Focus of our laboratory:

- **individual nodes** of gene network
- **quantitative** study of bacterial gene regulation
  - specificity and cross-talk in two-component signaling
  - combinatorial transcriptional control
  - translational control by small regulatory RNA
  - nonlinear proteolysis
- small regulatory circuits
- metabolic control and growth physiology
- synthetic genetic logic gates and circuits
- **directed evolution** of gene expression and regulation

⇒ from molecules to cellular physiology
Focus of our laboratory:

- individual nodes of gene network
- quantitative study of bacterial gene regulation
  - specificity and cross-talk in two-component signaling
  - combinatorial transcriptional control
  - translational control by small regulatory RNA
  - nonlinear proteolysis
- small regulatory circuits
- metabolism and growth control
- synthetic genetic logic gates and circuits
- directed evolution of gene expression and regulation
  ➞ from molecules to cellular physiology

Erel Levine (in March)
Quantitative characterization of the *lac* promoter

*lac* promoter of *E. coli*:
- best-studied system of molecular biology
  - all molecular components characterized
  - many mutants studied *in vivo*
  - most parameters measured *in vitro*
- exemplary model system of combinatorial gene regulation
  - involves activation, repression, and DNA looping

Quantitative confrontation of model and experiment
- applicability of the thermodynamic description of tsx control?
- can the *in vivo* behavior of a system
  be understood in terms of its parts?
Review of lactose utilization

- lac operon: pumps in lactose (LacY) and converts it to glucose (LacZ)
- lac promoter (Plac): express Lac only when lactose is present and glucose is absent

<table>
<thead>
<tr>
<th>IPTG</th>
<th>glucose</th>
<th>expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>high</td>
<td>OFF</td>
</tr>
<tr>
<td>low</td>
<td>low</td>
<td>OFF</td>
</tr>
<tr>
<td>high</td>
<td>high</td>
<td>OFF</td>
</tr>
<tr>
<td>high</td>
<td>low</td>
<td>ON</td>
</tr>
</tbody>
</table>

molecular ingredients:
- specific protein-DNA binding
- protein-protein interaction
- protein-mediated DNA looping

➤ theory: quantitative prediction of gene regulation by LacI, cAMP-Crp
Thermodynamic framework of gene regulation

[Shea & Ackers, JMB 1985]

gene expression \( \propto \) eq. promoter occupation probability \( P \) in the presence of A

\[
P([A],[RNAp]) = \frac{W(0,1) + W(1,1)}{W(0,0) + W(0,1) + W(1,0) + W(1,1)}
\]

define \( W(0,0)=1 \), then for activation

\[
W(0,1) = \frac{[RNAp]}{K_p}, \quad W(1,0) = \frac{[A]}{K_A}
\]

\[
W(1,1) = \omega_{A-p} \cdot \left(\frac{[A]}{K_A}\right) \cdot \left(\frac{[RNAp]}{K_p}\right)
\]

\[
P \approx \frac{[RNAp]}{K_p} \cdot \frac{1 + \omega_{A-p} [A]/K_A}{1 + [A]/K_A}
\]

(for typical weak promoters)
Thermodynamic framework of gene regulation

[Shea & Ackers, JMB 1985]

gene expression $\propto$ eq. promoter occupation probability $P$ in the presence of A

\[ P([A],[RNAp]) = \frac{W(0,1) + W(1,1)}{W(0,0) + W(0,1) + W(1,0) + W(1,1)} \]

define $W(0,0) = 1$, then for activation:

\[ W(0,1) = \frac{[RNAp]}{K_p}, \quad W(1,0) = \frac{[A]}{K_A} \]

\[ W(1,1) = \omega_{A-p} \cdot \left( \frac{[A]}{K_A} \right) \cdot \left( \frac{[RNAp]}{K_p} \right) \]

\[ P \approx \frac{[RNAp]}{K_p} \cdot \frac{1 + \omega_{A-p} [A]/K_A}{1 + [A]/K_A} \]

