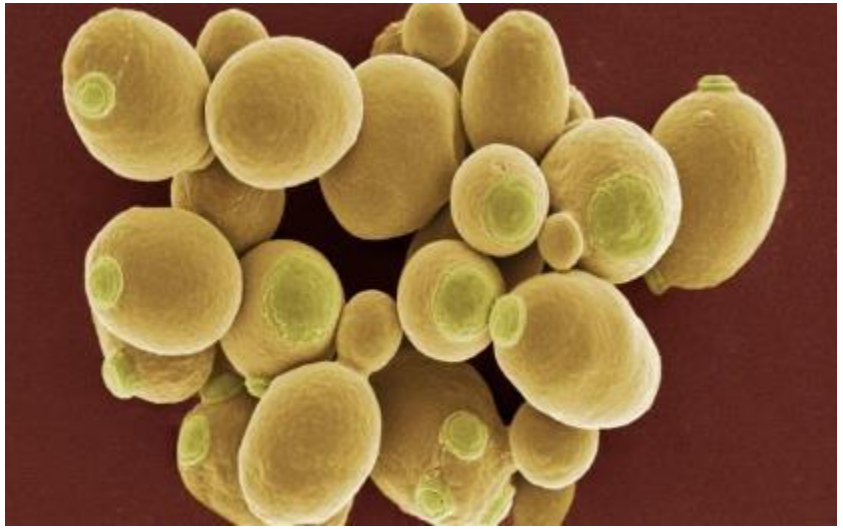
Cells as Chemical Factories

~6 Gb Synthesis For top 1000 small molecules drug intermediates

Artemisinic Acid (Antimalarial)



Paddon, Christopher J., et al. "High-level semi-synthetic production of the potent antimalarial artemisinin." *Nature* 496.7446 (2013): 528-532.



http://www.latonkorea.com/Plant.html

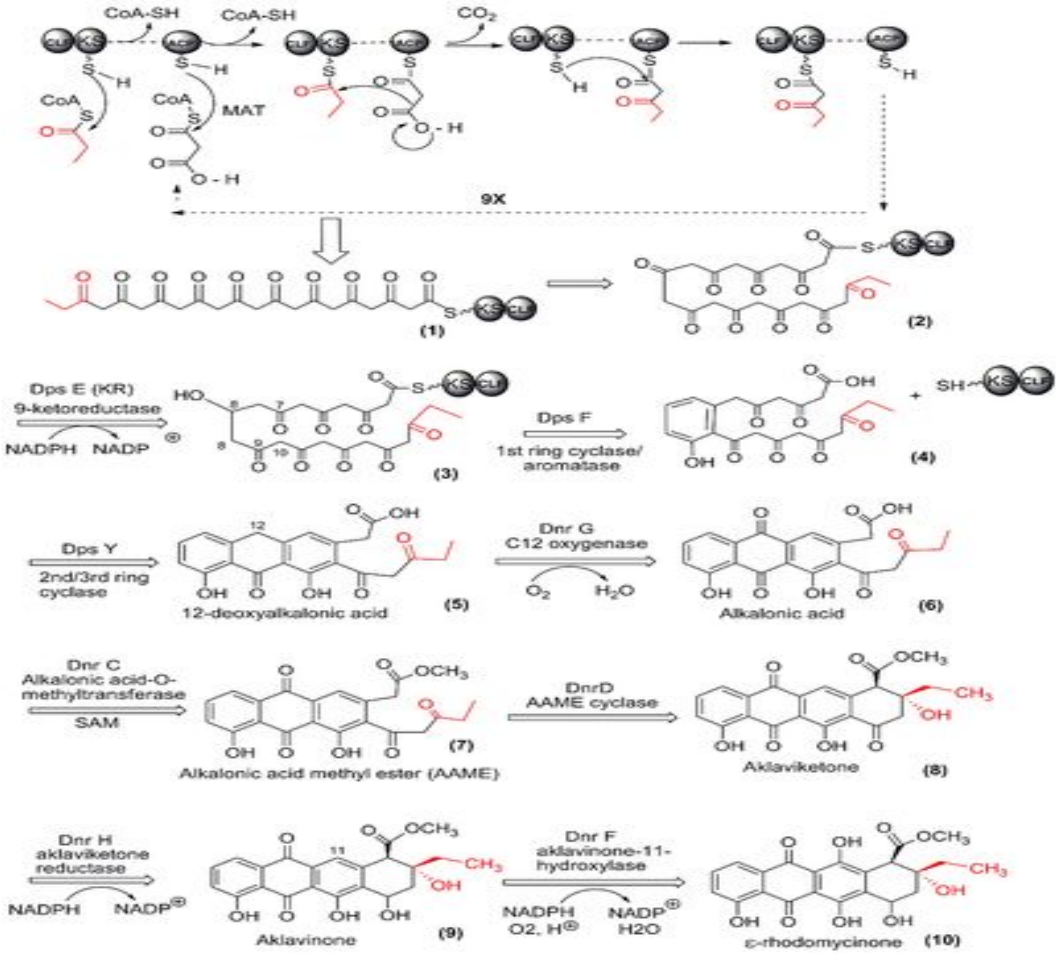
http://neweuropeans.org/eurowine-3-yeast-europe-and-economics/

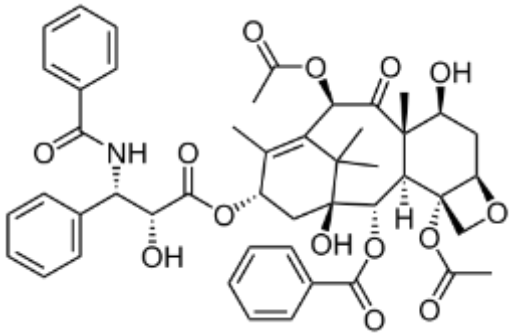
Therapeutics from Natural Product Pathways

~2/3 of Small Molecule Pharmaceuticals are Natural Products

~5 Gb Synthesis for Known PKS Pathway Modules

1000's of Annotated PKS's





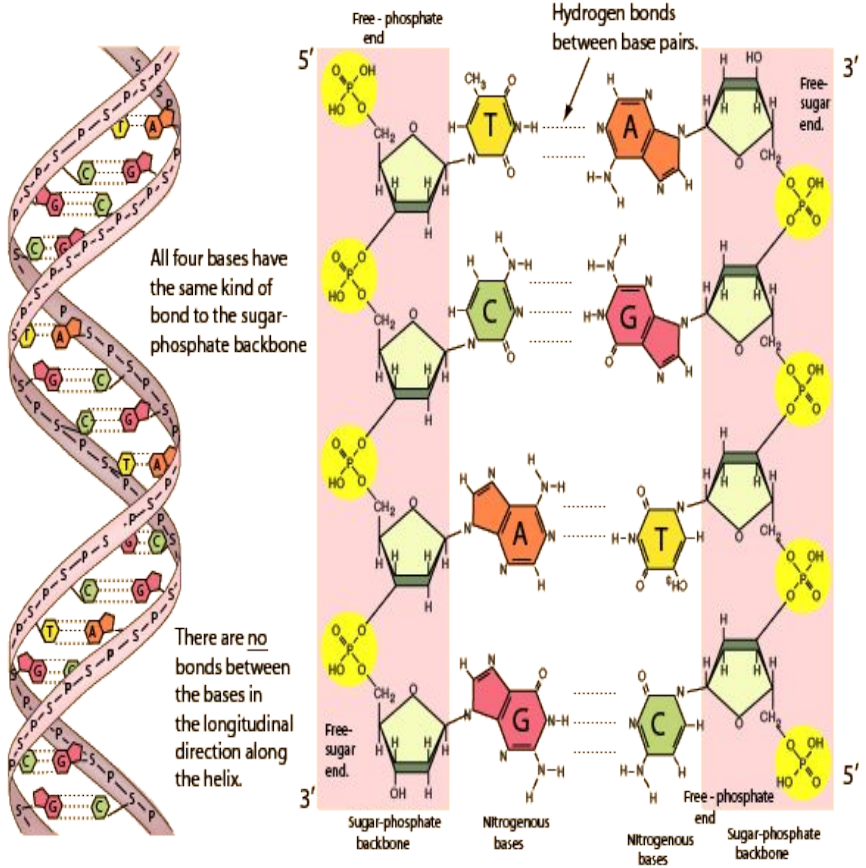
Paclitaxel (Taxol)

Pacific Yew

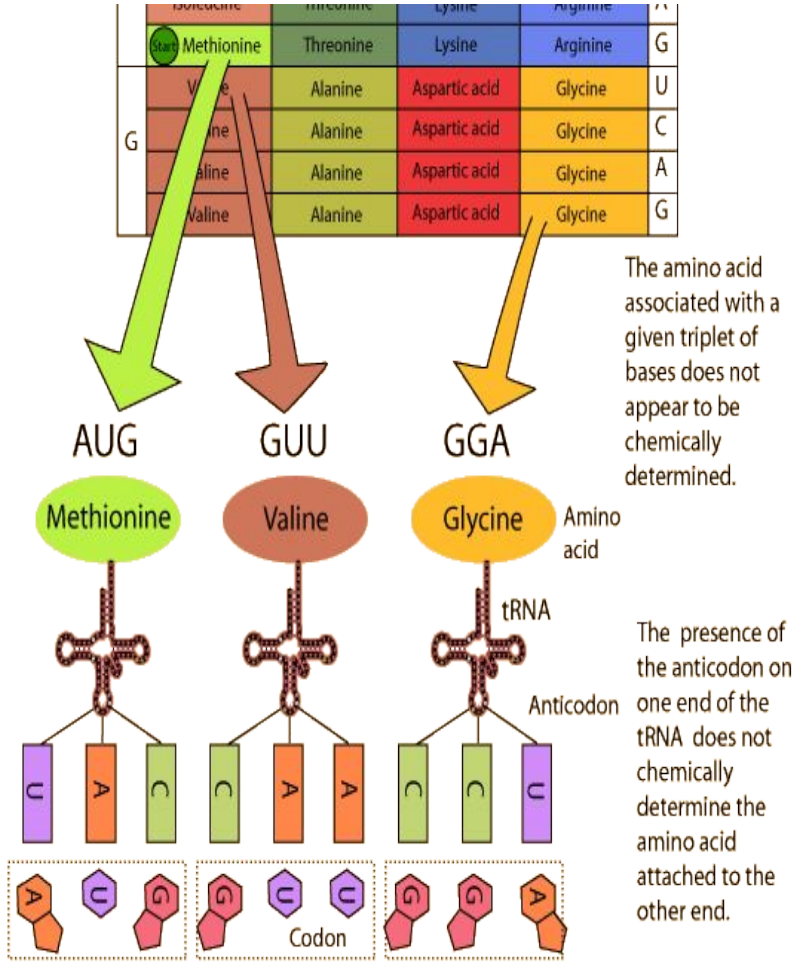
Recoding

The Genetic Code

Life's Operating System



Flattened segment of the double helix



mRNA transcribed code

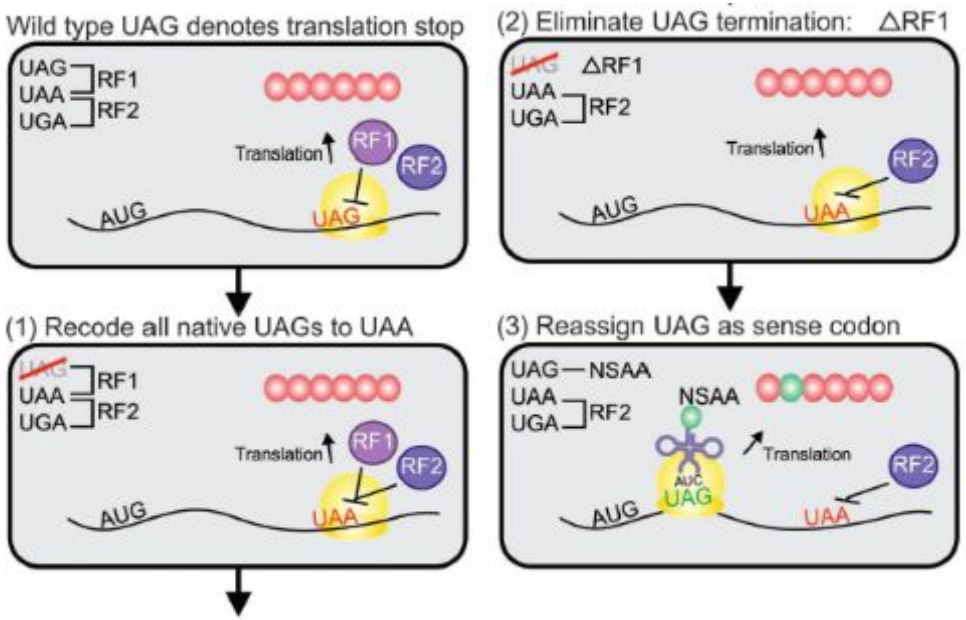
mit media lab

Science

Genomically Recoded Organisms Expand Biological Functions

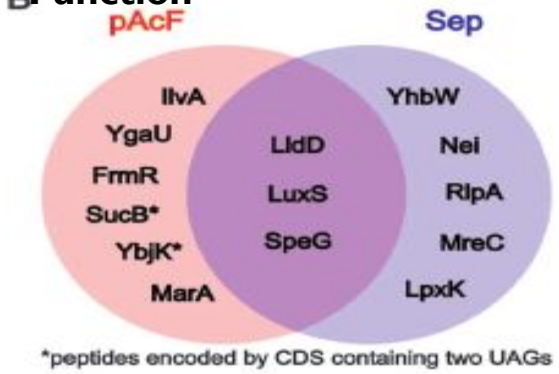
Marc J. Lajoie,[1,2] Alexis J. Rovner,[3,4] Daniel B. Goodman,[1,5] Hans-Rudolf Aerni,[4,6] Adrian D. Haimovich,[3,4] Gleb Kuznetsov,[1] Jaron A. Mercer,[7] Harris H. Wang,[8] Peter A. Carr,[9] Joshua A. Mosberg,[1,2] Nadin Rohland,[1] Peter G. Schultz,[10] Joseph M. Jacobson,[11,12] Jesse Rinehart,[4,6] George M. Church,[1,13*] Farren J. Isaacs[3,4*]

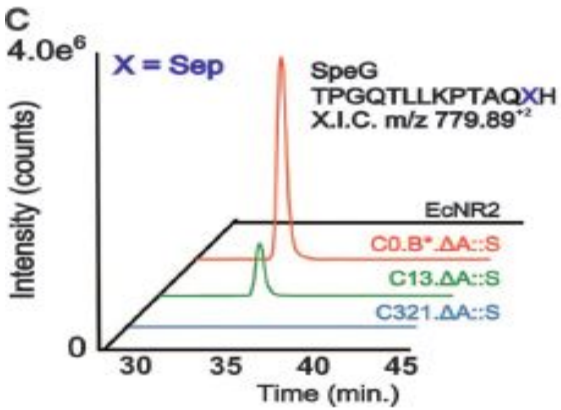
www.sciencemag.org SCIENCE VOL 342 18 OCTOBER 2013



