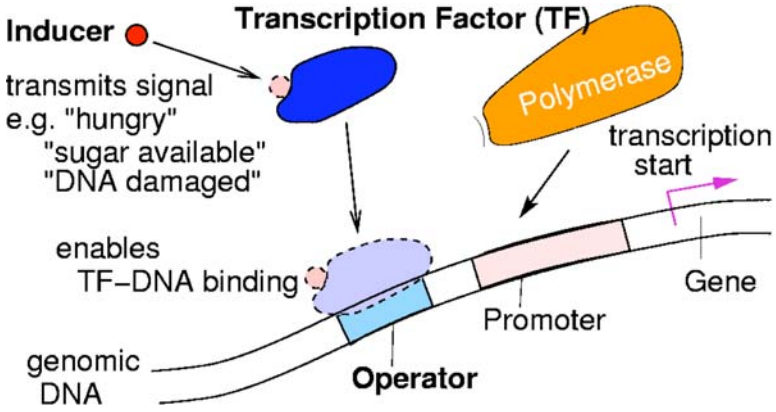
Transcriptional regulation in bacteria



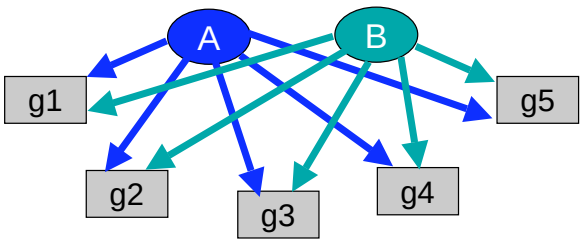
- activation of transcription factors (TF)
- binding of TF(s) to DNA targets (~10 bp)
- enhance transcription by recruiting RNA polymerase (RNAp) to its target ("promoter")

➔ one TF can regulate (activate or repress) multiple promoters

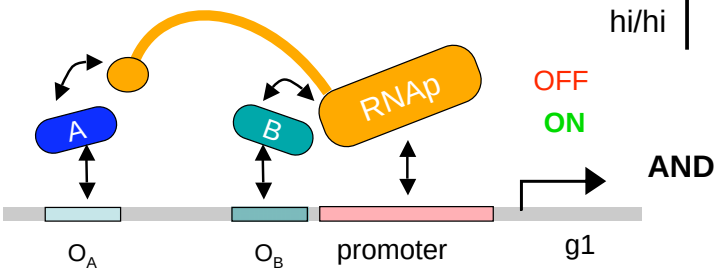
➔ some promoters regulated by multiple TFs

➔ regulated gene may encode another regulator

Illustration of combinatorial control:



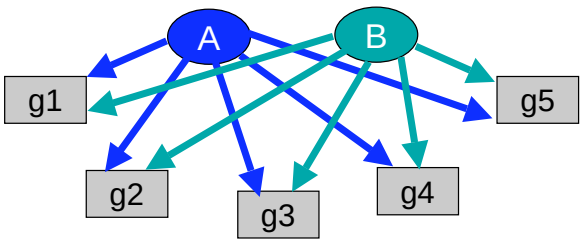
	AND	OR	NAND	XOR	EQ
A/B	g1	g2	g3	g4	g5
lo/lo	OFF	OFF	ON	OFF	ON
lo/hi	OFF	ON	ON	ON	OFF
hi/lo	OFF	ON	ON	ON	OFF
hi/hi	ON	ON	OFF	OFF	ON



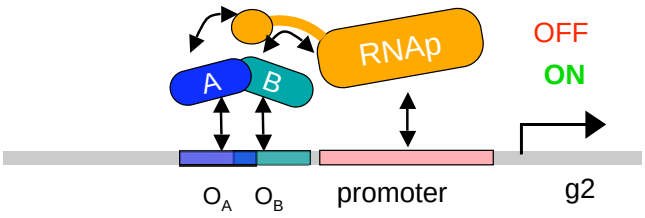
weak promoter;
weak operator sites

synergistic activation when **both** A and B are present
[c.f.: Hochschild et al, Busby et al]

Illustration of combinatorial control:



	AND	OR	NAND	XOR	EQ
A/B	g1	g2	g3	g4	g5
lo/lo	OFF	OFF	ON	OFF	ON
lo/hi	OFF	ON	ON	ON	OFF
hi/lo	OFF	ON	ON	ON	OFF
hi/hi	ON	ON	OFF	OFF	ON



weak promoter;
strong operator sites

mutual exclusion of A and B:
suppresses elevated promoter activity when both A and B are present

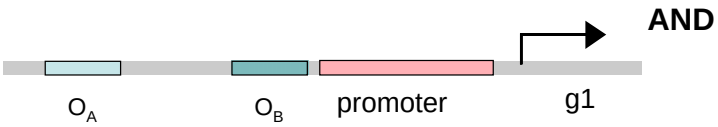
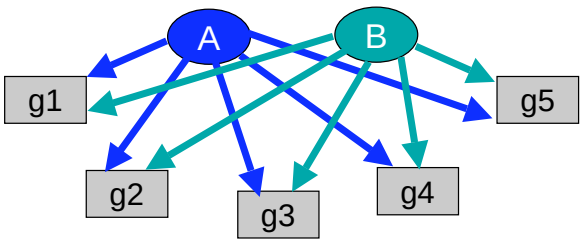
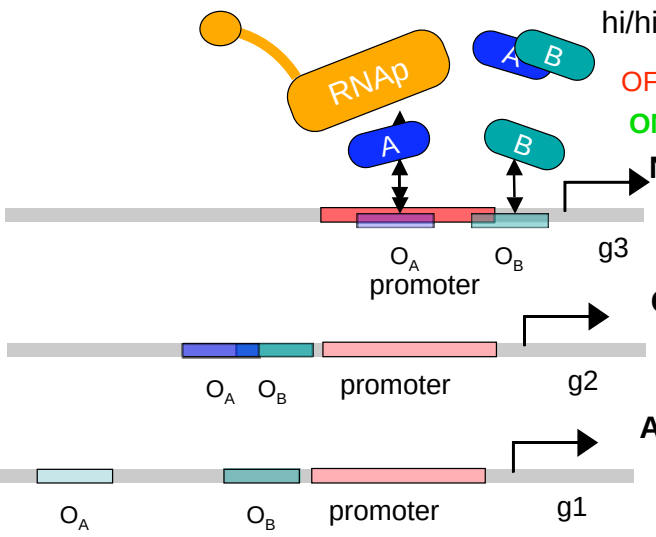


Illustration of combinatorial control:



	AND	OR	NAND	XOR	EQ
A/B	g1	g2	g3	g4	g5
lo/lo	OFF	OFF	ON	OFF	ON
lo/hi	OFF	ON	ON	ON	OFF
hi/lo	OFF	ON	ON	ON	OFF
hi/hi	ON	ON	OFF	OFF	ON



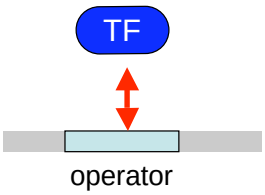
strong promoter;
weak operator sites

"collaborative repulsion" of RNAp by A and B without the need of direct cooperative interaction
[c.f. J. Widom]

Quantitative model:

programmable
molecular interaction

- specific protein-DNA binding:

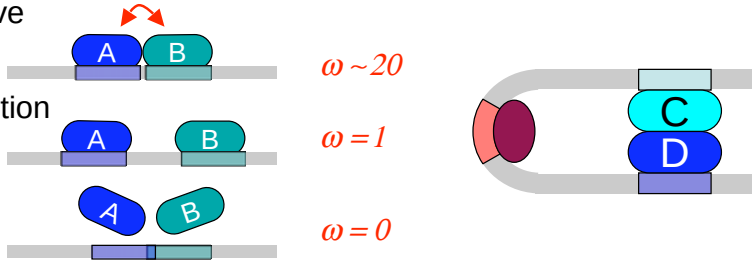


binding probability: $p = \frac{1}{1+e^{\beta(\varepsilon+\mu)}} \approx \frac{[TF]}{[TF]+K}$

K: effective binding const ***in vivo***
(i.e., w.r.t. genomic background)
tunable via choice of operator seq

- protein-protein interaction: $\omega = e^{-\beta\varepsilon_{\text{int}}}$ tunable via site placements

