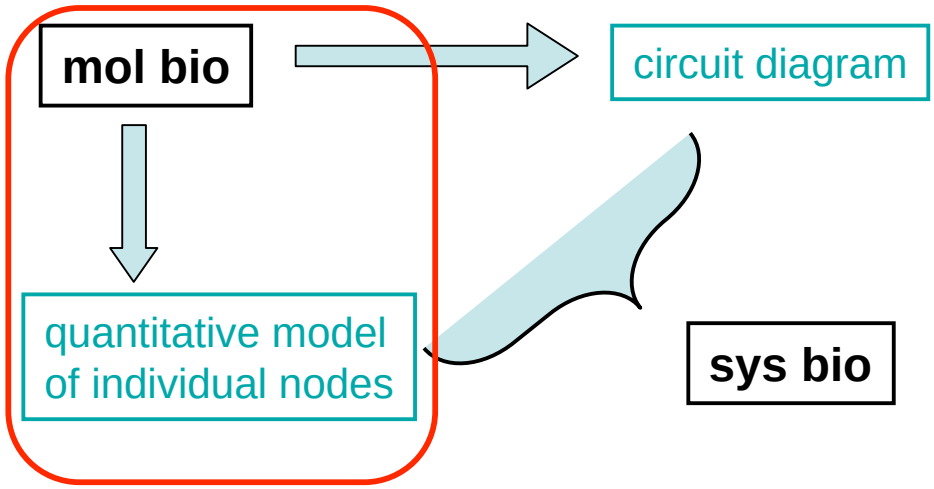
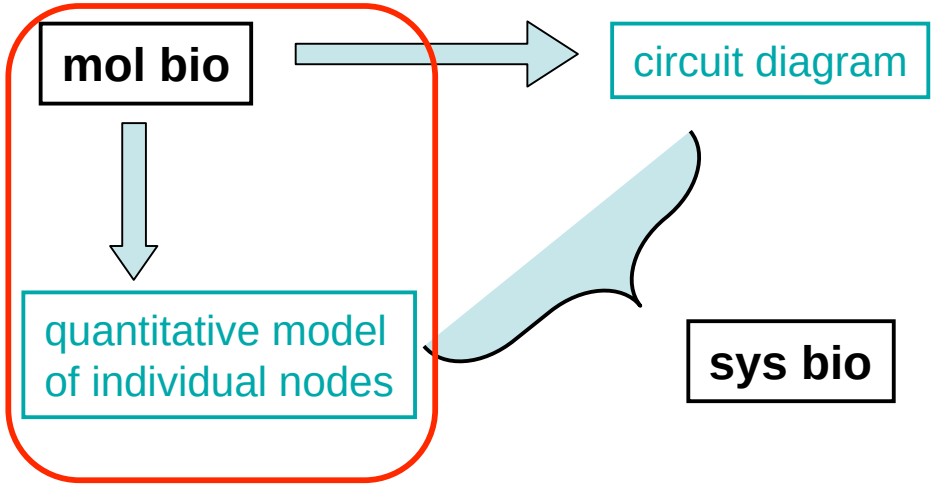


Focus of our laboratory:

- 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 ⟵ current focus
- 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 ← Erel Levine (in March)
  - nonlinear proteolysis
- small regulatory circuits
- metabolism and growth control ← current focus
- synthetic genetic logic gates and circuits
- directed evolution of gene expression and regulation

➔ from molecules to cellular physiology

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

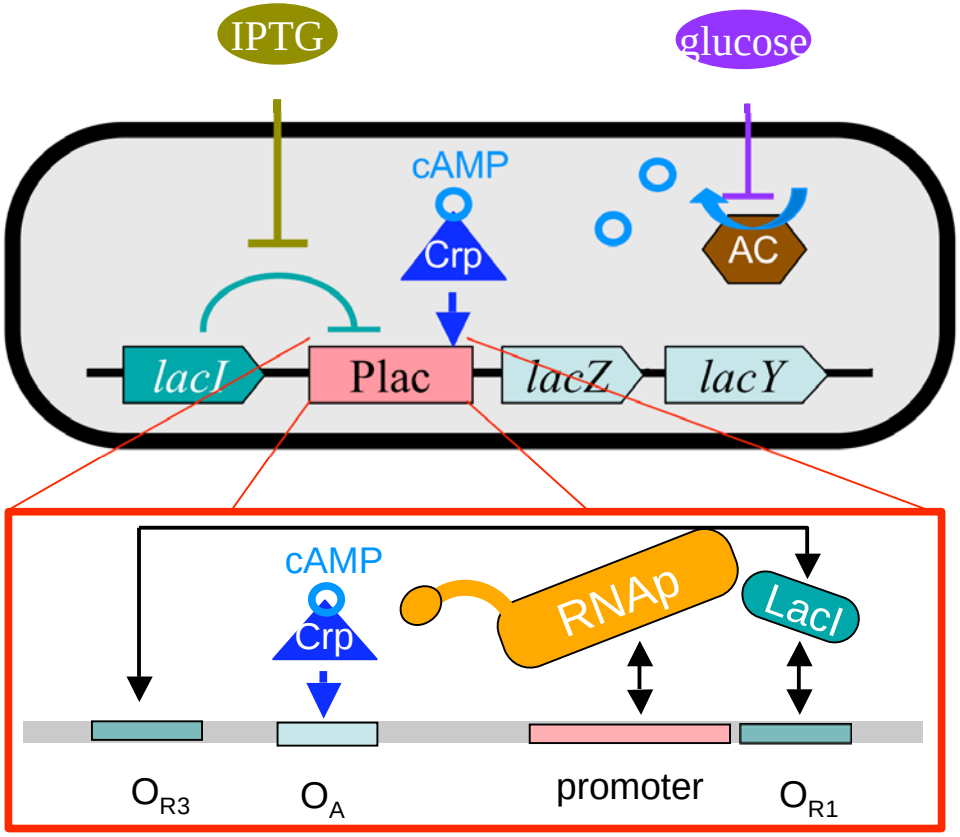
Quantitative confrontation of model and experiment

➔ 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



IPTG	glucose	expression
low	high	**OFF**
low	low	**OFF**
high	high	**OFF**
high	low	**ON**

molecular ingredients:

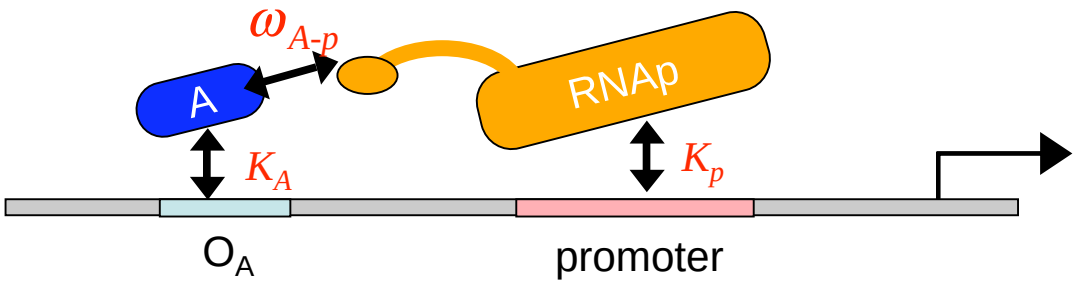
- specific protein-DNA binding
- protein-protein interaction
- protein-mediated DNA looping

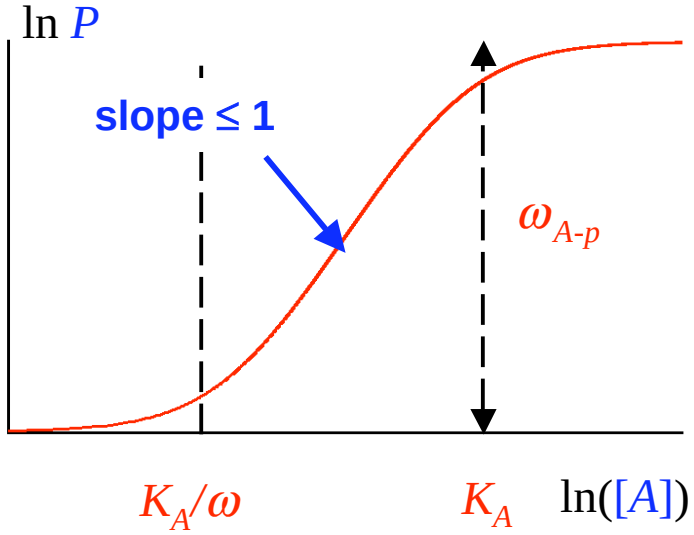
➔theory: quantitative prediction of gene regulation by LacI, cAMP-Crp

Thermodynamic framework of gene regulation

[Shea & Ackers, JMB 1985]

gene expression ∝ 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

$$W(0,1) = [RNAp]/K_p, \quad W(1,0) = [A]/K_A$$
$$W(1,1) = \omega_{A-p}\cdot([A]/K_A)\cdot([RNAp]/K_P)$$

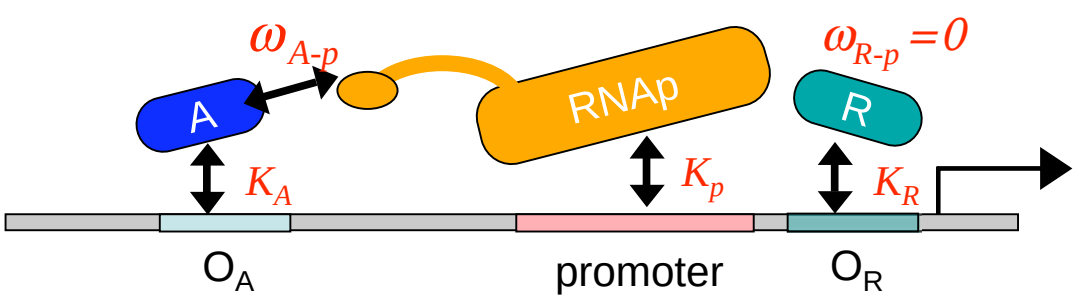
$$P \approx \frac{[RNAp]}{K_p}\cdot\frac{1+\omega_{A-p}[A]/K_A}{1+[A]/K_A}$$

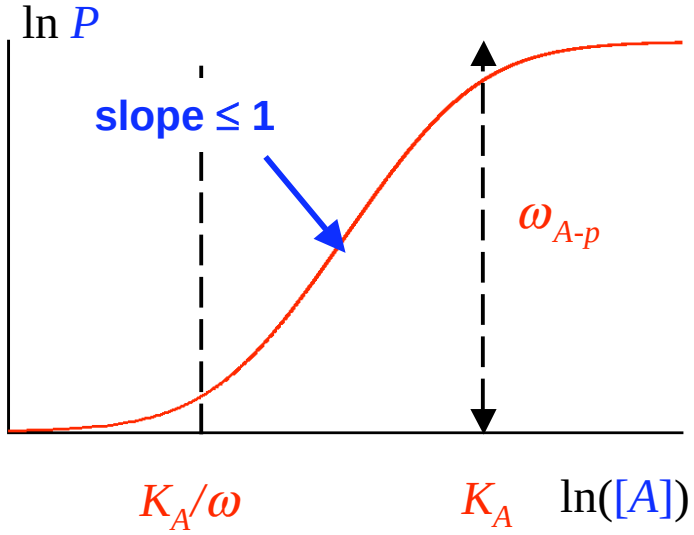
(for typical weak promoters)

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

$$W(0,1) = [RNAp]/K_p, \quad W(1,0) = [A]/K_A$$
$$W(1,1) = \omega_{A-p}\cdot([A]/K_A)\cdot([RNAp]/K_P)$$

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

(for typical weak promoters)

for repression, W(1, 1)=0

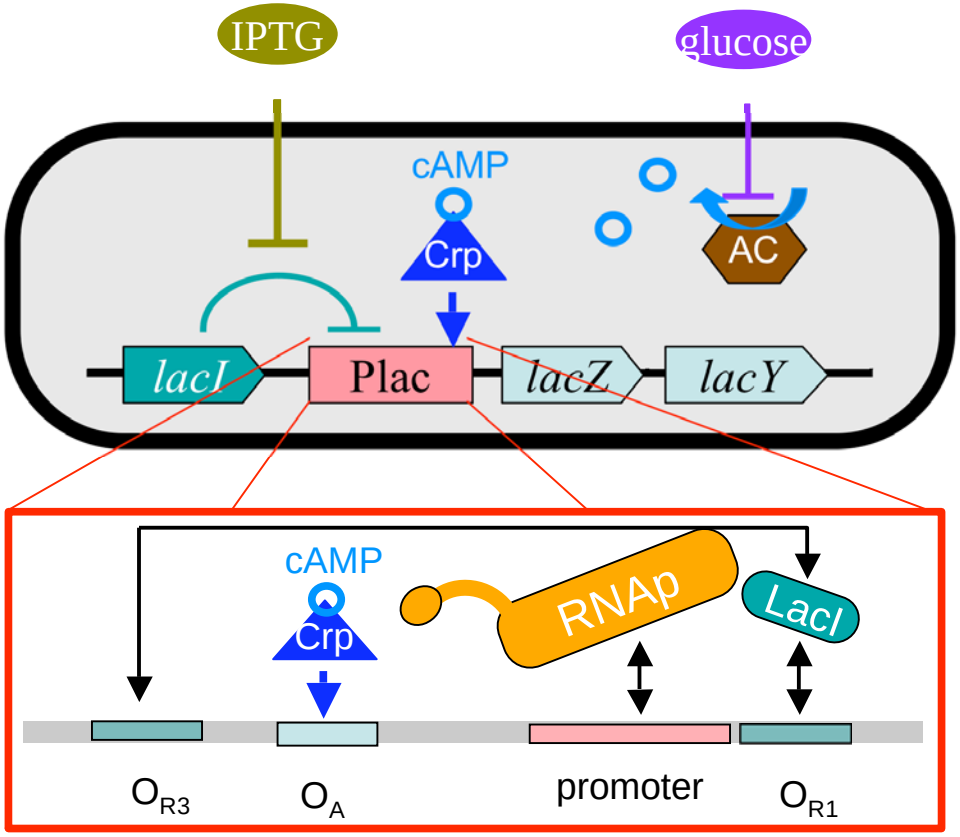
$$P \approx \frac{[RNAp]}{K_p}\cdot\frac{1}{1+[R]/K_R}$$

co-regulation multiplicative

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

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



IPTG	glucose	expression
low	high	**OFF**
low	low	**OFF**
high	high	**OFF**
high	low	**ON**

