Synthetic Biology: DNA Digital Storage, Computation and the Organic Computer

Alex Widdel

University of Minnesota, Morris

Outline

What is Synthetic Biology?

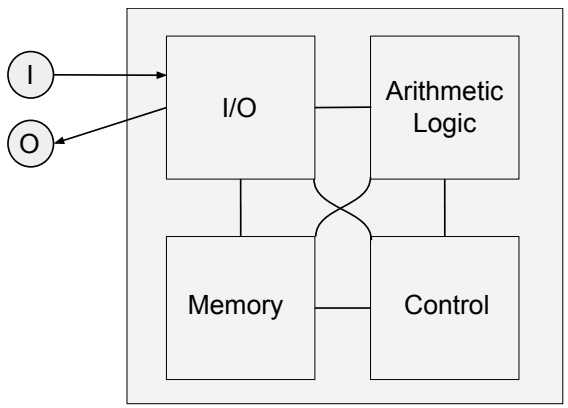
- Biology is viewed as technology.
- One central goal: construction of a universal bio-computer.
- A union of biology, computer science, and engineering.
- The interdisciplinary nature and youth of synthetic biology has led to debate over the term.

Background: Systems Architecture

Systems: Von Neumann Architecture

Four parts:

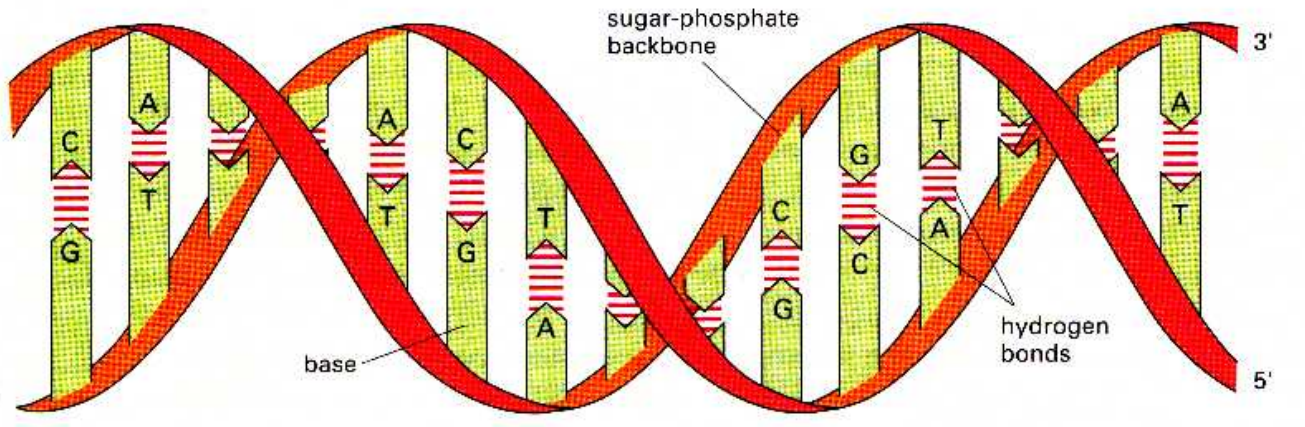
- Memory
- Input/output device (IO)
- Control Unit
- Arithmetic Logic Unit (ALU)