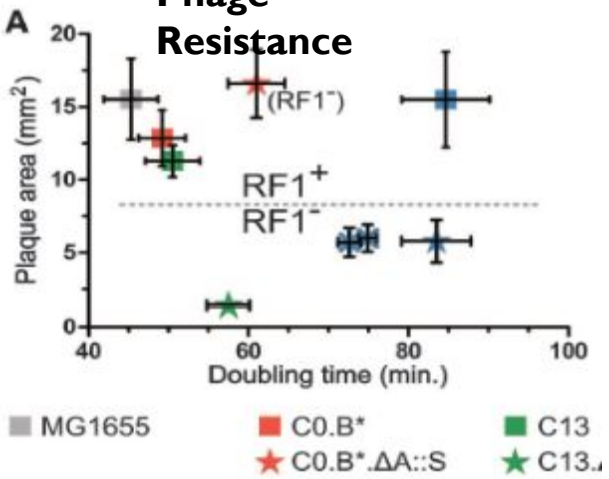
mit media lab

Conferring New Function





Phage Resistance



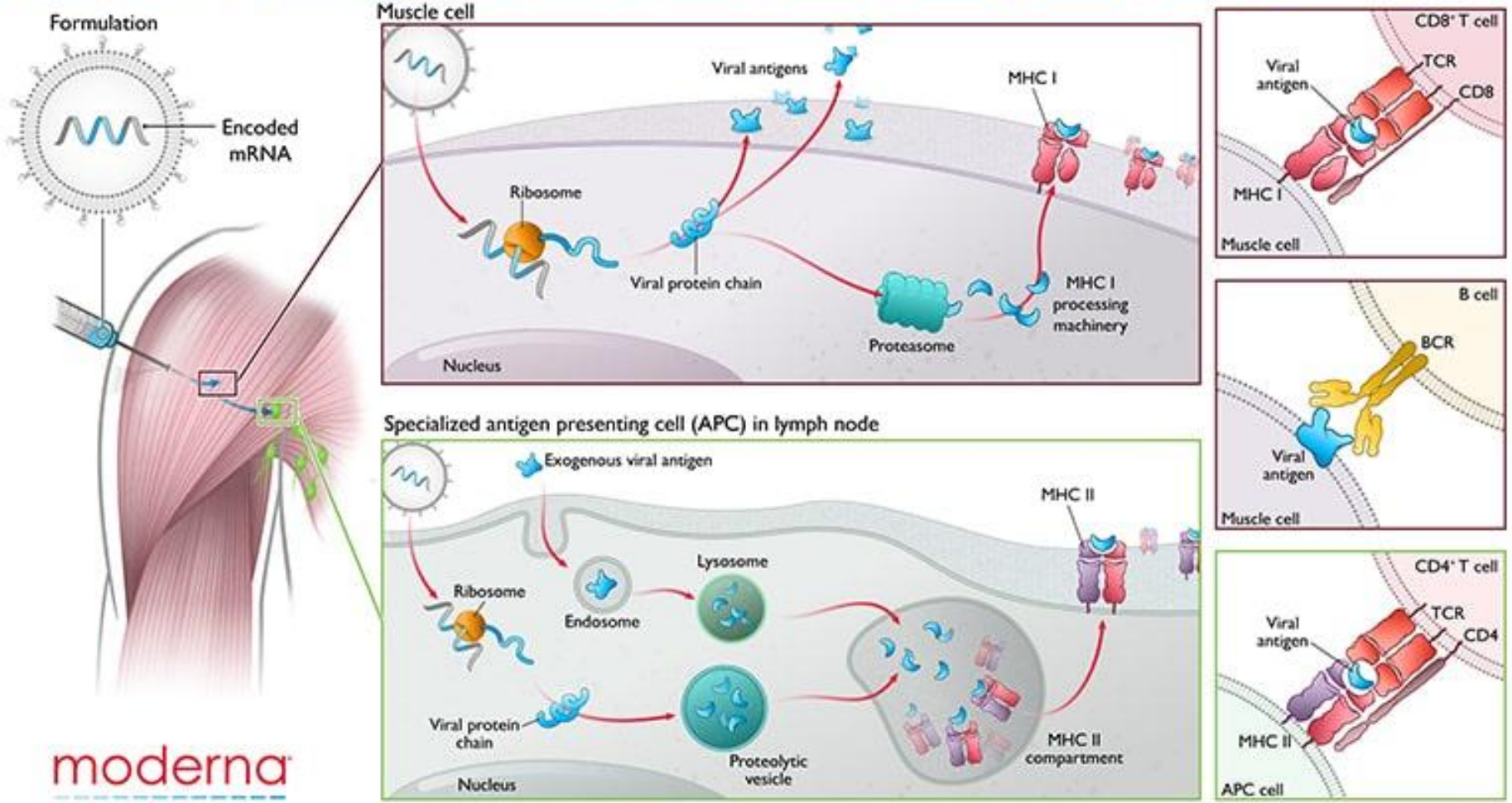
Building:

- **Therapeutics**
- **Nanostructures**
- **Nanomachines**
- **Data Storage**

With Nucleotides

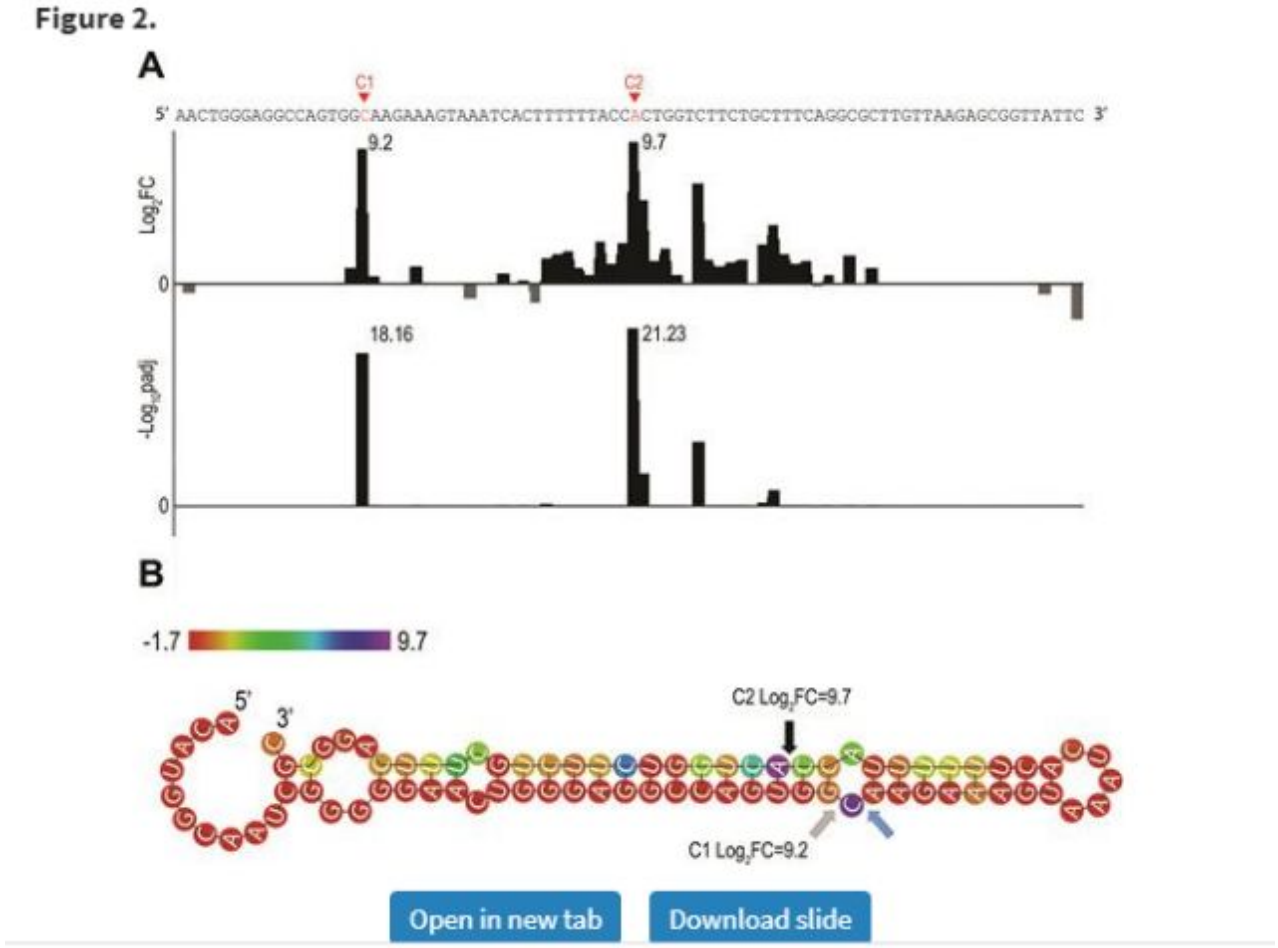
Moderna's mRNA Vaccine Approach

Closely mimics a native viral infection leading to B and T cell responses



moderna

RNA Cleavage Rules



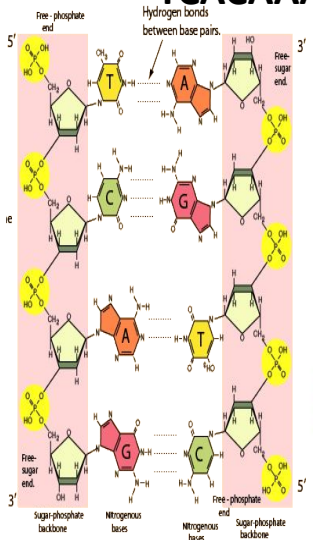
Altuvia Y, Bar A, Reiss N, Karavani E, Argaman L, Margalit H. In vivo cleavage rules and target repertoire of RNase III in Escherichia coli. Nucleic acids research. 2018 Nov 2;46(19):10380-94.

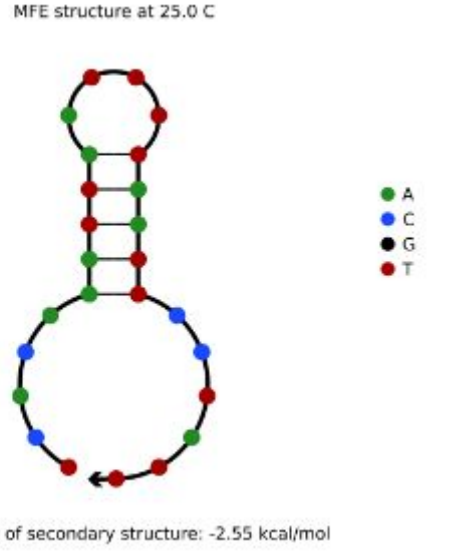
NUPACK nucleic acid package

GC = 10%

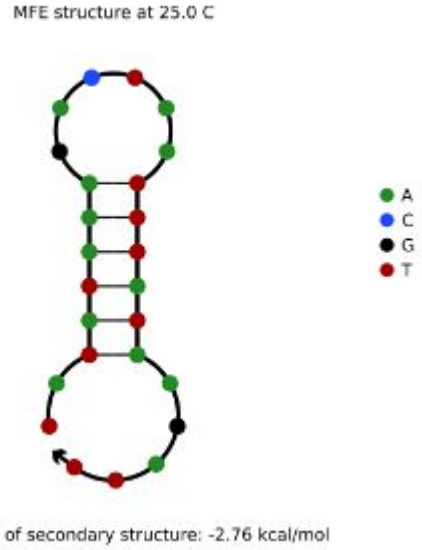
Minimum Free Energy Secondary Structures : 25^0 C

TCACAAATTAATTTTAATTCCTATT TATATAAAGACTAATTTATAAGATT ATCTAATATCTTAGTGATAAAATTT

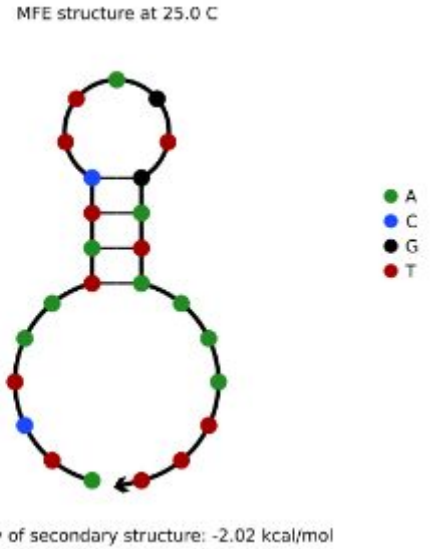




Free energy of secondary structure: -2.55 kcal/mol

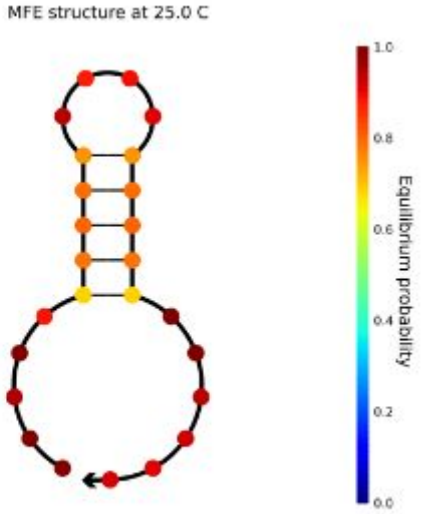


Free energy of secondary structure: -2.76 kcal/mol

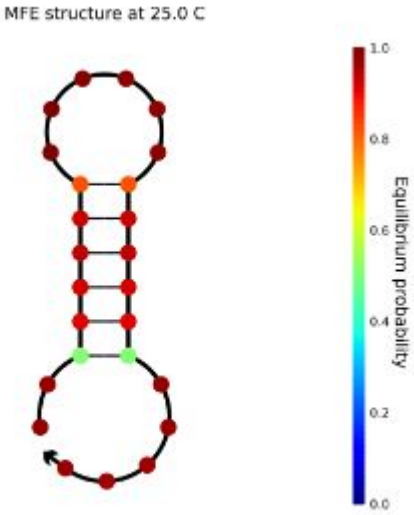


Free energy of secondary structure: -2.02 kcal/mol

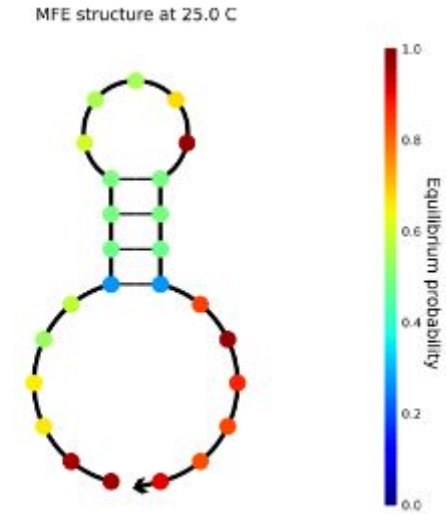
A/T ~ -1.2kcal/mol
G/C ~ -2.0 kcal/mol



Free energy of secondary structure: -2.55 kcal/mol



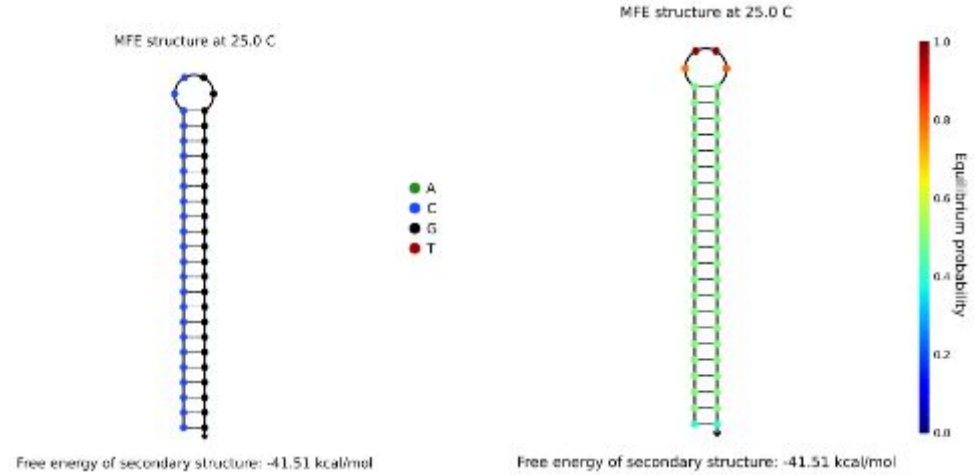
Free energy of secondary structure: -2.76 kcal/mol