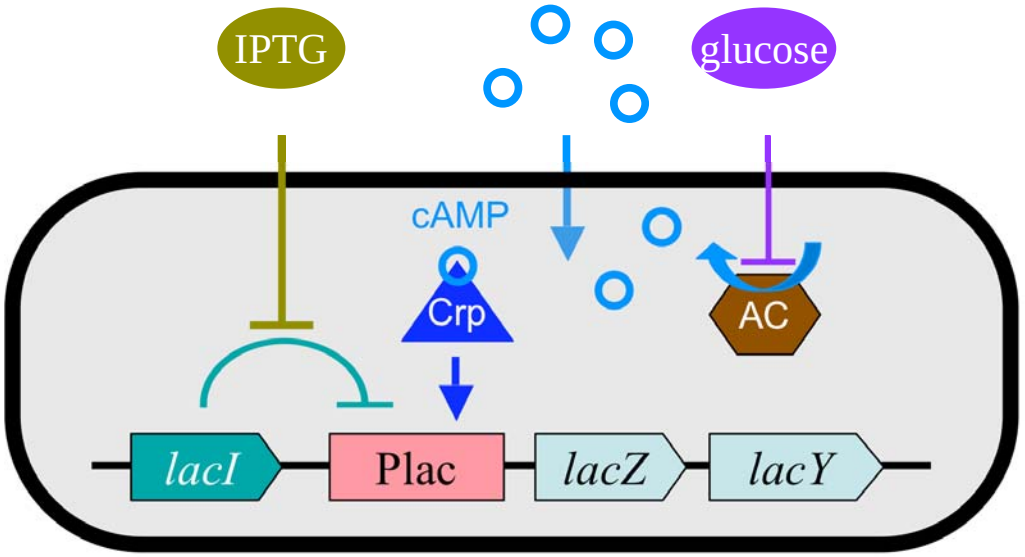
molecular ingredients:

- specific protein-DNA binding
- protein-protein interaction
- protein-mediated DNA looping

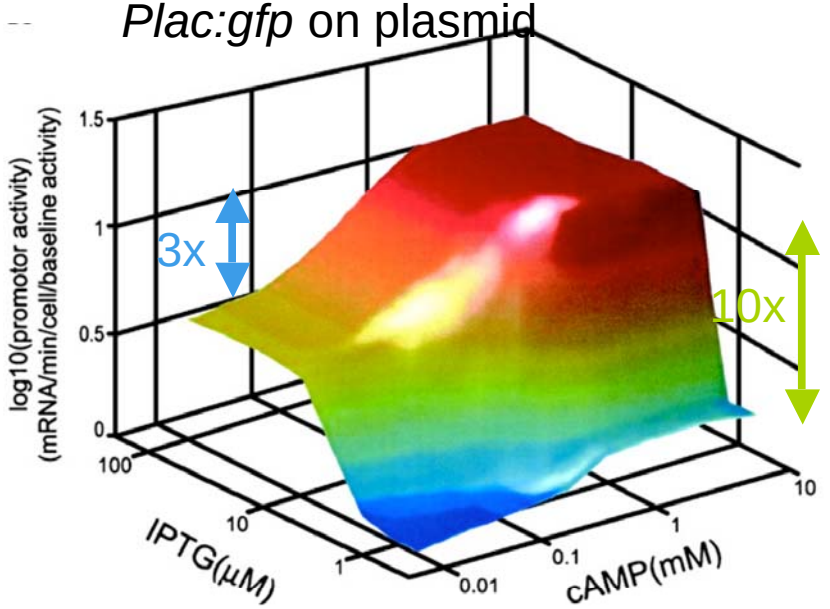
➔ theory: quantitative prediction of gene regulation by LacI, cAMP-Crp

➔ expt: characterize LacZ activity for different levels of regulatory proteins
-- control protein levels by varying the inducers (IPTG and cAMP)

Quantitative characterization



Plac:gfp on plasmid



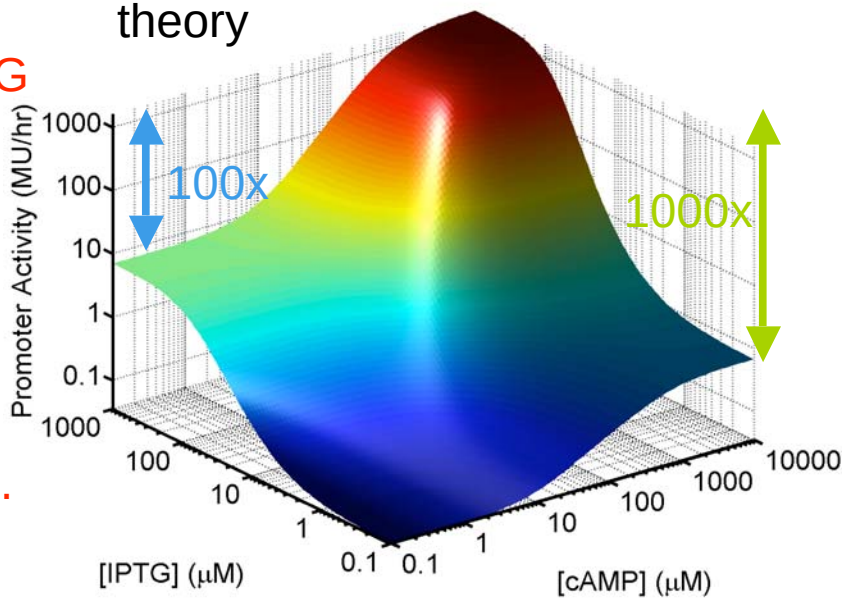
Previous expt: [Setty et al, PNAS, 2003]

Grow cells in medium with glucose, cAMP, IPTG
-- use glucose to suppress cAMP synthesis
-- control cAMP-level extracellularly

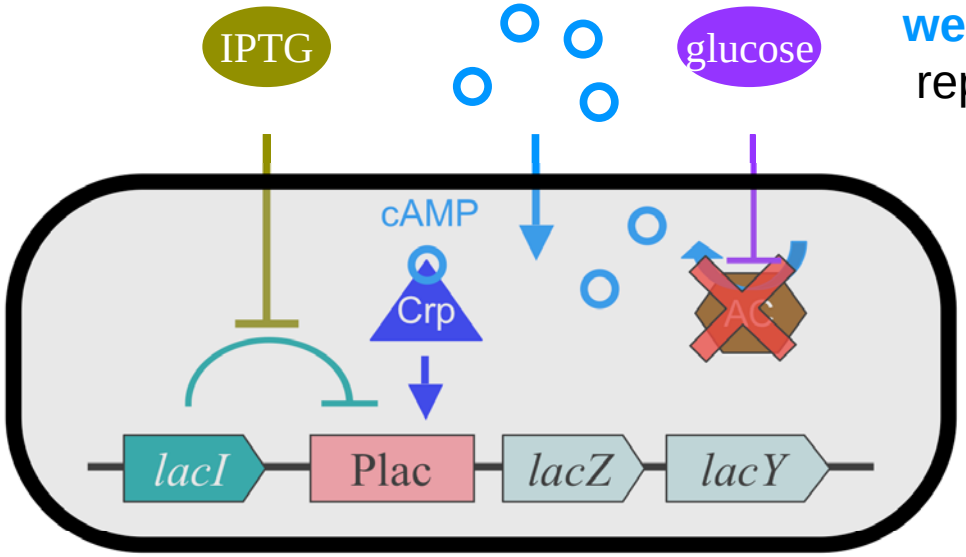
inconsistent with behavior of mutants:
Δ*lacI*: > 1000x; **Δ*crp* > 50x**

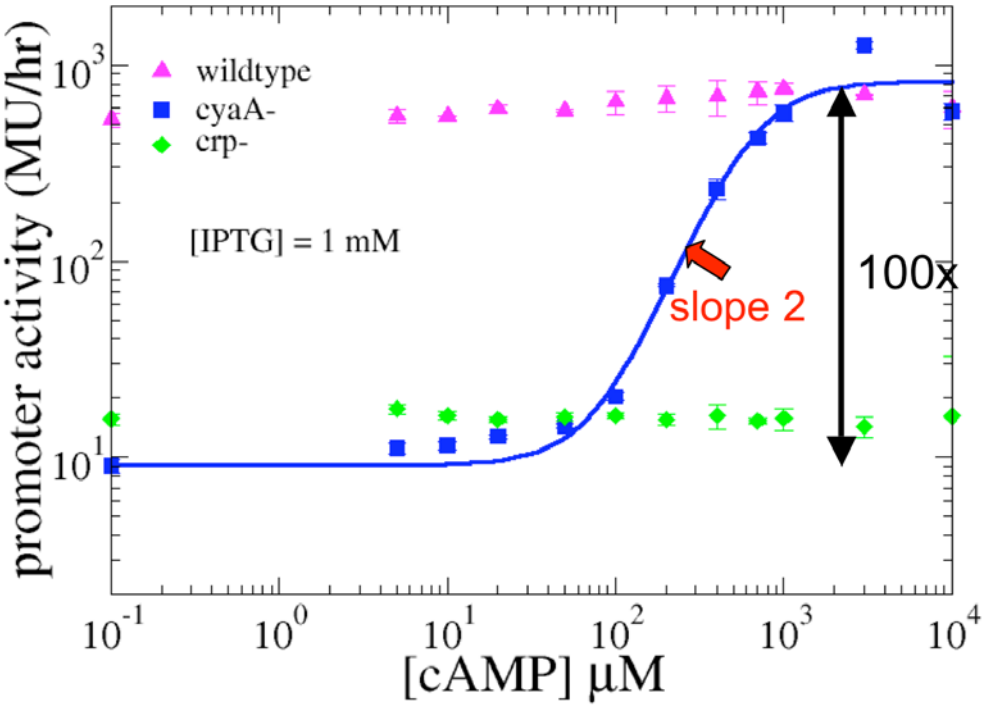
➔ possible problems: complex links between extracellular and intracellular inducer conc.

theory



Quantitative characterization of mutants





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

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

incompatible w/ biochem and thermodynamic model of tsx control

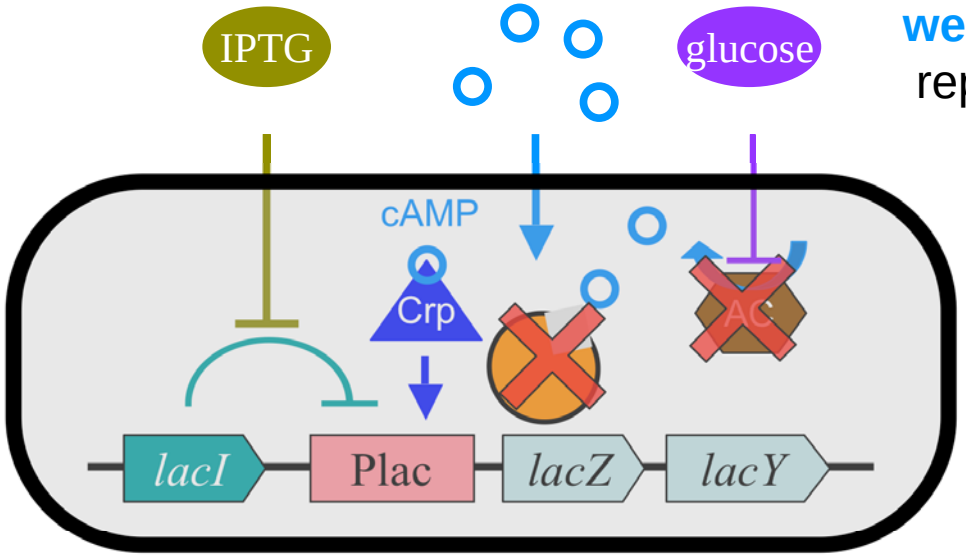
CRP dimer activated by binding of single cAMP molecule

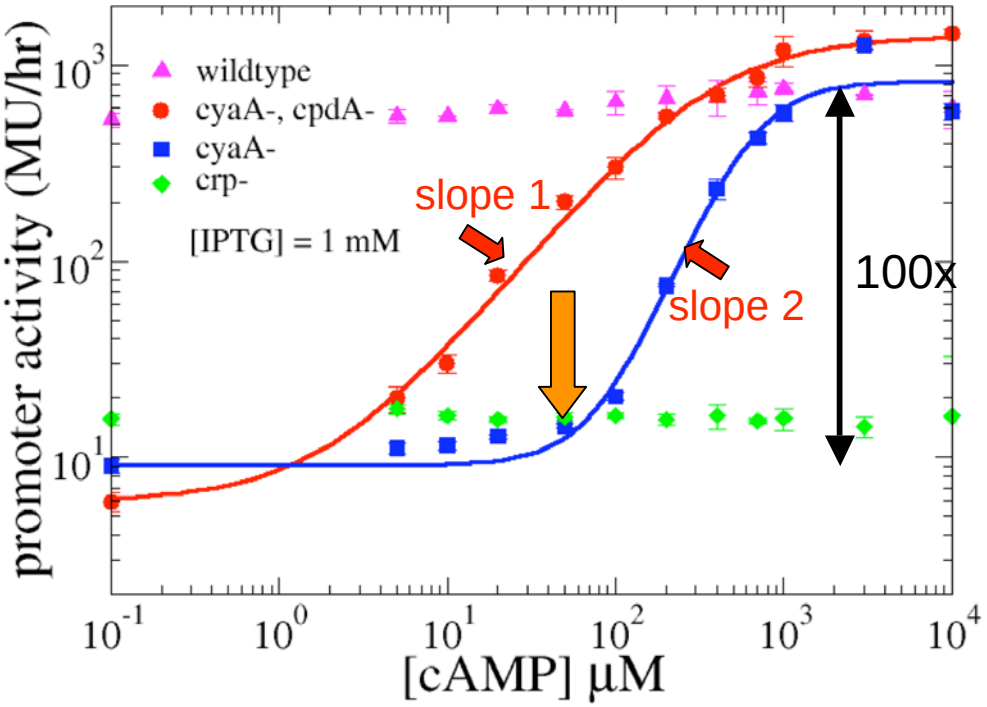
CRP_2 + cAMP ⇌ CRP_2:cAMP

(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 ***cyaA*** 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

