Bringing Wet & Dry Together…

Computer science tools and models

Deeper Understanding

Reliable Engineering
Outline

• Vision and Motivation
• Proto BioCompiler
• Calibrating Flow Cytometry
• Building EQuIP Models
• Prediction & Validation
Vision: WYSIWYG Synthetic Biology

Bioengineering should be like document preparation:
Why is this important?

• Breaking the complexity barrier:
  - Multiplication of research impact
  - Reduction of barriers to entry

*Sampling of systems in publications with experimental circuits

[Raytheon BBN Technologies]

DNA synthesis

<table>
<thead>
<tr>
<th>Year</th>
<th>Length in base pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>207</td>
</tr>
<tr>
<td>1980</td>
<td>2,100</td>
</tr>
<tr>
<td>1985</td>
<td>2,700</td>
</tr>
<tr>
<td>1990</td>
<td>7,500</td>
</tr>
<tr>
<td>1995</td>
<td>14,600</td>
</tr>
<tr>
<td>2000</td>
<td>32,000</td>
</tr>
<tr>
<td>2005</td>
<td>583,000</td>
</tr>
<tr>
<td>2010</td>
<td>1,080,000</td>
</tr>
</tbody>
</table>

Circuit size


*Max (18 months) vs. Moving average (18 months)

[Sampling & Purnick & Weiss, ‘09]
Why a tool-chain?

Organism Level Description

This gap is too big to cross with a single method!
Collaborators:

Ron Weiss

Douglas Densmore

A high-level program of a system that reacts depending on sensor output

```
(def simple-sensor-actuator ()
  (let ((x (test-sensor)))
    (debug x)
    (debug-2 (not x))))
```

Mammalian Target  E. coli Target

A Tool-Chain Example

Program instantiated for two target platforms

[expr]

Dox

blue not yellow

[expr]

Ara

green not red

Mammalian Target

E. coli Target

A Tool-Chain Example

Abstract genetic regulatory networks

Mammalian Target

E. coli Target

A Tool-Chain Example

Automated part selection using database of known part behaviors

Mammalian Target

E. coli Target

A Tool-Chain Example

Automated assembly step selection for two different platform-specific assembly protocols

Mammalian Target

E. coli Target

A Tool-Chain Example

Resulting cells demonstrating expected behavior

Uninduced

Uninduced

Induced

Induced

Mammalian Target

E. coli Target

If detect explosives:
emit signal
If signal > threshold:
glow red

Outline

• Vision and Motivation
• **Proto BioCompiler**
• Calibrating Flow Cytometry
• Building EQuIP Models
• Prediction & Validation
Focus: BioCompiler

Compilation & Optimization

Organism Level Description

High Level Description

Abstract Genetic Regulatory Network

DNA Parts Sequence

Assembly Instructions

Cells

High level simulator

Coarse chemical simulator

Detailed chemical simulator

Testing

If detect explosives: emit signal
If signal > threshold: glow red

Other tools aiming at high-level design: Cello, Eugene, GEC, GenoCAD, etc.

[Beal, Lu, Weiss, 2011]
Transcriptional Logic Computations

Stabilizes at decay = production
Motif-Based Compilation

- Operators translated to motifs:
Design Optimization

(def sr-latch (s r)
  (letfed+ (o boolean (not (or r o-bar)))
    (o-bar boolean (not (or s o))))
  o))

(green (sr-latch (aTc) (IPTG)))

Unoptimized: 15 functional units, 13 transcription factors
Design Optimization

(def sr-latch (s r)
  (letfeder+ ((o boolean (not (or r o-bar)))
    (o-bar boolean (not (or s o))))
  o))

(green (sr-latch (aTc) (IPTG)))

Final Optimized:
5 functional units
4 transcription factors

Unoptimized: 15 functional units, 13 transcription factors
Complex Example: 4-bit Counter

(4-bit-counter)

Optimized compiler already outperforms human designers

compiled for mammalian
Barriers & Emerging Solutions:

• Barrier: Availability of High-Gain Devices
  – Emerging Solution: combinatorial device libraries based on TALs, ZFs, miRNAs

• Barrier: Characterization of Devices
  – Emerging solution: TASBE characterization method

• Barrier: Predictability of Biological Circuits
  – Emerging solution: EQuiP prediction method
Outline

- Vision and Motivation
- Proto BioCompiler
- Calibrating Flow Cytometry \textit{TASBE Method}
- Building EQuIP Models
- Prediction & Validation