Free energy of secondary structure: -2.02 kcal/mol

Base Pairing Energies

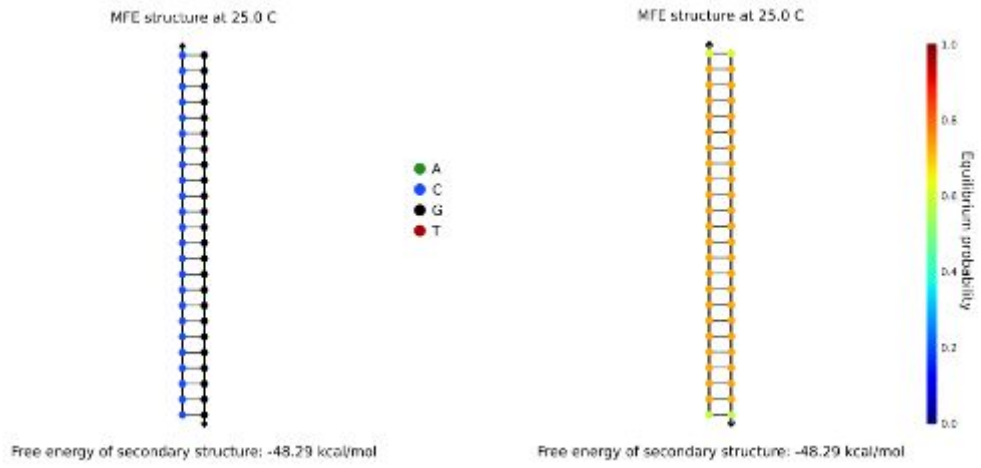
CCCCCCCCCCCCCCCCCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGG



G/C ~ -2.0 kcal/mol

CCCCCCCCCCCCCCCCCCCCCCCC

GGGGGGGGGGGGGGGGGGGGGGG

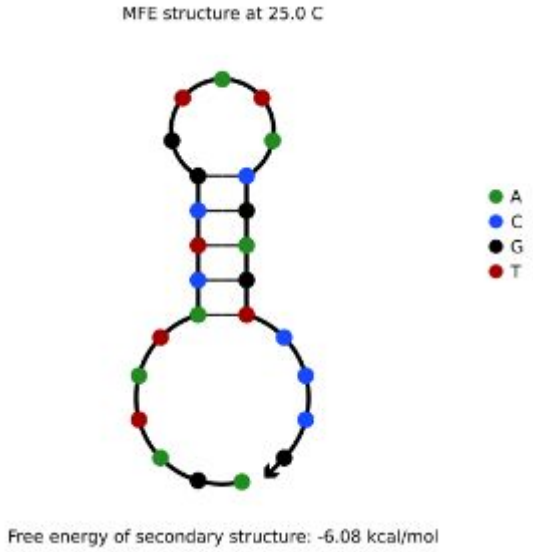


NUPACK nucleic acid package

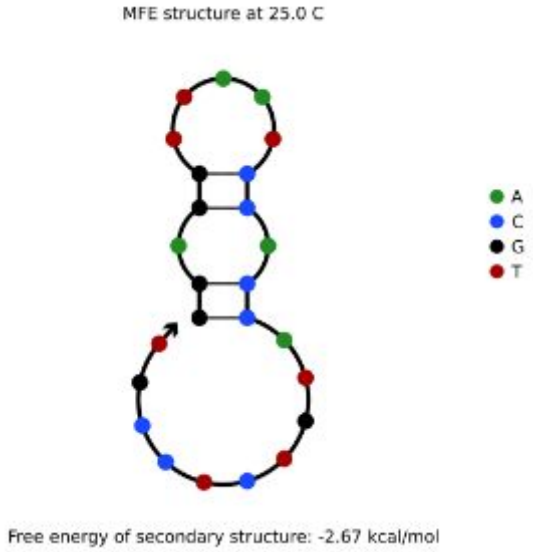
GC = 50%

Minimum Free Energy Secondary Structures : 25^0 C

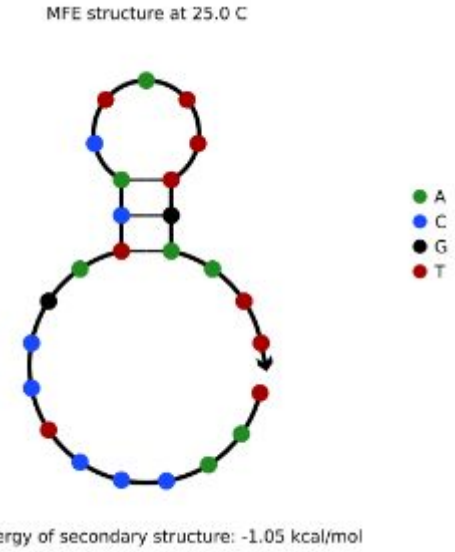
AGATATACTCGGTATACGAGTCCCG **GGAGGTTAATCCACCATGTCTCCGT** **TAACCCTCCGATCACTATTTGAATT**



Free energy of secondary structure: -6.08 kcal/mol



Free energy of secondary structure: -2.67 kcal/mol



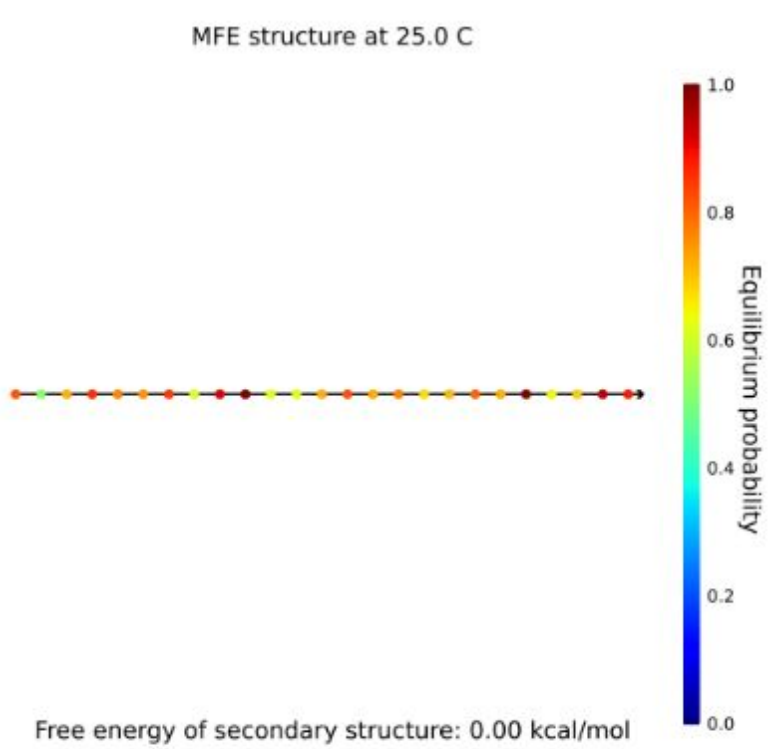
Free energy of secondary structure: -1.05 kcal/mol

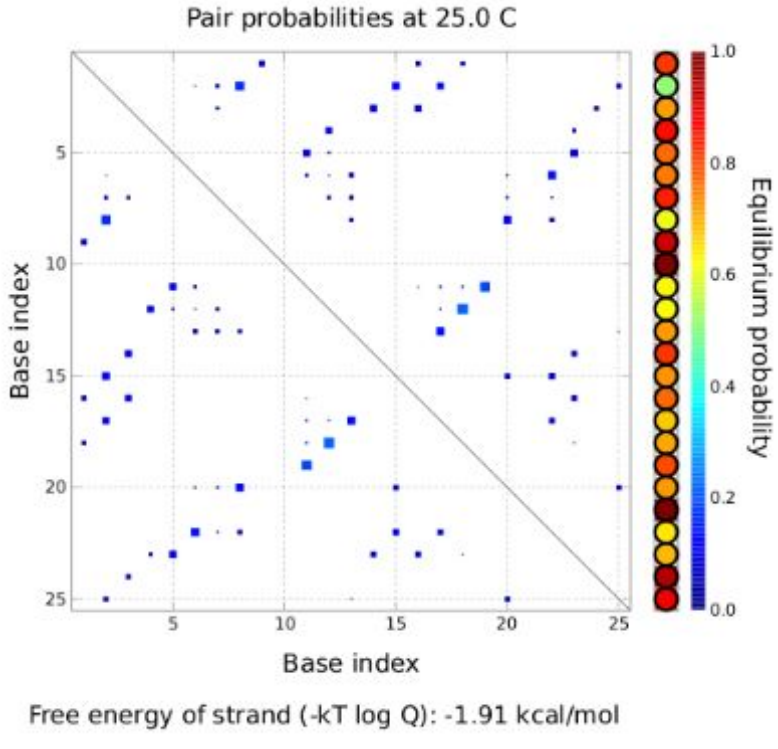
NUPACK nucleic acid package

GC = 50%

Minimum Free Energy Secondary Structures : 25^0 C

TCTAAGGGAATTCAGAGAACACTAG



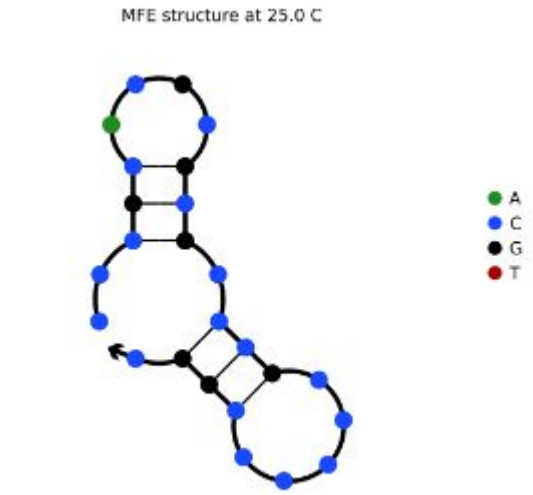


NUPACK nucleic acid package

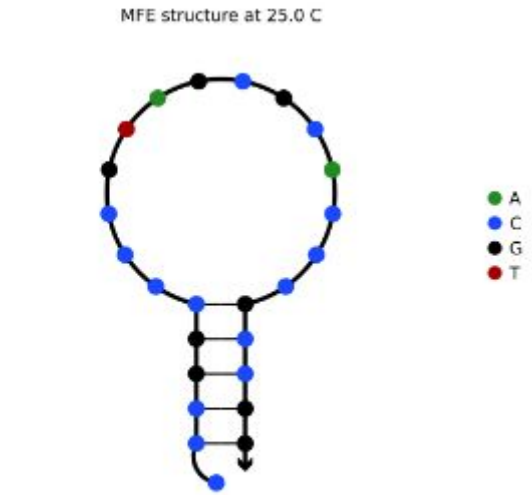
GC = 90%

Minimum Free Energy Secondary Structures : 25^0 C

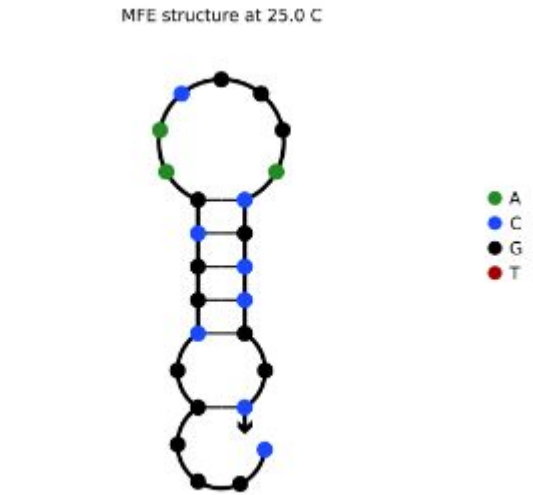
CCCGCACGCGCGCCCGCCCCCCGGC CCCGGCCCCGTAGCGCACCCGCCGG CGGGGGCGGCGAACGGGACGCCGGC



Free energy of secondary structure: -6.16 kcal/mol



Free energy of secondary structure: -5.97 kcal/mol

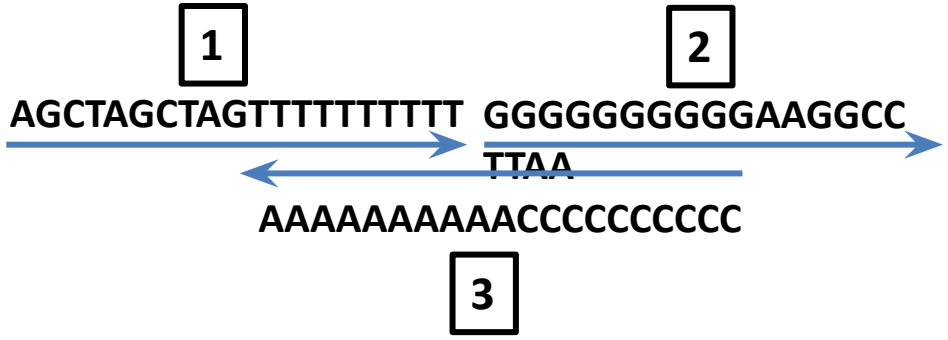


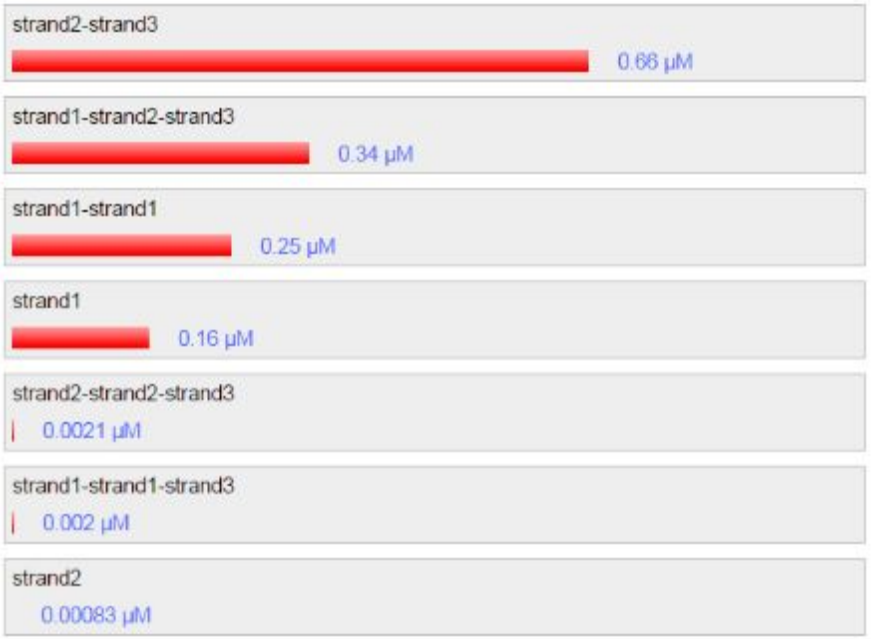
Free energy of secondary structure: -9.98 kcal/mol