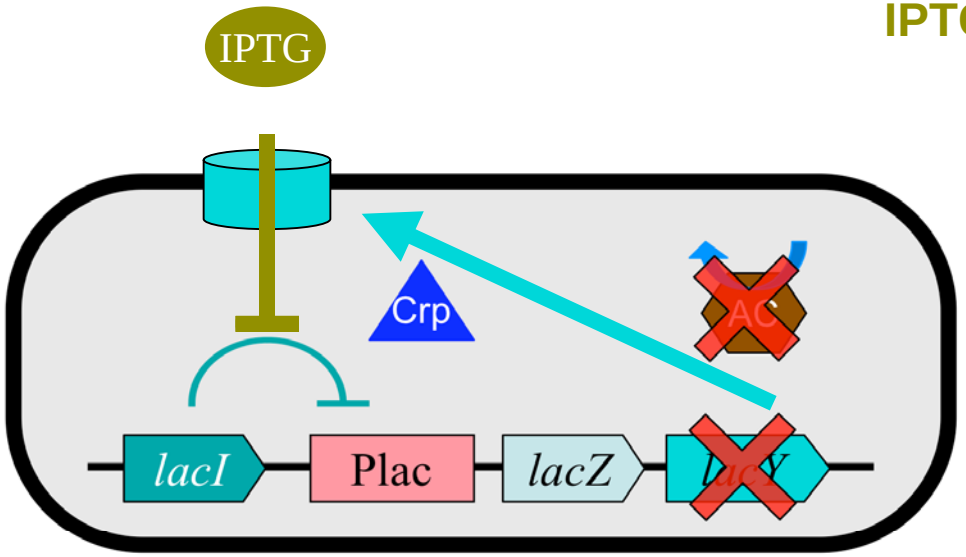
CRP_2 + cAMP ⇌ CRP_2:cAMP

(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



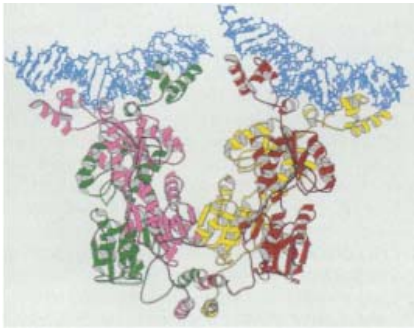
IPTG dependence: *cyaA-* cells with [cAMP]=0

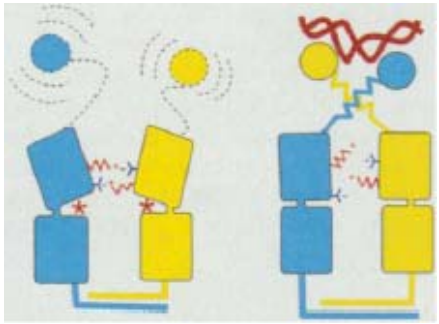
➔ very cooperative! (Hill coeff ≈ 4)

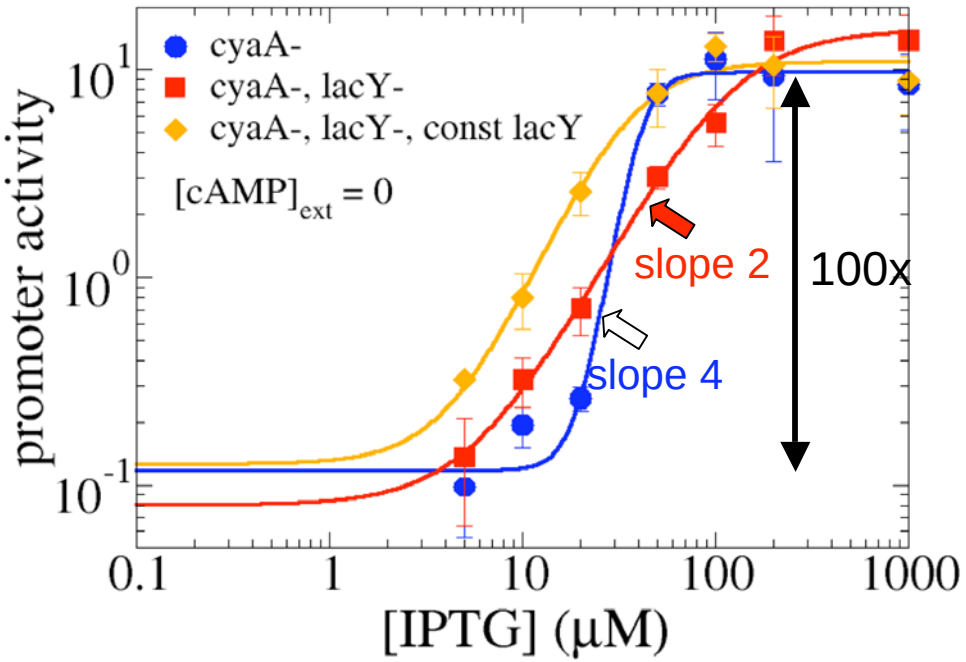
➔ delete *lacY* Hill coeff ≈ 2

➔ constitutive expression of LacY only shifted IPTG dependence

➔ Hill coeff = 2 widely cited in literature



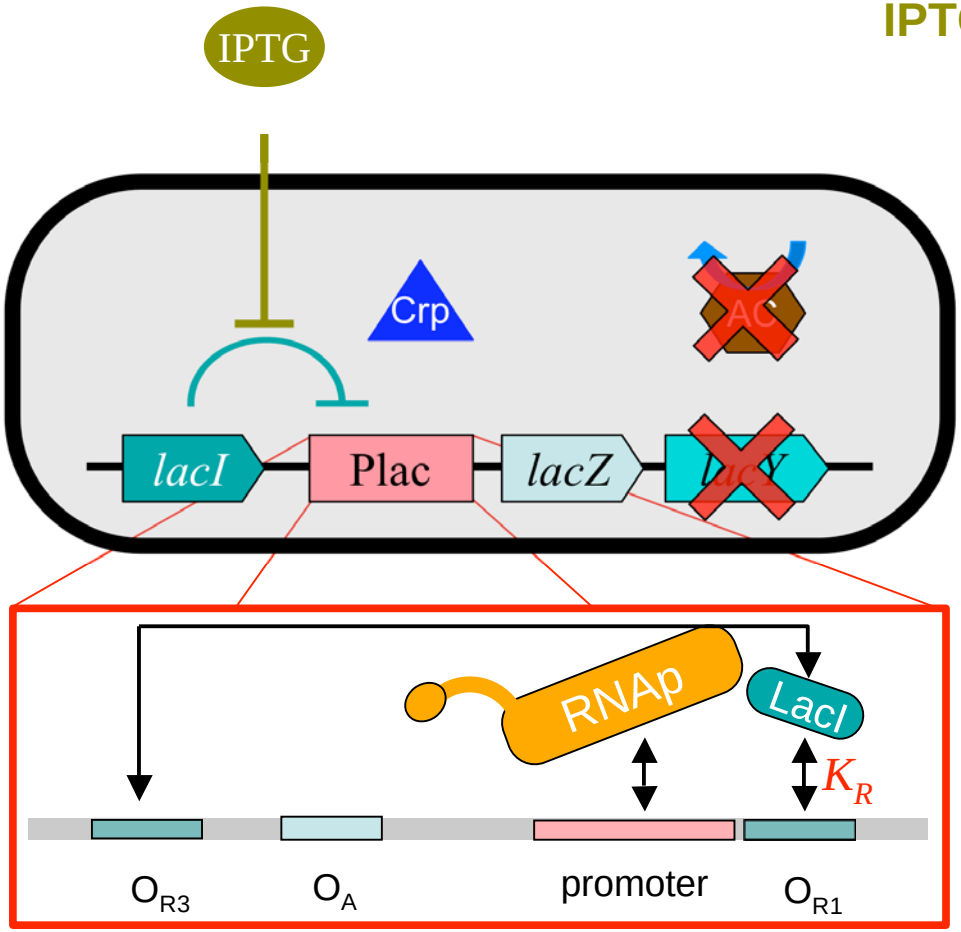




- LacI forms tetramer (dimer of dimers)
- strong coupling within each dimer and weak coupling between dimers

but… **Hill coeff = 2** is one of the many **pseudo-facts** regarding Lac

Quantitative characterization of mutants



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

IPTG dependence: *cyaA-* 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

  $LacI_4$ + IPTG $\rightleftharpoons$ $LacI_4$: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

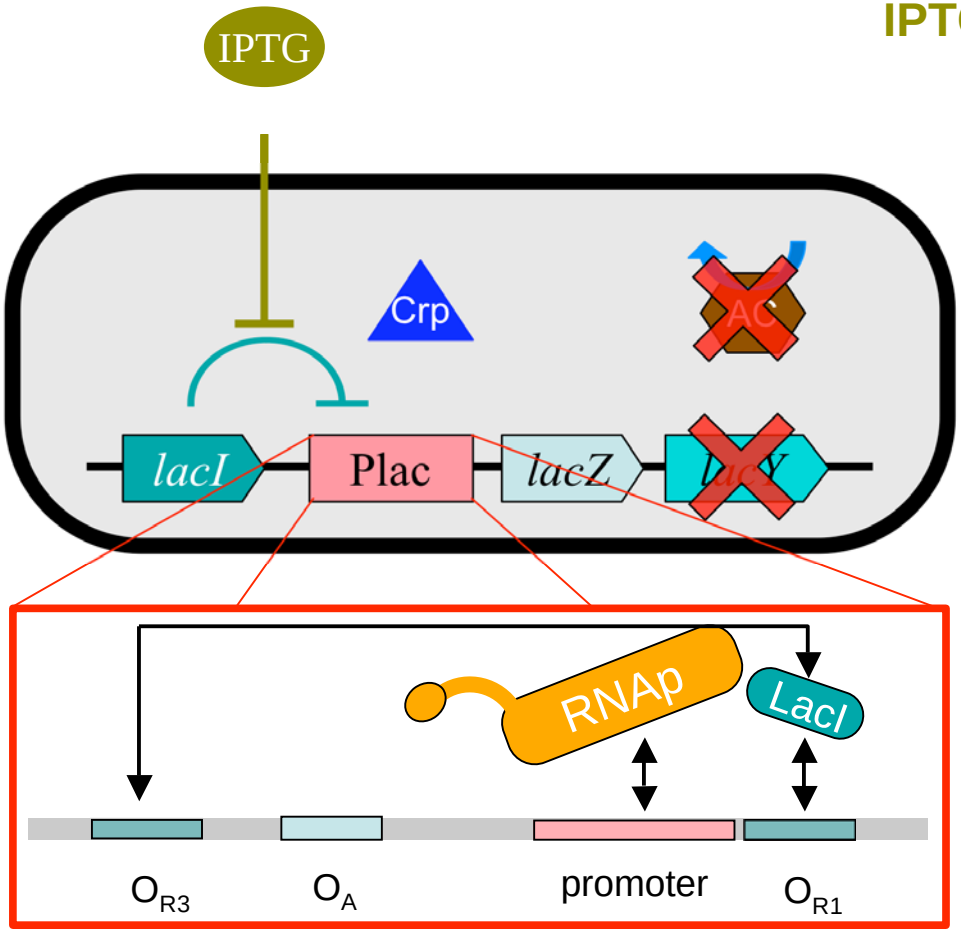
active repressors

$$[R] = \frac{2 \cdot [\mathrm{LacI}_4]_{\mathrm{total}}}{\left(1 + [\mathrm{IPTG}] / \mathrm{K}_{\mathrm{IPTG}}\right)^2}$$

simple repression

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

Quantitative characterization of mutants



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

➔ increase fold-repression by $\mathcal{L}_o$-fold

➔ effective Hill coeff (1.5 ~ 3) depends on $\mathcal{L}_o$

but value of $\mathcal{L}_o$ not known independently

IPTG dependence: *cyaA-* 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

  $LacI_4 + IPTG \rightleftharpoons LacI_4{:}IPTG$
- 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

active repressors $$[R] = \frac{2 \cdot [\mathrm{LacI}_4]_{\mathrm{total}}}{\left(1 + [\mathrm{IPTG}] / \mathrm{K}_{\mathrm{IPTG}}\right)^2}$$

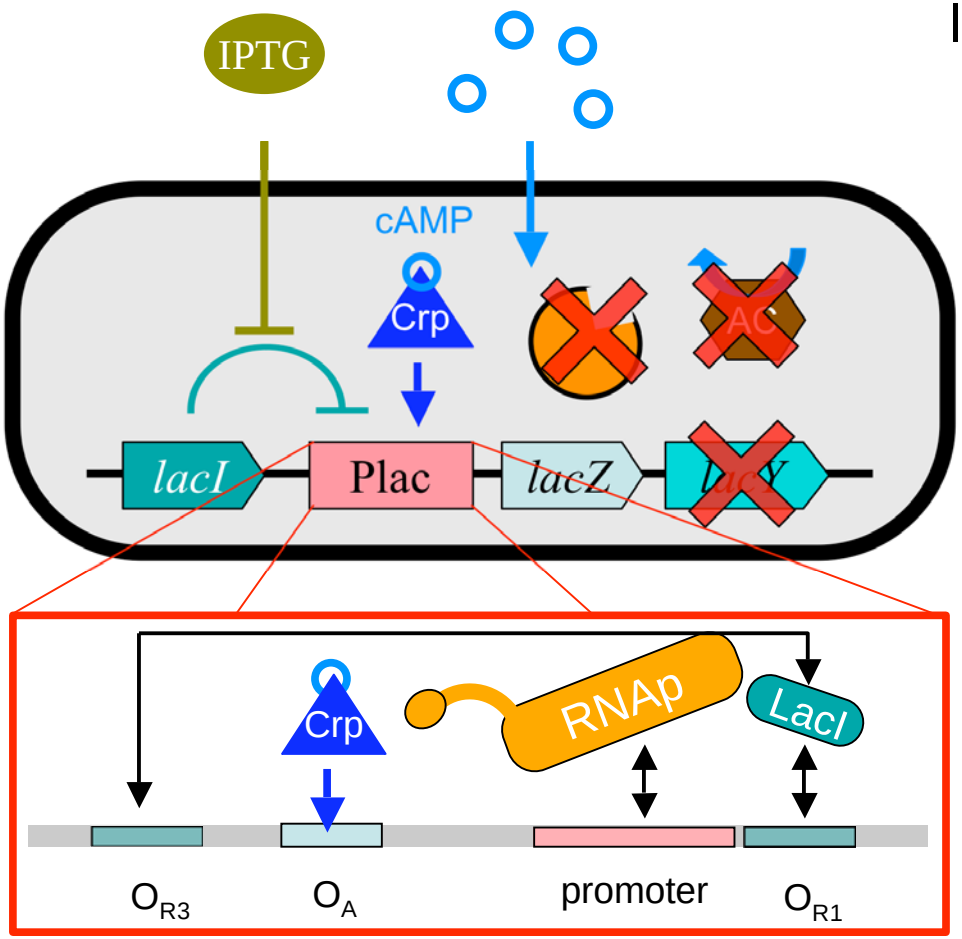
simple repression $$\text{tsx activity} \propto \frac{1}{1 + [R] / K_R}$$