Background: Biology

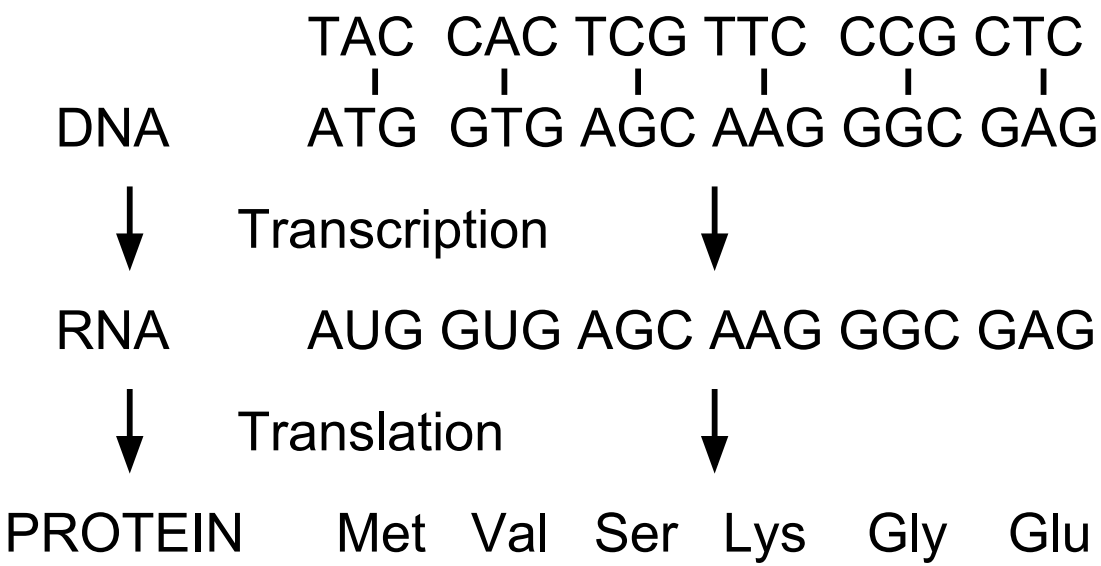
Biology: DNA and Protein Synthesis

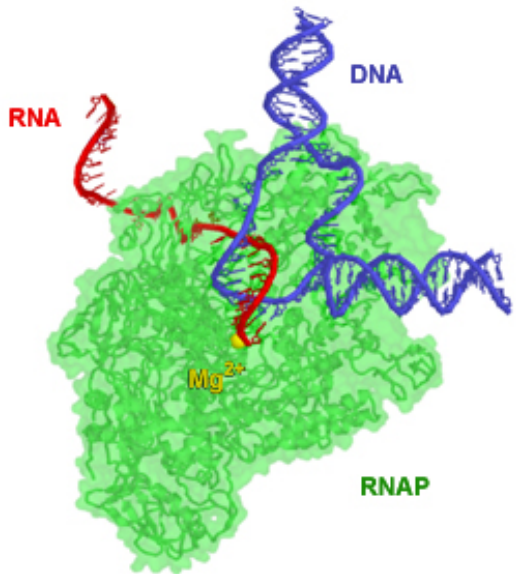
Structure of DNA:



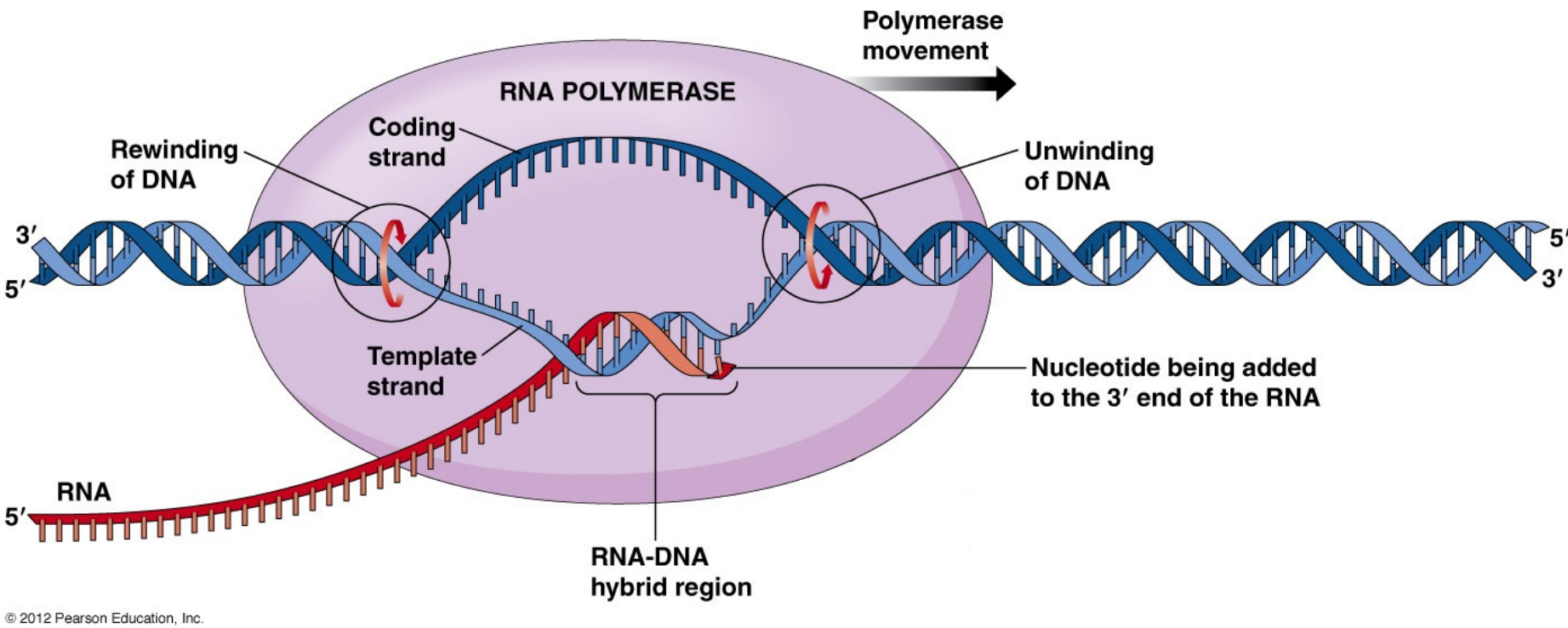
DNA can encode bits in a way that is compatible with the way computers store information. The equivalent to writing bits in a biological system is DNA synthesis, while the equivalent of reading bits is DNA sequencing.

