A Scalable Cellular Logic Technology Using Zinc-Finger Proteins

Christopher Batten, Ronny Krashinsky, Thomas Knight, Jr.

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
June 20, 2004
Synthetic Biology

• Synthetic biology hopes to bring engineering practices common in other engineering disciplines to the field of molecular genetics and thus create a novel nanoscale computational substrate

• Advantages
  – Tightly integrated biological inputs and outputs
  – Easily grow thousands of computational engines
  – Natural use of directed evolution

• Disadvantages
  – Speed is on the order of millihertz (tens of seconds)
  – Modest computational capability of each engine

Synthetic biology is not an attempt to replace silicon computing!
Synthetic Biology Applications

- Autonomous biochemical sensors
- Biomaterial manufacturing
- Programmed therapeutics
- Smart agriculture
- Engineered experimental systems for biologists

M. Elowitz and S. Leibler
A synthetic oscillatory network of transcriptional regulators
_Nature_, January 2000
Outline

• Background
  – Protein expression basics
  – Transcription-based cellular logic
  – Zinc-Finger Proteins (ZFPs)

• Proposed ZFP Logic Technology

• Evaluation
  – Analytical model
  – Simulation results

• Future Work and Conclusions
Protein Expression Basics

- RNA polymerase binds to promoter
- RNAP transcribes gene into messenger RNA
- Ribosome translates messenger RNA into protein
Protein Expression Basics

- RNA polymerase binds to promoter
- RNAP transcribes gene into messenger RNA
- Ribosome translates messenger RNA into protein
Protein Expression Basics

- RNA polymerase (RNAP) binds to promoter
- **RNAP transcribes gene into messenger RNA**
- Ribosome translates messenger RNA into protein
Protein Expression Basics

- RNA polymerase binds to promoter
- RNAP transcribes gene into messenger RNA
- Ribosome translates messenger RNA into protein
Regulation Through Repression

• Repressor proteins can bind to the promoter and block the RNA polymerase from performing transcription
• The DNA site near the promoter recognized by the repressor is called an **operator**
• The target gene can code for another repression protein enabling regulatory cascades
Transcription-Based Inverter

- Protein concentrations are analogous to electrical wires.
- Proteins are not physically isolated, so unique wires require unique proteins.

\[ \begin{array}{c}
1 & \quad 0 \\
& \quad \downarrow \\
& \quad \text{O}
\end{array} \quad \begin{array}{c}
\text{R} \\
\text{X}
\end{array} \quad \begin{array}{c}
0 & \quad 1 \\
& \quad \downarrow \\
& \quad \text{Z}
\end{array} \]
Simple Inverter Model

Chemical Equations

Repressor Binding \( R + O \leftrightarrow RO \quad K_{R+R} = \frac{(O)(R)}{(RO)} \)
Protein Synthesis \( O \rightarrow O + Z \quad k_x \)
Protein Decay \( Z \rightarrow k_{deg} \)

Total Concentration Equations

Total Operator \( (O_T) = (O) + (RO) \)
Total Repressor \( (R_T) = (R) + (RO) \approx (R) \quad if \quad (R_T) >> (O) \)

Transfer Function Derivation

\[
\frac{(O)}{(O_T)} = \frac{(O)}{(O) + (RO)} = \frac{1}{1 + (RO)/(O)} = \frac{1}{1 + (R)/K_{R+R}}
\]

\[
\frac{d(Z)}{dt} = k_x \cdot (O) - k_{deg} \cdot (Z) = 0 \quad at \ equilibrium
\]

\[
(Z) = \frac{k_x}{k_{deg}} \cdot \frac{(O_T)}{1 + (R)/K_{R+R}}
\]
Simple Inverter Model

Chemical Equations

Repressor Binding \( R + O \leftrightarrow RO \) \( K_{R+R} = \frac{(O)(R)}{(RO)} \)

Protein Synthesis \( O \rightarrow O + Z \) \( k_x \)

Protein Decay \( Z \rightarrow k_{\text{deg}} \)