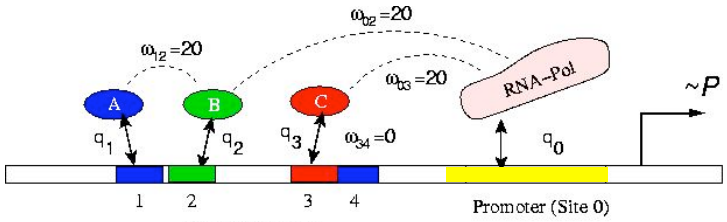
cooperative $\omega \sim 20$
no interaction $\omega = 1$
repulsive $\omega = 0$



- DNA looping tunable via placement of DNA-bending proteins
- applicability of thermodynamics

In general:

- binding sites: $j = \{1, \ldots, L\}$
- binding constants: K_j
- TF conc: $[TF_{\alpha(j)}]$



Let $\sigma_j = \{0,1\}$ represent the occupation state of each site j

➔ weight for TF binding: $W_{\text{unocc}}\left[\sigma_1,\ldots,\sigma_L\right] = \prod_{j=1}^{L}\left([TF_{\alpha(j)}]/K_j\right)^{\sigma_j} \cdot \prod_{i<j}\omega_{i,j}^{\sigma_i\sigma_j}$

➔ weight for promoter occupation:

$$W_{\text{occ}}\left[\sigma_1,\ldots,\sigma_L\right] = \frac{[RNAp]}{K_p}\prod_{j=1}^{L}\left[1-\sigma_j\,\delta\left(\omega_{j-p},0\right)\right]\cdot\left[1+\sum_{j=1}^{L}\sigma_j\,\omega_{j-p}\right]\cdot W_{\text{unocc}}$$

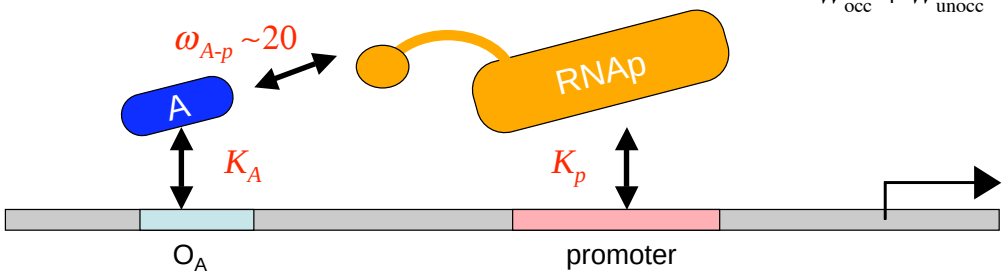
promoter occupation probab. ∝ gene expression [Shea & Ackers, '86]

➡ $$P\left\{\left[TF_\alpha\right] \mid K_j, \omega_{i,j}\right\} = \frac{W_{\text{occ}}}{W_{\text{occ}}+W_{\text{unocc}}}$$

Thermodynamic framework of gene regulation

[Shea & Ackers, JMB 1985]

gene expression ∝ eq. promoter occupation probability $P = \frac{W_{occ}}{W_{occ} + W_{unocc}}$



simple activation:

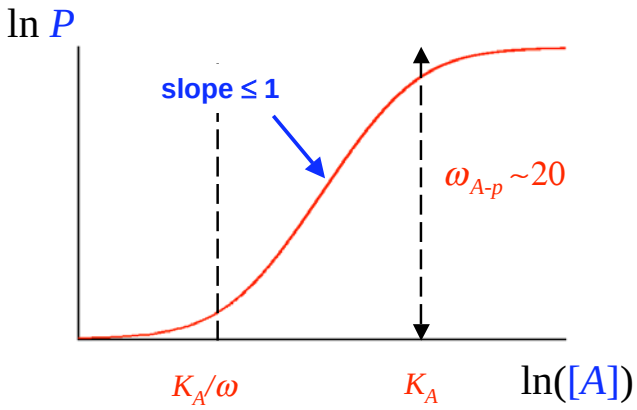
$$W_{unocc} = 1 + \frac{[A]}{K_A}$$

$$W_{occ} = \frac{[RNAp]}{K_p} + \omega_{A-p} \cdot \frac{[A]}{K_A} \cdot \frac{[RNAp]}{K_p}$$

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

(for typical weak promoters)

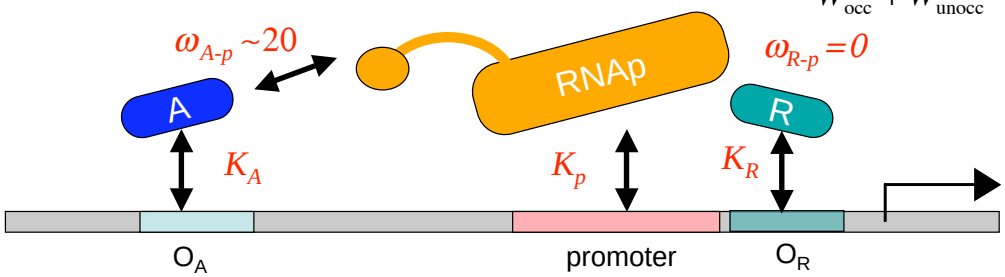
Note: max "fold-change" fixed by $\omega_{A\text{-}p}$



Thermodynamic framework of gene regulation

[Shea & Ackers, JMB 1985]

gene expression ∝ eq. promoter occupation probability $P = \frac{W_{occ}}{W_{occ} + W_{unocc}}$



simple activation:

$$W_{unocc} = 1 + \frac{[A]}{K_A}$$

$$W_{occ} = \frac{[RNAp]}{K_p} + \omega_{A-p} \cdot \frac{[A]}{K_A} \cdot \frac{[RNAp]}{K_p}$$

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

(for typical weak promoters)

Note: max "fold-change" fixed by $\omega_{A\text{-}p}$

simple repression:

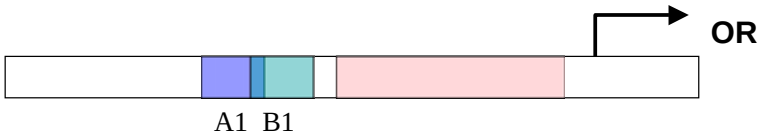
$$P \approx \frac{[RNAp]}{K_p} \cdot \frac{1}{1 + [R]/K_R}$$

co-regulation

multiplicative

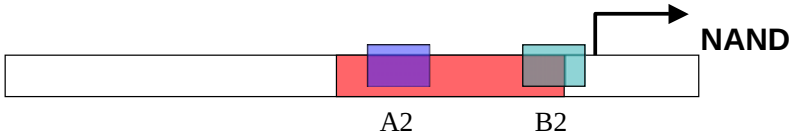
$$P \propto \frac{1 + \omega_{A-p}[A]/K_A}{1 + [A]/K_A} \cdot \frac{1}{1 + [R]/K_R}$$

Quantitative design of combinatorial control



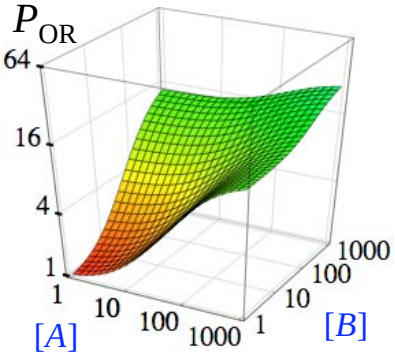
-- weak promoter
$[RNAp]/K_p = 0.001$
-- strong A and B sites
$K_{A1} = K_{B1} = 200$

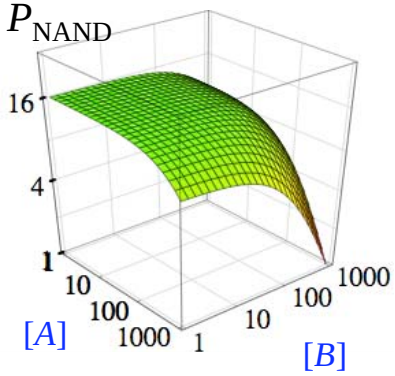
-- TF-RNAp interaction
$\omega_{A1\text{-}p} = \omega_{B1\text{-}p} = 20$
$\omega_{A1\text{-}B1} = 0$