First, some metrology…

Unit mismatch!
How Flow Cytometry Works

Challenges:
- Autofluorescence
- Variation in measurements
- Spectral overlap
- Time Contamination
- Lots of data points!
- Different protein fluorescence
- Individual cells behave (very) differently
Fluorescent Beads $\rightarrow$ Absolute Units

Run beads every time: flow cytometers drift up to 20 percent!
Also can detect instrument problems, mistakes in settings
Compensating for Autofluorescence

Negative control used for this
Compensating for Spectral Overlap

Strong positive control used for each color

Note: only linear when autofluorescence subtracted
Translating Fluorescence to MEFL

- Only FITC channel (e.g. GFP) goes directly
- Others obtained from triple/dual constitutive controls
- Must have exact same constitutive promoter!
- Must have a FITC control protein!
Outline

• Vision and Motivation
• Proto BioCompiler
• Calibrating Flow Cytometry
• Building EQuIIP Models
• Prediction & Validation

[Davidsohn et al., IWBDA, 2013]
TASBE Characterization Method

Transcript cotransfection of 5 plasmids
Calibrated flow cytometry
Analysis by copy-count subpopulations
Multi-plasmid cotransfection!?!?

- Avoids all problems with adjacency, plasmid size, sequence validations

- Variation appears to be independent

Relative Noise Model

- mKate from EBFP2
- EYFP from EBFP2
- EBFP2 from mKate
- EYFP from mKate
- EBFP2 from EYFP
- mKate from EYFP
Result: Input/Output Relations

R1 = TAL14

R1 = TAL21
Expression Dynamics

Fraction Active

Mean Expression

Results $\rightarrow$ division rate, mean expression time, production scaling factor
EQuIP model

Model = first-order discrete-time approximation
Outline

• Vision and Motivation
• Proto BioCompiler
• Calibrating Flow Cytometry
• Building EQuIP Models
• Prediction & Validation

[Davidsohn et al., IWBDA, 2013]
EQuIP Prediction

\[
\text{Regulated Production} \quad \Delta \text{Output} = \int \text{Input(t)} \, dt
\]

\[
\text{Loss} \quad \lambda = \int \Delta \text{Output} \, dt
\]

\[
\text{Regulated Production} \quad \text{Output(t)} = \int \text{Input(t)} \, dt
\]
Incremental Discrete Simulation

[TAL21] → [TAL14] → [TAL14] → [OFP] → [OFP]

State
Production
Loss
Production
Loss
Production
Loss
Production

Time

hour 1

hour 2

hour 46
High Quality Cascade Predictions

1.6x mean error on 1000x range!

TAL14 → TAL21

Circles = EQuIP predictions
Crosses = Experimental Data

<table>
<thead>
<tr>
<th>Cascade</th>
<th>Mean Prediction Error</th>
<th>+/- Induction</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAL14-TAL21</td>
<td>1.20x</td>
<td>2.0x</td>
</tr>
<tr>
<td>TAL14-LmrA</td>
<td>1.81x</td>
<td>1.8x</td>
</tr>
<tr>
<td>TAL21-TAL14</td>
<td>1.30x</td>
<td>1.7x</td>
</tr>
<tr>
<td>TAL21-LmrA</td>
<td>1.56x</td>
<td>2.2x</td>
</tr>
<tr>
<td>LmrA-TAL14</td>
<td>1.75x</td>
<td>1.5x</td>
</tr>
<tr>
<td>LmrA-TAL21</td>
<td>1.74x</td>
<td>1.1x</td>
</tr>
</tbody>
</table>
Summary

Automation supports design and debugging of biological devices, sensors, actuators, circuits
  – BioCompiler automates regulatory network design
  – TASBE method calibrates flow cytometry data
  – Cotransfected test circuits give good models
  – EQuiP accurately predicts cascade behavior from models of individual repressors

Looking forward:
• Real measurements → good engineering
• Bigger, better circuits on more platforms

[multiple manuscripts in preparation]
Acknowledgements:

Raytheon BBN Technologies
Aaron Adler
Joseph Loyall
Rick Schantz
Fusun Yaman

MIT
Ron Weiss
Jonathan Babb
Noah Davidsohn
Ting Lu

Boston University
Douglas Densmore
Evan Appleton
Swapnil Bhatia
Traci Haddock
Chenkai Liu
Viktor Vasilev

DARPA
Characterization Tools Online!

https://synbiotools.bbn.com/

- On first use, you will have to terms of service
- Your data is secure, and can’t be shared on site.
- FireFox recommended; Chrome has an image-display bug.

Register: individual accounts or group account?
Anonymous access also available (but not private)