Total Concentration Equations

Total Operator \( (O_T) = (O) + (RO) \)

Total Repressor \( (R_T) = (R) + (RO) \approx (R) \) if \( (R_T) >> (O) \)

Transfer Function Derivation

\[
\frac{(O)}{(O_T)} = \frac{(O)}{(O) + (RO)} = \frac{1}{1 + \frac{(RO)}{(O)}} = \frac{1}{1 + \frac{(R)}{K_{R+R}}}
\]

\[
\frac{\text{d}(Z)}{\text{d}t} = k_x \cdot (O) - k_{\text{deg}} \cdot (Z) = 0 \text{ at equilibrium}
\]

\[
(Z) = \frac{k_x}{k_{\text{deg}}} \cdot \frac{(O_T)}{1 + \frac{(R)}{K_{R+R}}}
\]
Cooperativity

- Cooperative DNA binding is where the binding of one protein increases the likelihood of a second protein binding
- Cooperativity adds more non-linearity to the system
  - Increases switching sensitivity
  - Improves robustness to noise
Cooperative Inverter Model

**Chemical Equations**

Coop Binding: \( R + R + O \Leftrightarrow R_2O \)  \( K_{R2O} = \frac{(O)(R)^2}{(R_2O)} \)

Protein Synthesis: \( O \rightarrow O + Z \)  \( k_x \)

Protein Decay: \( Z \rightarrow \)  \( k_{deg} \)

**Total Concentration Equations**

Total Operator: \( (O_T) = (O) + (R_2O) \)

Total Repressor: \( (R_T) = (R) + 2\cdot(R_2O) \approx (R) \)  \( if \ (R_T) \gg (O) \)

**Transfer Function Derivation**

\[
\frac{(O)}{(O_T)} = \frac{(O)}{(O) + (RO)} = \frac{1}{1 + (RO)/(O)} = \frac{1}{1 + (R)^2/K_{R20}}
\]

\[
\frac{d(Z)}{dt} = k_x \cdot (O) - k_{deg} \cdot (Z) = 0 \ at \ equilibrium
\]

\[
(Z) = \frac{k_x}{k_{deg}} \cdot (O) = \frac{k_x}{k_{deg}} \cdot \frac{(O)}{1 + (R)^2/K_{R20}}
\]

Cooperative Non-Linearity
Cooperative Inverter Model

Chemical Equations

Coop Binding \[ R + R + O \leftrightarrow R_2O \]

Protein Synthesis \[ O \rightarrow O + Z \]

Protein Decay \[ Z \rightarrow k_{\text{deg}} \]

Total Concentration Equations

Total Operator \[ (O_T) = (O) + (R_2O) \]

Total Repressor \[ (R_T) = (R) + 2\cdot(R_2O) \approx (R) \text{ if } (R_T) >> (O) \]

Transfer Function Derivation

\[ \frac{(O)}{(O_T)} = \frac{(O)}{(O) + (RO)} = \frac{1}{1 + (RO)/(O)} = \frac{1}{1 + (R)^2/K_{R2O}} \]

\[ \frac{d(Z)}{dt} = k_x \cdot (O) - k_{\text{deg}} \cdot (Z) = 0 \text{ at equilibrium} \]

\[ (Z) = \frac{k_x}{k_{\text{deg}}} \cdot \frac{(O_T)}{(O)} = \frac{k_x}{k_{\text{deg}}} \cdot \frac{1}{1 + (R)^2/K_{R2O}} \]

Cooperative Non-Linearity
Cellular Logic Summary

- Current systems are limited to less than a dozen gates
  - Three inverter ring oscillator [Elowitz00]
  - RS latch [Gardner00]
  - Inter-cell communication [Weiss01]

- A natural repressor-based logic technology presents serious scalability issues
  - Scavenging natural repressor proteins is time consuming
  - Matching natural repressor proteins to work together is difficult