- include DNA looping in model

$$[R] \rightarrow [R] + \frac{\mathcal{L}_o \cdot [\mathrm{LacI}_4]_{\mathrm{total}}}{\left(1 + [\mathrm{IPTG}] / \mathrm{K}_{\mathrm{IPTG}}\right)^4}$$

$\mathcal{L}_o$: local increase of [LacI] due to looping

Quantitative characterization of mutants



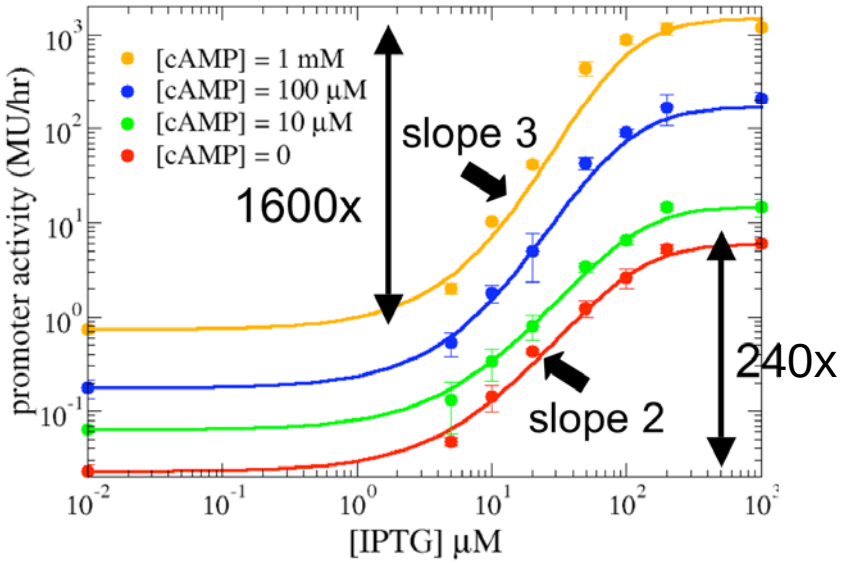
auxiliary Lac operators stabilize LacI-O1 binding via **DNA looping** [Muller-Hill]

➔ increase fold-repression by $\mathcal{L}_0$-fold

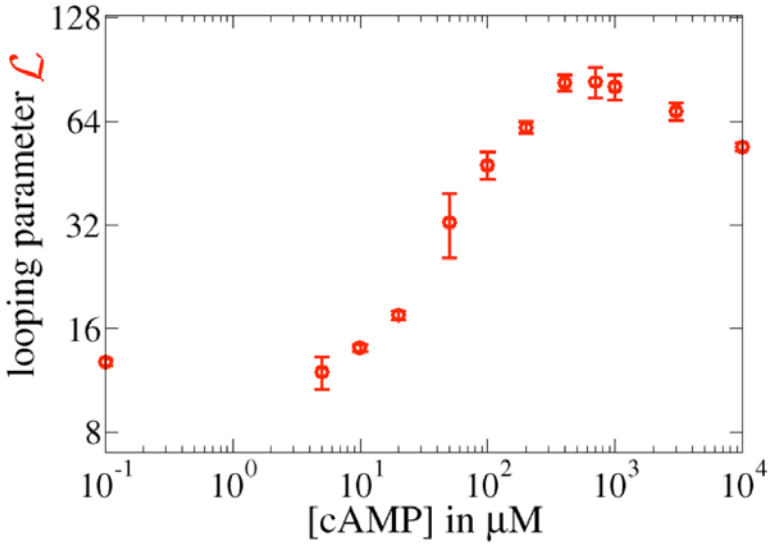
➔ effective Hill coeff (1.5 ~ 3) depends on $\mathcal{L}_0$

but value of $\mathcal{L}_0$ not known independently

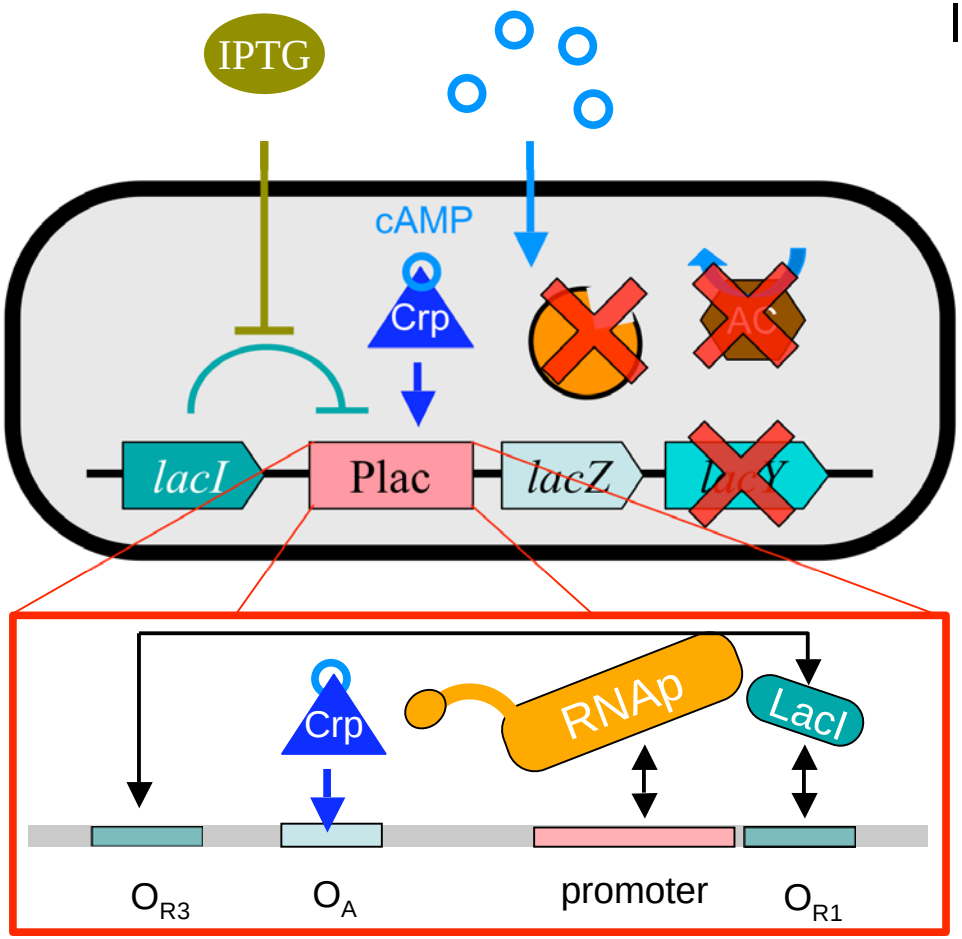
looping model w/ $\mathcal{L}_0 \approx 12$, $2[\mathrm{LacI}_4]/K_R = 20$



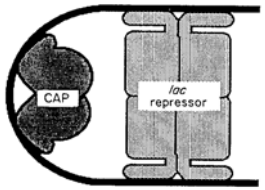
➔ single parameter $\mathcal{L}_0$ fits both fold-repression and slope



