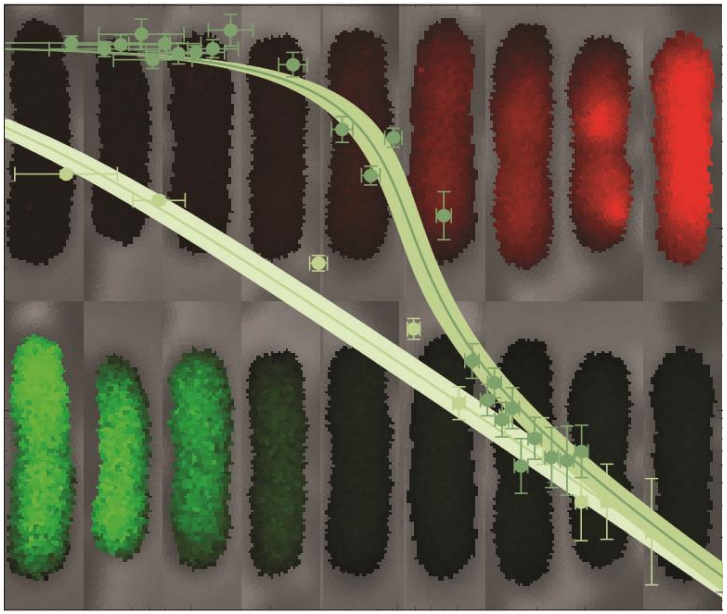
Statistical mechanics of genetic regulation

Willem Kegel

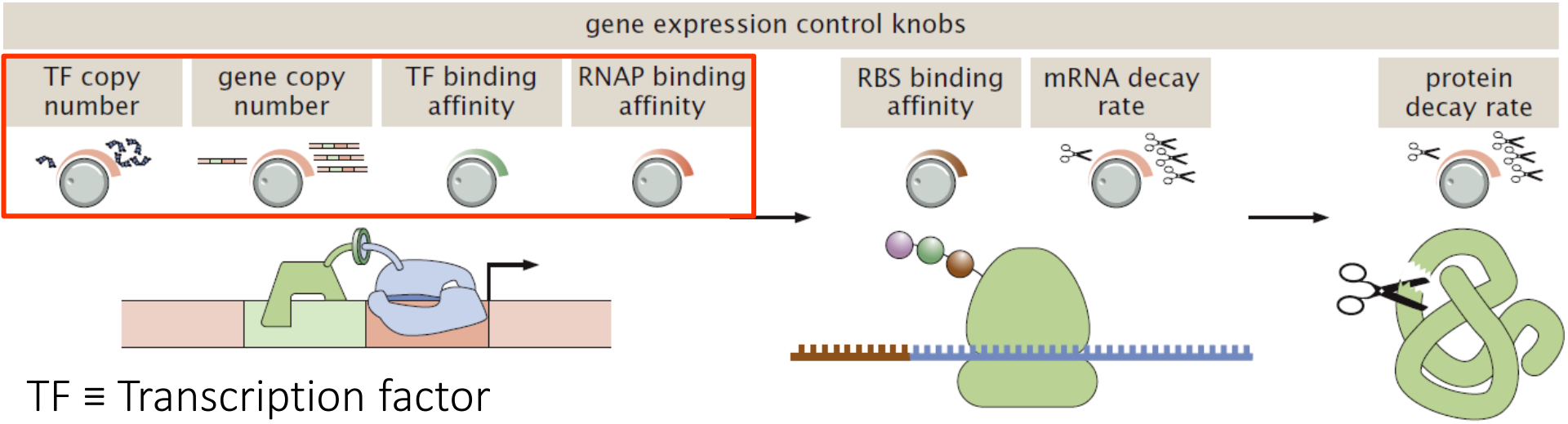
Universiteit Utrecht



THE BOLTZMANN GENOME

Varenna July 2015

Central Dogma of Molecular Biology



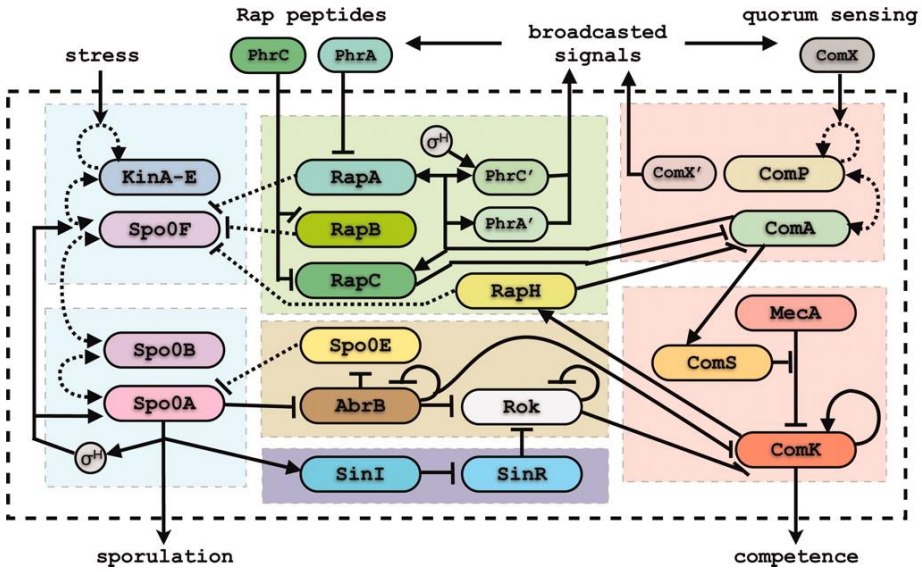
TF ≡ Transcription factor

DNA ⟶ mRNA ⟶ Protein(s)

Genetic regulatory networks

Gene circuits

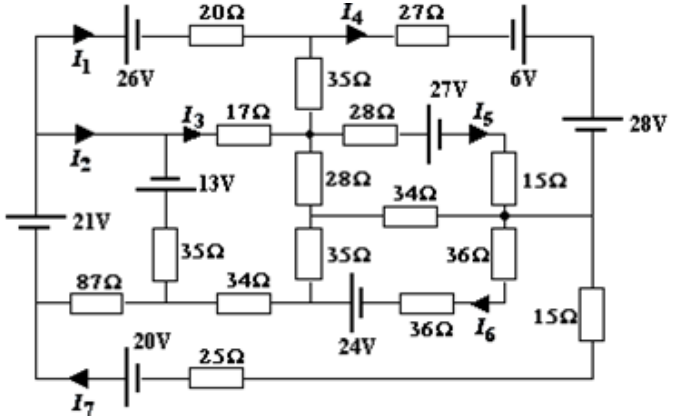
- Models of On / Off states of genes are used to describe genetic regulatory networks.
- Understanding of processes are dominated by qualitative "cartoons".