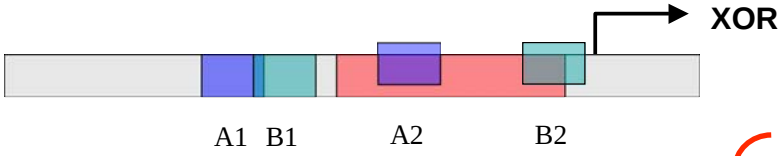
-- strong promoter
$[RNAp]/K_p = 0.02$
-- strong A and B sites
$K_{A2} = K_{B2} = 120$

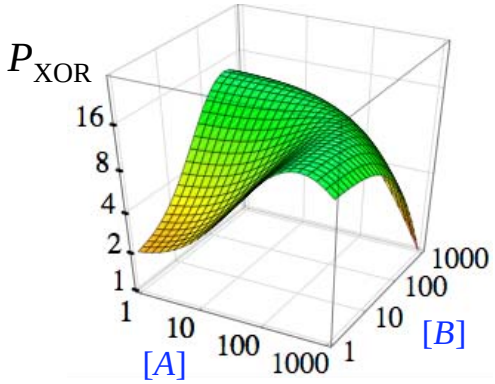
-- TF-RNAp interaction
$\omega_{A2\text{-}p} = \omega_{B2\text{-}p} = 0$



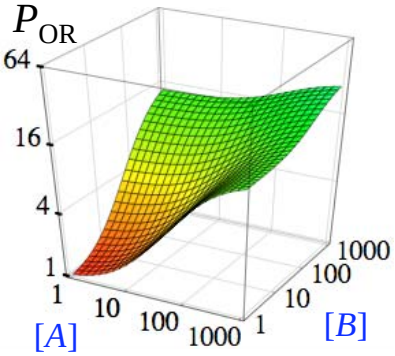


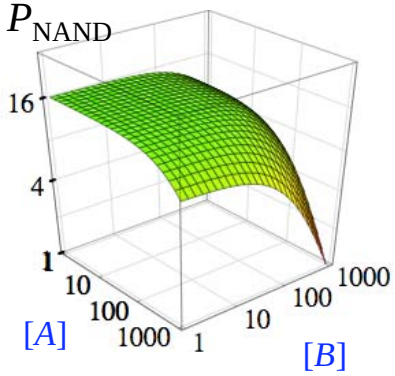
Quantitative design of combinatorial control





x



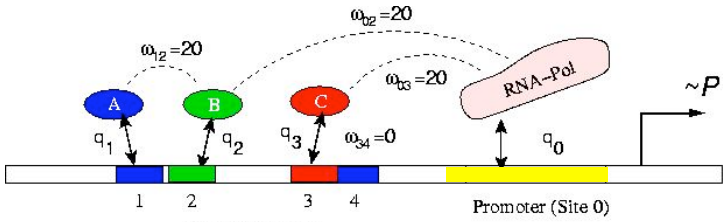


-- weak promoter
$[RNAp]/K_p = 0.001$
-- operator affinities
$K_{A1} = K_{B1} = 200$
$K_{A2} = K_{B2} = 620$

-- TF-RNAp interaction
$\omega_{A1\text{-}B1} = 0$
$\omega_{A2\text{-}p} = \omega_{B2\text{-}p} = 0$

In general:

- binding sites: $j = \{1, \ldots, L\}$
- binding constants: K_j
- TF conc: $[TF_{\alpha(j)}]$



Let $\sigma_j = \{0,1\}$ represent the occupation state of each site j

➔ weight for TF binding: $$W_{\text{unocc}}\left[\sigma_1,\ldots,\sigma_L\right] = \prod_{j=1}^{L}\left([TF_{\alpha(j)}]/K_j\right)^{\sigma_j} \cdot \prod_{i<j}\omega_{i,j}^{\sigma_i\sigma_j}$$

➔ weight for promoter occupation:

$$W_{\text{occ}}\left[\sigma_1,\ldots,\sigma_L\right] = \frac{[RNAp]}{K_p}\prod_{j=1}^{L}\left[1-\sigma_j\,\delta\left(\omega_{j-p},0\right)\right]\cdot\left[1+\sum_{j=1}^{L}\sigma_j\,\omega_{j-p}\right]\cdot W_{\text{unocc}}$$

promoter occupation probab. ∝ gene expression [Shea & Ackers, '86]

➡ $$P\left\{\left[TF_\alpha\right] \mid K_j, \omega_{i,j}\right\} = \frac{W_{\text{occ}}}{W_{\text{occ}} + W_{\text{unocc}}}$$

Task: find $\{K_j, \omega_{i,j}\}$ to implement the desired $P\{[TF_\alpha]\}$

➔ **Molecular Boltzmann machine!**

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

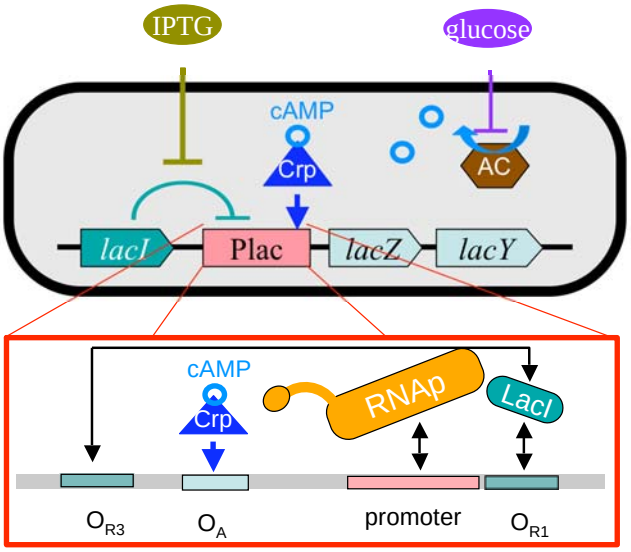
Quantitative confrontation of model and experiment

➔ 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**

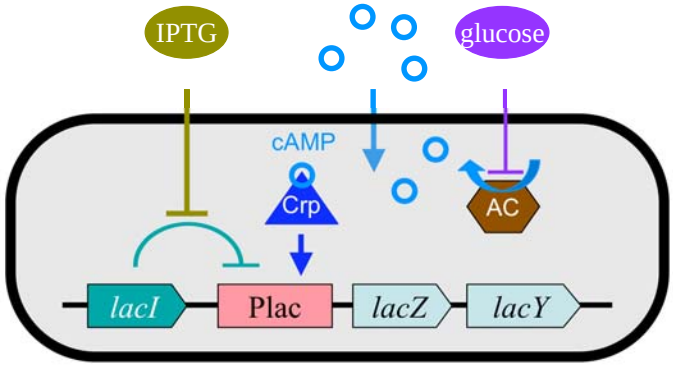
molecular ingredients:

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

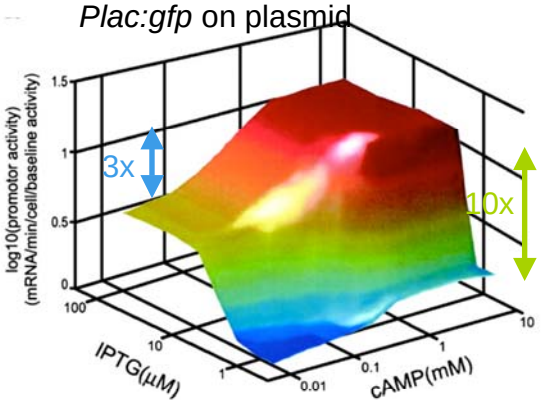
➔theory: quantitative prediction of gene regulation by LacI, cAMP-Crp

➔ 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



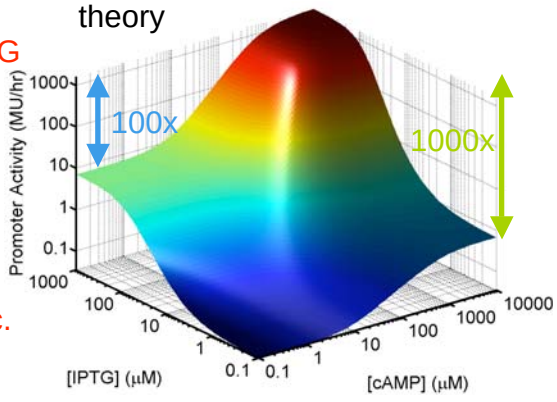
Previous expt: [Setty et al, PNAS, 2003]