Quantitative characterization of mutants



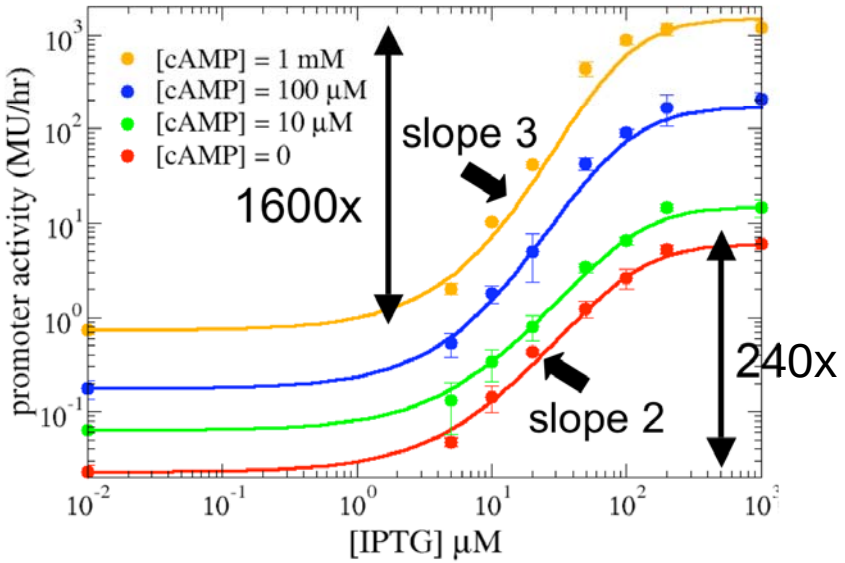
Crp-dependence of DNA looping



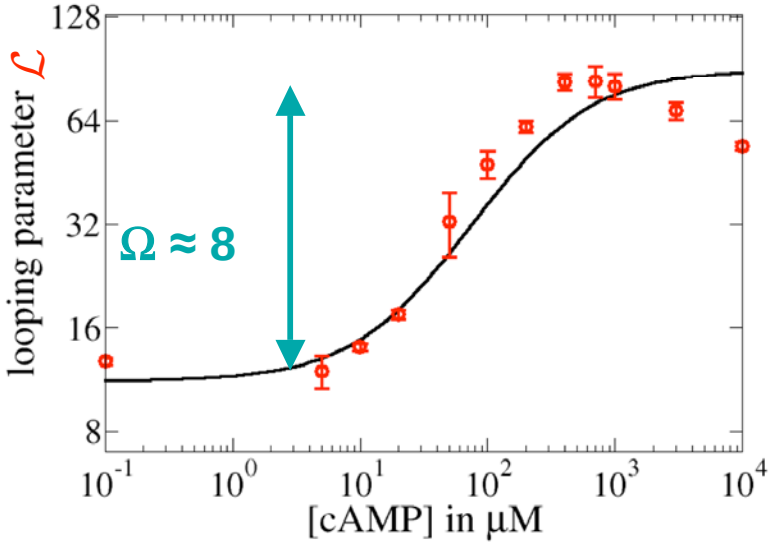
Fried et al, 84;
Balaeff et al, 04

in vitro study found coop. factor **Ω = 4 ~12**

looping model w/ $\mathcal{L}_0 \approx 12$, $2[\mathrm{LacI}_4]/K_R = 20$

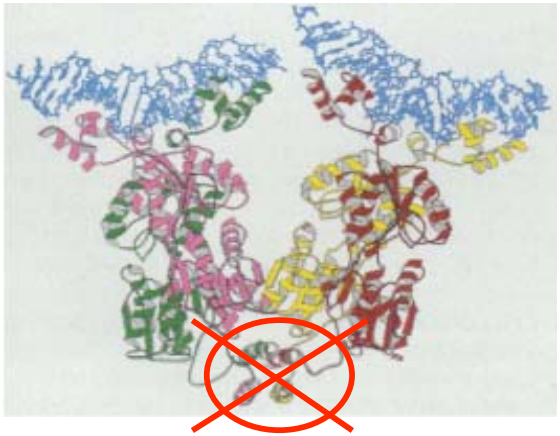


➔ single parameter $\mathcal{L}_0$ fits both fold-repression and slope

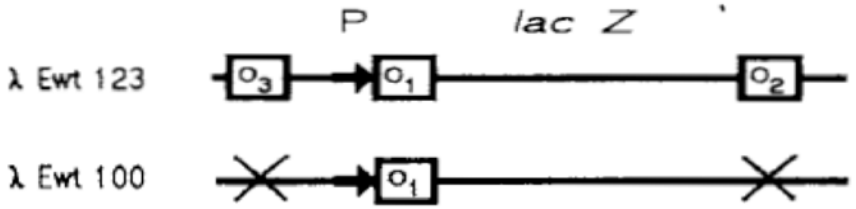


Direct probe of DNA looping *in vivo*

Use dimeric LacI mutant



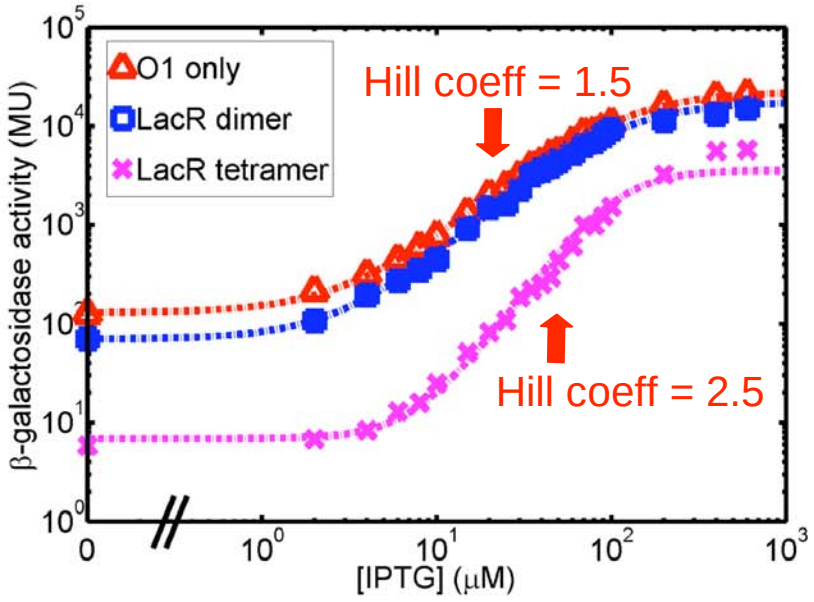
remove auxiliary operators

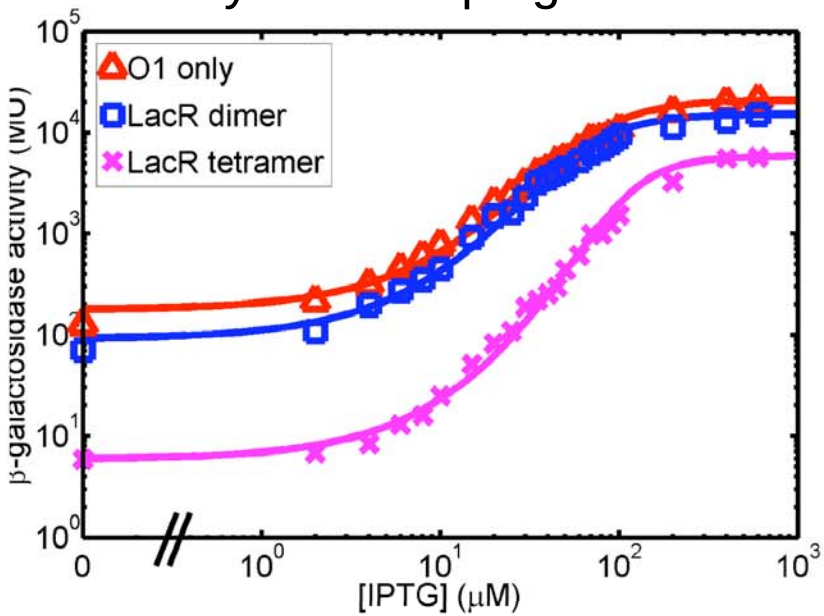


➔ cooperativity in IPTG response requires DNA looping (Lac tetramer + auxiliary ops)

[Oehler & Muller-Hill, 06]

data well-fitted by DNA looping model

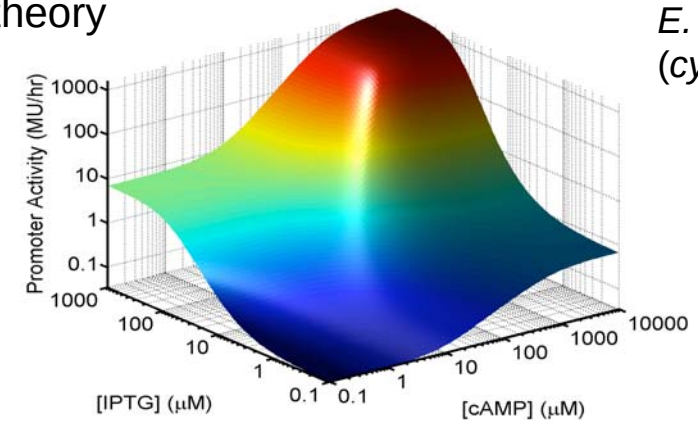




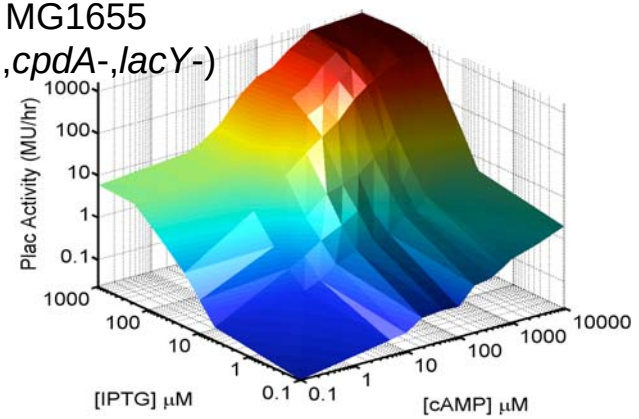
➔ IPTG-LacI-operator interaction same as *in vitro*

Summary

theory



E. coli MG1655
(*cyaA*-,*cpdA*-,*lacY*-)



- **main findings for the *lac* promoter:**
  - Crp enhances DNA looping
  - abrupt IPTG response despite non-cooperative LacI-IPTG interaction;

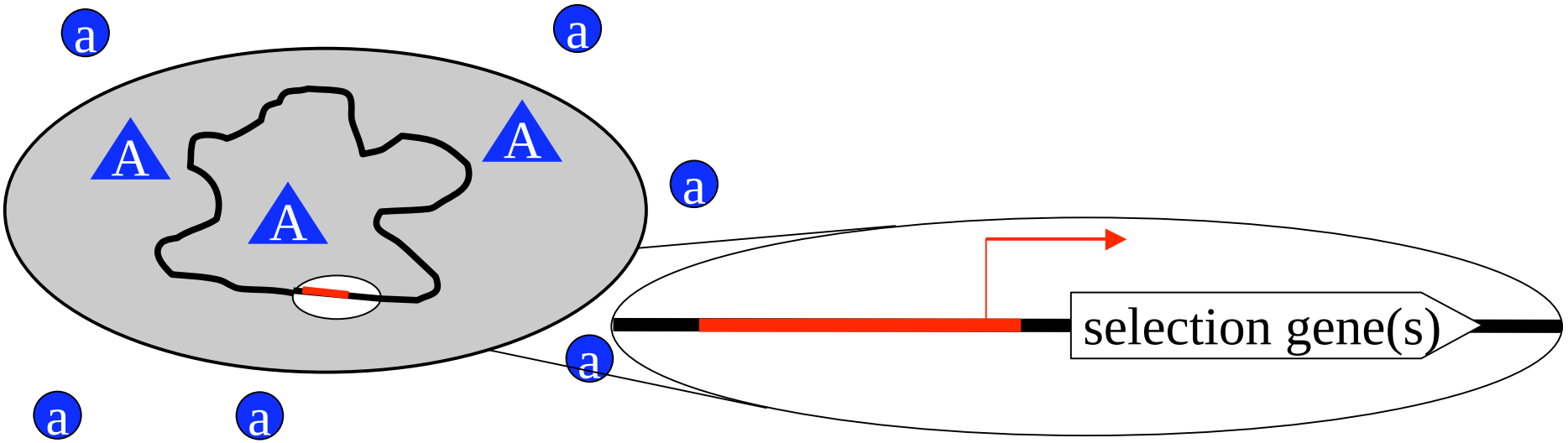
  ➔ **suggests physiological role of Crp-cAMP as enhancer of repression**
  - mechanism of Crp-LacI interaction?
  - coop cAMP response due to PDE; physiological function? mechanism?
- **general lessons for quantitative systems biology:**
  - hidden interaction abound even for the “best studied” system
  - 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

***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**

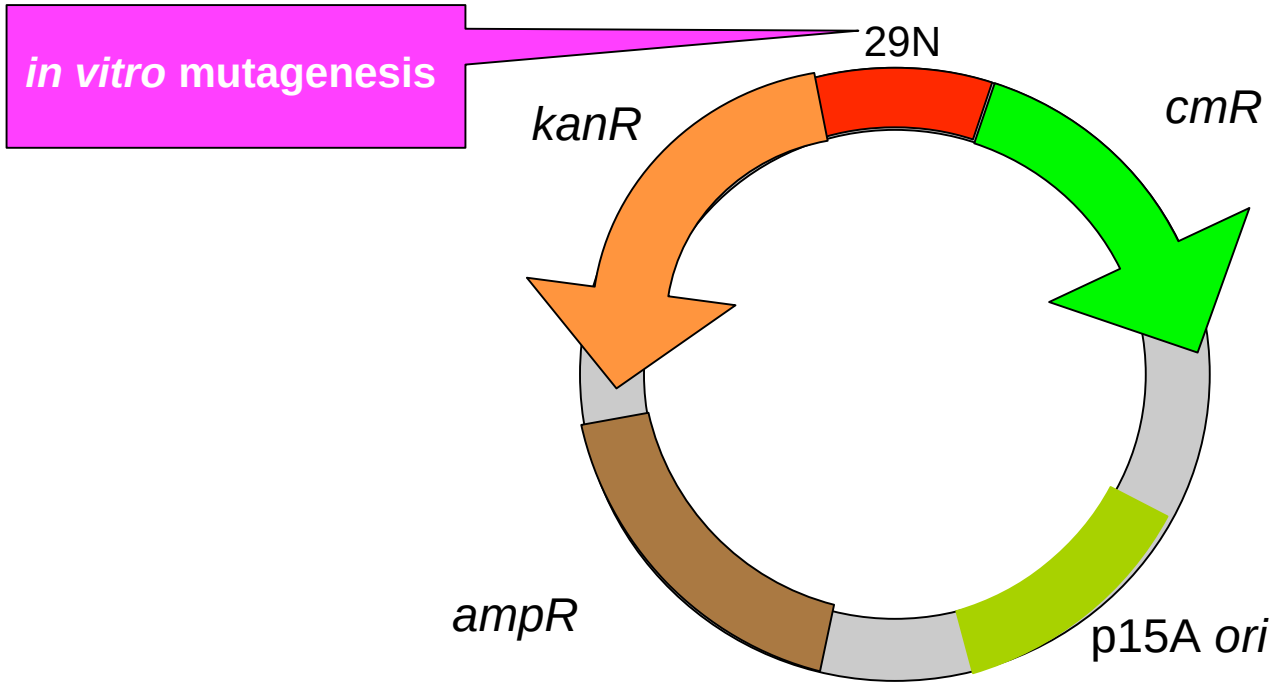
Selectable output:
-- gene product lethal if drug 1 present
-- gene product essential if drug 2 present

Defined region of mutagenesis

e.g, inverter gate

[a]	drug	gene
lo	1	**OFF**
hi	2	**ON**

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 σ^{70} core promoter:

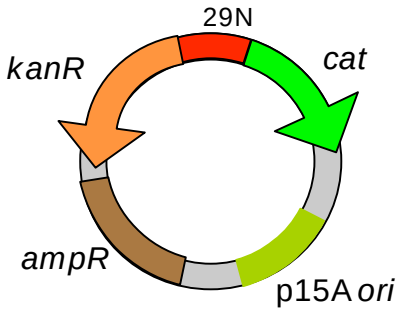
-10/-35 hexamers: **TTGACA**<-- 17nt -->**TATAAT**
extended -10: TG**TG**N**TATAAT**

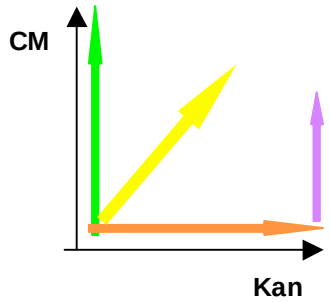
➔ two selection genes: divergent overlapping promoters possible?

➔ dependence on evolutionary path?

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**





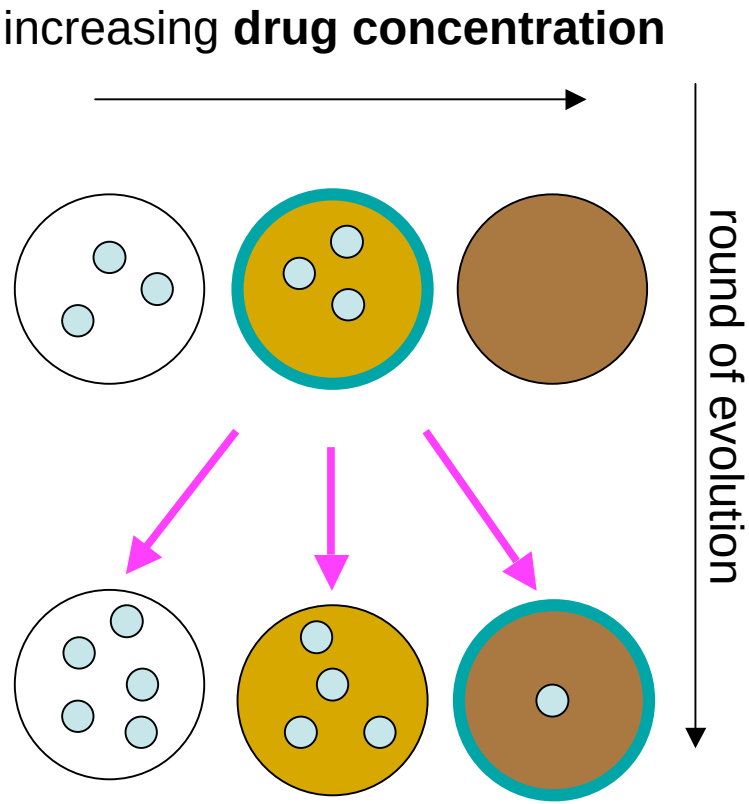
- **selection**
  - grow on plates with various drug conc (CM and/or kan)
  - collect several hundred clones with the highest drug resistance

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

- **selection**

...