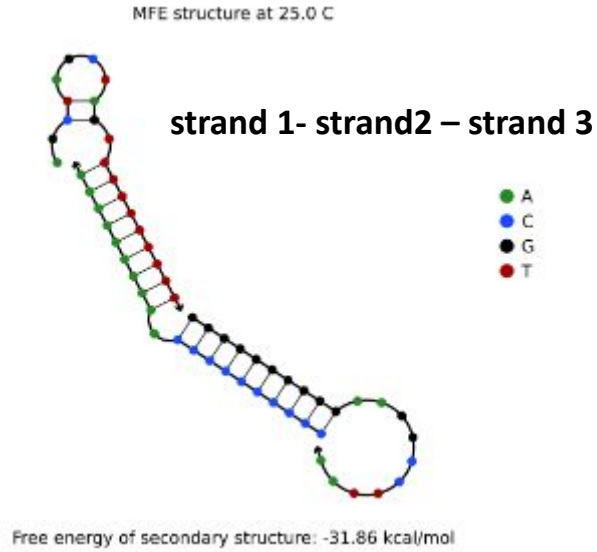
NUPACK nucleic acid package

Putting The Pieces Together

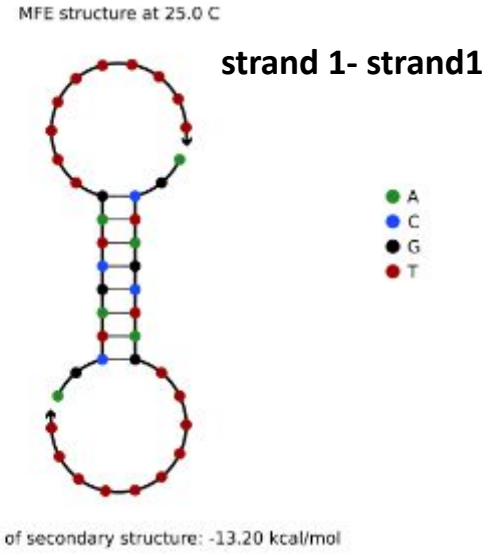
Minimum Free Energy Secondary Structures : 25^0 C