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

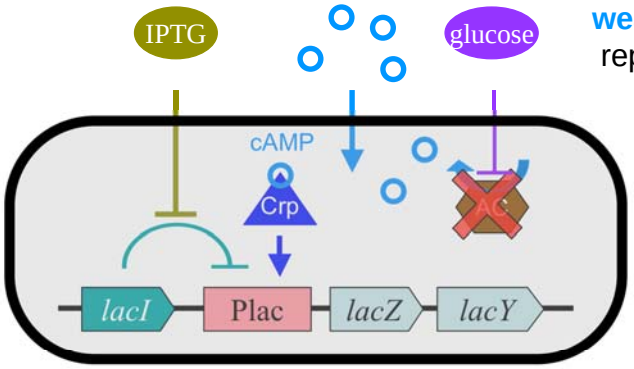
inconsistent with behavior of mutants:
Δ*lacI*: > 1000x; **Δ*crp* > 50x**

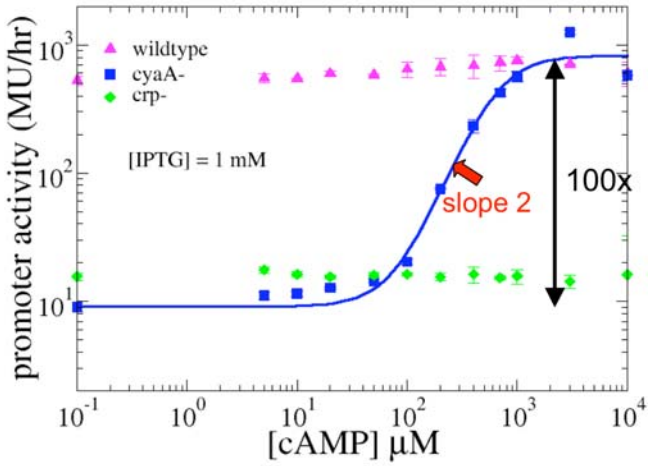
➔ possible problems: complex links between extracellular and intracellular inducer conc.

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

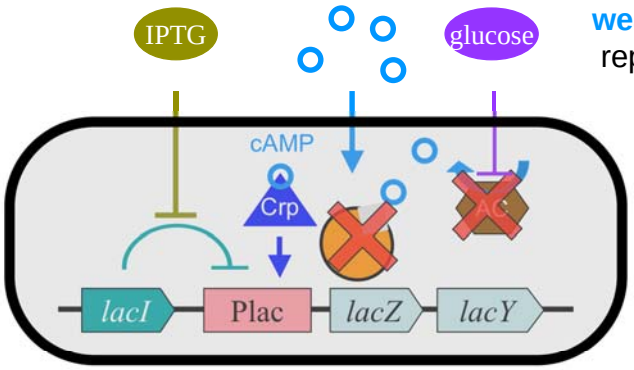
CRP dimer activated by binding of single cAMP molecule

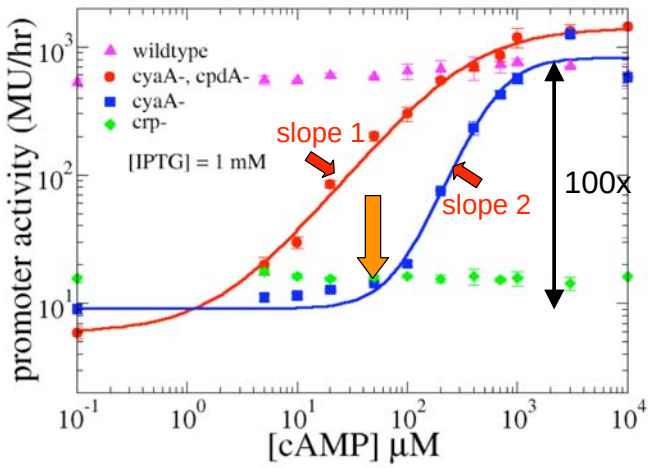
$CRP_2 + cAMP \rightleftharpoons CRP_2{:}cAMP$

(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

$CRP_2 + cAMP \rightleftharpoons CRP_2{:}cAMP$

(expect Hill coeff = 1)

in vitro biochem irrelevant?
other effects exerted by CRP-cAMP?

➔ cAMP degraded by **PDE** (*cpdA*)

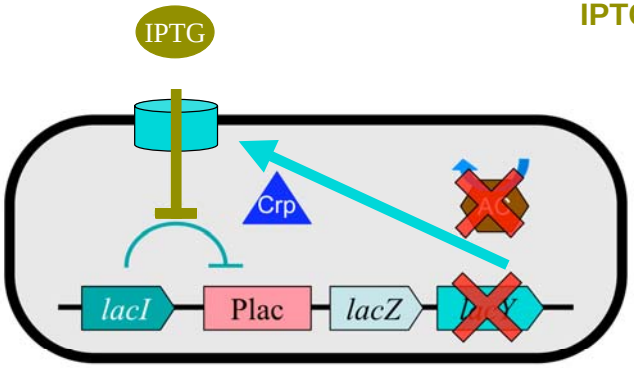
➔ effect of *cpdA* deletion?

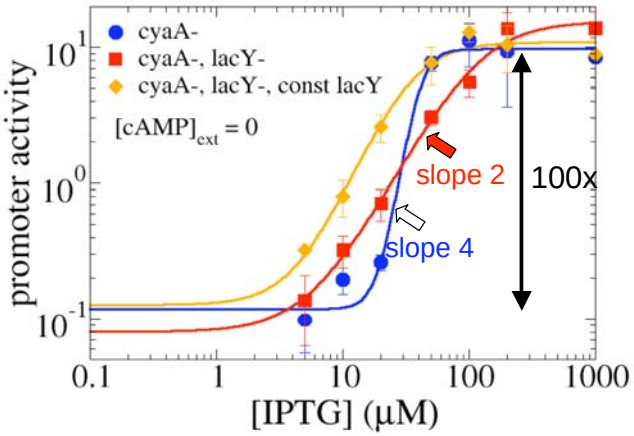
➔ Hill coeff ≈ 1, agrees with model

➔ role of PDE: no known phenotype

➔ mechanism of cooperativity?

Quantitative characterization of mutants

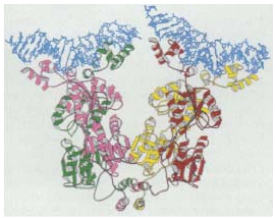


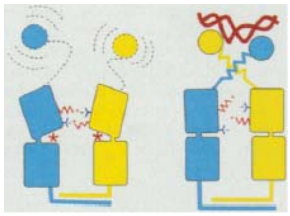


IPTG dependence: *cyaA-* cells with [cAMP]=0
➔ very cooperative! (Hill coeff ≈ 4)

➔ delete *lacY* Hill coeff ≈ 2
➔ constitutive expression of LacY
only shifted IPTG dependence

➔ Hill coeff = 2 widely cited in literature

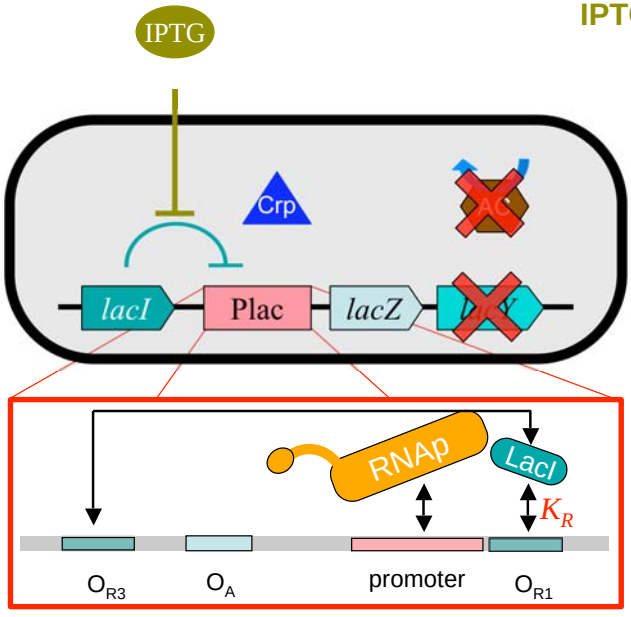




- LacI forms tetramer (dimer of dimers)
- strong coupling within each dimer and weak coupling between dimers

