D. Combinatorial transcriptional control

Complex bacterial promoters with multiple inputs:

[Diagram showing complex bacterial promoters with multiple inputs]

Complex eukaryotic promoter

[Diagram showing complex eukaryotic promoter with multiple inputs]
Mechanisms of complex transcriptional control?

• specific, complex protein-protein interaction
  … but different TFs can work together to implement different functions
  ⇒ combinatorial control favors simpler, less specific interaction

• alternative: regulated recruitment  [Ptashne & Gann ’97]
  – simple, glue-like interaction between TFs/RNAP
  – arrange DNA binding sites/strengths to accomplish desired control functions
  … but how to implement? and what are the limitations?

Theory of Combinatorial Transcription Control

[Buchler et al, PNAS ’03]
Theory of Combinatorial Transcription Control

[Buchler et al, PNAS ’03]

quantitative description:
• occupation of site \( j \): \( \sigma_j = \{0,1\} \)
• TF-operator interaction: \( K_j = \exp(-\beta G_j) = 1 \sim 1000 \) nM
• TF-TF and TF-RNAP interaction:
  
  - regulated recruitment [M. Ptashne]
  
  - long-distance interaction possible via DNA looping
• promoter activity \( \approx \) equilibrium promoter occupation prob \( P(\{\sigma_j\}) = \{\sigma_p\} \)

\[ Z = \sum_{\{\sigma\}} \prod_{j=1}^N \left( \frac{[TF_j]/K_j}{\sigma_j} \right)^{\sigma_j} \prod_{i<j} \omega_{i,j}^{\sigma_i \sigma_j} \]

\[ \approx \text{programmable molecular Boltzmann machine!} \]

What kind of control functions \( P(\{\sigma_j\}) \) are implementable via the appropriate choices of \( \{K_j, \omega_{i,j}\} \)?

1. non-interacting

simple activation:
\[ W_{on} = 1 + \frac{[A]}{K_A} \]
\[ W_{off} = \frac{[RNAP]}{K_p} + \omega_{A-p} \cdot \frac{[A]}{K_A} \cdot \frac{[RNAP]}{K_p} \]
\[ P = \frac{W_{on}}{W_{off}} = \frac{[RNAP]}{K_p} \cdot \frac{1 + \omega_{A-p} [A]/K_A}{1 + [A]/K_A} \]
(for typical weak promoters)

simple repression:
\[ P = \frac{[RNAP]}{K_p} \cdot \frac{1}{1 + [R]/K_R} \]

co-regulation multiplicative
\[ P = \frac{1 + \omega_{A-p} [A]/K_A}{1 + [A]/K_A} \cdot \frac{1}{1 + [R]/K_R} \]

[example: lac promoter (details later)]
2. Synergistic activation

RNAp can simultaneously contact two TFs (e.g., Crp and Fnr)

statistical weight $W$ for each configuration $\{\sigma_A, \sigma_B, \sigma_p\}$, with $q_X = [X]/K_X$

$$W_{\text{off}} = \begin{cases} 
W(0,0,0) = 1 \\
W(1,0,0) = q_A \\
W(0,1,0) = q_B \\
W(1,1,0) = q_A \cdot q_B \\
W(0,0,1) = q_p \\
W(1,0,1) = \omega_A \cdot q_p \\
W(0,1,1) = \omega_B \cdot q_p \\
W(1,1,1) = \omega_A \cdot q_A \cdot q_p 
\end{cases}$$

consider $\omega = \omega_A \cdot \omega_B$

$$\mathcal{P}(\{A\},\{B\}) = \frac{W}{W_{\text{off}}} = \frac{1 + \omega_A q_A \cdot 1 + \omega_B q_B}{1 + q_A + q_B}$$

3. Competitive (or “independent”) activation

$$\mathcal{P}(\{A\},\{B\}) = \frac{1}{1 + q_A + q_B}$$

4. Cooperative activation

could work as AND or OR by choosing $K_A$, $K_B$

$$\mathcal{P}(\{A\},\{B\}) = \frac{1 + \omega_A q_A \cdot 1 + \omega_B q_B}{1 + q_A + q_B}$$
5. Cooperative repression

![Diagram of cooperative repression]

- **Promoter** $O_A$, $O_B$
- **RNAp**
- **Note:** Direct interaction between A and B not necessary (cf. "collaborative competition")
- **Logic:** $\approx$ NAND

6. Competitive repression

![Diagram of competitive repression]

- **Promoter**
- **Logic:** $\approx$ NOR

---

More complex control function, e.g., XOR?

<table>
<thead>
<tr>
<th>A/B</th>
<th>AND</th>
<th>OR</th>
<th>NAND</th>
<th>XOR</th>
<th>EQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>lo/lo</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td>lo/hi</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
</tr>
<tr>
<td>hi/lo</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
</tr>
<tr>
<td>hi/hi</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
</tr>
</tbody>
</table>

Cannot be implemented by overlapping A and B sites

![Diagram of XOR function]

**XOR**$(A,B) = (A \text{ OR } B) \text{ AND NOT}(A \text{ AND } B)$

**Gene cascade**

- **Problems:**
  - Need a gene for each intermediate result
  - Multiple rounds of gene expression: noise + delay
  - Synchronization difficult
  - Amplification nontrivial

![Diagram of gene cascade]
More complex control function, e.g., **XOR**?