D. Schultz, P. G. Wolynes, E. B. Jacob and J. N. Onuchic, PNAS *106 (2009).*

Electronic Circuits

- Quantitatively well understood
- Input-output depends on small, well known set of parameters

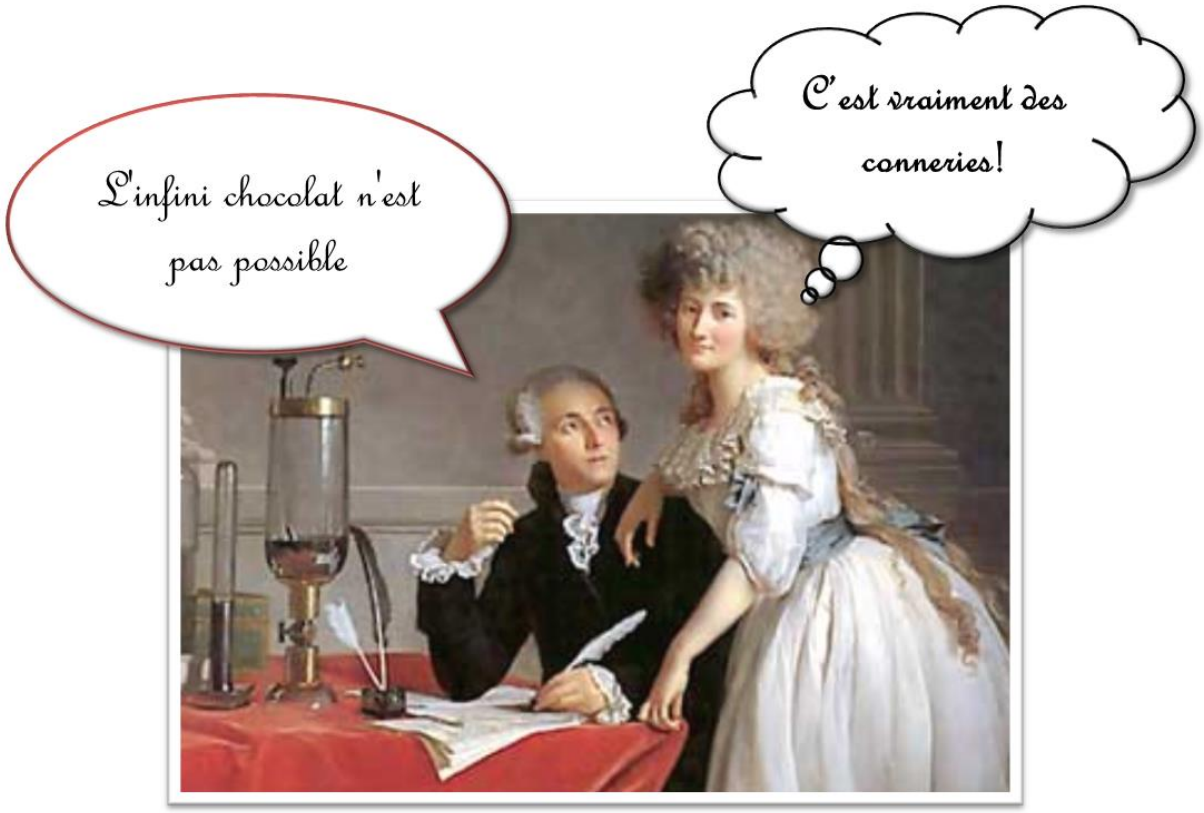


Is it possible to get a more quantitative description of cellular decision making?

Theory as prejudice

Prejudice in chemistry:

Mass is conserved



Antoine Lavoisier (1743–1794)

Prejudice in solid-state physics:

Dulong – Petit law (1819)

Prejudice: heat capacity of ALL solids = 3Nk, independent of temperature ('classical').

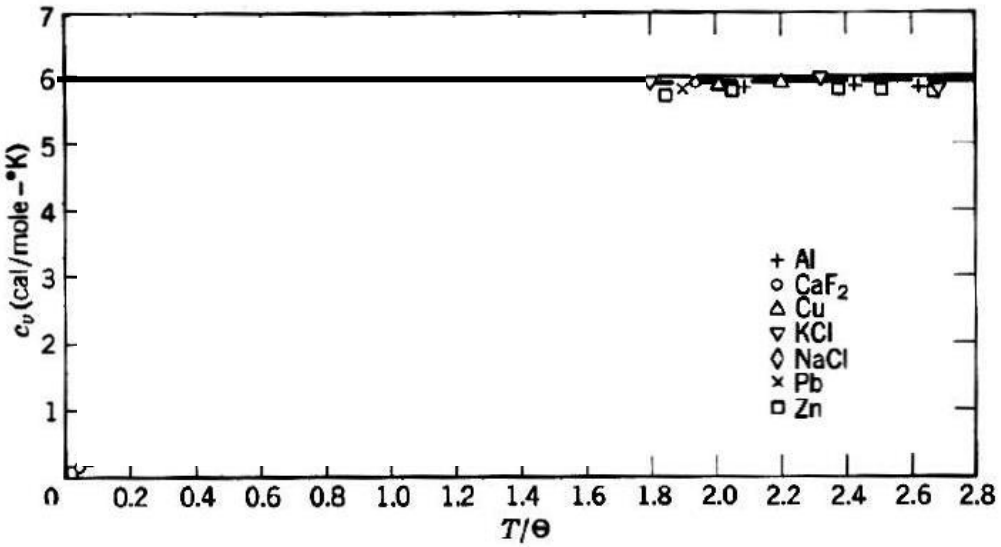


Figure 11-5 The measured specific heat at constant volume, as a function of temperature, for several materials. Horizontal line I represents the Dulong-Petit law, and curve II represents the predictions of the Debye theory.

Prejudice in solid-state physics:

Dulong – Petit law (1819)

Prejudice: heat capacity of ALL solids = 3Nk, independent of temperature ('classical').