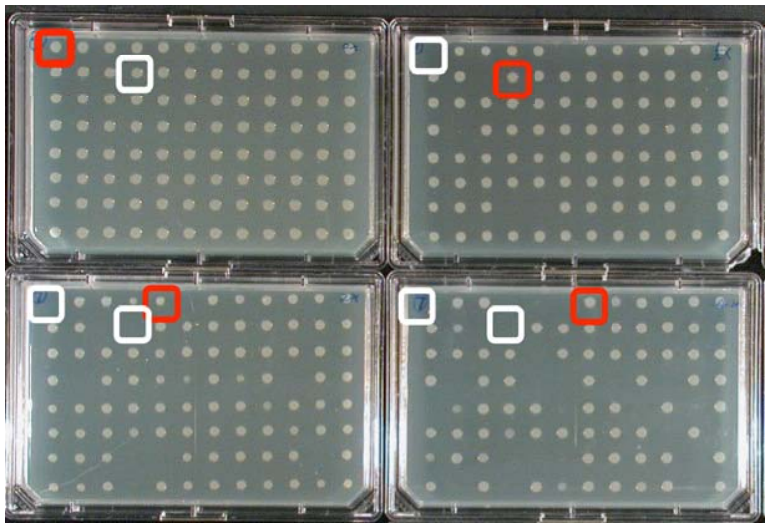
all intermediate clones “saved” for future analysis



Semi-quantitative phenotype assay

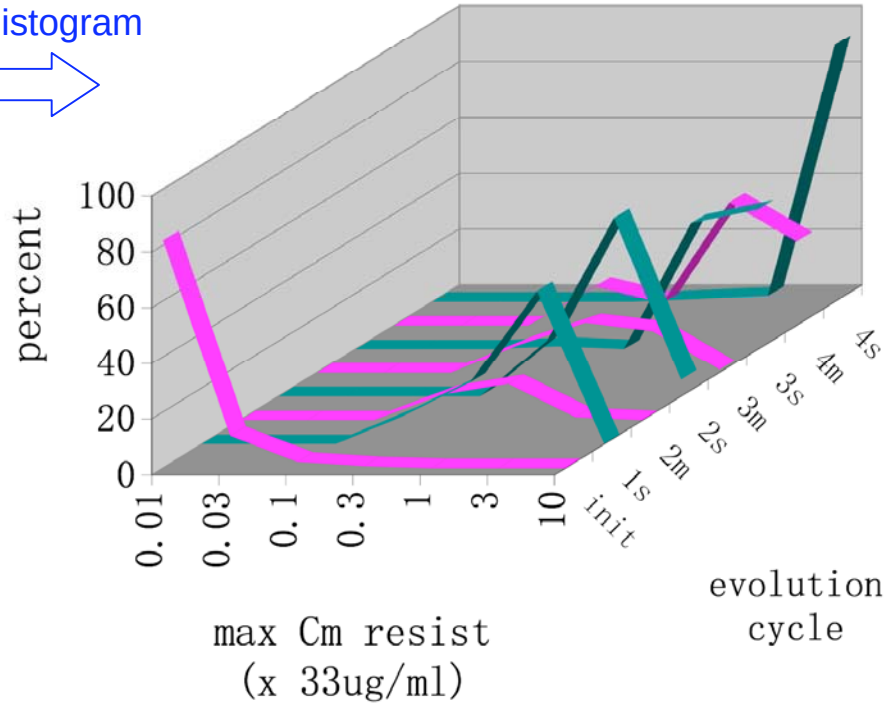
Characterize distribution of phenotypes at each stage of evolution

Cm = 0 Cm = 10 μg/ml



Cm = 33 μg/ml Cm = 100 μg/ml

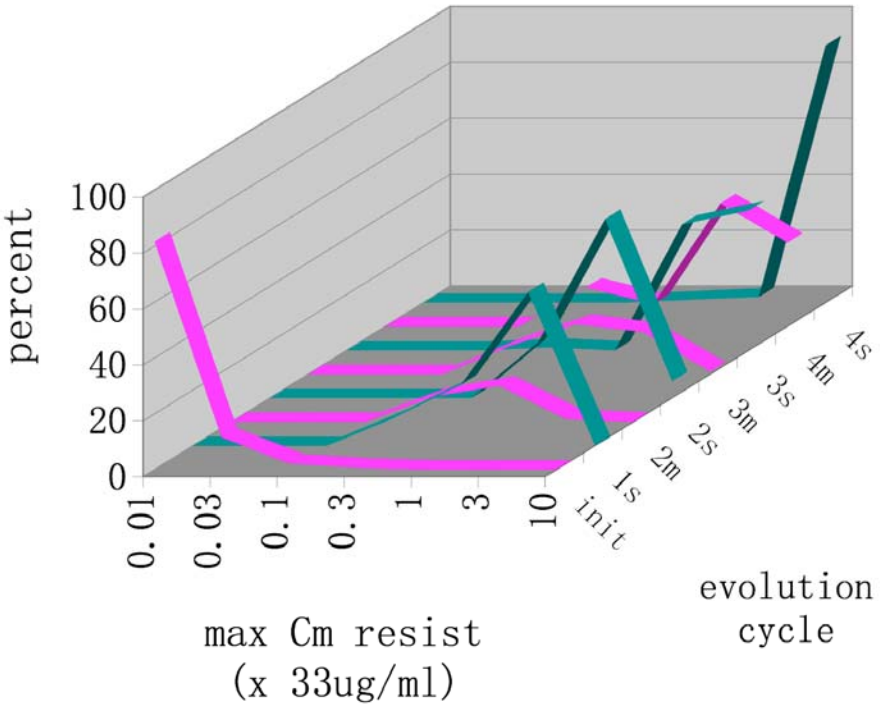
make histogram

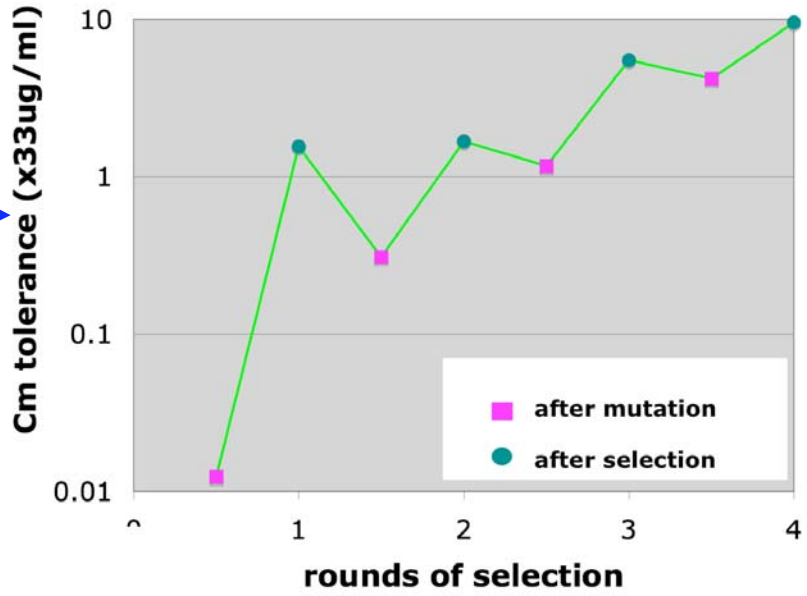


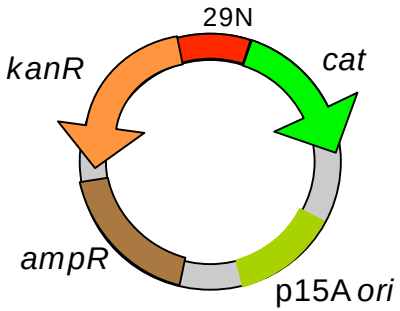
- collect 96 clones
- grow on agar plates with different drug conc
- identify max drug resistance

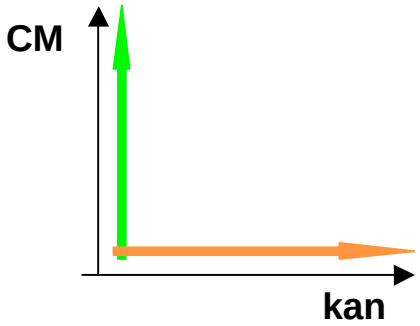
▢ : Max drug resistance for the clone

Evolution in single direction: phenotype

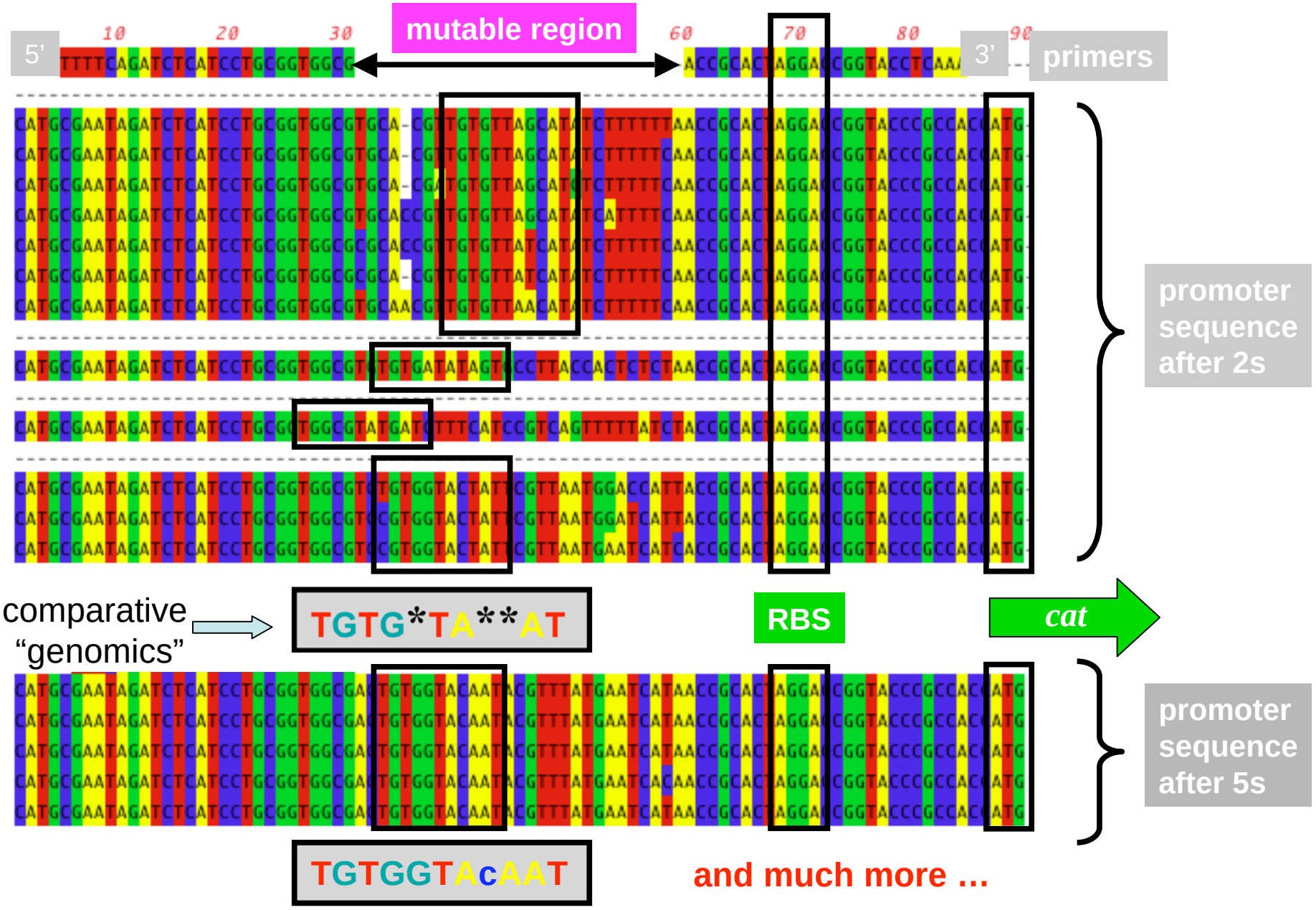








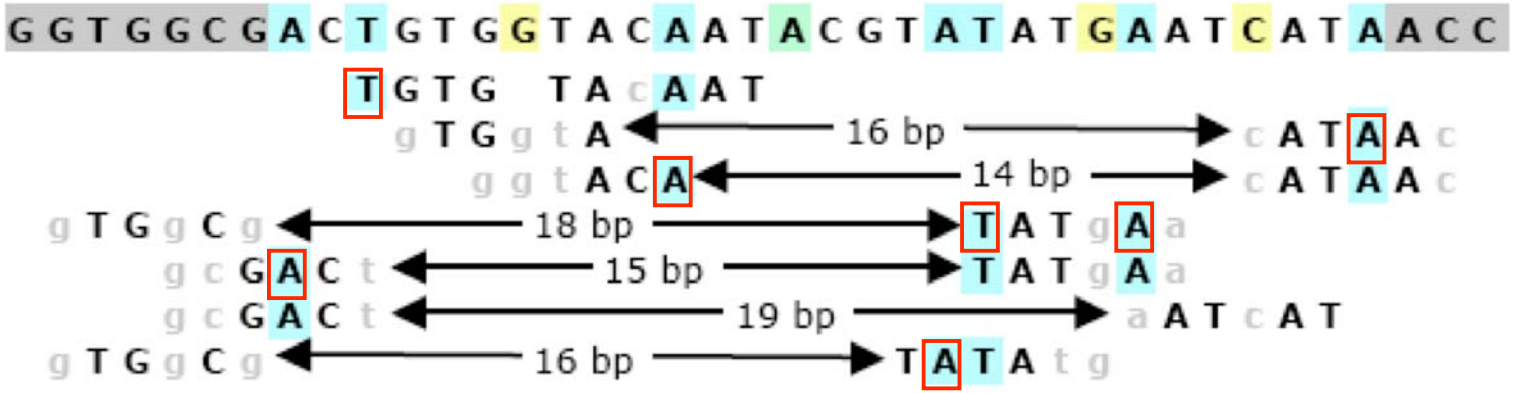
Evolution in single direction (CM): genotype



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?