<table>
<thead>
<tr>
<th>A/B</th>
<th>AND</th>
<th>OR</th>
<th>NAND</th>
<th>XOR</th>
<th>EQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>lo/lo</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td>lo/hi</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
</tr>
<tr>
<td>hi/lo</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
</tr>
<tr>
<td>hi/hi</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
</tr>
</tbody>
</table>

**XOR**(A,B) = (A OR B) AND NOT(A AND B)

**Problem:**
- Need dedicated component
- Lose combinatorial control
  - E.g., can’t implement **AND** elsewhere

Allosteric or co-factor mediated

[Diagram showing allosteric or co-factor mediated control]

---

More complex control function, e.g., **XOR**?

<table>
<thead>
<tr>
<th>A/B</th>
<th>AND</th>
<th>OR</th>
<th>NAND</th>
<th>XOR</th>
<th>EQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>lo/lo</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td>lo/hi</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
</tr>
<tr>
<td>hi/lo</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
</tr>
<tr>
<td>hi/hi</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
</tr>
</tbody>
</table>

**XOR**(A,B) = (A OR B) AND NOT(A AND B)

**Regulated recruitment**

- Integrates **OR** and **NAND** into a single regulatory region
- No need for special proteins
- Modular and evolvable

[Diagram showing regulated recruitment]
More complex control function, e.g., XOR?

Alternative implementations exist, e.g.,

Gene expression \( \propto P_{p1}(A,B) + P_{p2}(A,B) \) \( \approx \) XOR
XOR promoter in bacteria?

Crp: senses carbon shortage
NtrC~P: senses nitrogen shortage
function of glnHPQ gene product: Glutamine transporter

Glutamine (C₅H₁₀N₂O₃)

use as carbon source under carbon shortage
use as nitrogen source under nitrogen shortage

➔ need quantitative characterization of such promoters
➔ design of synthetic promoters using exogenous regulators

EQ gate?

<table>
<thead>
<tr>
<th>A/B</th>
<th>Eq</th>
</tr>
</thead>
<tbody>
<tr>
<td>lo/lo</td>
<td>ON</td>
</tr>
<tr>
<td>lo/hi</td>
<td>OFF</td>
</tr>
<tr>
<td>hi/lo</td>
<td>OFF</td>
</tr>
<tr>
<td>hi/hi</td>
<td>ON</td>
</tr>
</tbody>
</table>

➔ possible solution: interaction at a distance
e.g., DNA-looping via dimers (AraC, GalR, MelR,…)

heterodimers: [A. Hochschild et al]
E.g., **distal repression** by heterodimer pair R and R’

allow control by **TF**

Implementation of **EQ gate**: 

```
A1 R' B1 A & NOT B  B2 R' A2 B & NOT A
```

```
A' B' A & NOT B' B & NOT A
```

**Effective cascade w/o new genes!**

---

**Generalized control architecture w/ multiple TF’s**

**Output = [...] OR [...] & NOT [...] & NOT [...] & NOT [...]**

phenotype: **dominant repression**

**Conjunctive Normal Form**

---

**Output = [...] & [...] OR [...] OR [...] OR [...]**

phenotype: **enhancer autonomy**

**Disjunctive Normal Form**

→ all logic functions reducible to minimal CNF or DNF
→ **schemata** for constructing arbitrary regulatory logic functions
(>programmable molecular computer!)
Molecular computer • symmetric interaction • hidden units (cofactors)

output promoter activity ($P$)

eurons TF binding sites ($j$)

input TF concentration ($n$)

firing threshold binding strength ($K_j$)

synapse TF-TF interaction ($\omega_{i,j}$)

=> single node of GRN is already a network!
• symmetric interaction
• "hidden units" (cofactors)

Molecular Boltzmann machine!
"learning" ↔ evolution of regulatory sequences

Summarize:
A large variety of control functions $P([TF_j])$ may be implementable via appropriate choices of $\{K_j, \omega_{i,j}\}$, i.e., via regulatory sequences alone = programmable molecular computer

=> synthesize "desired" transcriptional "logic gates"

=> "breed" regulatory sequences to implement desired control functions

Potential application: cell-specific gene expression profiling

<table>
<thead>
<tr>
<th>gene 1</th>
<th>gene 2</th>
<th>gene 3</th>
<th>cell X</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell A</td>
<td>cell B</td>
<td>cell C</td>
<td>cell X</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

=> cell type discrimination: use multiple traits

=> cell X revealed by a reporter gene driven by designed regulatory sequences

=> cell X eliminated by promoter activating apoptosis

=> targeted delivery not required (~ smart bomb!)
Ingredients for complex transcription control

• programmable protein-DNA interaction
• weak, glue-like interaction between nearby proteins
• long-distance activation/repression
• insulation of gene regulatory control

Eukaryotes: formalism as phenomenological model

• generic cooperative interaction
  mediated by nucleosomes
  [Polach & Widom, ’96]
  ➔ physical attraction not necessary

• short-range “quenching” [Amosti, Levine]

• distal repression via recruitment
  of chromatin modification agents

• insulating elements: crucial for minimizing cross talk

similar ingredients but superior implementation platform