- Sophisticated synthetic biological systems require a scalable cellular logic technology with good cooperativity
  - Zinc-finger proteins can be engineered to create many unique proteins relatively easily
  - Zinc-finger proteins can be fused with dimerization domains to increase cooperativity
  - A cellular logic technology of only zinc-finger proteins should hopefully be easier to characterize
Single Zinc-Finger Structure

- Alpha Helix
- DNA Three Base Recognition Region
- Two Beta Sheets
- Zinc Atom
Poly-Finger ZFPs

A.C. Jamieson, J.C. Miller, and C.O. Pabo.
Drug discovery with engineered zinc-finger proteins.
_Nature Reviews Drug Discovery_, May 2003
Engineering ZFPs

- Early hopes for a code to simply map amino-acid residues to DNA bases have not materialized [Choo94]
- Some success has been had engineering ZFP fingers to recognize GNNG sequences [Dreier00, Segal99]
- These GNNG fingers can then be easily composed into poly-finger ZFPs
- Recent work has broadened these techniques to include ANNA fingers [Dreier01]

We are nearing the point where an appropriate poly-finger ZFP can be easily composed from a library of fingers to recognize almost any DNA sequence.
Engineering ZFP Dimers

- Dimerization is the natural phenomenon where two proteins bind together
- Dimerization is a form of cooperative DNA binding and increases cooperativity
- Two-finger ZFPs have been fused to GCN4 leucine zipper dimerization domains to create cooperative ZFP DNA binding proteins [Wolfe00]
Proposed ZFP Logic Technology

• Use two-finger ZFPs fused to a GCN4 leucine zipper as basic repressor monomer
• Each gate/wire has a unique engineered ZFP
• Why two-finger monomers?
  – Recognizes 6 base pairs permitting an encoding space suitable for hundreds of gates
  – Specificity suitable for *E. coli* genome
  – Affinity suitable for biologic circuit dynamics
• Since all gates have identical leucine zipper dimerization domains, monomers from different gates could dimerize causing **inter-gate interference**
Proposed ZFP Logic Technology

- Leucine Zipper
- A1
- A2
- ZFP
- ZFP

Pr

ZFRP Gene A

TTGACA

-35

N17

TATAAT

-10

N5-7

ZFRP Gene Z

Leucine Zipper

Z1

ZFP

ZFP

Z2
Proposed ZFP Logic Technology

Leucine Zipper

Dimerization

ZFP Operator (12 Bases)

ZFRP Gene A

ZFRP Gene Z

Pr
Proposed ZFP Logic Technology

Leucine Zipper

Dimerization with Interference Protein

Interference From Other Gates

ZFP Operator (12 Bases)

Dimerization

Pr

ZFRP Gene A

ZFRP Gene Z
**Analytical Model**

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Equilibrium Constant</th>
<th>Equation</th>
<th>Remote Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dimerization</strong></td>
<td>( K_{R+R} )</td>
<td>( \frac{(R)^2}{(R_2)} )</td>
<td>( e^{\frac{E_{\text{dim}}}{RT}} )</td>
</tr>
<tr>
<td><strong>Dimer Binding</strong></td>
<td>( K_{R_2+O} )</td>
<td>( \frac{(O)(R_2)}{(R_2O)} )</td>
<td>( e^{\frac{2E_{\text{op}}}{RT}} )</td>
</tr>
<tr>
<td><strong>Monomer Binding</strong></td>
<td>( K_{R+R} )</td>
<td>( \frac{(O)(R)}{(OR)} )</td>
<td>( e^{\frac{E_{\text{op}}}{RT}} )</td>
</tr>
<tr>
<td><strong>Monomer Binding</strong></td>
<td>( K_{R+R} )</td>
<td>( \frac{(O)(R)}{(RO)} )</td>
<td>( e^{\frac{E_{\text{op}}}{RT}} )</td>
</tr>
<tr>
<td><strong>Cooperative Binding</strong></td>
<td>( K_{OR+R} )</td>
<td>( \frac{(OR)(R)}{(R_2O)} )</td>
<td>( e^{\frac{E_{\text{op}+E_{\text{dim}}}}{RT}} )</td>
</tr>
<tr>
<td><strong>Cooperative Binding</strong></td>
<td>( K_{RO+R} )</td>
<td>( \frac{(RO)(R)}{(R_2O)} )</td>
<td>( e^{\frac{E_{\text{op}+E_{\text{dim}}}}{RT}} )</td>
</tr>
<tr>
<td><strong>Protein Synthesis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Protein Decay</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dimerization</strong></td>
<td>( K_{X+X} )</td>
<td>( \frac{(X)^2}{(X_2)} )</td>
<td>( e^{\frac{E_{\text{dim}}}{RT}} )</td>
</tr>
<tr>
<td><strong>Inter-Gate Interference</strong></td>
<td>( K_{X+R} )</td>
<td>( \frac{(X)(R)}{(XR)} )</td>
<td>( e^{\frac{E_{\text{dim}}}{RT}} )</td>
</tr>
</tbody>
</table>