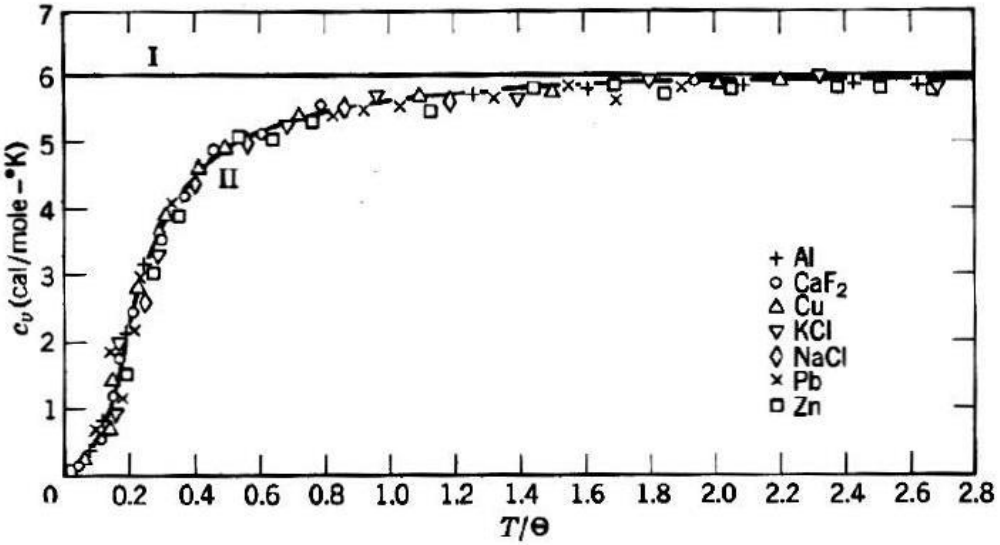


Figure 11-5 The measured specific heat at constant volume, as a function of temperature, for several materials. Horizontal line I represents the Dulong-Petit law, and curve II represents the predictions of the Debye theory.

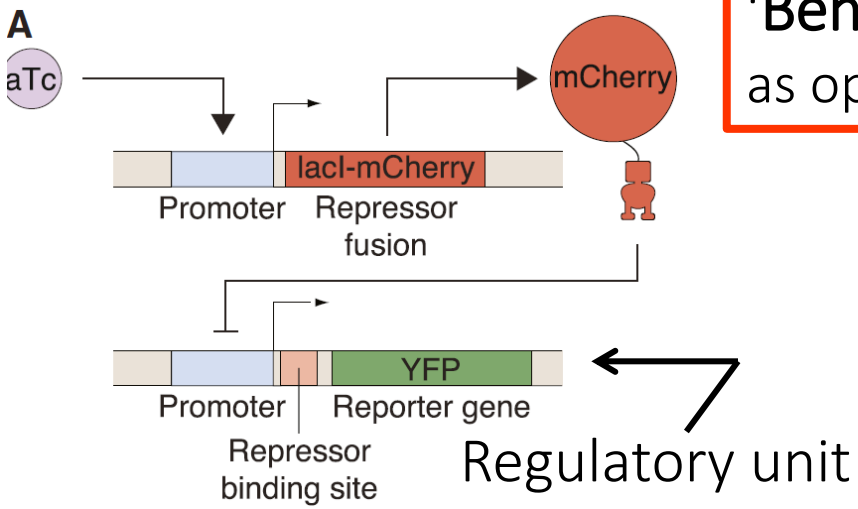
Debye model (1912): phonons in a box -

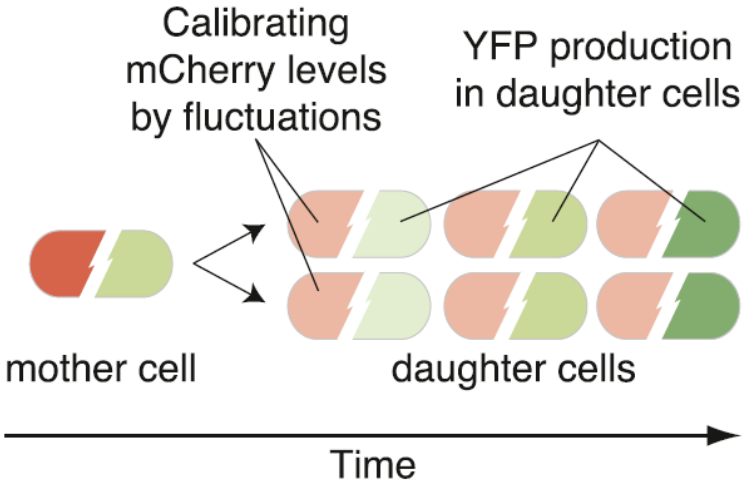
Solid-state equivalent of Planck's law of black body radiation -> fundamental NEW concept ('QM')

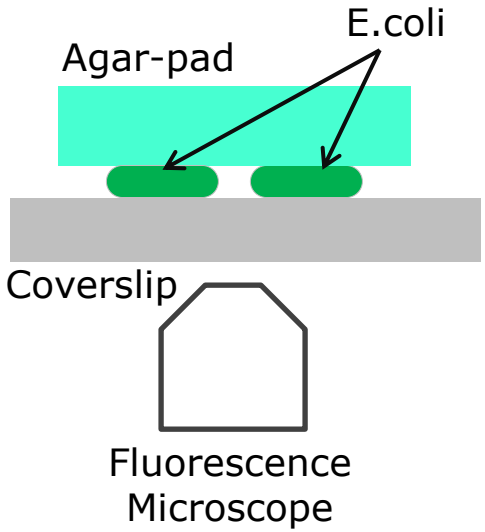
Prejudice: heat capacity of ALL solids obey universal behavior when properly scaled.

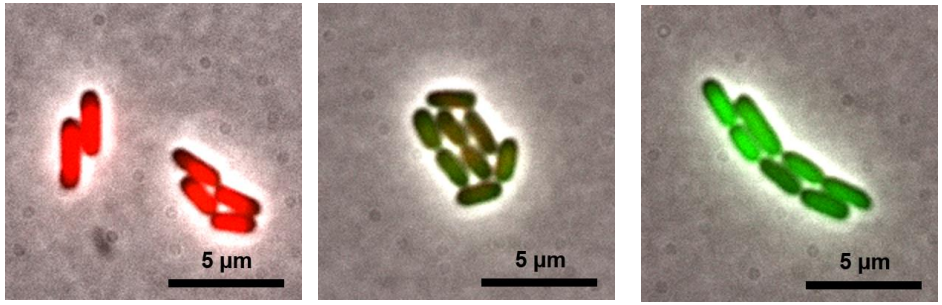
Quantitatively measuring gene expression *in vivo*

'Bend nature to test models' -
as opposed to 'Bend models to fit nature'









R. C. Brewster*, F. M. Weinert*, H. G. Garcia, D. Song, M. Rydenfelt and R. Phillips, *Cell 156 (2014).*

Biology & thermodynamics / statistical mechanics

R, repressors, P RNAP,

(RNAP = RNA polymerase -> provides first step in gene expression if adsorbed onto promoter)

Experiments : $$\text{Fold-change} = \frac{\text{gene expression rate in presence of Regulation Factor}}{\text{gene expression rate in absence of Regulation Factor}}$$