(for typical weak promoters)

for repression, $W(1,1) = 0$:

\[ P \approx \frac{[RNAp]}{K_p} \cdot \frac{1}{1 + [R]/K_R} \]

co-regulation:

\[ P \propto \frac{1 + \omega_{A-p} [A]/K_A}{1 + [A]/K_A} \cdot \frac{1}{1 + [R]/K_R} \]

multiplicative
Review of lactose utilization

- lac operon: pumps in lactose (LacY) and converts it to glucose (LacZ)
- lac promoter (Plac): express Lac only when lactose is present and glucose is absent

<table>
<thead>
<tr>
<th>IPTG</th>
<th>glucose</th>
<th>expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>high</td>
<td>OFF</td>
</tr>
<tr>
<td>low</td>
<td>low</td>
<td>OFF</td>
</tr>
<tr>
<td>high</td>
<td>high</td>
<td>OFF</td>
</tr>
<tr>
<td>high</td>
<td>low</td>
<td>ON</td>
</tr>
</tbody>
</table>

molecular ingredients:
- specific protein-DNA binding
- protein-protein interaction
- protein-mediated DNA looping

theory: quantitative prediction of gene regulation by LacI, cAMP-Crp
expt: characterize LacZ activity for different levels of regulatory proteins
   -- control protein levels by varying the inducers (IPTG and cAMP)
Quantitative characterization

Previous expt: [Setty et al, PNAS, 2003]
Grow cells in medium with glucose, cAMP, IPTG
-- use glucose to suppress cAMP synthesis
-- control cAMP-level extracellularly

inconsistent with behavior of mutants:
\( \Delta lacI: > 1000x; \quad \Delta crp > 50x \)

\( \Rightarrow \) possible problems: complex links between extracellular and intracellular inducer conc.
**Quantitative characterization of mutants**

**weak cAMP dependence:** glucose-mediated repression of AC activity may be incomplete

- delete \( cyaA \) gene (encoding \( AC \))
- find \( \sim 100x \) change in LacZ activity
- Hill coeff \( \approx 2 \)

incompatible w/ biochem and thermodynamic model of tsx control

CRP dimer activated by binding of single cAMP molecule

\[
CRP_2 + cAMP \rightleftharpoons CRP_2\cdot cAMP
\]

(expect Hill coeff = 1)

*in vitro* biochem irrelevant? other effects exerted by CRP-cAMP?
Quantitative characterization of mutants

**weak cAMP dependence:** glucose-mediated repression of AC activity may be incomplete

- delete *cyA* gene (encoding AC)
- find ~100x change in LacZ activity
- Hill coeff ≈ 2

incompatible w/ biochem and thermodynamic model of tsx control

CRP dimer activated by binding of single cAMP molecule

$$\text{CRP}_2 + \text{cAMP} \rightleftharpoons \text{CRP}_2:\text{cAMP}$$

(expect Hill coeff = 1)

in *vitro* biochem irrelevant?

other effects exerted by CRP-cAMP?

- cAMP degraded by PDE (*cpdA*)
- effect of *cpdA* deletion?
- Hill coeff ≈ 1, agrees with model
- role of PDE: no known phenotype
- mechanism of cooperativity?
Quantitative characterization of mutants

**IPTG dependence:** cyaA- cells with [cAMP]=0

- very cooperative! (Hill coeff ≈ 4)

- delete lacY  
  Hill coeff ≈ 2

- constitutive expression of LacY  
  only shifted IPTG dependence

- Hill coeff = 2 widely cited in literature

- LacI forms tetramer (dimer of dimers)
- strong coupling within each dimer and weak coupling between dimers

**but...** Hill coeff = 2 is one of the many **pseudo-facts** regarding Lac
Quantitative characterization of mutants

IPTG dependence: *cyA*-cells with [cAMP]=0 ➔ very cooperative!
- LacI forms tetramer (dimer of dimers)
- strong coupling within each dimer and weak coupling between dimers
- LacI$_4$-IPTG binding non-cooperative
  \[
  \text{LacI}_4 + \text{IPTG} \rightleftharpoons \text{LacI}_4:\text{IPTG}
  \]
- weakly cooperative in the presence of operator DNA (Hill coeff = 1.4 ~ 1.6)
  \[\text{Hill coeff} = \frac{1}{K_R} + 1\]
  [Matthews lab, '85]