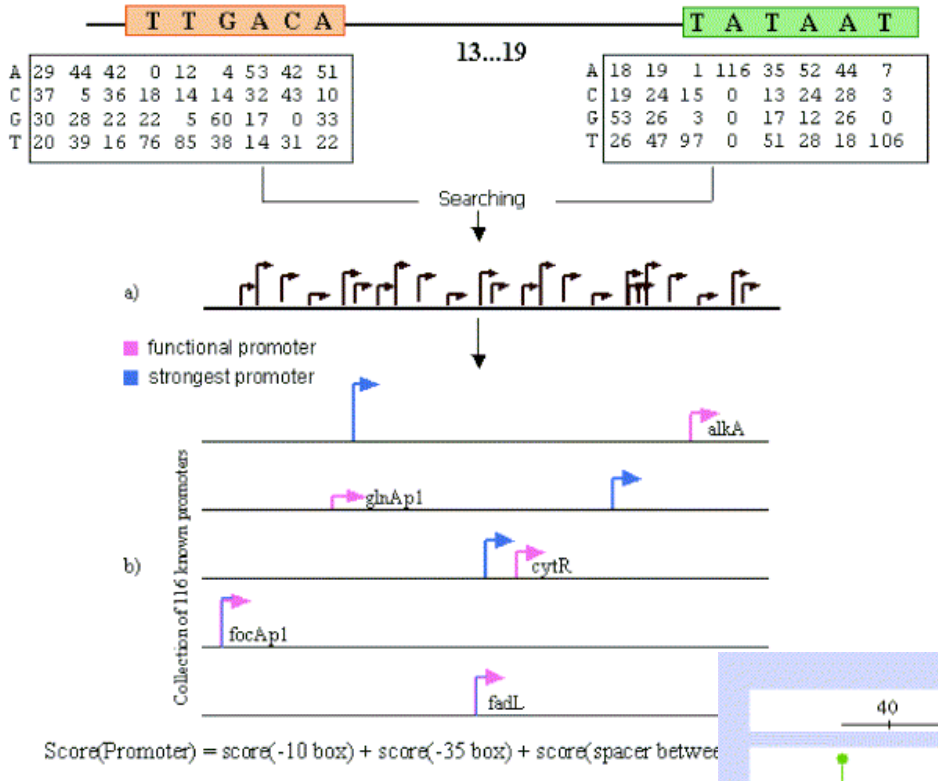
-- stronger expression from multiple promoters?

-- robustness to mutation provided by multiple copies?

Benefit: makes subsequent evolution of activators/repressors easier

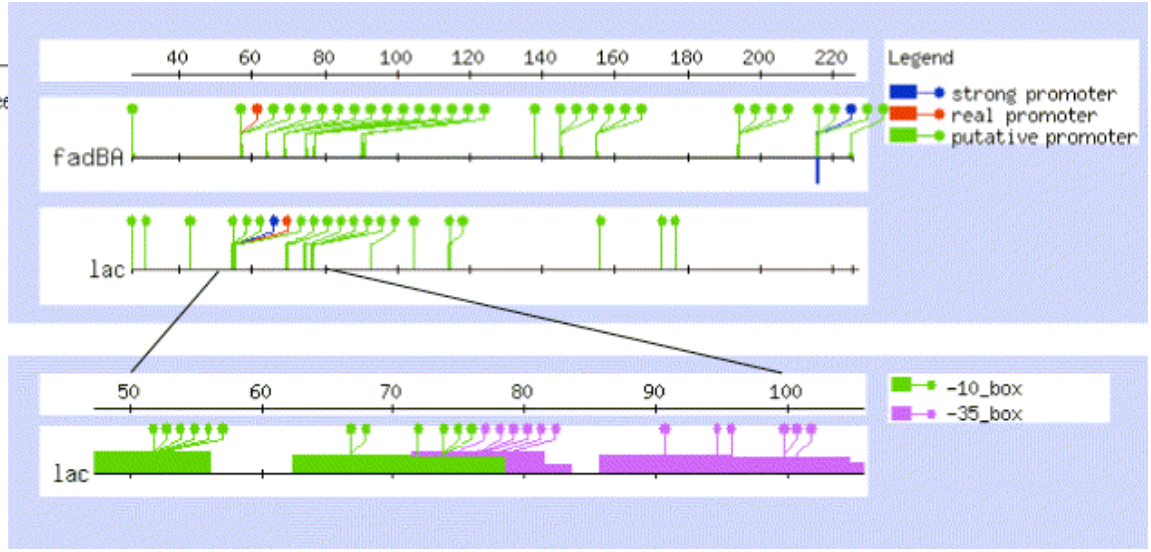
Multiple promoters seen in bioinfo studies

[Huerta & Collado-vides, 03]

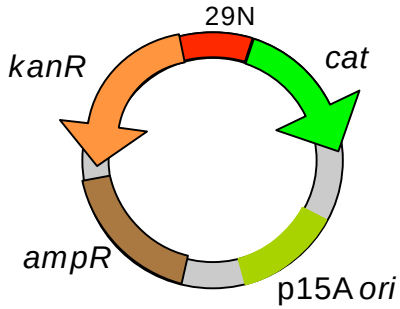


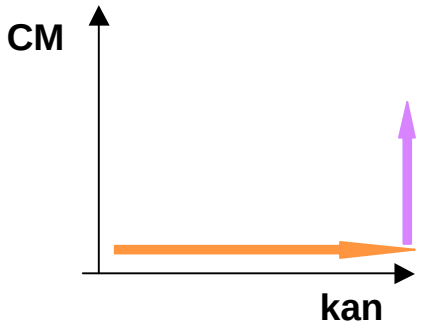
- avg of 38 putative promoter signals observed in a typical 250bp region upstream of gene start;

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

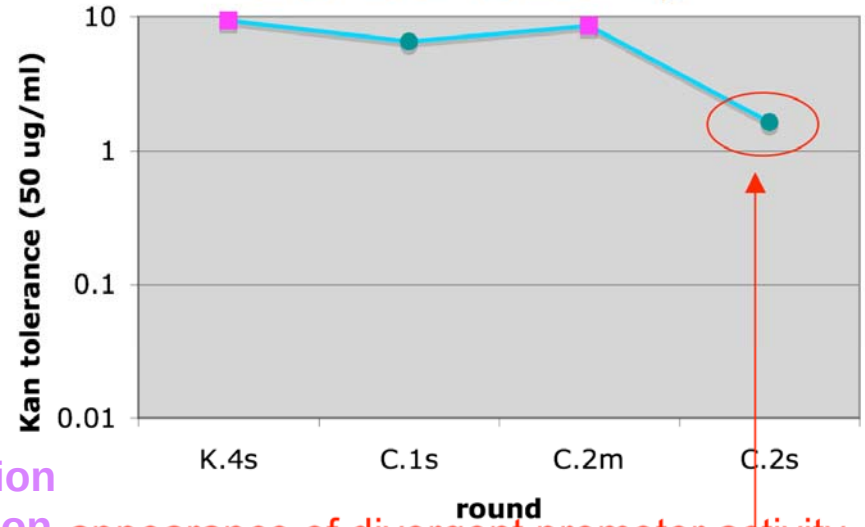


Reversal of evolution direction: phenotype

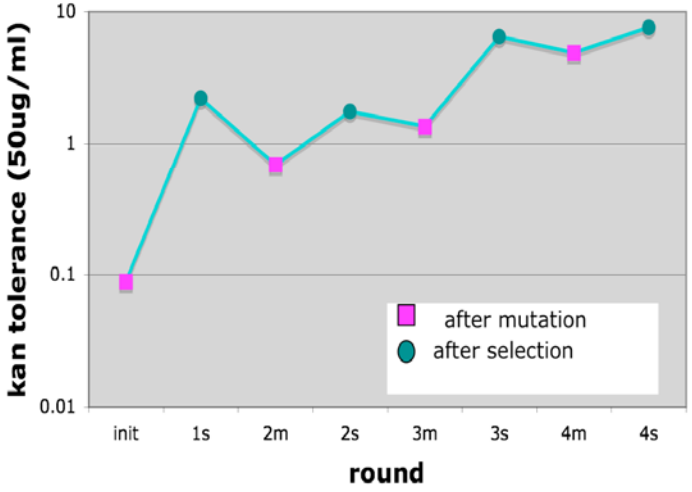




Kan resistance decreasing

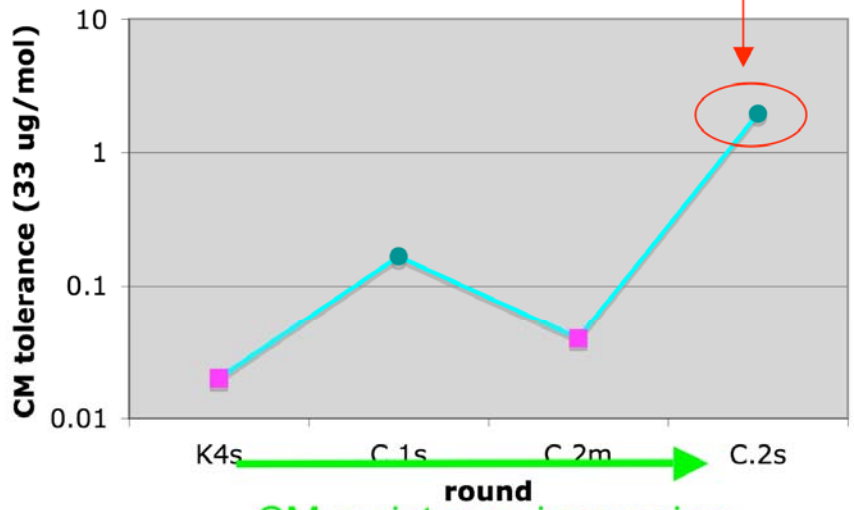


4 rounds of selection in Kan direction



revert selection
to CM direction

appearance of divergent promoter activity



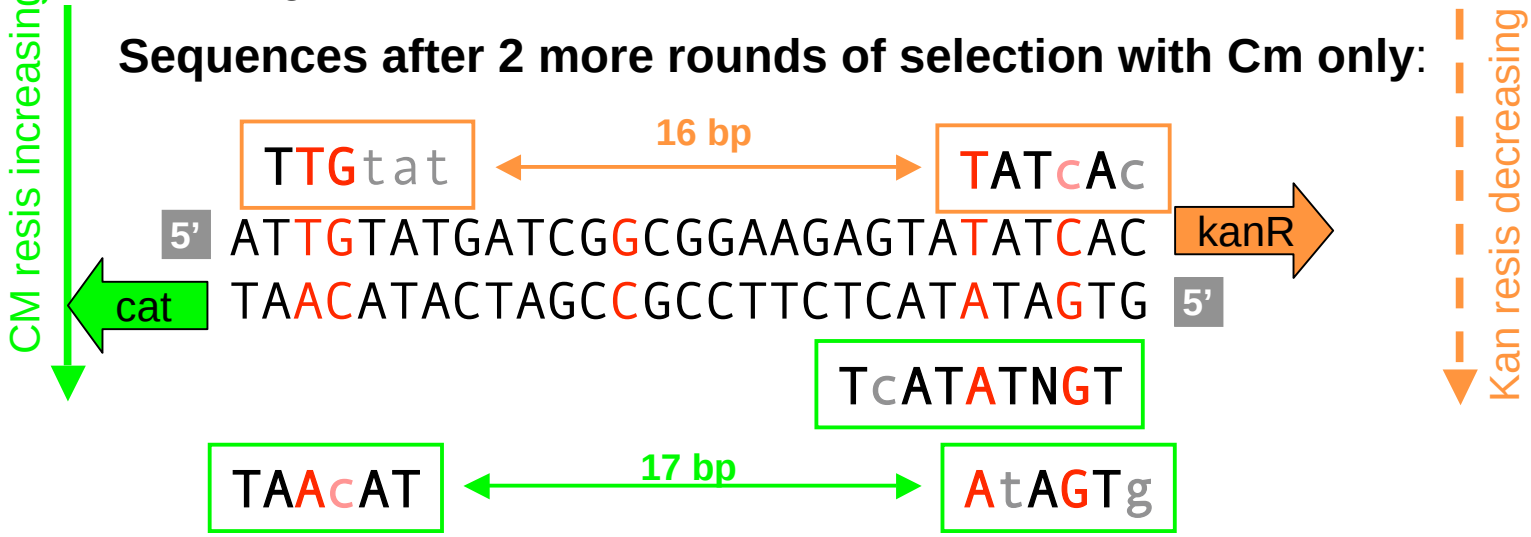
CM resistance increasing

Reversal of evolution direction: genotype

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

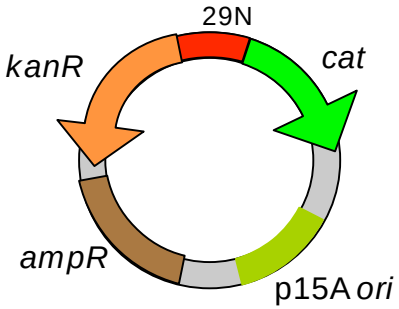
-- one **extended -10 motif** and one **-35/-10 motif** in the Kan direction
-- no significant motifs detected in the Cm direction

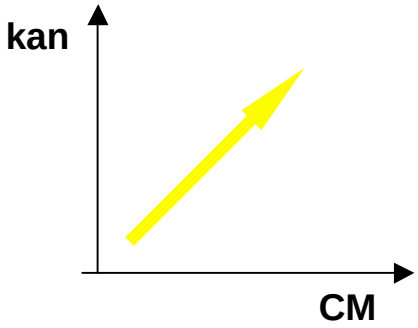
Sequences after 2 more rounds of selection with Cm only:



-- one **extended -10** and one standard **-35/-10 motifs** in Cm direction
-- weakened **-10/-35 motif** and lost extended -10 and in the Kan direction