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 ⇌ $LacI_4$:IPTG
- weakly cooperative in the presence of operator DNA (Hill coeff = 1.4 ~ 1.6)

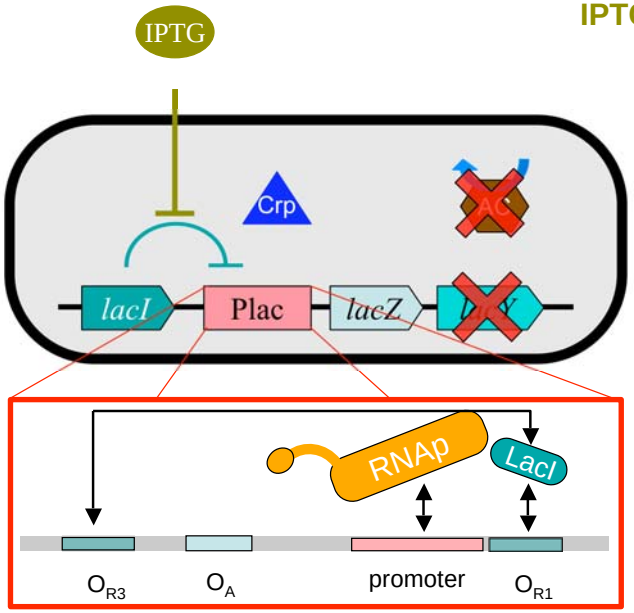
[Matthews lab, '85]

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

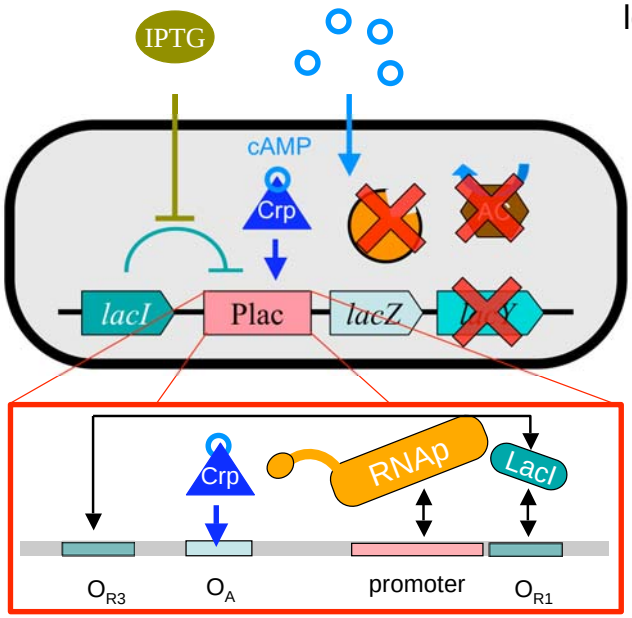
simple repression $$\text{tsx activity} \propto \frac{1}{1 + [R] / K_R}$$

- 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



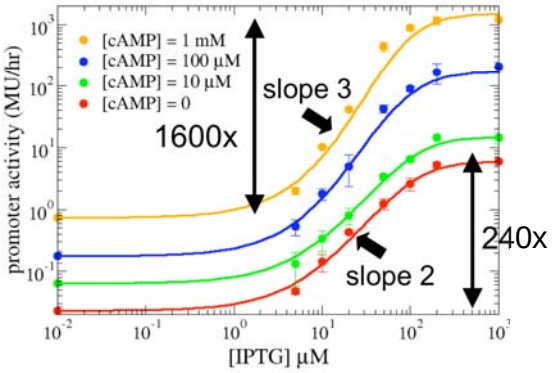
auxiliary Lac operators stabilize LacI-O1 binding via **DNA looping** [Muller-Hill]

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

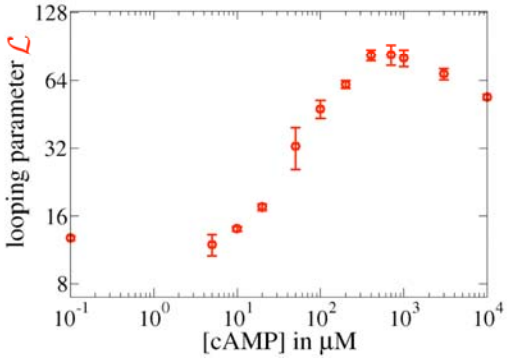
➔ effective Hill coeff (1.5 ~ 3) depends on $\mathcal{L}_o$

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

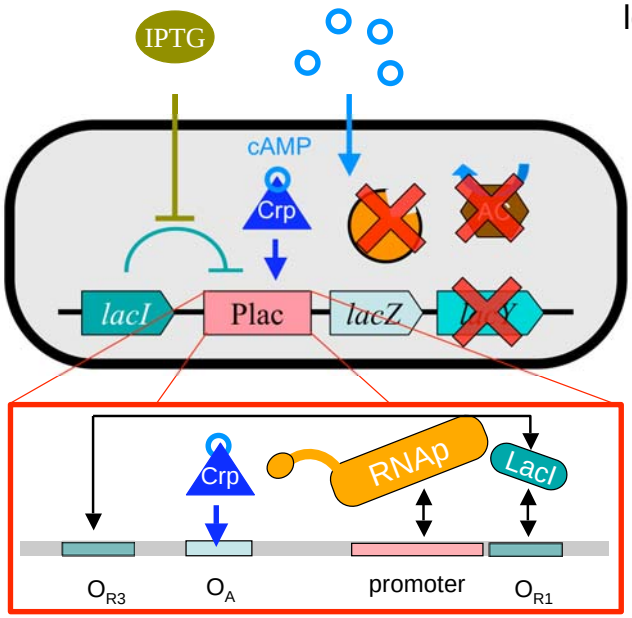
looping model w/ $\mathcal{L}_0 \approx 12$, $2[\mathrm{LacI}_4]/K_R = 20$



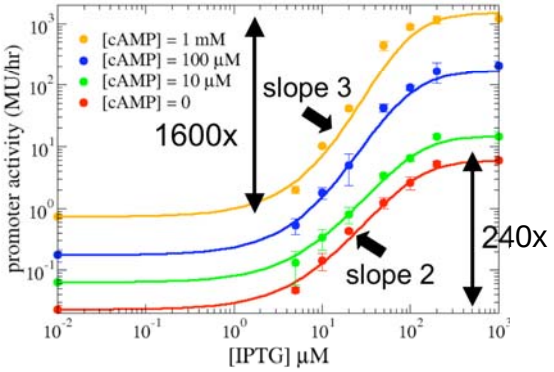
➔ single parameter $\mathcal{L}_0$ fits both fold-repression and slope



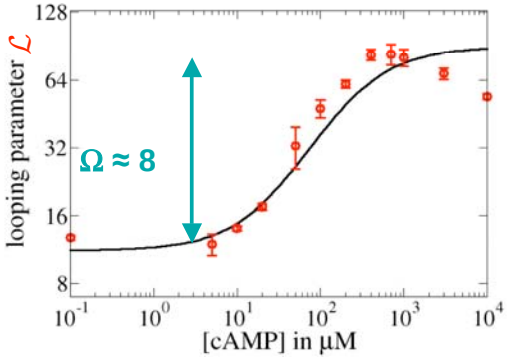
Quantitative characterization of mutants



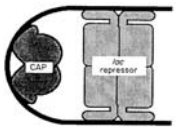
looping model w/ $\mathcal{L}_0 \approx 12$, $2[\mathrm{LacI}_4]/K_R = 20$



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



Crp-dependence of DNA looping

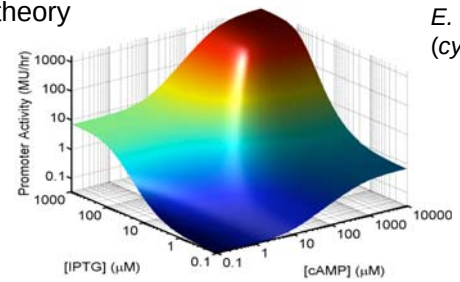


Fried et al, 84;
Balaeff et al, 04

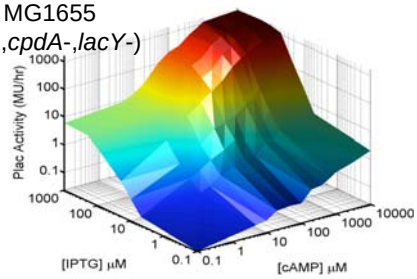
in vitro study found coop. factor **Ω = 4 ~12**