\( K \): Equilibrium dissociation constant  
\( k \): Dynamic rate constant  
\( E \): Binding energy or change in potential energy caused by the reaction  
  More negative \( E \) means the reaction is more likely to occur
Dimerization and Operator Energy

Leucine Zipper

Interference From Other Gates

$E_{\text{dim}}$

$E_{\text{op}}$

ZFRP Gene A

ZFRP Gene Z
Percent Operator Bound

- For **very low dimerization energies**, system approaches uncooperative repressor monomer system
- For **very high dimerization energies**, system approaches uncooperative covalently bonded repressor system
- For **moderate dimerization energies**, the system is cooperative i.e. the slope of the curve is steeper than for the uncooperative systems
Cooperativity

\[ E_{op} = [-1: -15] \text{ kcal} \]
Inter-Gate Interference

\[ E_{\text{dim}} = -2 \text{ kcal}, K_{\text{dim}} = 3.6 \times 10^{-2} \text{ M} \]

\[ E_{\text{dim}} = -6 \text{ kcal}, K_{\text{dim}} = 4.5 \times 10^{-5} \text{ M} \]

\[ E_{\text{dim}} = -4 \text{ kcal}, K_{\text{dim}} = 1.3 \times 10^{-3} \text{ M} \]

\[ E_{\text{dim}} = -8 \text{ kcal}, K_{\text{dim}} = 1.6 \times 10^{-6} \text{ M} \]

Operator % Bound vs. Repressor Concentration (M)
Desired Dimerization Energy

• Tradeoffs in setting the dimerization energy
  – Stronger dimerization energy increases cooperativity
  – Stronger dimerization energy increases inter-gate interference

We desire the weakest dimerization energy which still achieves the maximum cooperativity

![Graph showing cooperativity and repressor concentration](image-url)
Transfer Curve and Interference

- No Interference
- $X_T = 1 \times 10^{-7}$ M
- $X_T = 5 \times 10^{-4}$ M
- $X_T = 1 \times 10^{-3}$ M

Output Protein Concentration (M)

Repressor Concentration (Rt) (M)
Transfer Curve and Interference

Max output protein concentration per gate is $5 \times 10^{-7}$ M

Inter-gate interference must be below $10^{-4}$ M
Transfer Curve and Interference

Max output protein concentration per gate is $5 \times 10^{-7}$ M

To first order, could have $\frac{10^{-4}}{5 \times 10^{-7}} \approx 200$ gates

Inter-gate interference must be below $10^{-4}$ M
Transfer Curve and Cooperativity
Future Work

• Model and Design Improvements
  – Model system transient response
  – Model stochastic effects
  – Design a system with increased cooperativity

• Implementation
  – Simple test circuits to investigate use of two finger ZFP dimer as a cooperative repressor in *E. coli*
  – Engineered zinc-finger system with heterodimers to implement more complex logic gates
Conclusions

• Current natural repressor-based biological circuits are limited to less than a dozen gates
• A cellular logic technology based on zinc-finger proteins should enable hundreds of gates
• Careful engineering of the dimerization energy can help mitigate inter-gate interference without sacrificing cooperativity