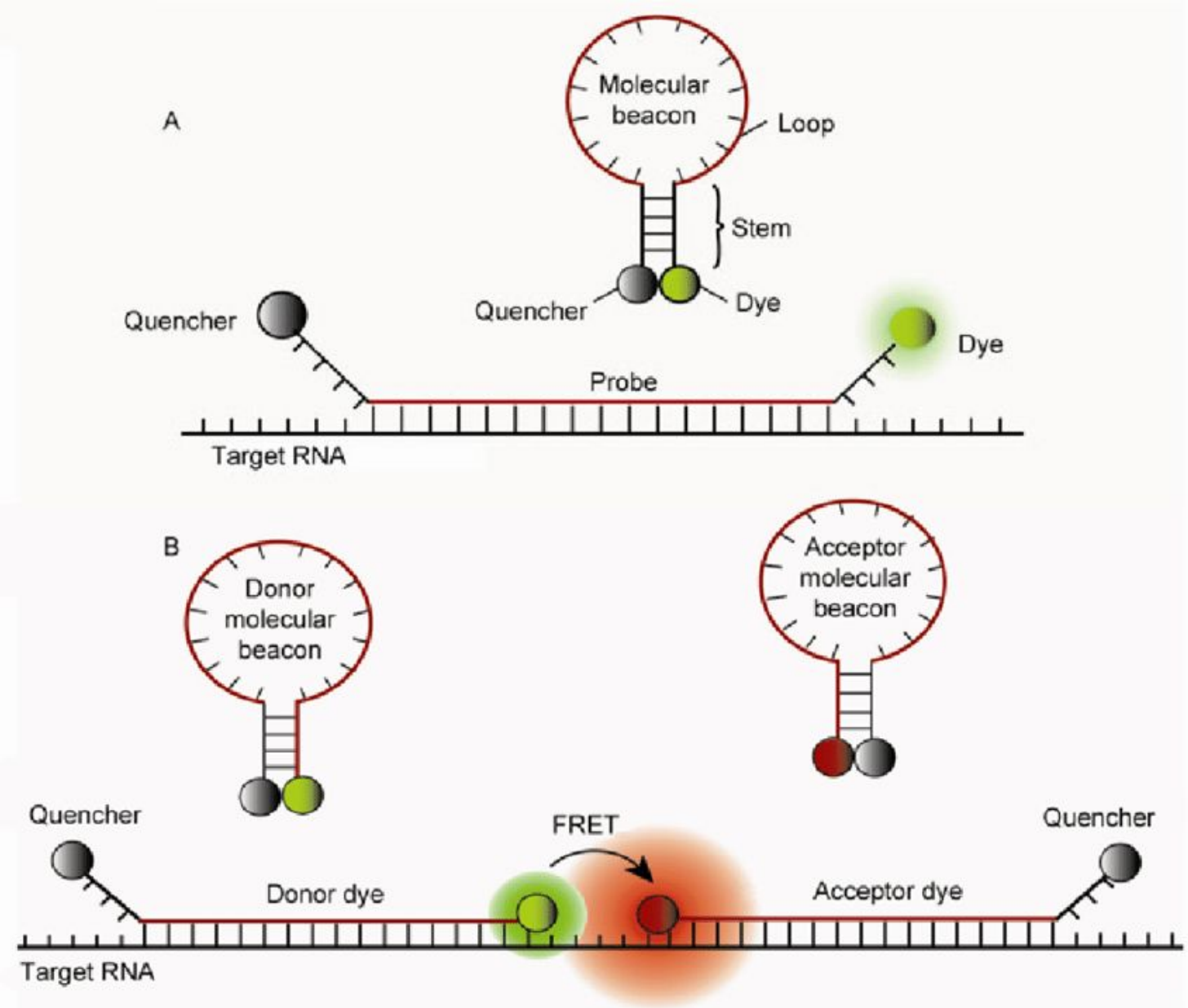


strand 1- strand2 – strand 3



strand 1- strand1

Molecular Beacon

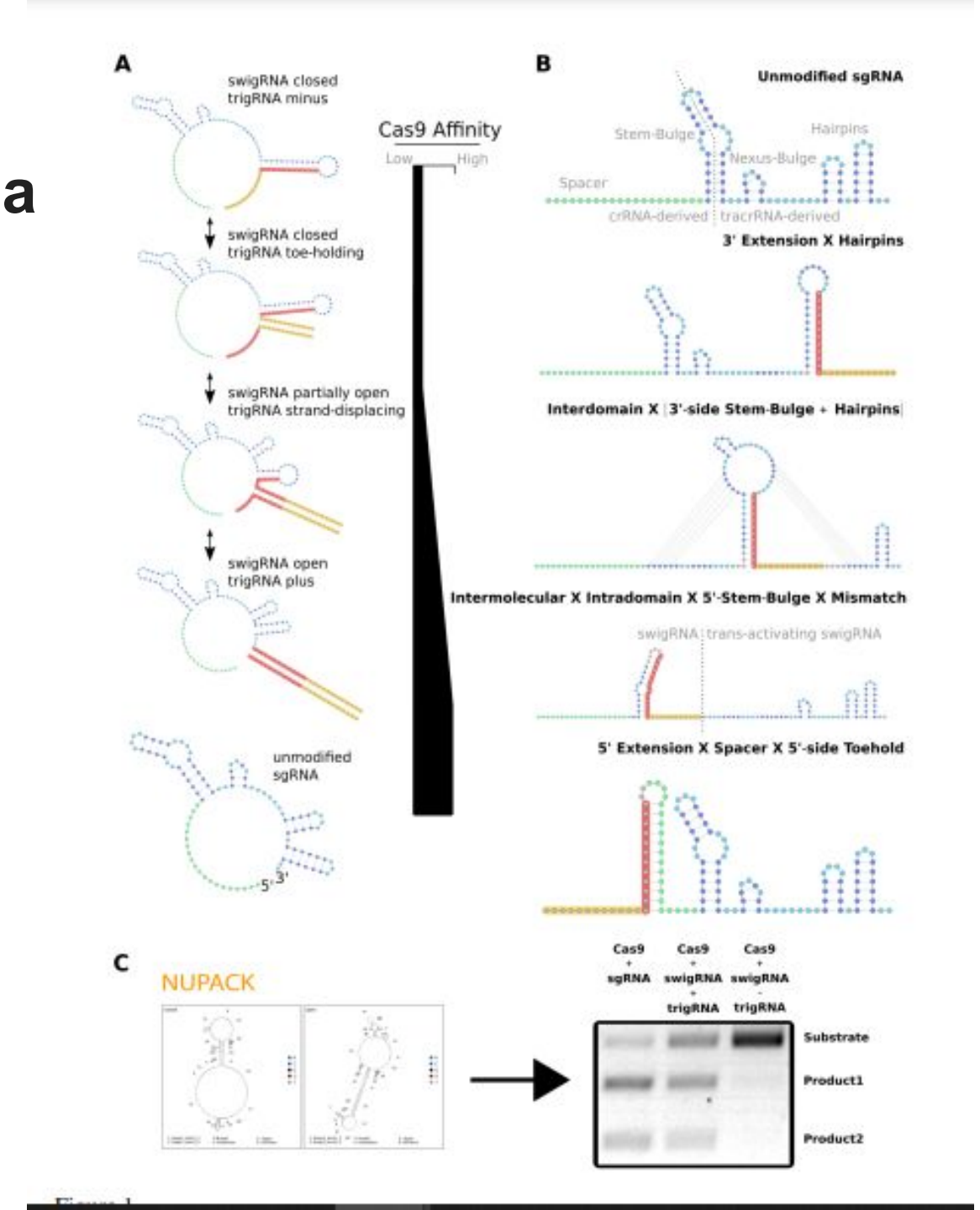


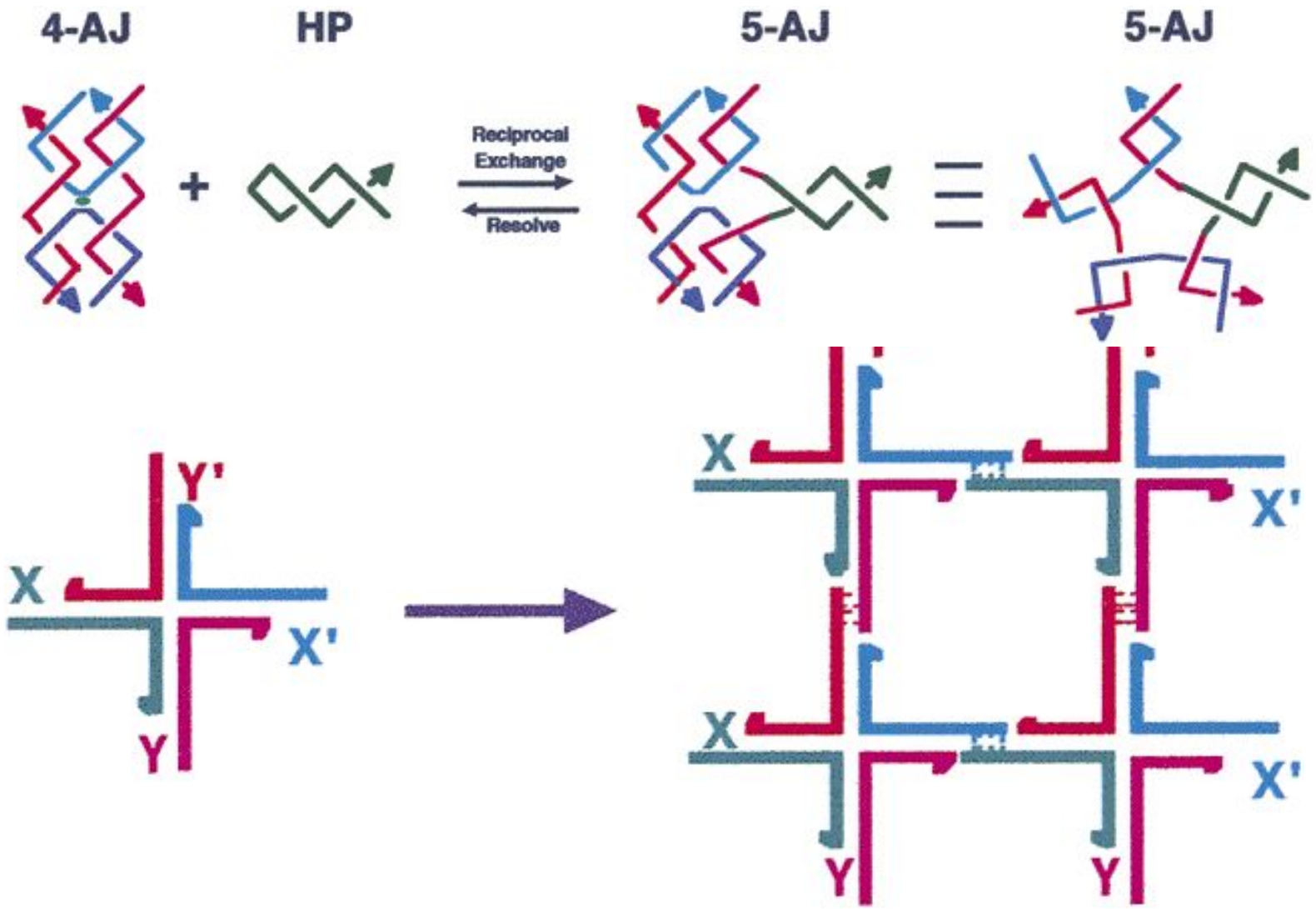
bioRχiv

THE PREPRINT SERVER FOR BIOLOGY

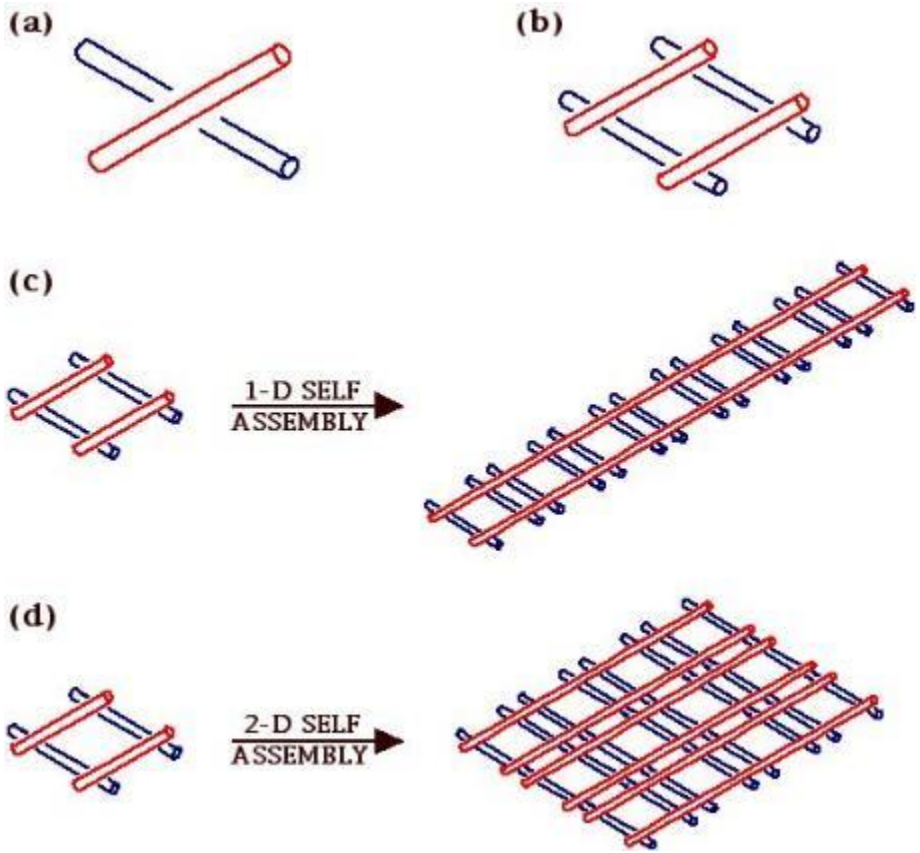
ssrna/dna-sensors via embedded strand-displacement programs in crispr/cas9 guides

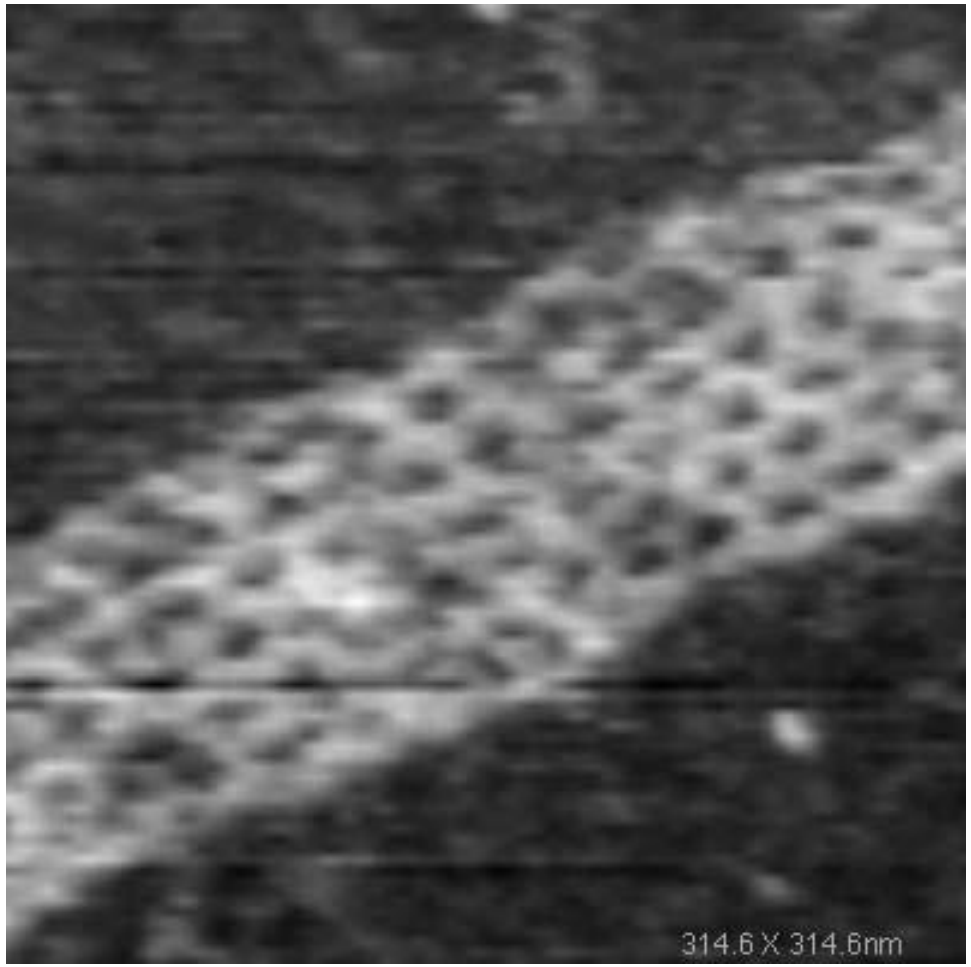
Jakimo, N., Chatterjee, P. and Jacobson, J.M., 2018.. *bioRxiv*, p.264424.





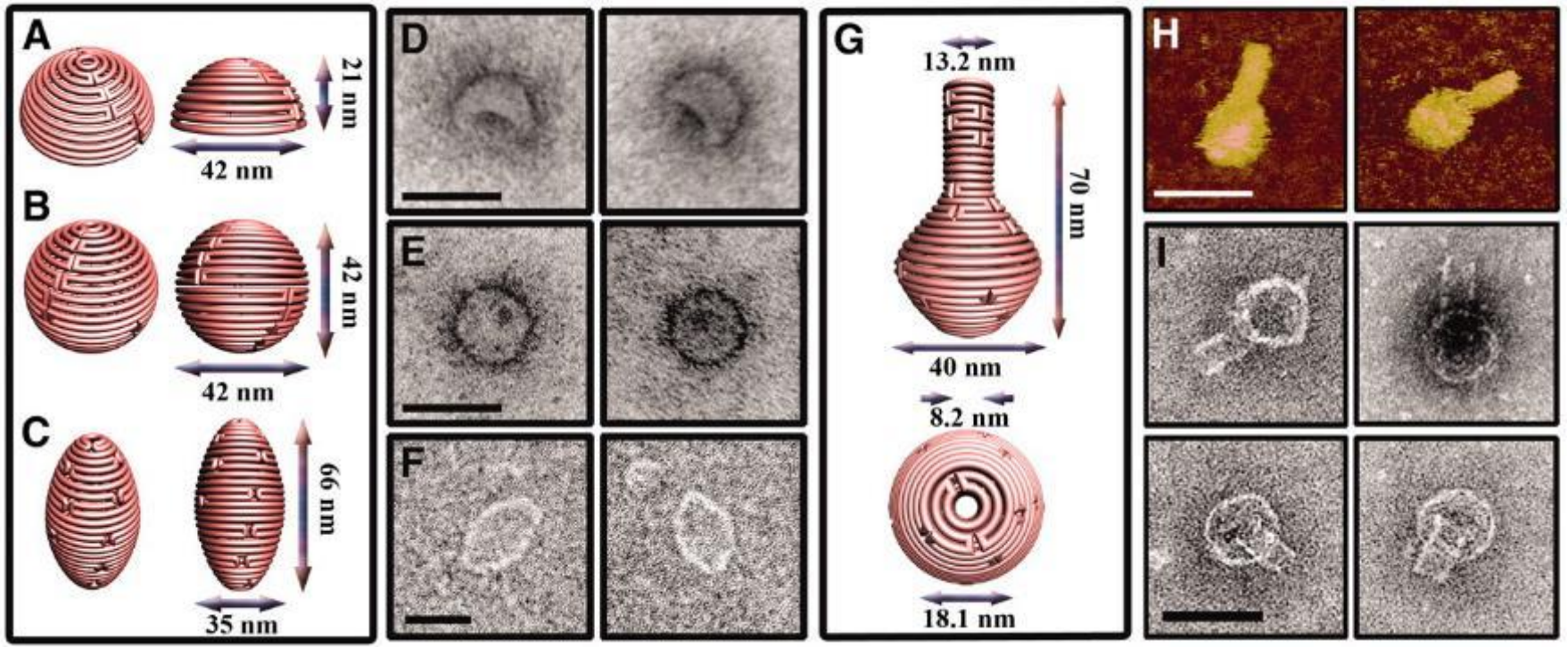
Nano Letters, **1** (1), 22 -26, 2001. 10.1021/nl000182v S1530-6984(00)00182-X



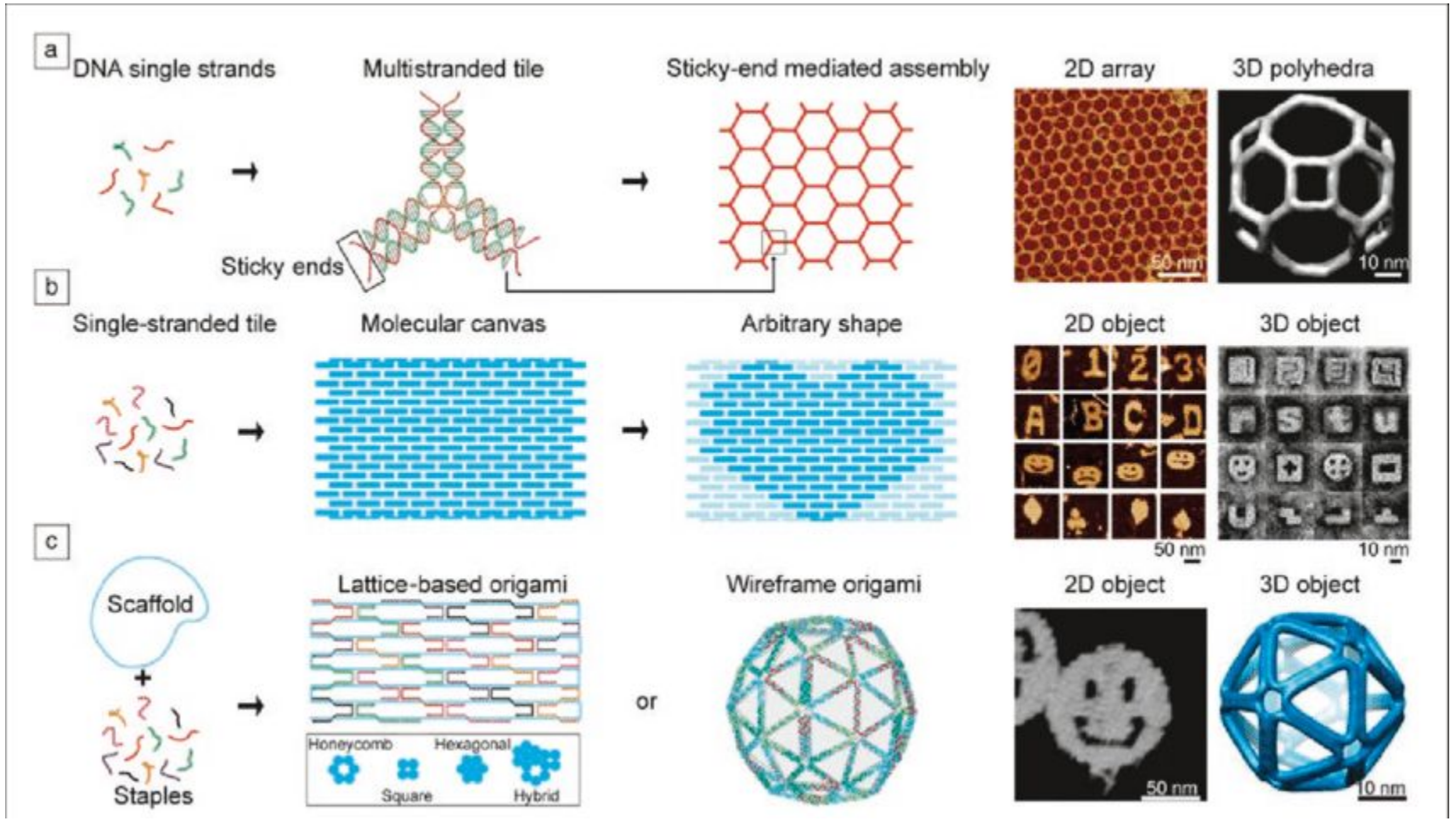


http://seemanlab4.chem.nyu.edu/HJ.arrays.html

3D DNA Origami

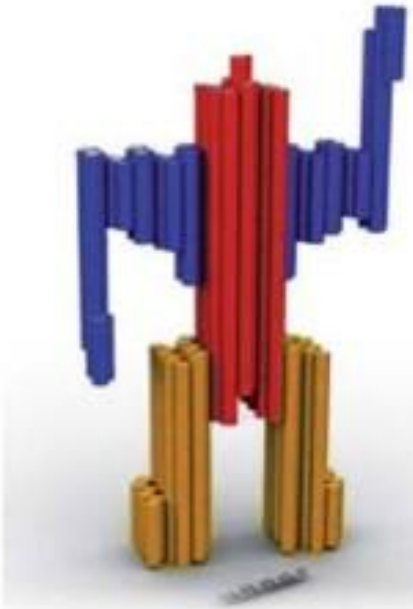


Science 15 April 2011:
Vol. 332 no. 6027 pp.
342–346
DOI: 10.1126/science.120299
8