➔ neither monomers of LacI dimer can bind IPTG for specific binding to Lac ops

auxiliary Lac operators stabilize LacI-O1 binding via **DNA looping** [Muller-Hill]

**Expression formula**

\[ [R] = \frac{2 \cdot [\text{LacI}_4 \text{ total}]}{(1 + [\text{IPTG}] / K_{\text{IPTG}})^2} \]

**R epression**

tsx activity \( \propto \frac{1}{1 + [R] / K_R} \)
Quantitative characterization of mutants

**IPTG dependence:** cyA- cells with [cAMP]=0 ➔ very cooperative!

- LacI forms tetramer (dimer of dimers)
- strong coupling within each dimer and weak coupling between dimers
- LacI<sub>4</sub>-IPTG binding non-cooperative
  \[ \text{LacI}_4 + \text{IPTG} \rightleftharpoons \text{LacI}_4:\text{IPTG} \]
  - weakly cooperative in the presence of operator DNA (Hill coeff = 1.4 ~ 1.6)
  
  [Matthews lab, ’85]

➔ **neither** monomers of LacI dimer can bind IPTG for specific binding to Lac ops

\[ [\text{R}] = \frac{2 \cdot [\text{LacI}_4]_{\text{total}}}{(1 + [\text{IPTG}] / K_{\text{IPTG}})^2} \]

active repressors

\[ \text{tsx activity} \propto \frac{1}{1 + [R] / K_R} \]

simple repression

- include DNA looping in model

\[ [\text{R}] \to [\text{R}] + \frac{L_0 \cdot [\text{LacI}_4]_{\text{total}}}{(1 + [\text{IPTG}] / K_{\text{IPTG}})^4} \]

$L_0$: local increase of [LacI] due to looping

auxiliary Lac operators stabilize LacI-O1 binding via **DNA looping** [Muller-Hill]

➔ increase fold-repression by $L_0$-fold

➔ effective Hill coeff (1.5 ~ 3) depends on $L_0$

but value of $L_0$ not known independently
Quantitative characterization of mutants

looping model w/ $L_0 \approx 12$, $2[LacI]/K_R = 20$

auxiliary Lac operators stabilize LacI-O1 binding via DNA looping [Muller-Hill]

- increase fold-repression by $L_0$-fold
- effective Hill coeff (1.5 ~ 3) depends on $L_0$

but value of $L_0$ not known independently
Quantitative characterization of mutants

looping model w/ $\mathcal{L}_0 \approx 12$, $2[LacI_4]/K_R = 20$

Crp-dependence of DNA looping

Fried et al, 84; Balaeff et al, 04

in vitro study found coop. factor $\Omega = 4 \sim 12$
Direct probe of DNA looping \textit{in vivo}

Use dimeric LacI mutant

remove auxiliary operators

\[ \text{Hill coeff } = 2.5 \]

\[ \text{Hill coeff } = 1.5 \]

data well-fitted by DNA looping model

\[ \Rightarrow \text{ cooperativity in IPTG response requires DNA looping (Lac tetramer + auxiliary ops)} \]

[Oehler & Muller-Hill, 06]

\[ \Rightarrow \text{ IPTG-LacI-operator interaction same as \textit{in vitro}} \]
Summary

- **main findings for the lac promoter:**
  - Crp enhances DNA looping
  - abrupt IPTG response despite non-cooperative LacI-IPTG interaction;
  - suggests physiological role of Crp-cAMP as enhancer of repression
  - mechanism of Crp-LacI interaction?
  - coop cAMP response due to PDE; physiological function? mechanism?