Summary

theory



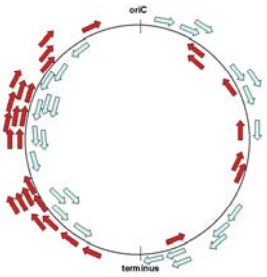
E. coli MG1655
(*cyaA*-,*cpdA*-,*lacY*-)

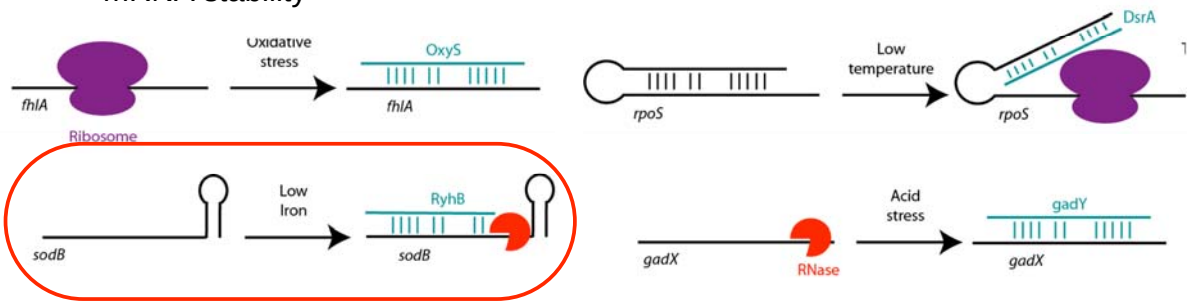


- **main findings for the *lac* promoter:**
  - Crp enhances DNA looping
  - abrupt IPTG response despite non-cooperative LacI-IPTG interaction;
  - ➔ **suggests physiological role of Crp-cAMP as enhancer of repression**
  - mechanism of Crp-LacI interaction?
  - coop cAMP response due to PDE; physiological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

small RNA mediated gene regulation

- over 60 small regulatory RNAs found in *E. coli*
- typical length: 100-150nt
- *cis*- or *trans*- acting
- various mechanisms of gene regulation
  - tsx termination (plasmid copy # control)
  - translational initiation
  - mRNA stability





- predominant role of sRNA in mediating stress responses

Q: Significant functional differences between sRNA- and protein- mediated gene regulation?

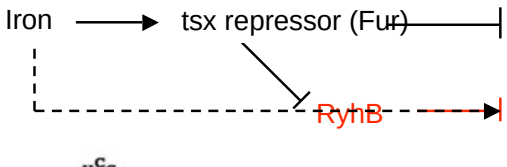
- faster?
  … but change in protein level dominated by rate of protein synthesis and turnover
- more economical?
  … but there are sRNA with dedicated protein regulators
- evolutionary convenience?
  … then why protein regulators?

Our study: unique properties in gene silencing by sRNA
threshold response, hierarchical cross-talk, anomalous noise profile
– establish through theoretical/expt'l study of exemplary system (RyhB/sodB)

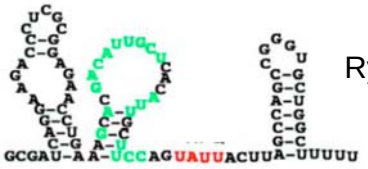
➔ sRNA are potent regulators suitable for specific tasks

sRNA regulation in iron metabolism

[S. Gottesman Lab]



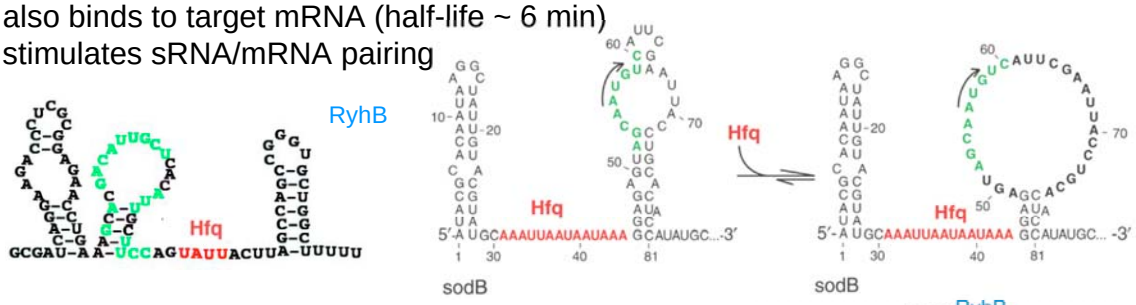
Gene	Fold Change (*ryhB -/+*)
acnA	1.7
bfr	2.4
fumA	7.6
sdhD	6.1
sodB	19.3



RyhB: well-characterized example of the most abundant class of sRNA known

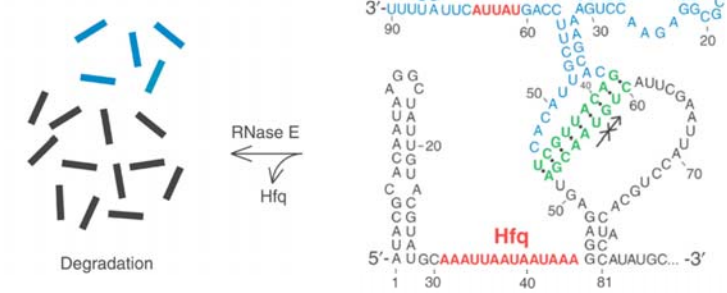
RyhB binds to the translational initiation regime of its target mRNA

- Hfq required for gene regulation by RyhB
  - highly abundant RNA-binding protein (30,000-60,000 / cell)
  - bind to and stabilizes RyhB (half-life > 30 min)
  - also binds to target mRNA (half-life ~ 6 min)
  - stimulates sRNA/mRNA pairing

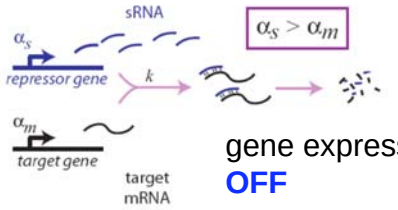


➔ rapid degradation of mRNA/sRNA
(RyhB half-life ~ 3 min)

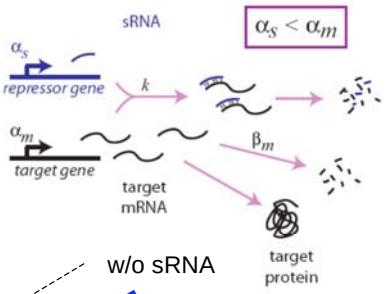
➔ stoichiometric rather than catalytic



Simple kinetic model of sRNA-mediated gene silencing

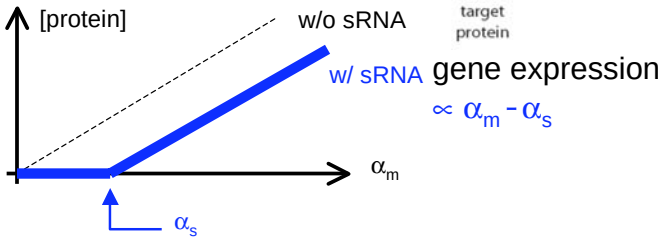


gene expression **OFF**



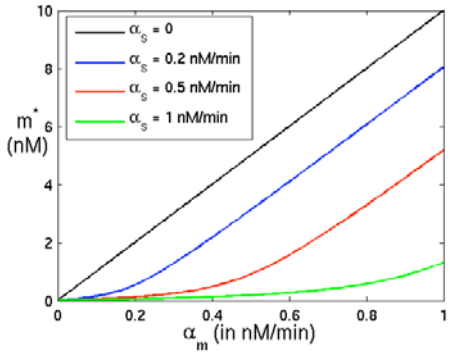
- qualitative expectation:

  threshold-linear response
  - tight repression for $\alpha_m \lesssim \alpha_s$
  - weak repression for $\alpha_m \gg \alpha_s$



- quantitative prediction:

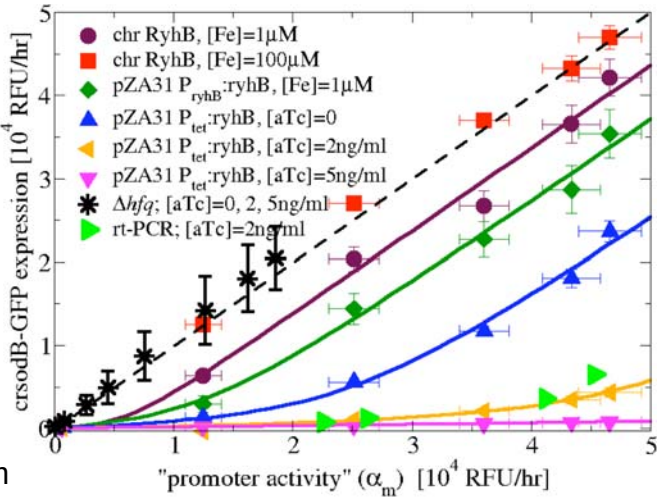
  rate constants:
  -- half-lives:
  $\beta_m^{-1} \approx 6$ min; $\beta_s^{-1} \approx 30$ min
  -- co-degradation: (~diffusion limited)
  $k^{-1} \sim 50$ nM-min



Expt'l characterization

plasmid-encoded inducible GFP reporte
translationally fused with truncated *sodB*

various sources of *ryhB*

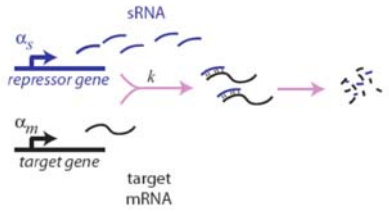


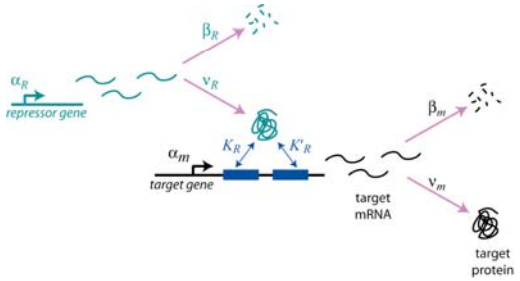
α_m ~ expression of target in *ΔryhB* cells

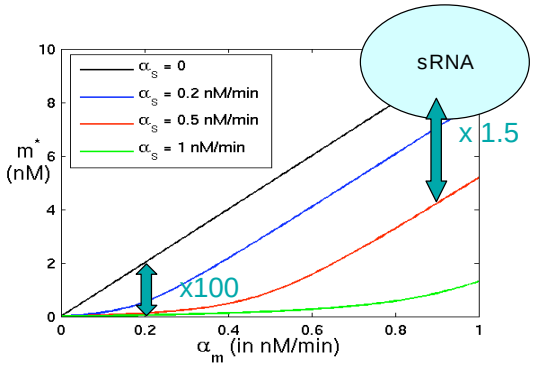
- native source of *ryhB*:
  -- some repression in Fe-poor medium
  -- repression abolished in Fe-rich medium (*ryhB* expression repressed)
- increased repression by P_{ryhB}:*ryhB* on plasmid (p15A *ori*)
- strong repression by P_{tet}:*ryhB* on plasmid (p15A *ori*)
  ➔ degree of repression increases with inducer level

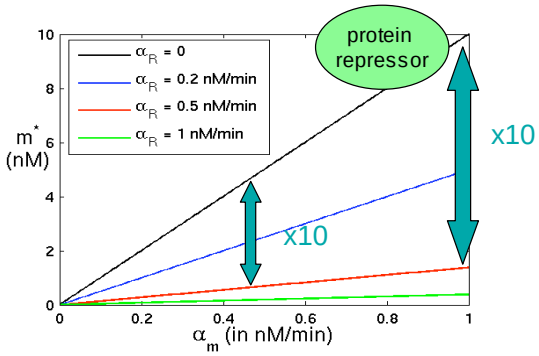
➔ one-parameter fit: agreement with the expected threshold-linear form
  – *ryhB*-dependent effects abolished in ***Δhfq*** strains
  – similar results obtained by rt-PCR

Compare to protein regulators









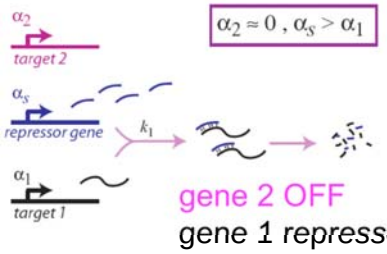
sRNA-mediated tsl regulation

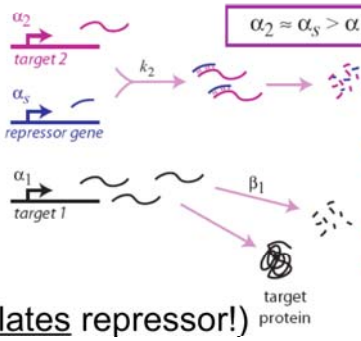
➔ fold-repression depends sensitively on α_m

protein-mediated tsx regulation

➔ fold repression independent of promoter activity (α_m)