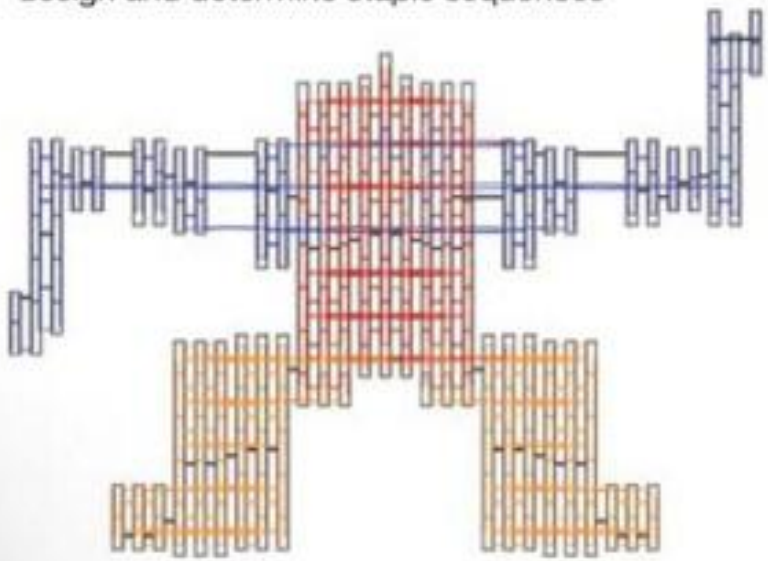


https://www.researchgate.net/figure/Summary-of-DNA-tile-and-DNA-origami-a-Multistranded-DNA-tile-Each-tile-is-composed-of_fig1_321675577

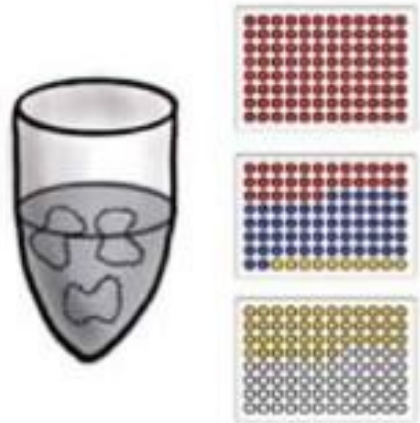
Step 1: conceive a target shape



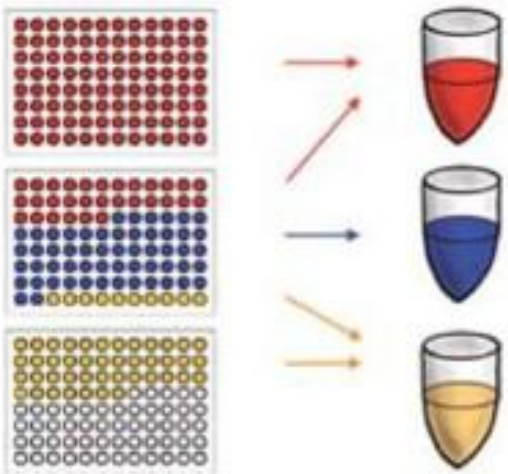
Step 2: design scaffold-staple layout, evaluate design and determine staple sequences



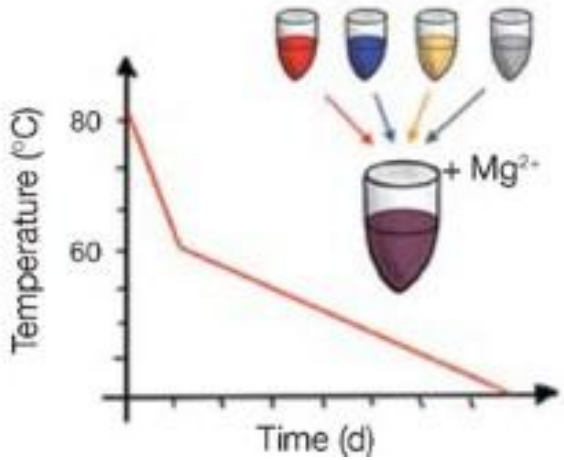
Step 3: prepare scaffold DNA and synthesize staple oligonucleotides



Step 4: pool staple oligonucleotides

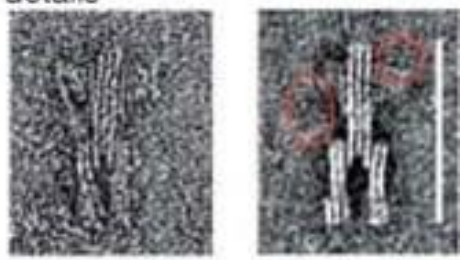


Step 5: run molecular self-assembly reactions



Step 6: analyze folding quality and purify

Step 7: analyze structural details



https://www.nature.com/articles/nmeth.1570

Algorithmic Assembly

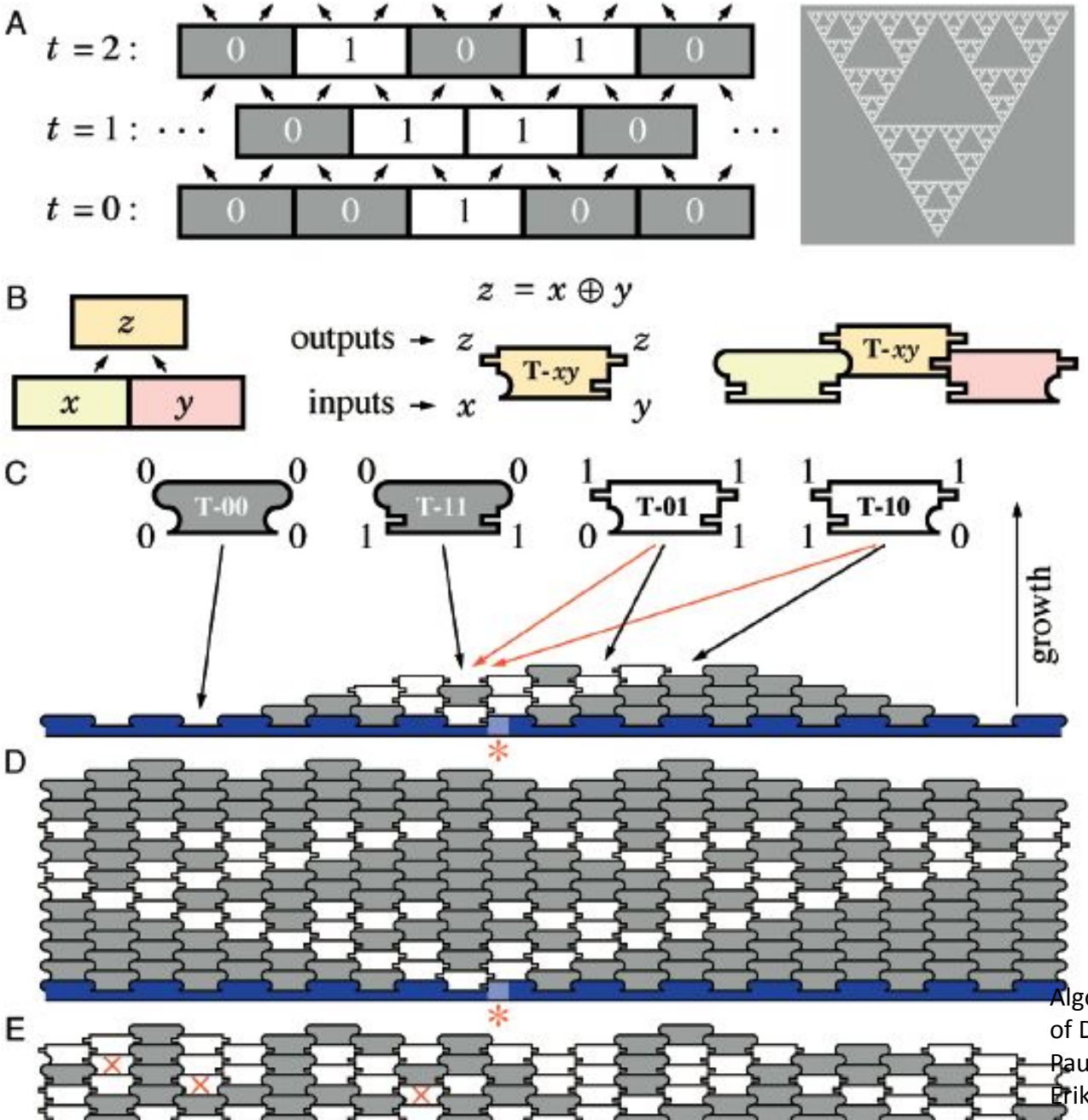


Figure 1. The XOR Cellular Automaton Its Implementation by Tile-Based S Assembly

(A) Left: three time steps of its execu drawn as a space–time history. C update synchronously according XOR by the equation shown. Cell even time steps are interleaved v those at odd time steps; arrows sl propagation of information. Right: Sierpinski triangle.

(B) Translating the space–time hist into a tiling. For each possible in pair, we generate a tile T-xy that b the inputs represented as shapes on lower half of each side and the outpu shapes duplicated on the top half of e side.

(C) The four Sierpinski rule tiles, T T-11, T-01, and T-10, represent the f entries of the truth table for X $0 \oplus 0 = 0$, $1 \oplus 1 = 0$, $0 \oplus 1 = 1$, a $1 \oplus 0 = 1$. Lower binding domains the sides of tiles match input from layer below; upper binding dom provide output to both neighbors the layer above. Semicircular dom represent '0' and rectangular doma '1'. Tiles that output '0' (T-00 and T are gray, and we refer to them as '0' t Tiles that output '1' (T-01 and T-10) white and are referred to as '1' t Initial conditions for the computat are provided by a nucleating struct (blue). Red asterisks indicate sites on nucleating structure that bear a binding domain; elsewhere, sites h all '0' binding domains. Black arr indicate associations that would f two bonds; red arrows, associations would form one bond.

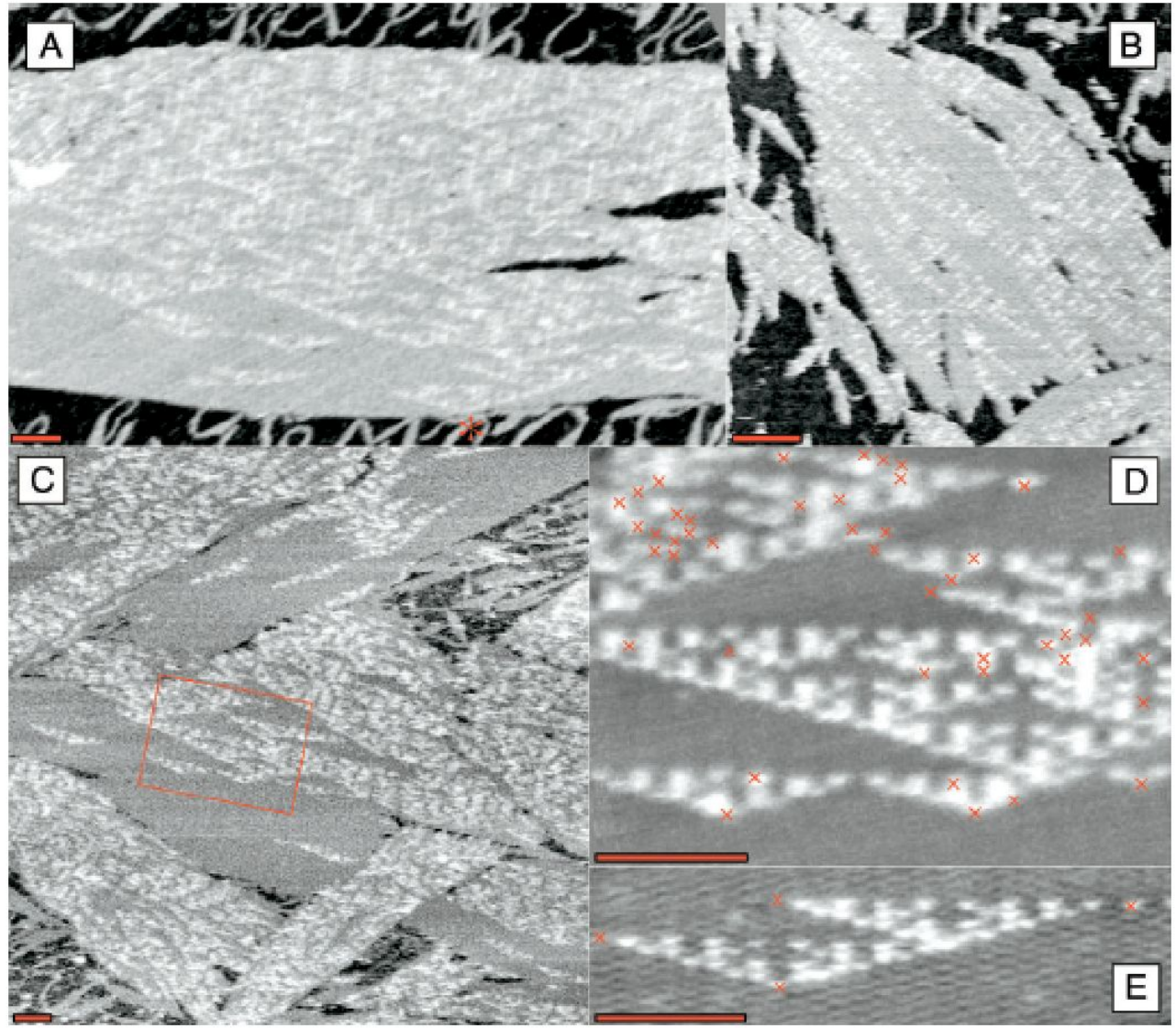
(D) Error-free growth results in Sierpinski pattern.