- **general lessons for quantitative systems biology:**
  - hidden interaction abound even for the “best studied” system
  - pseudo-facts abound even for the best known components
  - quantitative description of in vivo biology is possible
  - need **solid, qualitative** knowledge of the components (e.g., Hill coeff)
  - **(semi) quantitative** characterization generates spectrum of phenotypes
  - provides clues for identifying unknown components and mechanisms
  - provides phenomenological description of Plac for high-level studies

---

*E. coli* MG1655 (*cyaA-, cpdA-, lacY-*)
**de novo evolution** of regulatory sequences

want gene expression only in the presence of inducer “a”

Steady level of regulatory protein A

TF activation controlled thru inducer a

Selectable output:

-- gene product lethal if drug 1 present
-- gene product essential if drug 2 present

Defined region of mutagenesis

<table>
<thead>
<tr>
<th>[a]</th>
<th>drug</th>
<th>gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>lo</td>
<td>1</td>
<td>OFF</td>
</tr>
<tr>
<td>hi</td>
<td>2</td>
<td>ON</td>
</tr>
</tbody>
</table>
Directed evolution of core promoters

- evolve promoters from random sequences in a tight space (29 nt) using **mutagenic PCR**
- **select** for cells with increasing resistant to Cm
- expect two variants of the $\sigma^{70}$ core promoter:
  - -10/-35 hexamers: **TTGACA**<-- 17nt -->**TATAAT**
  - extended -10: **TGTGNTATAAT**
- two selection genes: **divergent overlapping promoters possible?**
- dependence on evolutionary path?

*in vitro* mutagenesis
Evolution procedure

• initial
  – initial population: random library of 29mer ligated into selection plasmid
  – transform plasmid in E. coli (TOP10) cells; transformation efficiency ~$10^4$ indept clones

• selection
  – grow on plates with various drug conc (CM and/or kan)
  – collect several hundred clones with the highest drug resistance

• mutagenesis
  – plasmid prep
  – mutagenic PCR of insert seq (substitution freq ~5%/base)
  – re-clone into initial vector, and re-transform into initial strain

• selection
  …

all intermediate clones “saved” for future analysis
Semi-quantitative phenotype assay

Characterize distribution of phenotypes at each stage of evolution

- collect 96 clones
- grow on agar plates with different drug conc
- identify max drug resistance

☐: Max drug resistance for the clone
Evolution in single direction: phenotype

max Cm resist (x 33ug/ml) evolution cycle

Cm tolerance (x33ug/ml)

- after mutation
- after selection

kanR 29N cat
ampR p15A ori

CM

kan
Evolution in single direction (CM): genotype

TGTG

**TA**

AT

5' 10 20 30 60 70 80 3' primers

mutable region

comparative "genomics"

TGTG*TA**AT

RBS

cat

promoter sequence after 2s

promoter sequence after 5s

TGTGGTAcAAT

and much more …
Degeneracy of evolved promoter (Cm direction)

after 1st round (Cm resistance = 1 x 33ug/ml)

after 5th round (Cm resistance > 10 x 33ug/ml)

- up to 7 partial promoter motifs packed in 29-nt region + flanking regions
- (almost) every fixed mutation attributable to additional motif(s)

Why?
-- stronger expression from multiple promoters?
-- robustness to mutation provided by multiple copies?

Benefit: makes subsequent evolution of activators/repressors easier
Multiple promoters seen in bioinfo studies

- avg of 38 putative promoter signals observed in a typical 250bp region upstream of gene start;

- in 50% of these regions, the “real” promoter is not the highest scoring promoter

[Huerta & Collado-vides, 03]
Reversal of evolution direction: phenotype

4 rounds of selection in Kan direction

Kan resistance decreasing
CM resistance increasing
appearance of divergent promoter activity

revert selection to CM direction

CM tolerance (33 ug/mol)
K.4s C.1s C.2m C.2s

CM tolerance (50 ug/ml)
K.4s C.1s C.2m C.2s

Kan tolerance (50 ug/ml)
init 1s 2m 2s 3m 3s 4m 4s

after mutation
after selection
Reversal of evolution direction: genotype

Sequences obtained after 4th round of selection with Kan only:

\[ \text{aTGAtA} \quad \text{17 bp} \quad \text{cATtAc} \]

\[ \text{TGNTATgAT} \]

\[ \text{ATGATATGATCGACGGAAGAGTACATTAC} \quad \text{kanR} \]
\[ \text{TACTATACTAGCTGCCTTCTCATGTAATG} \quad 5' \]

-- one \text{extended -10 motif} and one \text{-35/-10 motif} in the Kan direction
-- no significant motifs detected in the Cm direction