Biological Information Flow: Transcription and Translation





Transcription



DNA Storage: Writing/Reading/Copying Data

Writing: DNA Synthesis

Encoding Data into DNA

- Low-level representation of data is in bit form.
- Two encoding schemes: one bit or two bits per base.

2-bit scheme:

Binary Sequence	Base
00	T
01	G
10	C
11	A

1-bit scheme:

Binary Sequence	Base
1	A or T
0	G or C

Writing: DNA Synthesis

"HELLO WORLD" converted to binary: 01001000 01100101 01101100 01101100 01101111 00100000 01010111 01101111 01110010 01101100 01100100
Binary to DNA:
GTCTGCGGGCATGCATGCAATCTTGGGAGCAAGATCGCATGCG

Empirically achieved storage density of 5.5 petabits per mm^3 (10^8 times higher than the best disk drives), or 700 terabytes per gram. Theoretical maximum of 455 exabytes of raw data.

Unnatural Base Pairs

- Synthesized DNA strands can include nucleotides not found in nature.
- Two artificially-created nucleotides (d5SICSTP and dNaMTP or X and Y for short) have been incorporated into a partially-synthetic E.coli.
- UBP also allow for multiple alternative encoding schemes.

Reading Data: DNA Sequencing

- Processing requires amplification of DNA.
- Sequencing is done by application of fluorescent dyes.
- A laser excites the dyes and a photograph is taken to determine ATCG content.

Copying DNA: Polymerase Chain Reaction (PCR)

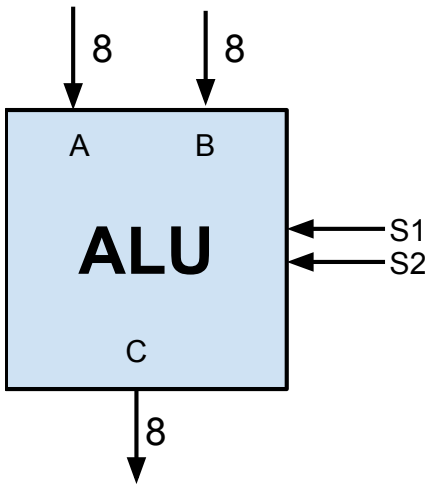
- 1 million copies yielded after 20 cycles, 1 billion after 30.
- Roughly 2 hours processing time required.
- About 1 in 10,000 bases are duplicated incorrectly per cycle.
- Typical desktop HDD has transfer rate up to 1030 Mbit/s.

Building a Biological CPU

The Bio-Computer: Components

Each unit (in a traditional computer) is made up of many circuits, which are turned on (1) or off (0) by switches. Operations on these logic inputs are performed by logic gates to produce a logic output. This makes switches and logic gates the basic atomic units of the ALU and control unit.