(E) Error-prone growth from a nucl ing structure with three '1' domains. crosses indicate four mismatch error

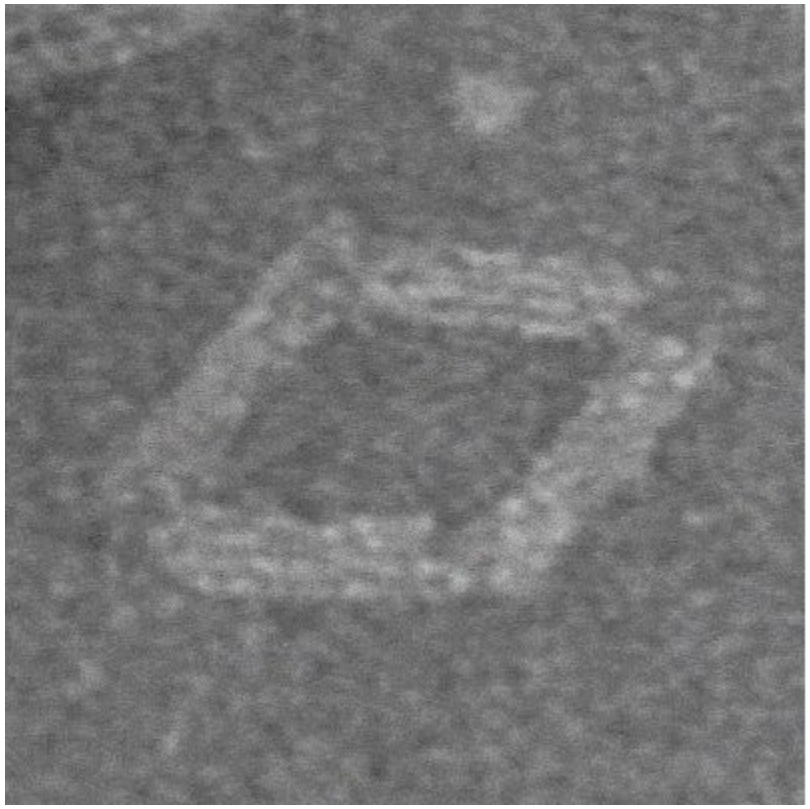
DOI: 10.1371/journal.pbio.0020424.g

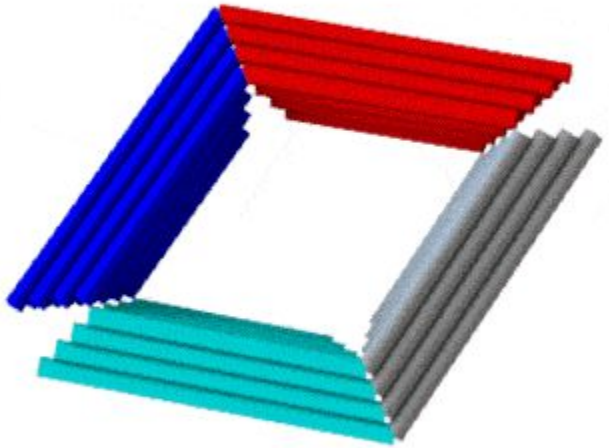
Algorithmic Self-Assembly
of DNA Sierpinski Triangles

Paul W. K. Rothemund1,2, Nick Papadakis2,
Erik Winfree1,2*


A
B
C
D
E

DNA NANO- MACHINES





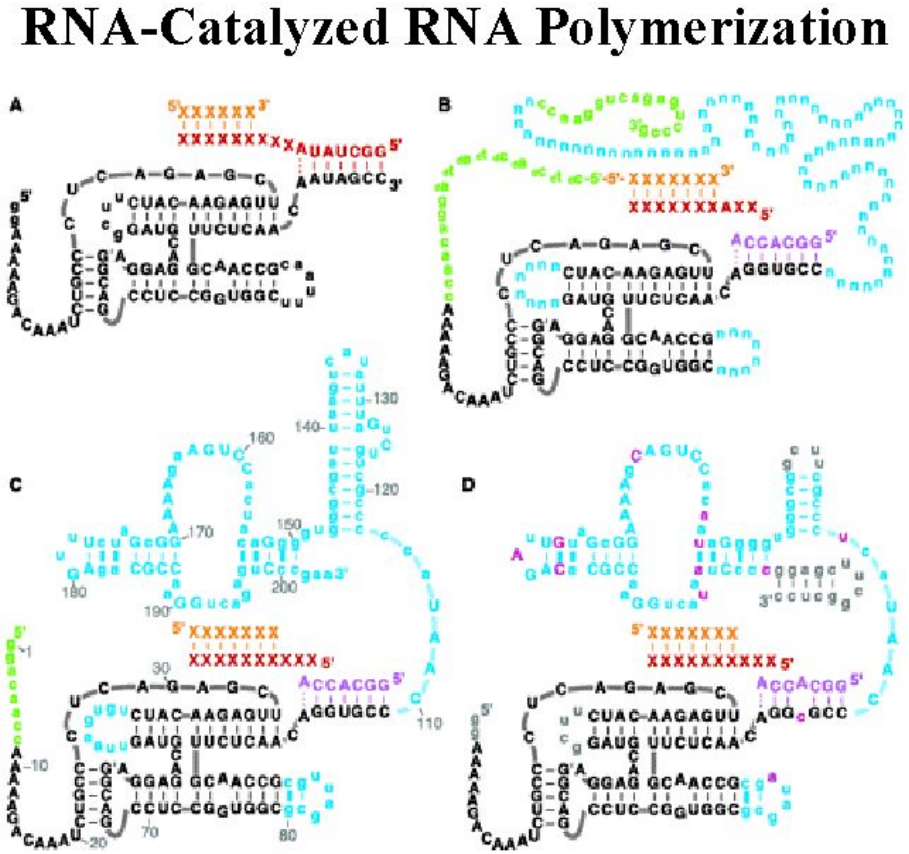
Programmable motion of DNA origami mechanisms

1. Alexander E. Marras, Lifeng Zhou, Hai-Jun Su, and Carlos E. Castro*Department of Mechanical and Aerospace Engineering, The Ohio State University, Columbus, OH 43210*

https://www.pnas.org/content/112/3/713

Information Rich Replication

(Non-Protein Biochemical Systems)

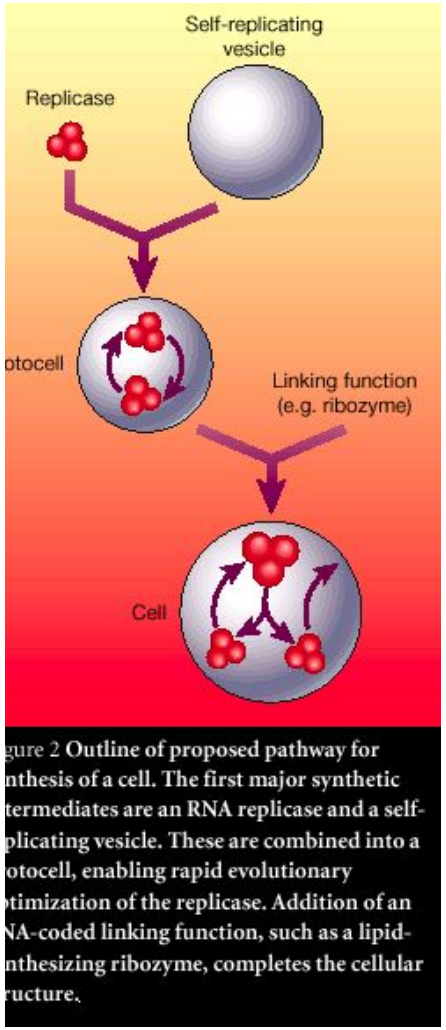


14 base extension. Effective Error Rate: ~ $1:10^3$

RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension

Science 2001 May 18; 292: 1319-1325

Wendy K. **Johnston**, Peter J. Unrau, Michael S. Lawrence, Margaret E. Glasner, and David P. Bartel



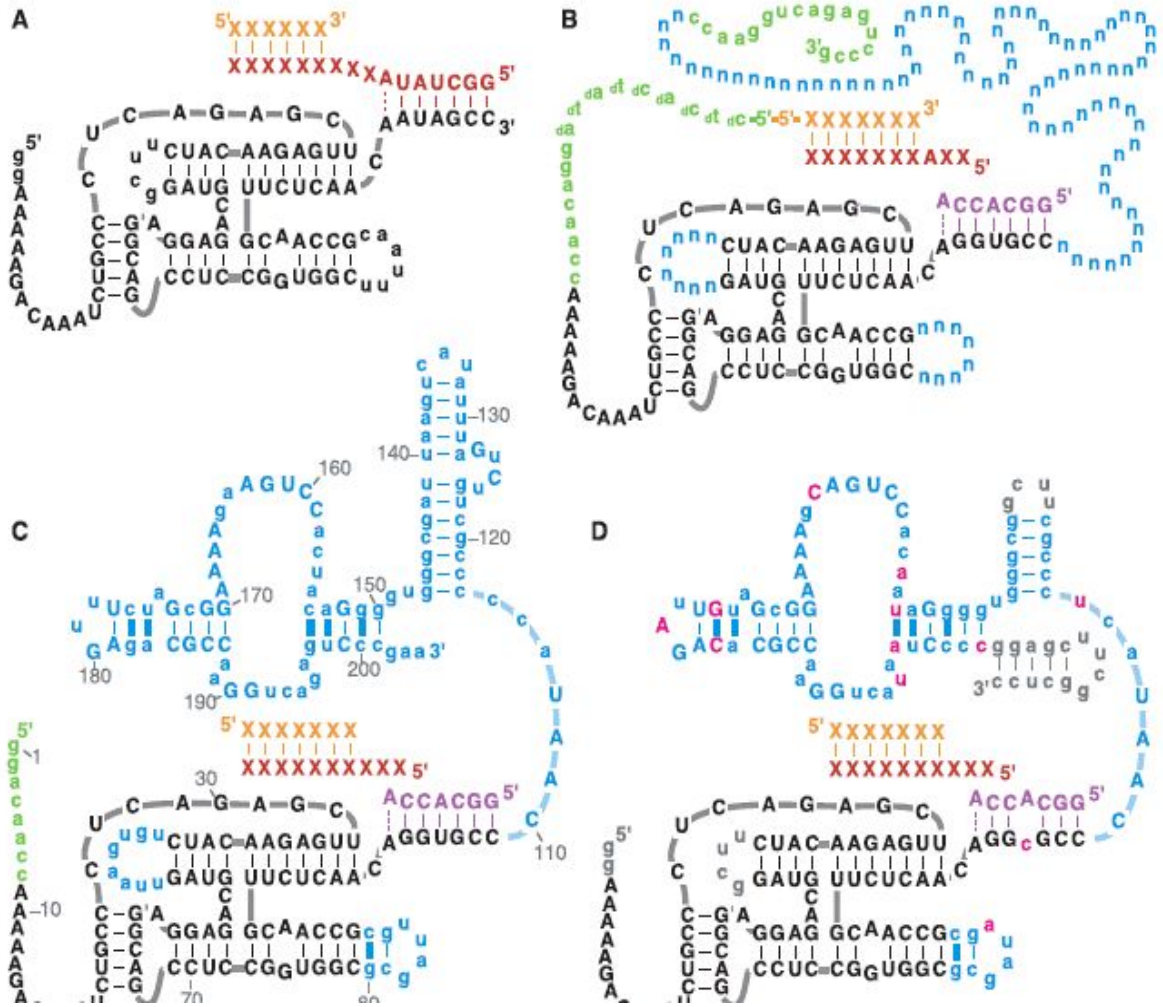
J. Szostak, Nature,409, Jan. 2001

RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension

Wendy K. Johnston, Peter J. Unrau,* Michael S. Lawrence, Margaret E. Glasner, David P. Bartel†

SCIENCE VOL 292 18 MAY 2001

Minimum ribozyme length 165 nt
14 nt extension
Error rate: 1%
Extension rate : 1nt per 153 minutes