Evolution in both directions: phenotype

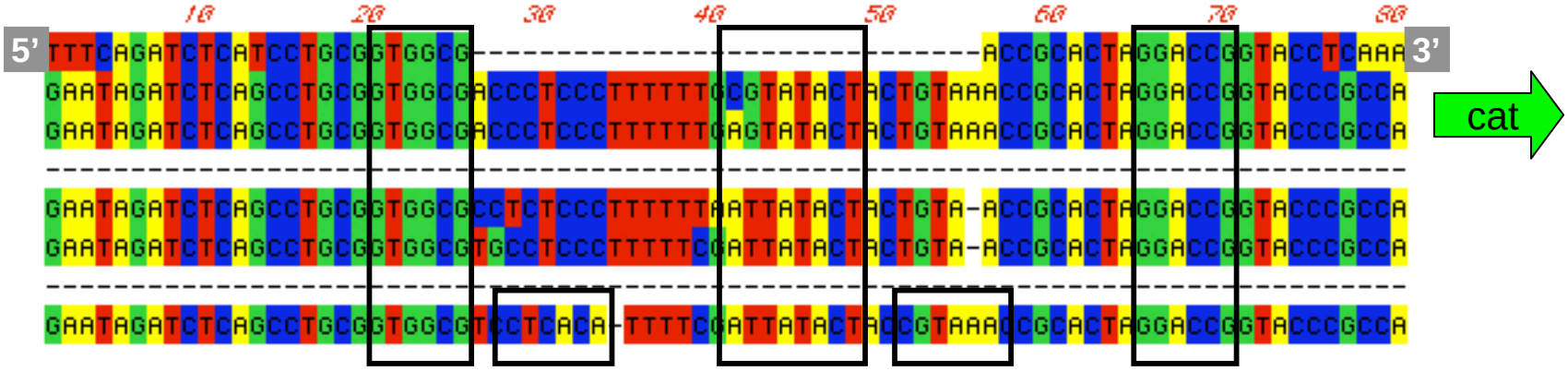




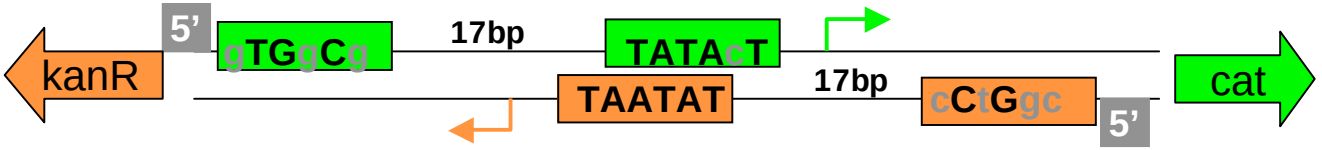
➔ evolution slightly slower than that driven in single direction (5 vs 4 rounds)

Evolution in both directions: genotype (5 rounds)

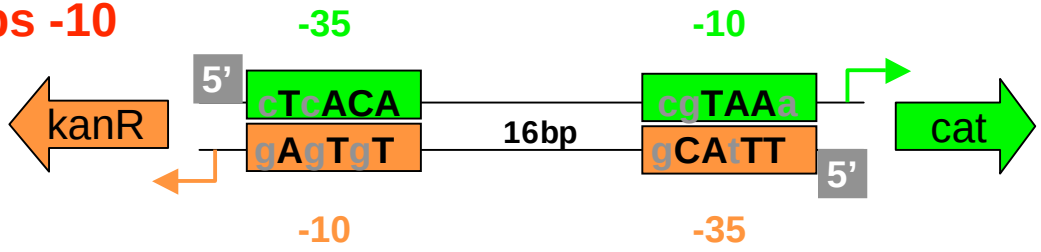
➔ **found two types of overlapping motifs:**



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

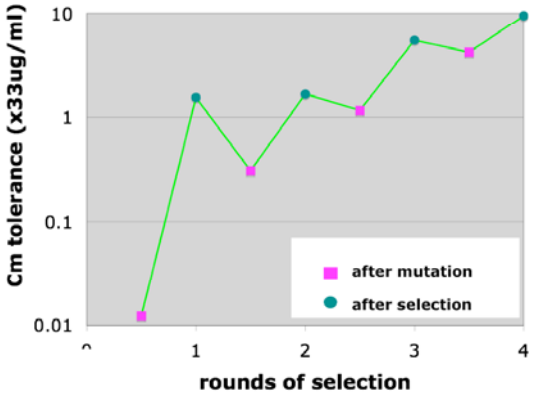


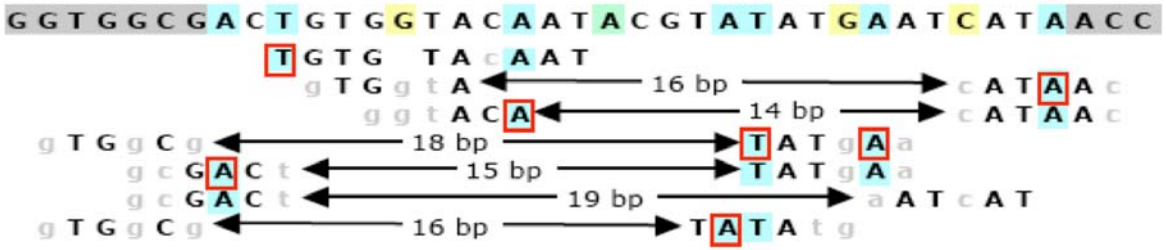
-35 overlaps -10



Summary: promoters are flexible!

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

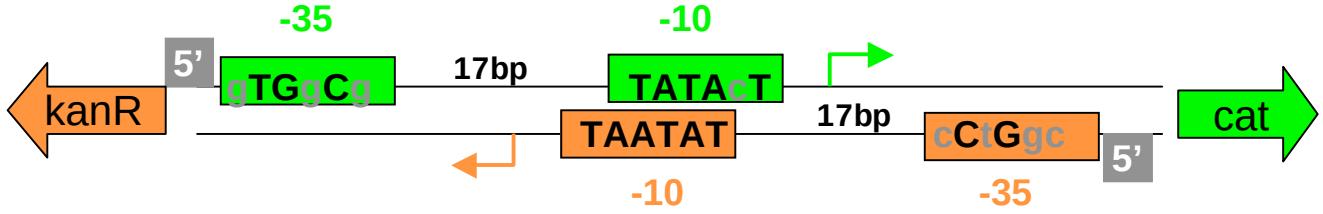




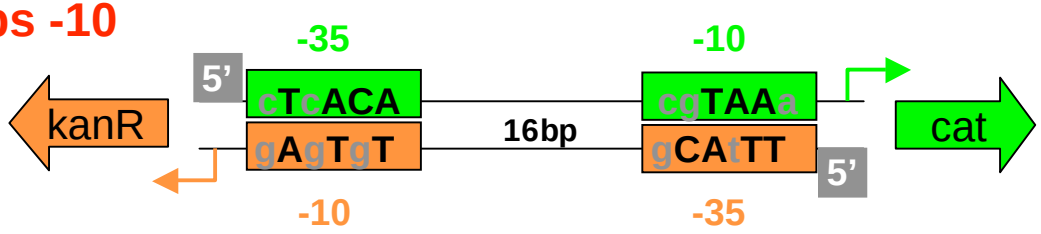
- **Reversal:**
  - existing promoter evolve quickly to reverse direction by few mutations
  - reduction of promoter activity in the reverse direction important (occlusion)

- **Divergent overlapping promoters:**

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



-35 overlaps -10

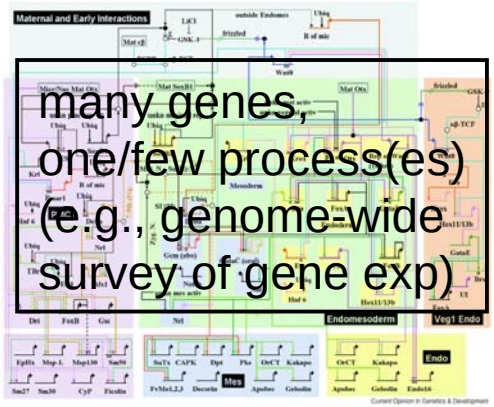


From molecules to system-level functions

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

high throughput methods

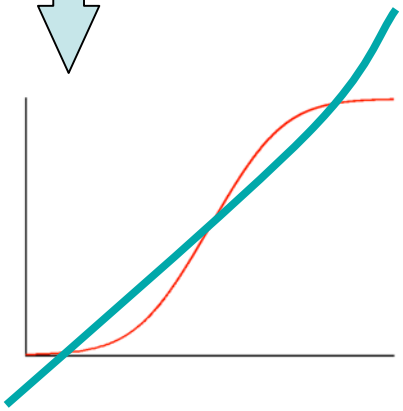
bioinformatic analysis



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

who talks to whom

quantitative analysis
of individual nodes
and small circuits

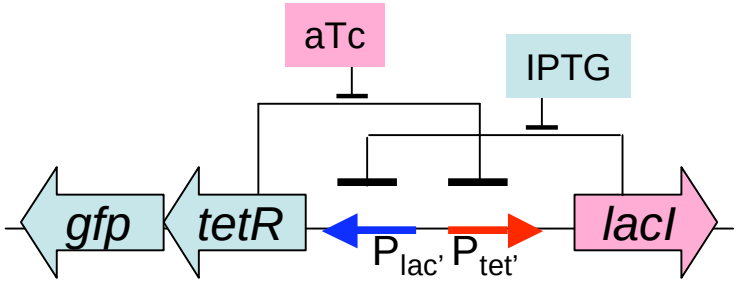


how they talk

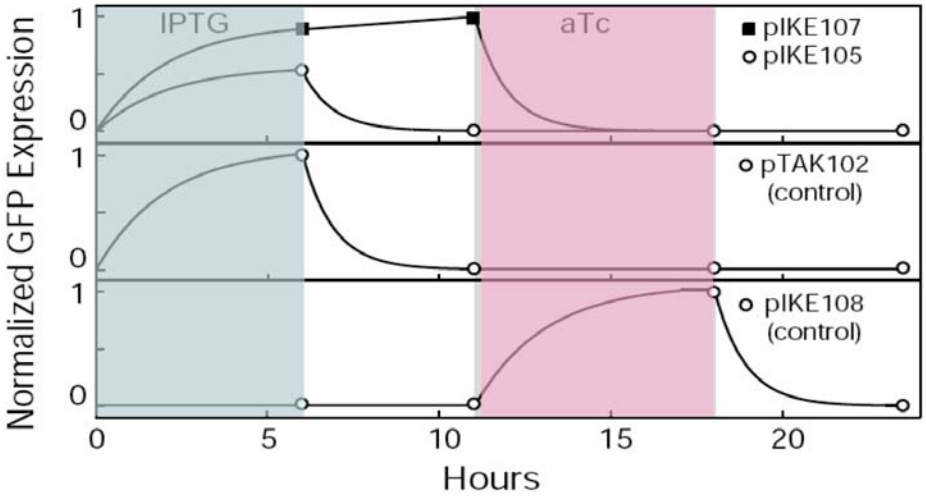
qualitative system-level properties
depend **quantitatively** on
the degree of regulation

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

Synthetic genetic switch



[Gardner, Cantor and Collins, Nature 2000]



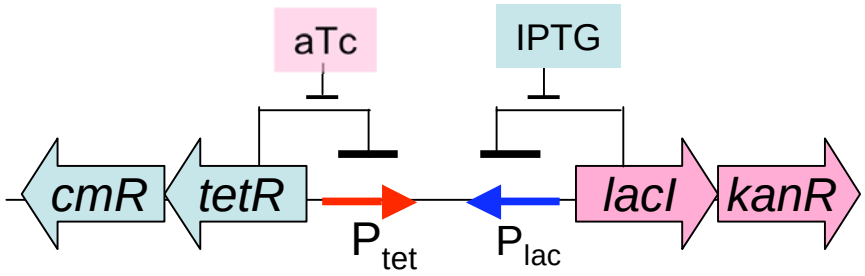
- induction time to switch: ~ 6 hrs (several cell divisions)
- **slow** speed possibly due to passive dilution

➔"speed limit of gene regulation" [Rosenfeld et al, Science, 2005]

Natural switches (e.g., phage lambda)
- induction time to switch: < 10 min
- ingredients for fast speed
  - – proteolysis
  - – auto-activation and repression

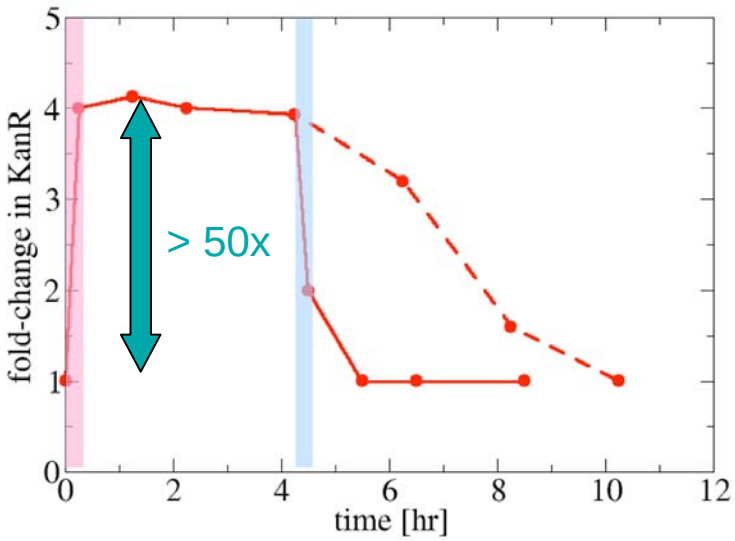
Q: faster switch using the same components?

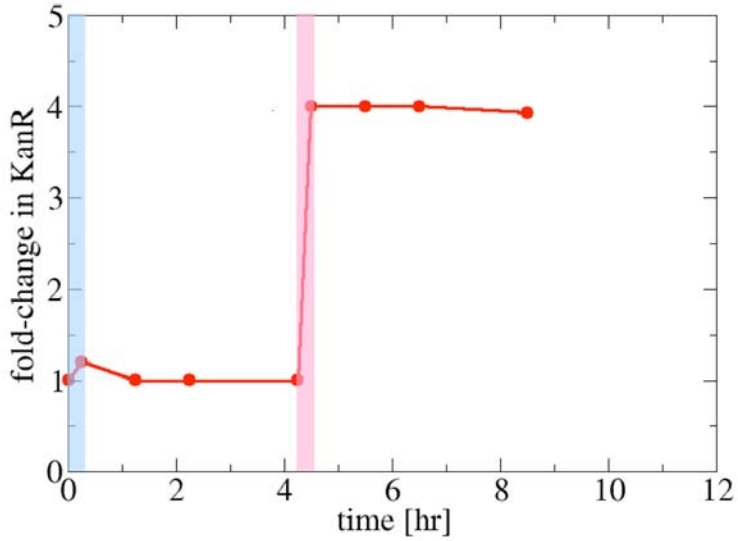
Alternative switch: face-to-face promoter construct



generate variants, screen for desired phenotype

	aTc	IPTG
KanR	growth	
CmR		growth





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

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*