Simple Arithmetic Logic Unit



S1	S2	Operation
0	0	C = A XOR B
0	1	C = A + 1
1	0	C = NOT A
1	1	C = 0

A	0	1	1	1	0	0	1	1
B	1	1	1	0	0	1	1	0
A XOR B	1	0	0	1	0	1	0	1

Systems: Von Neumann Bio-Computer

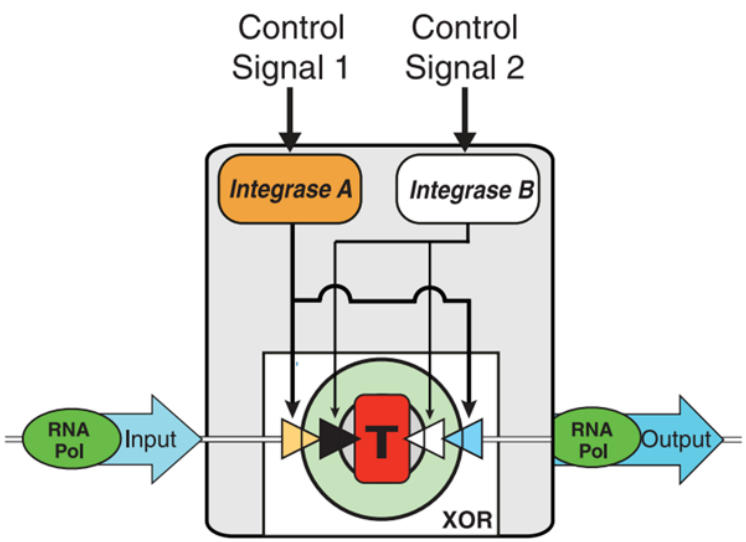
- BioBrick are the basic unit in the three levels of synthetic biology: parts, devices, and systems.
- BioBricks can be considered as active elements generating signals (proteins) when stimulated by a control signal (a protein of a certain shape).
- Structurally, a BioBrick is a specialized DNA strand.
- BioBricks convert inputs to outputs via transcription and translation.

Execution of Logic in Biological Systems

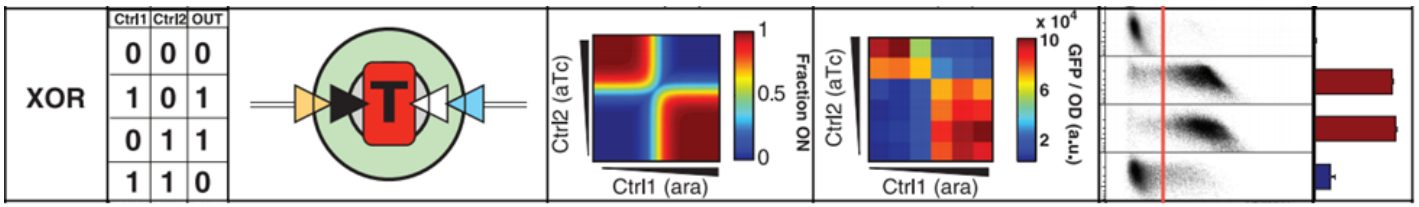
- Key component: transcriptor.
- Gates based off of transcriptors are called Boolean Integrase Logic (BIL) gates.
- Single-layer architecture.
- Flip with the introduction of an enzyme (specialized protein) control signal.

Transcriptor Structure: XOR Gate

- Two nested sets of recombination sites flank a transcription terminator T.
- The terminator exists in either it's un-inverted, transcription-blocking state or an inverted, transcription-allowing state.
- The presence of an integrase flips the terminator.