Ligase Core Ribozyme

Primer

Template

Random Sequence

UPPERCASE (AGTC)
= Conserved Residues
from Round to Round

Ribozyme-Catalyzed Transcription of an Active Ribozyme

Aniela Wochner, James Attwater, Alan Coulson, Philipp Holliger*

SCIENCE VOL 332 8 APRIL 2011

95 nt base template dependent Extension
24 nt minizyme

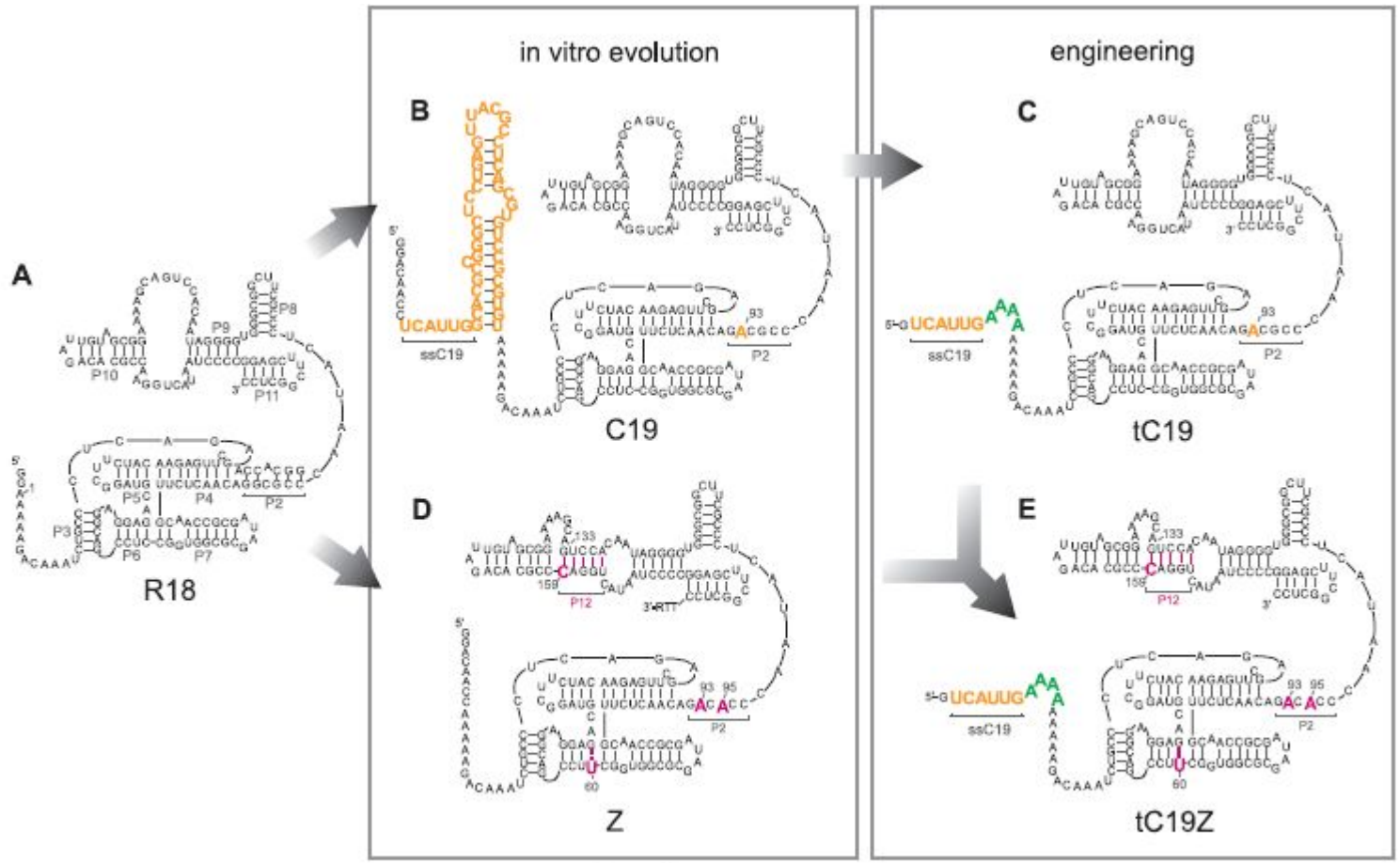
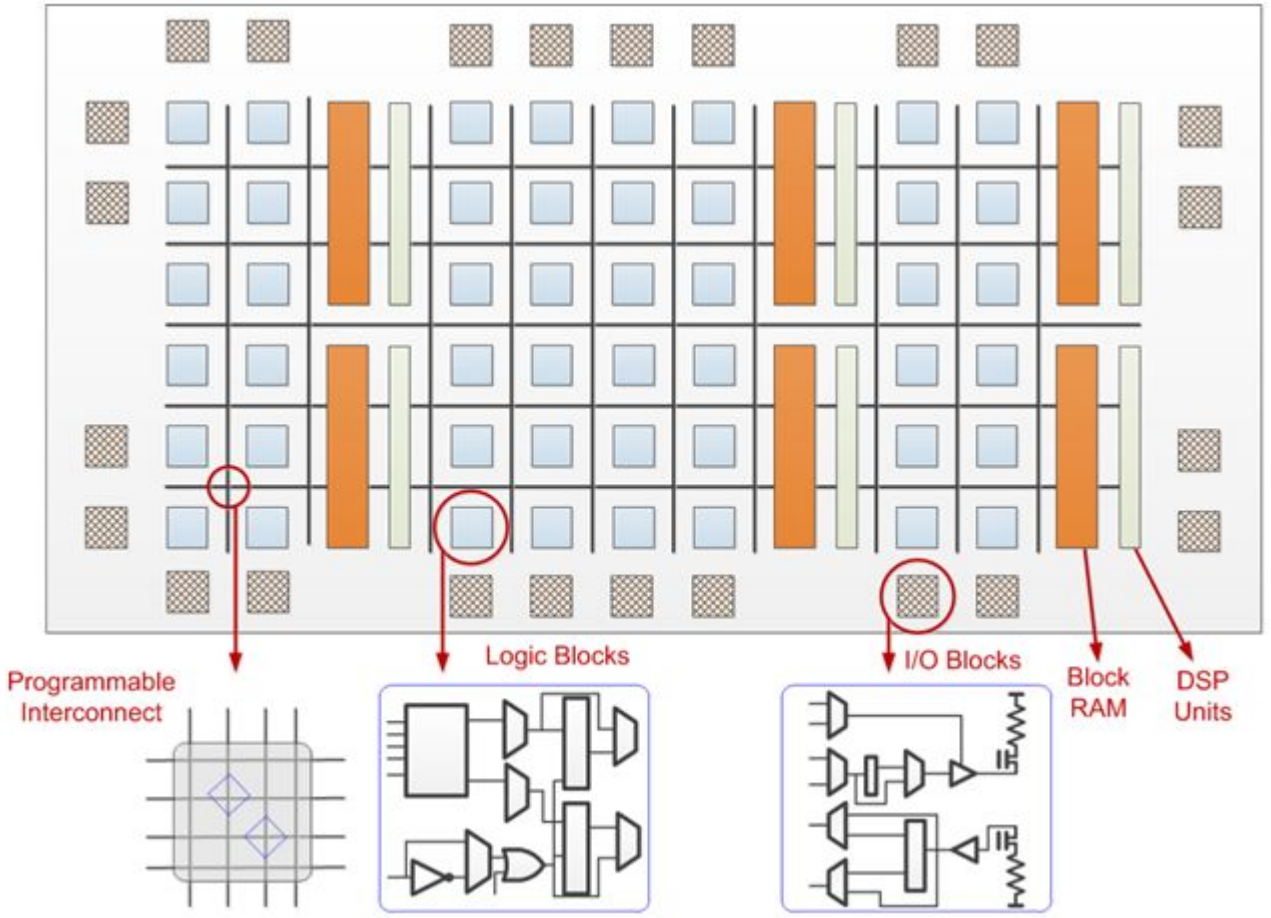


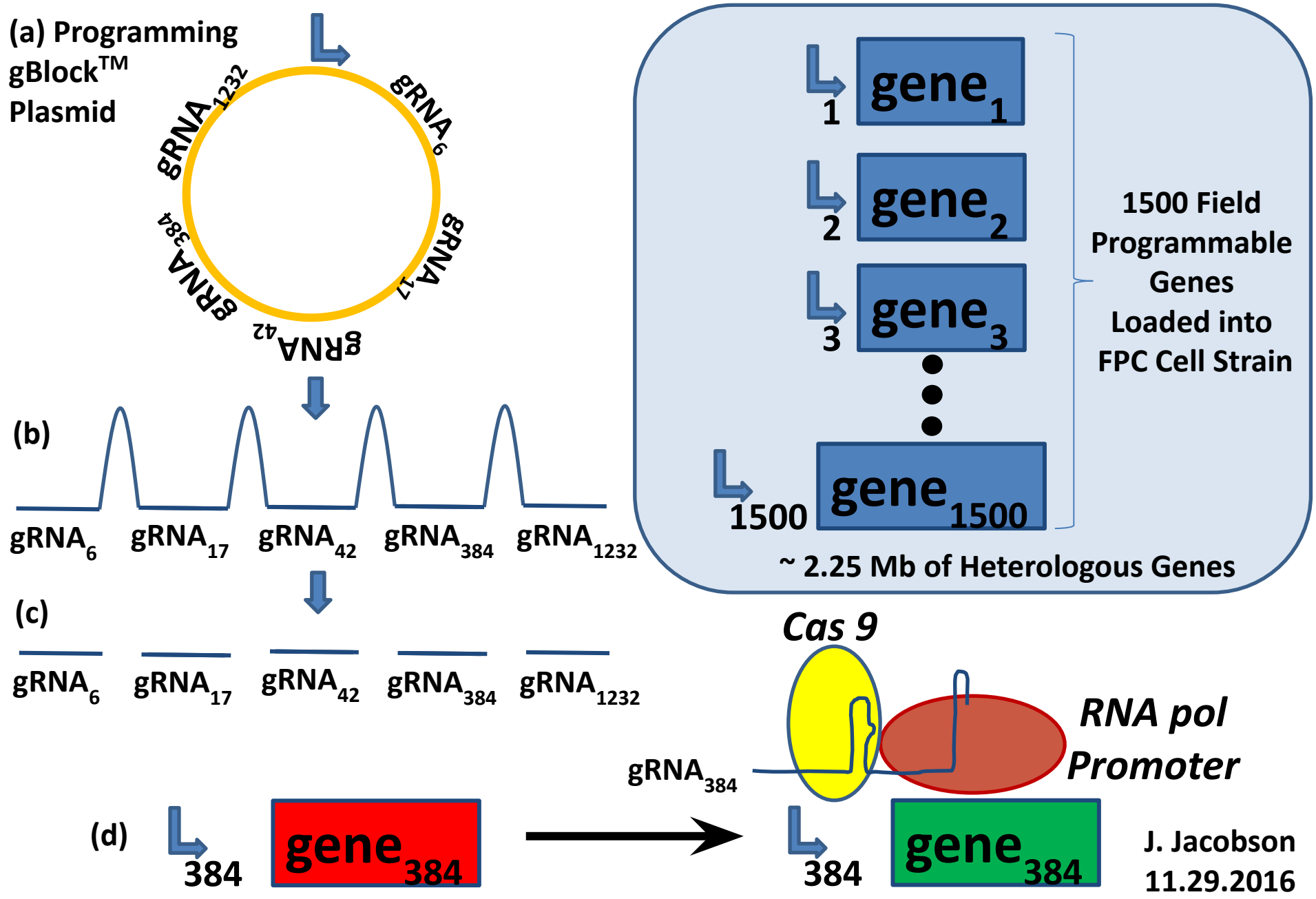
Fig. 1. Evolved and engineered ribozymes. Secondary structures of (**A**) R18 (*7*); (**B**) C19, as predicted by mfold (*27*); (**C**) tC19; and proposed secondary structures for (**D**) Z; and (**E**) tC19Z. Mutations isolated from the Z selection are depicted in magenta, sequences isolated from the C19 selection in orange, and engineered residues in green. The A159C mutation is shown to reshape the processivity domain by stabilizing helix P12. [Run-through transcript, 3′-RTT (table S1).]

FPGA (Field Programmable Gate Array)



Can we make an FPGA for

bioFPGA (Field Programmable Gene Array)



BACKUP

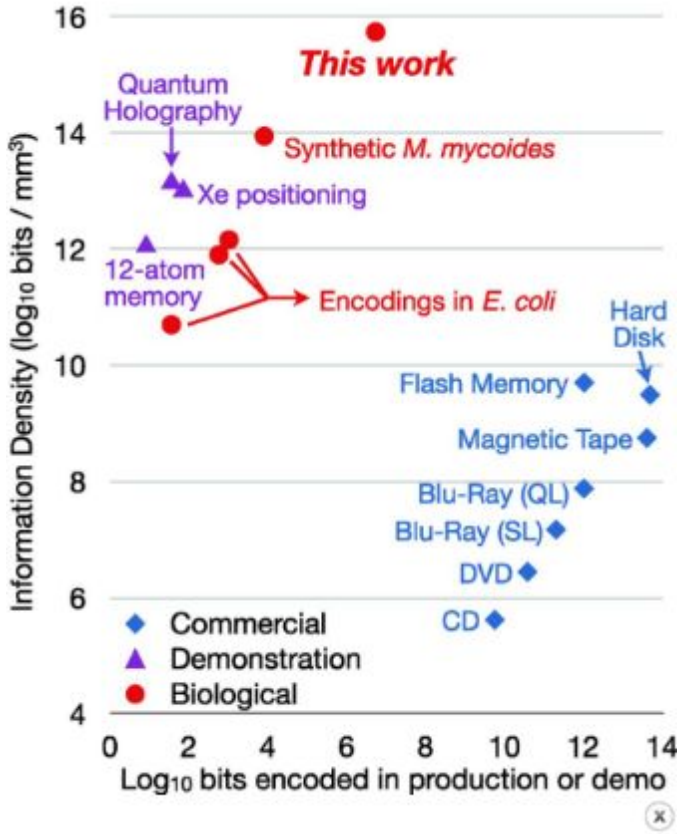
DNA Data Storage

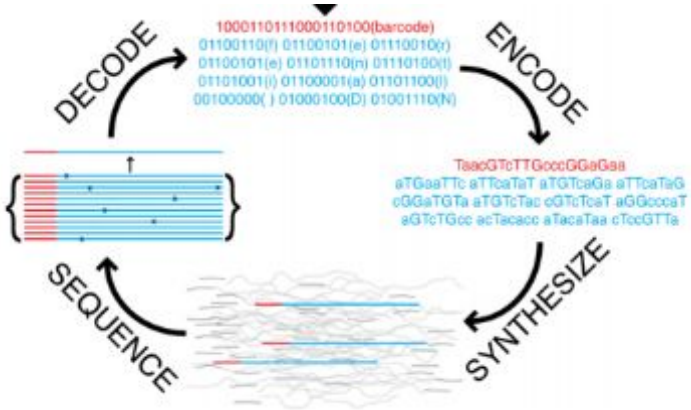
980,000 Sq. ft

=

~ 10 Petabits

$12,400 / MB





Flash: 64GB = $10
DNA Chip: 12 MB
~ $20
10^4x COST

Table S1. Previous DNA Storage works. Bolded lines are included in Figure 1 as representative sequences. Other DNA information works are included here for reference. We do not include works related to DNA computation or associative memory where bits are encoded for purposes other than storage considerations (these works usually encode only modest amounts of information).

Year	Bits Encoded	Ref	Storage Mechanism	Described Usage	Notes
1988	**35**	**(6)**	**Plasmid/E. coli**	**Art**	**Encoded image; Microvenus**
1998	718	see note	Plasmid/E. coli	Art	Text from Bible (Genesis) http://www.ekac.org/geninfo.html
1999	138	(17)	DNA Microdot	Encrypted Message	"JUNE 6 INVASION:NORMANDY"
2001	**561**	**(2)**	**Plasmid/E.coli**	**Archival Storage**	**Lines from Dickens**
2003	1106	(18)	Plasmid/E.coli	Archival Storage	Parts of "It's a Small World"
2005	**1007**	**(19)**	**Plasmid/E.coli**	**Art**	**Encoded poem "Tomten"**
2007	120	(20)	E.coli genome	Archival Storage	"E=mc^2 1905!"; use multiple encodings to correct errors
2008	12	(21)	plasmid	Archival Storage	Use restriction fragment length to encode data
2009	1688	(22)	Plasmid/E.coli	Archival Storage	Text/Music/Image
2010	**7920**	**(7)**	**Mycoplasma genome**	**Watermark**	**Watermarking of synthetic genome**

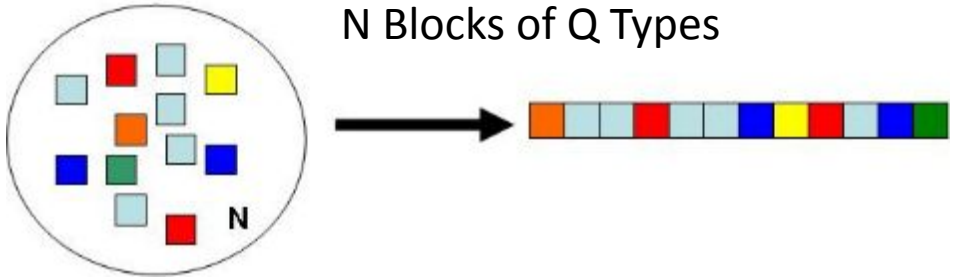
Fabricational Complexity

THE CENTER FOR BITS AND ATOMS
Massachusetts Institute of Technology

Application: Why Are There 20 Amino Acids in Biology?

(What is the right balance between Codon code redundancy and diversity?)

Question: Given N monomeric building blocks of Q different types, what is the optimal number of different types of building blocks Q which maximizes the complexity of the ensemble of all possible constructs?



The complexion for the total number of different ways to arrange N blocks of Q different types (where each type has the same number) is given by:

$$W = \frac{N!}{\prod_i n_i!} = \frac{N!}{(N/Q)!^Q}$$

And the complexity is: $F(N,Q) \approx N\ln(N) - Q^*\left[(N/Q)\ln(N/Q) - N/Q\right]$

For a given polymer length N we can ask which Q* achieves the half max for complexity such that:

$$F(N,Q^*) = 0.5F(N,N)$$