Theory :

$$\text{Fold-change} = \frac{\langle P_s \rangle}{\langle P_s(R=0) \rangle}$$

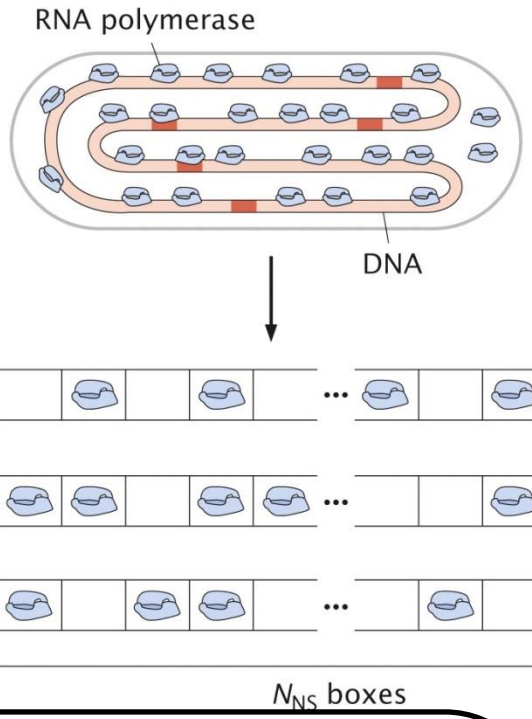
average # adsorbed RNAP onto promoter sites in the presence of repressors

average # adsorbed RNAP onto promoter sites in the absence of repressors

Adsorption of repressor **->** blocks promoter availability for RNAP

Calculate Fold-change via $\langle P_s \rangle$, $\langle P_s (R=0) \rangle$ from statistical mechanics

The Boltzmann g(e)nome



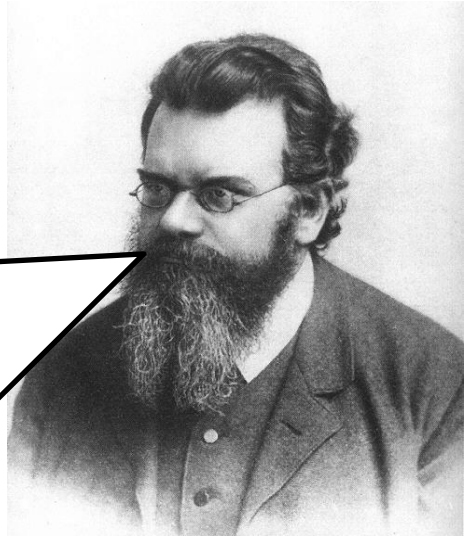
RNAP, R mainly sit on 'non-specific' DNA sequences

'complexions'

a-priori equal probabilities:

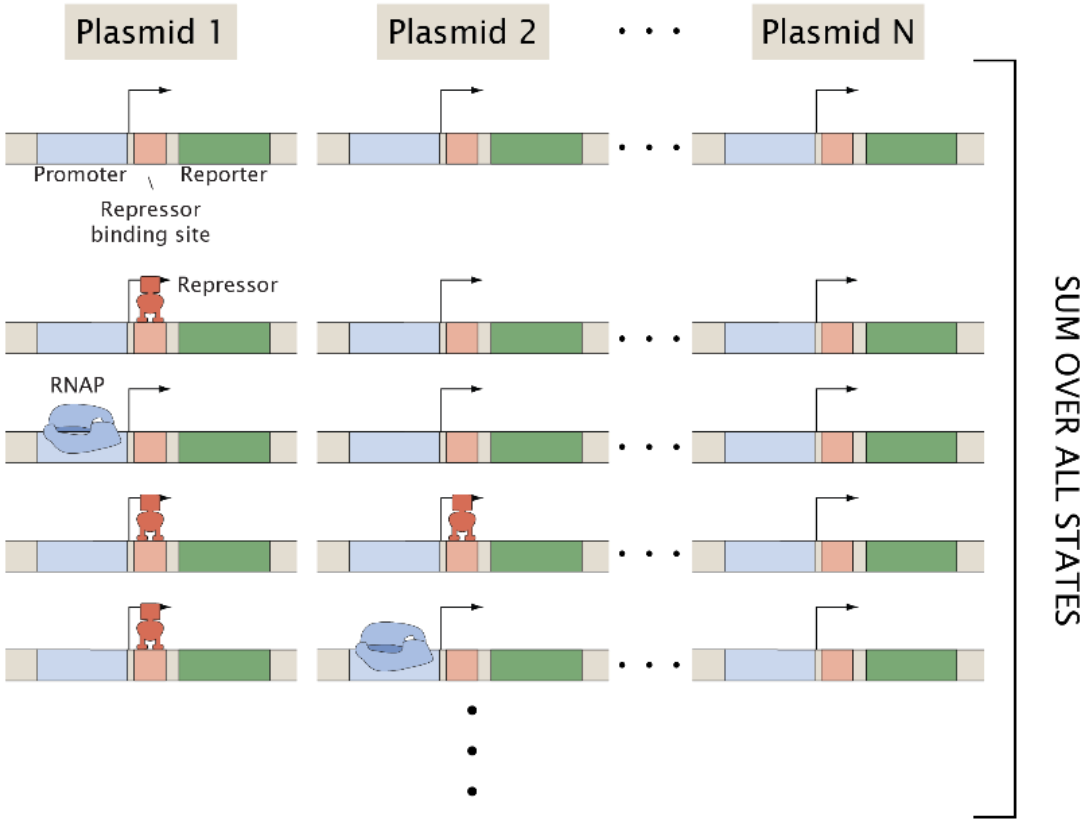
$$p_{\text{bound}} = \frac{\sum_{\text{states}} \left(\text{promoter} \right)}{\sum_{\text{states}} \left(\right) + \sum_{\text{states}} \left(\right)}$$

Figure 6.12 Physical Biology of the Cell, 2ed. (© Garland Science 2013)



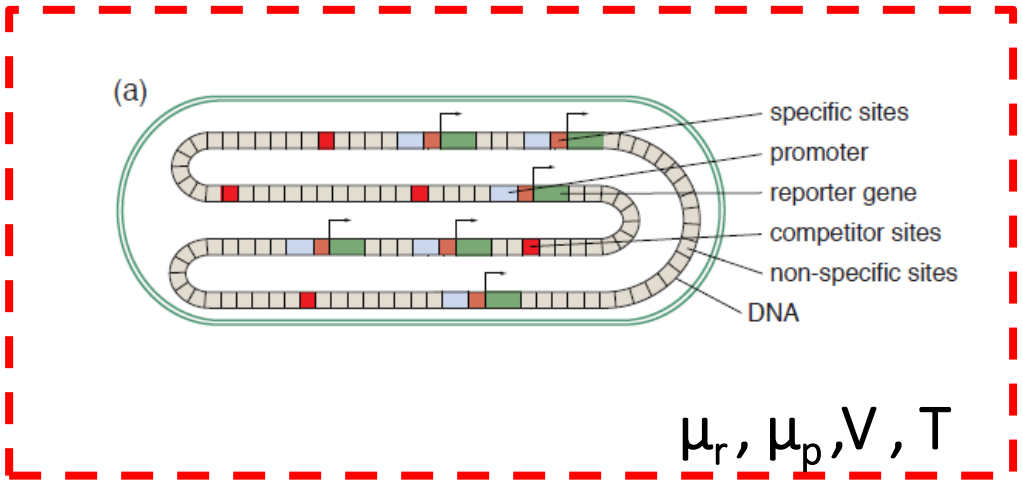
- Regulatory proteins typically (roughly 75%) regulate more than one gene