Sequences after 2 more rounds of selection with Cm only:

\[ \text{TTGtat} \quad \text{16 bp} \quad \text{TATcAc} \]

\[ \text{ATTGTATGATCGGCGGAAGAGTATATCAC} \quad \text{kanR} \]
\[ \text{TAACATACTAGCGCCTTCTCATATAGTG} \quad 5' \]

-- one \text{extended -10} and one standard \text{-35/-10 motifs} in Cm direction
-- weakened \text{-10/-35 motif} and lost extended -10 and in the Kan direction
Evolution in both directions: phenotype

- evolution slightly slower than that driven in single direction (5 vs 4 rounds)
Evolution in both directions: genotype (5 rounds)

- found two types of overlapping motifs:

-10 overlaps -10 (with -35 on flanking sequences)

-35 overlaps -10
Summary: promoters are flexible!

• **Single direction:** multiple promoters in confined space

• **Reversal:**
  – existing promoter evolve quickly to reverse direction by few mutations
  – reduction of promoter activity in the reverse direction important (occlusion)

• **Divergent overlapping promoters:**
  
  -10 overlaps -10 (with -35 on flanking sequences)

  ![Diagram of divergent overlapping promoters]

  -35 overlaps -10

  ![Diagram of divergent overlapping promoters]
From molecules to system-level functions

**traditional mol bio:**
one gene, one process (e.g., A activates B)

quantitative analysis of individual nodes and small circuits

how they talk

**high throughput methods**

bioinformatic analysis

who talks to whom

**systems biology:**
many components and processes (e.g., predictive modeling of cell and multi-cellular organisms)

**many genes, one/few process(es)** (e.g., genome-wide survey of gene exp)

**qualitative** system-level properties depend **quantitatively** on the degree of regulation
Natural switches (e.g., phage lambda)

• induction time to switch: < 10 min
• ingredients for fast speed
  – proteolysis
  – auto-activation and repression

• induction time to switch: ~ 6 hrs (several cell divisions)
• slow speed possibly due to passive dilution

⇒ “speed limit of gene regulation”  
[Rosenfeld et al, Science, 2005]

Synthetic genetic switch

Q: faster switch using the same components?
**Alternative switch: face-to-face promoter construct**

- **cmR** to **tetR**
- **P<sub>tet</sub>** → **lacI** → **kanR**

- **aTc** and **IPTG**

<table>
<thead>
<tr>
<th></th>
<th>aTc</th>
<th>IPTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>KanR</td>
<td>growth</td>
<td></td>
</tr>
<tr>
<td>CmR</td>
<td>growth</td>
<td></td>
</tr>
</tbody>
</table>

- Induction time needed for switching ~ 15 min (fast)
- Stability: 6-8 hours
- Large fold-change in induction (LacZ and GFP activity)
- Fast switch also in the reverse direction

Generate variants, screen for desired phenotype

> 50x
Acknowledgement

• theory
  – Nicolas Buchler, Ulrich Gerland, Tom Kuhlman (combinatorial tsx control)
  – Erel Levine, Matt Scott (sRNA-mediated regulation)
  – Bob White, Kay Hamacher (two-component signaling)
  – Ulrich Gerland, Weiqun Peng (molecular evolution)
  – Peter Lenz, Erel Levine (metabolic pathways)
  – Eddie Mateescu (growth control)

• experiment
  – Tom Kuhlman, Zhongge Zhang (lac promoter)
  – Tom Kuhlman, Erel Levine, Min Huang, Zhongge Zhang (sRNA)
  – Sabrina Li, Shumo Liu, Robert Yee (promoter evolution, fast switch)
  – Dalai Yan, Joseph Tian (nitrogen assimilation, synthetic promoter)
  – Hendrik Szurmant (two-component signaling)

• biology collaborators
  – Bill Loomis, Milton Saier, Lin Chao (UCSD)
  – Jim Hoch (Scripps), Sydney Kustu (Berkeley), Yiping Wang (Peking U.)

• support
  – NIH, NSF, DOE, BWF

related publications:  http://matisse.ucsd.edu/~hwa/pub