multiple targets

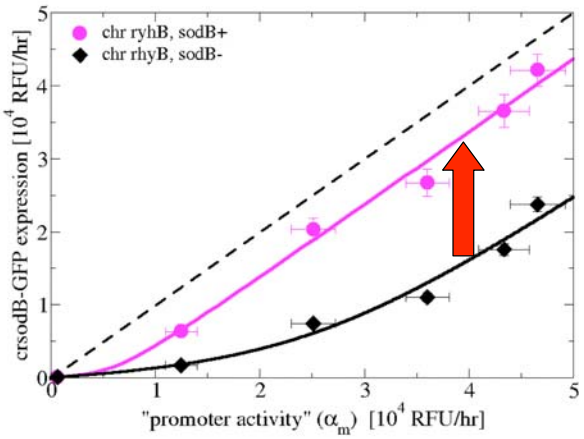


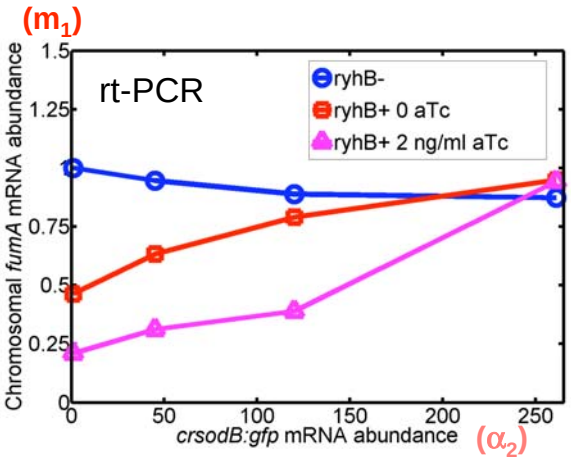


gene 2 ON

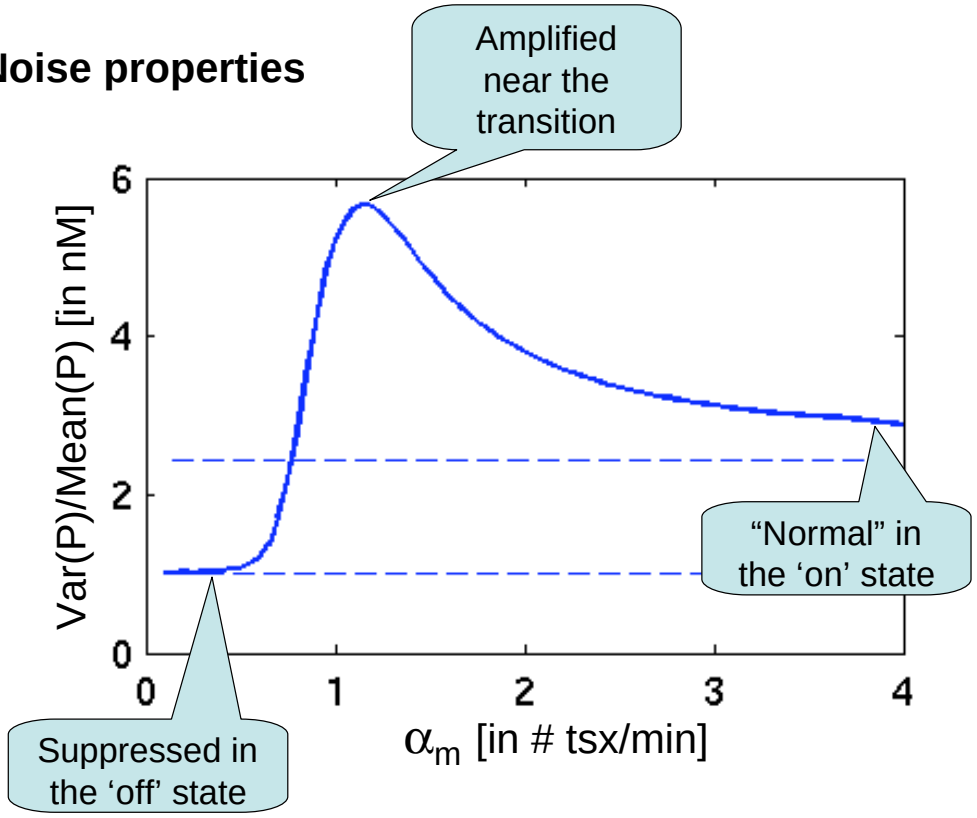
repression of gene 1 relieved by gene 2

➔ built in cross talk (gene 2 regulates repressor!)

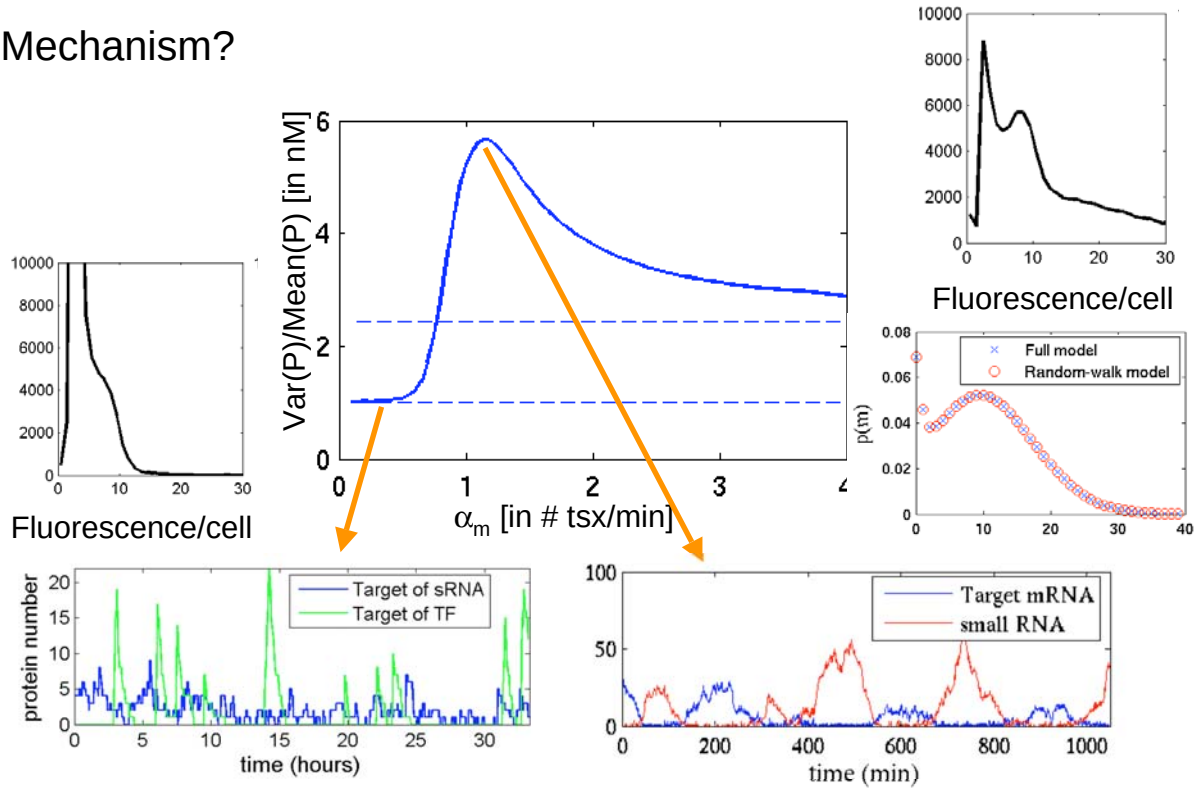




Noise properties

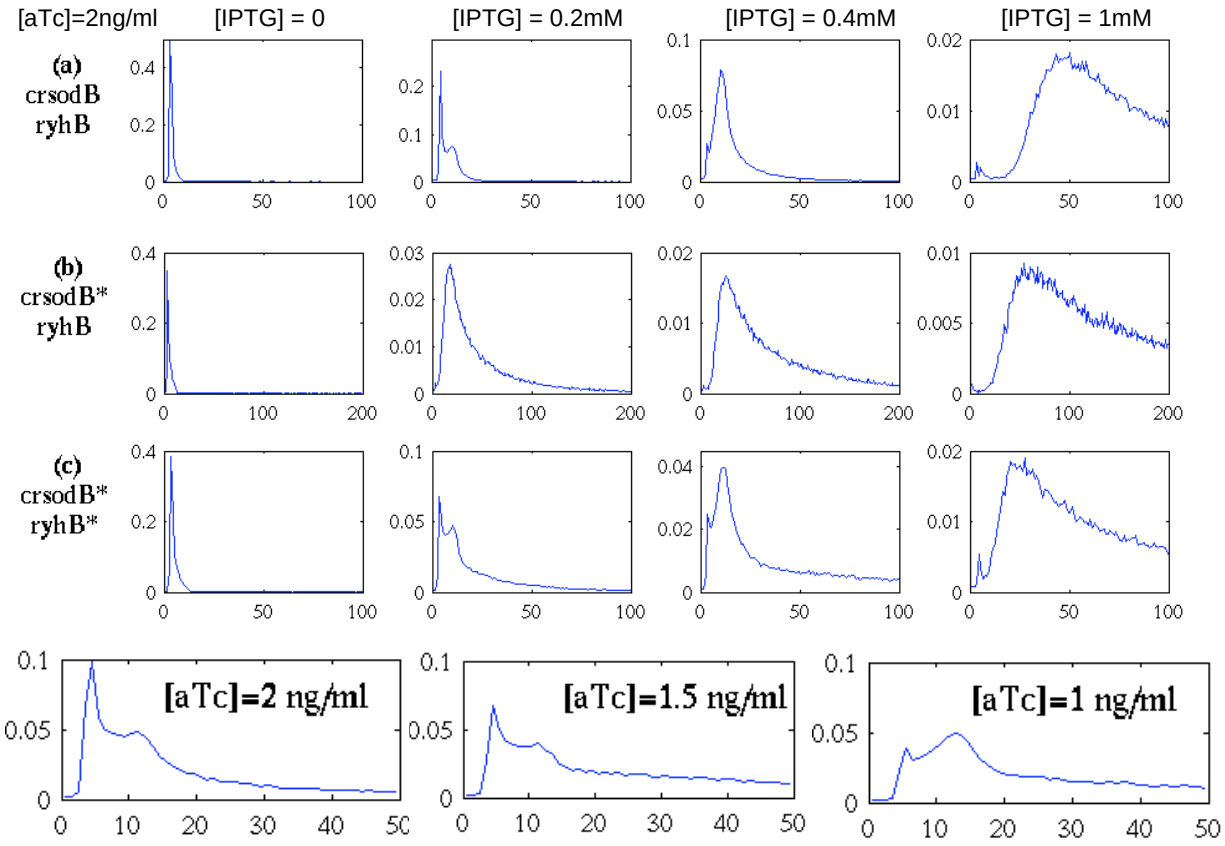


Mechanism?



➔ protein regulation more bursty due to translation of leaky transcripts

➔ fluctuation-induced bistability at the "transition"



Summary: Unique features of sRNA-mediated control

- threshold-linear response
  - threshold set by α_s(rather than [s])
  - strong repression over small changes in α_m- α_s
- multiple targets: hierarchical cross-talk
- noise characteristics
  - suppression in the low state
  - fluctuation-induced bistability at the transition
- kinetics: fast recovery from repression
- robustness of genetic circuits involving sRNA