- Genes often exist in high copy number on:
  - multiple identical chromosomal copies,
  - plasmids
  - viral vectors

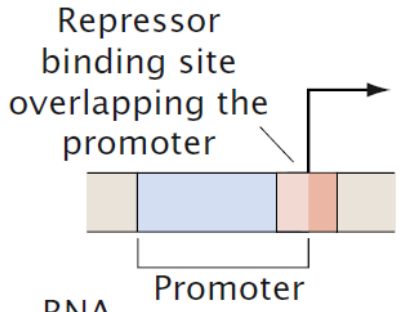
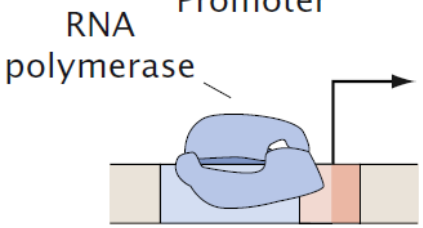
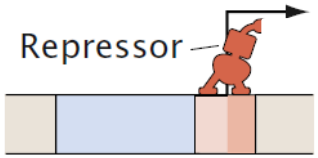
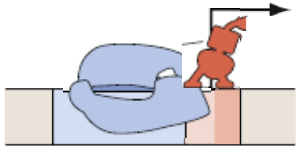


Genetic regulation in the grand ensemble

Take genome in equilibrium with (virtual) repressor and RNAP bath of constant chemical potentials μ_r, μ_p (not constant NUMBERS)



- Consider SPECIFIC binding sites, weight binding of component i onto site j by $\lambda_i e^{-\beta\varepsilon_{ij}} = e^{-\beta(\varepsilon_{ij}-\mu_i)}$
- Solve for Lagrange multiplier $\lambda_i = e^{\beta\mu_i}$ afterwards by applying appropriate conservation relation – distribution of P, R over any number and type of binding sites is taken care of by ultimate value of λ_i

STATE	WEIGHT
	1
	$\lambda_p e^{-\beta\varepsilon_p}$
	$\lambda_r e^{-\beta\varepsilon_s}$
	0

Fugacities $\lambda_p = e^{\beta\mu_p}$ $\lambda_r = e^{\beta\mu_r}$

$$\Xi = \sum_{p=0}^{1} \sum_{r=0}^{1-p} \lambda_p^p \lambda_r^r Z(p,r) = 1 + \lambda_p x_p + \lambda_r x_s \qquad x_i = e^{-\beta\varepsilon_i}$$

$$\text{Fold-change} = \frac{\langle P_s \rangle}{\langle P_s(R=0) \rangle}$$

$$< P_s >= \frac{\lambda_p e^{-\beta\varepsilon_p}}{1+\lambda_p e^{-\beta\varepsilon_p} + \lambda_r e^{-\beta\varepsilon_s}}$$

$$< P_s(R=0) >= \frac{\lambda_p e^{-\beta\varepsilon_p}}{1+\lambda_p e^{-\beta\varepsilon_p}}$$

Using $|\varepsilon_p| << |\varepsilon_s|$ -> weak promoter limit $\lambda_p e^{-\beta\epsilon_p} \ll 1$

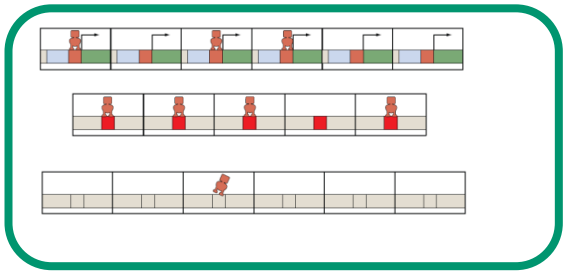
→ $$\text{Fold-change} = \frac{1}{1 + \lambda_r e^{-\beta\epsilon_s}}$$

Generalization to any # of genes, competing sites etc --- implicit in fugacity λ_r !

RNAP does not appear in the equation

F. Weinert, R. Brewster, M. Rydenfelt, R. Phillips, WKK, Phys Rev Lett **113** (2014)

Add ‘competing’ reservoirs, each with N sites and repressor adsorption energy ε: non-specific, competitor, …



Specific: $\varepsilon = \varepsilon_s$
$N = N_s$

$$\langle R_s \rangle = N_s \frac{\lambda_r e^{-\beta\epsilon_s}}{1 + \lambda_r e^{-\beta\epsilon_s}}$$

Non-specific: set $\varepsilon = 0$
$N = N_{ns}$

$$\langle R_{ns} \rangle = N_{ns} \frac{\lambda_r}{1 + \lambda_r}$$

Compatitor (decoy) sites:
$\varepsilon = \varepsilon_c$; $N = N_c$

$$\langle R_c \rangle = N_c \frac{\lambda_r e^{-\beta\epsilon_c}}{1 + \lambda_r e^{-\beta\epsilon_c}}$$