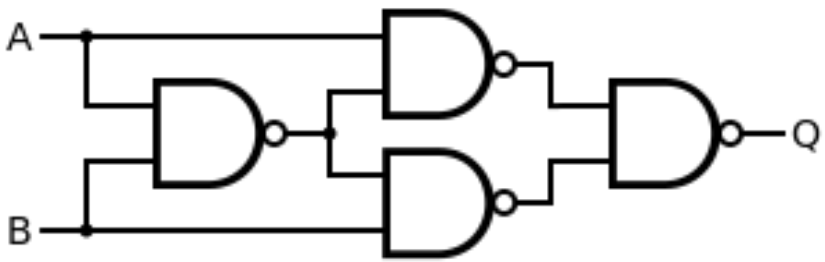


Gate Architecture/Control Signal Thresholds

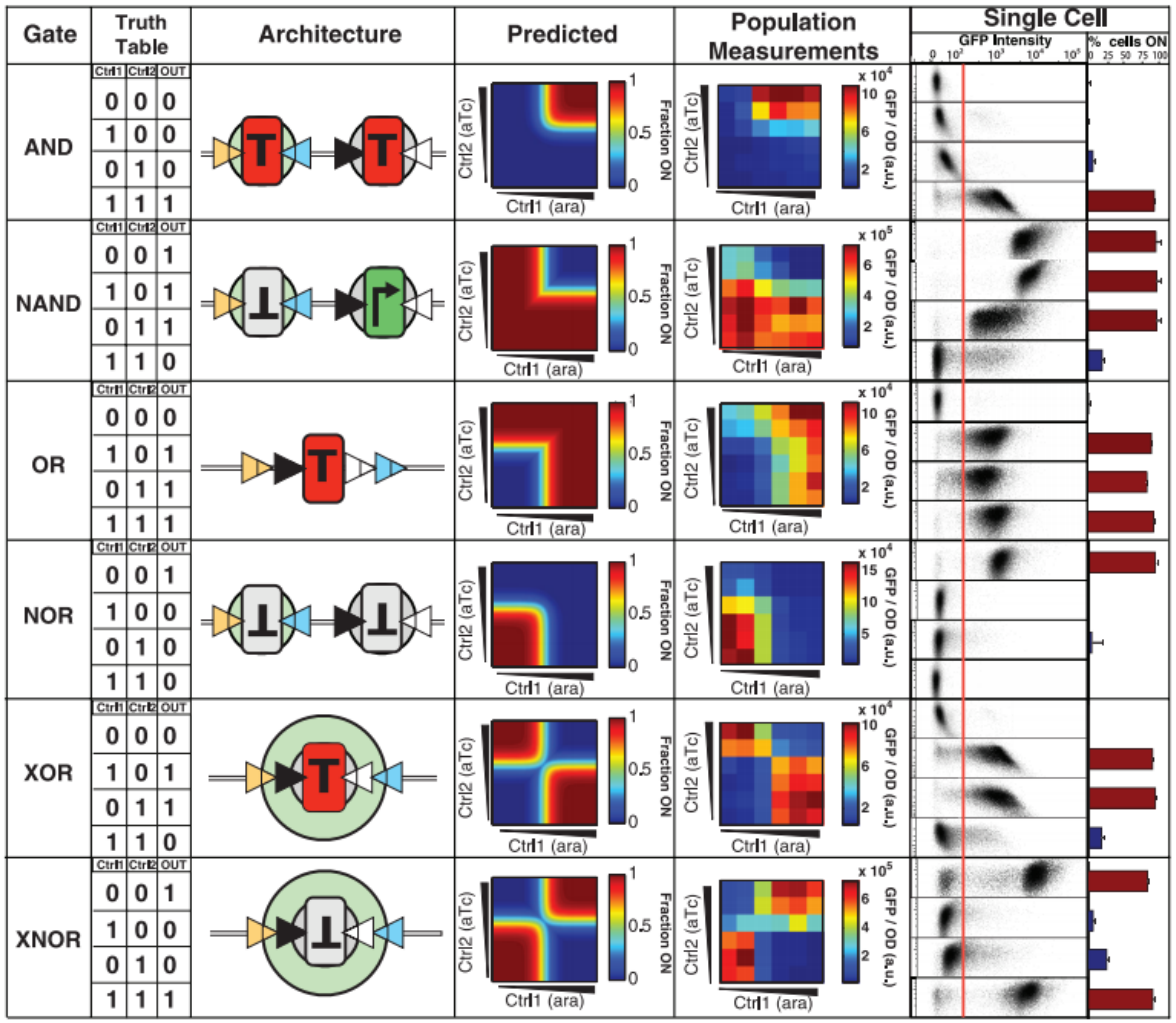


T: Asymmetric Transcription Terminator
Control 1 (ara): arabinose
Control 2 (aTc): anhydrous tetracycline

XOR Gate in Transistor-Transistor Logic



BIL Gates


Gate
Truth Table
Architecture
Predicted
Population Measurements
Single Cell
GFP Intensity
% cells ON
AND
NAND
OR
NOR
XOR
XNOR
Ctrl1 Ctrl2 OUT
0 0 0
1 0 0
0 1 0
1 1 1
Ctrl1 Ctrl2 OUT
0 0 1
1 0 1
0 1 1
1 1 0
Ctrl1 Ctrl2 OUT
0 0 0
1 0 1
0 1 1
1 1 1
Ctrl1 Ctrl2 OUT
0 0 1
1 0 0
0 1 0
1 1 0
Ctrl1 Ctrl2 OUT
0 0 0
1 0 1
0 1 1
1 1 0
Ctrl1 Ctrl2 OUT
0 0 1
1 0 0
0 1 0
1 1 1
Ctrl2 (aTc)
Ctrl1 (ara)
Fraction ON
GFP / OD (a.u.)

Concluding Thoughts

Proof of concept accomplished:

- Digital DNA storage
- Cell-cell Communication
- Transcriptor Logic

The next great challenge is the organization of BioBricks into larger, integrated systems. Existing BioBricks and transcriptors are slow, taking up to an hour to process input and generate an output. However, computing from within living systems can be very valuable.

References I


G. Moe-Behrens.
The Digital Microprocessor, or How to Build a Computer With Biological Parts.
Computational and Structural Biotechnology Journal, 2009.


G. M. Church, Y. Gao, S. Kosuri.
Next-generation Digital Information Storage in DNA.
Science, 337(6102):1628, 2012.


J. Bonnet, P. Yin, M. E. Ortiz, P. Subsoontorn, D. Endy.
Amplifying Genetic Logic Gates.
Science, 340(6132):599-603, 2013.

References II

D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. Foster, C. I. R., F. E. Romesberg.
A Semi-Synthetic Organism With an Expanded Genetic Alphabet.
509(7500):388, 2014.