One λ that rules them all: value follows from repressor conservation

$$R = \langle R_s \rangle + \langle R_c \rangle + \langle R_{ns} \rangle$$

Leads to $a\lambda_r^3 + b\lambda_r^2 + c\lambda_r - R = 0$

3 reservoirs -> cubic equation

$$a = e^{\beta\epsilon_c} e^{\beta\epsilon_s} N_{ns}$$

$$b = (e^{\beta\epsilon_c} + e^{\beta\epsilon_s}) N_{ns} + e^{\beta\epsilon_c} e^{\beta\epsilon_s} (N_s + N_c - R)$$

$$c = N_{ns} + e^{\beta\epsilon_c}(N_c - R) + e^{\beta\epsilon_s}(N_s - R)$$

With (real, +) solution

$$\lambda_r = \Delta_+ + \Delta_- - \frac{b}{3a}$$

$$\Delta_\pm = \left(C_2 \pm \sqrt{C_1^3 + C_2^2}\right)^{1/3}, \; C_1 = (c/3a) - (b/3a)^2$$

$$C_2 = (bc/6a^2) + (R/2a) - (b/3a)^3$$

binding sites, binding energies, # repressors

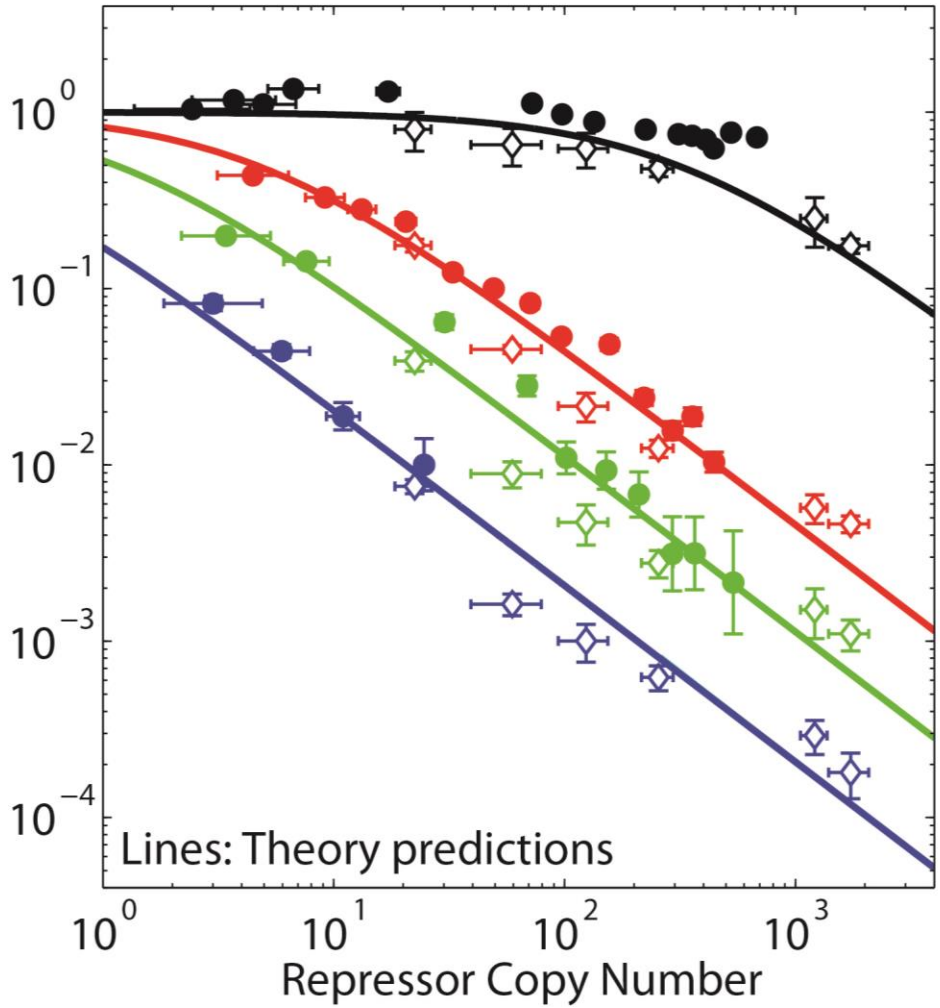
reflected in value of Lagrange multiplier λ_r

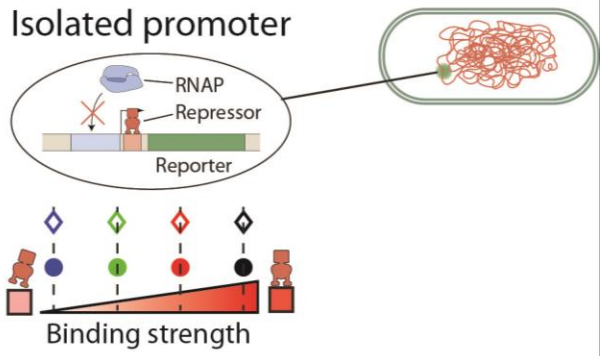
⟶ Still true that

$$\text{Fold-change} = \frac{1}{1 + \lambda_r e^{-\beta\epsilon_s}}$$

All scenario's are described quantitatively by the theory – NO free parameters!

$$\text{Fold-change} = \frac{1}{1 + \lambda_r e^{-\beta \epsilon_s}}$$

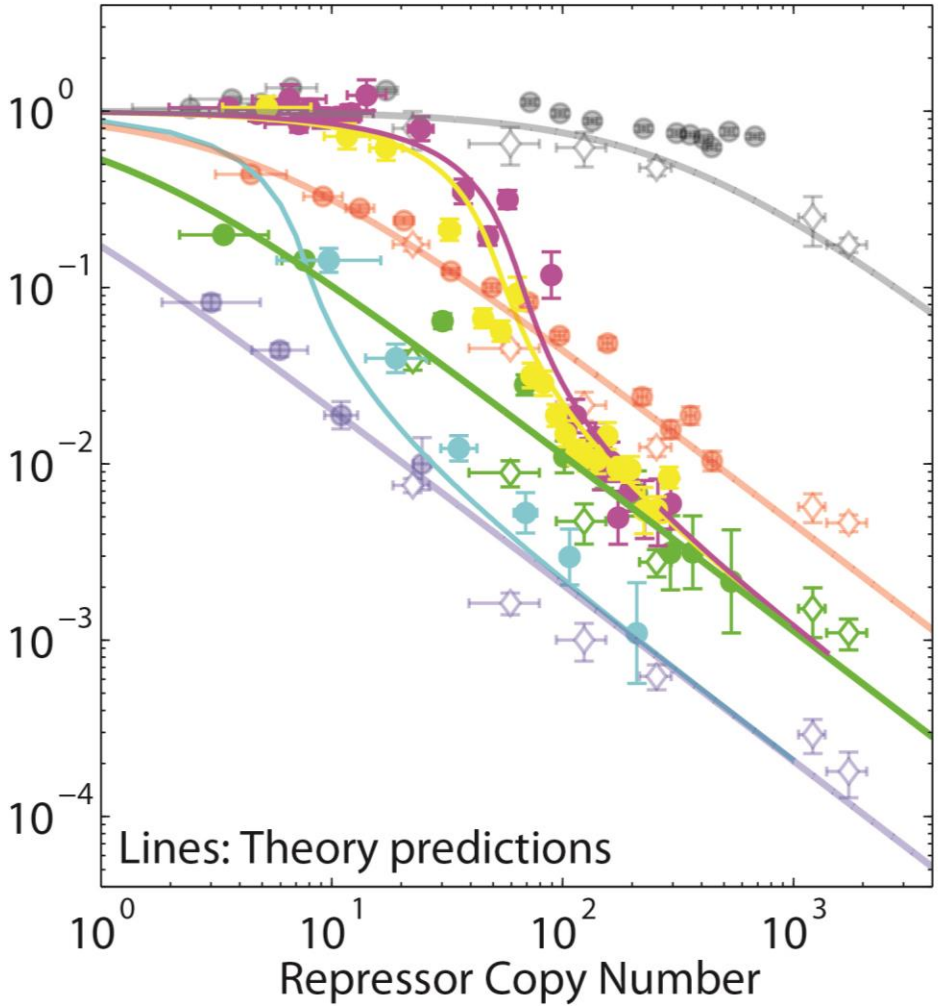


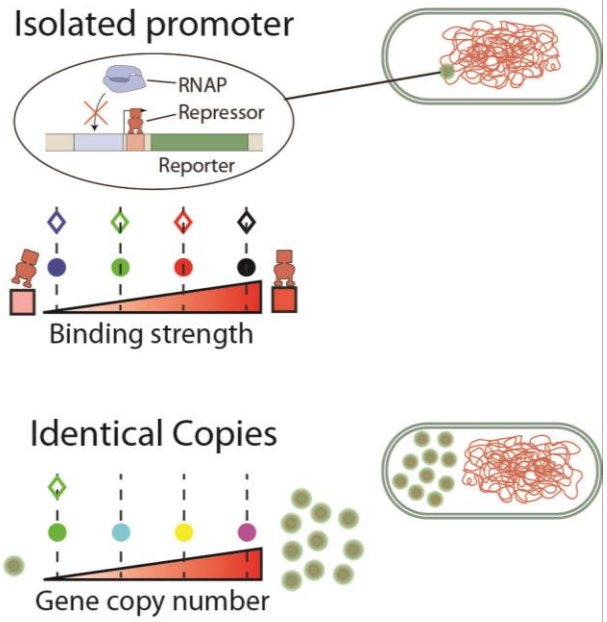


Weinert, F. M., Brewster, R. C., Rydenfelt, M., Phillips, R., & Kegel, W. K. (**2014**) *Physical review letters*, 113(25), 258101.

$$\text{Fold-change} = \frac{1}{1 + \lambda_r e^{-\beta \epsilon_s}}$$

All scenario's are described quantitatively by the theory – NO free parameters!

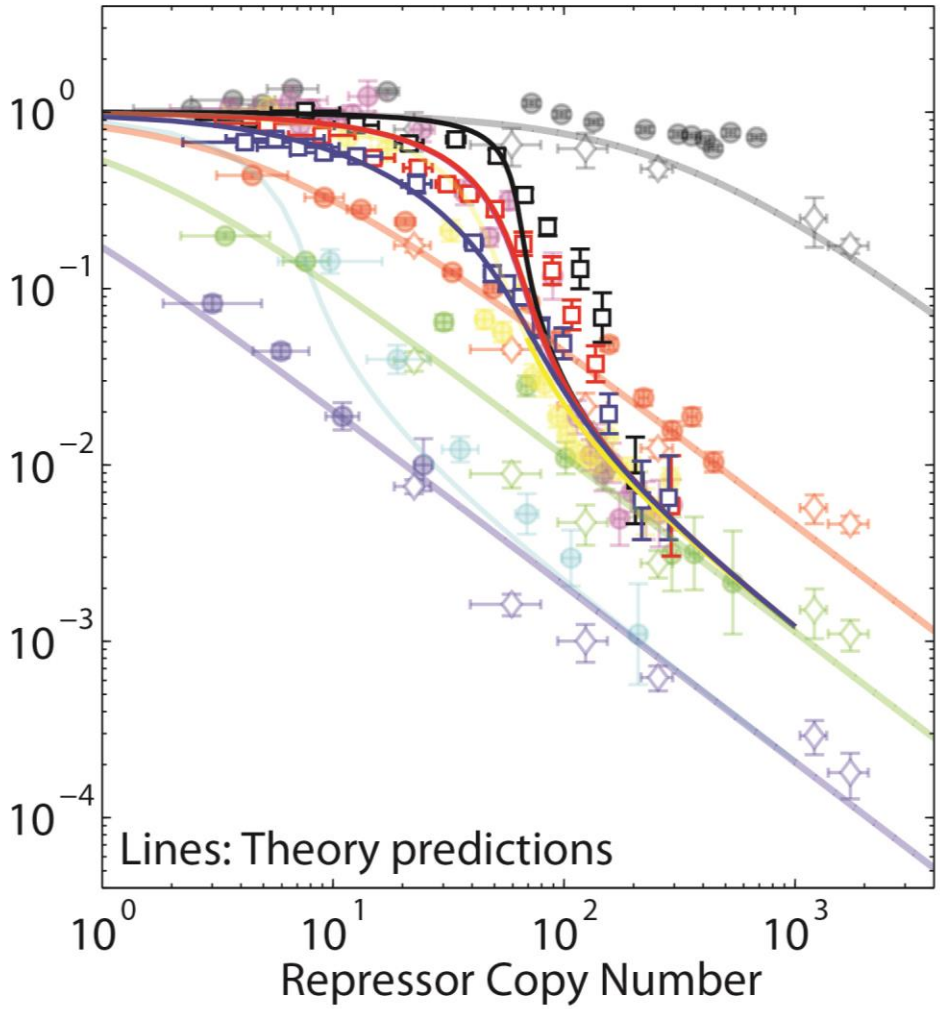


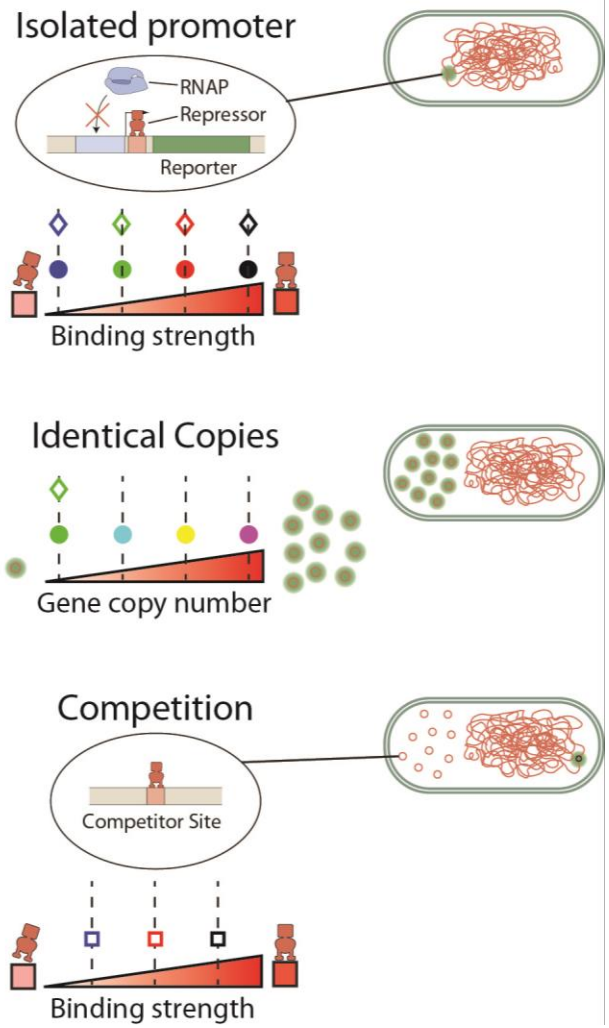


Weinert, F. M., Brewster, R. C., Rydenfelt, M., Phillips, R., & Kegel, W. K. (**2014**) *Physical review letters*, 113(25), 258101.

$$\text{Fold-change} = \frac{1}{1 + \lambda_r e^{-\beta \epsilon_s}}$$

All scenario's are described quantitatively by the theory – NO free parameters!

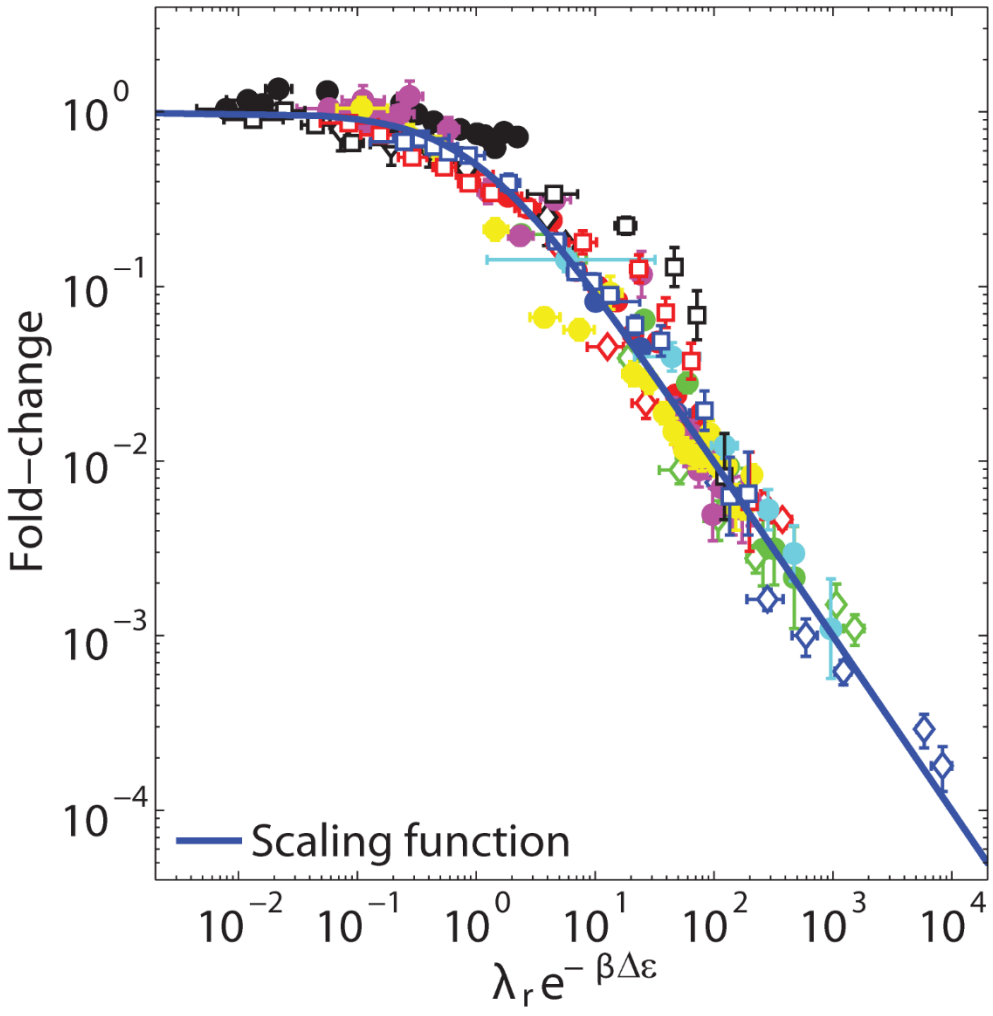


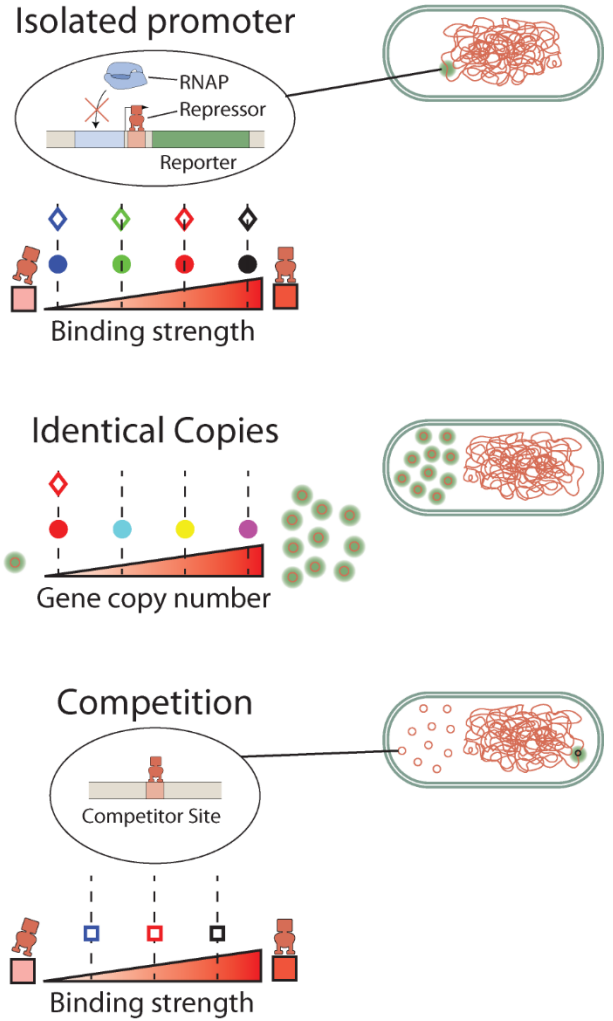


Weinert, F. M., Brewster, R. C., Rydenfelt, M., Phillips, R., & Kegel, W. K. (**2014**) *Physical review letters*, 113(25), 258101.

All data collapse onto the scaling function

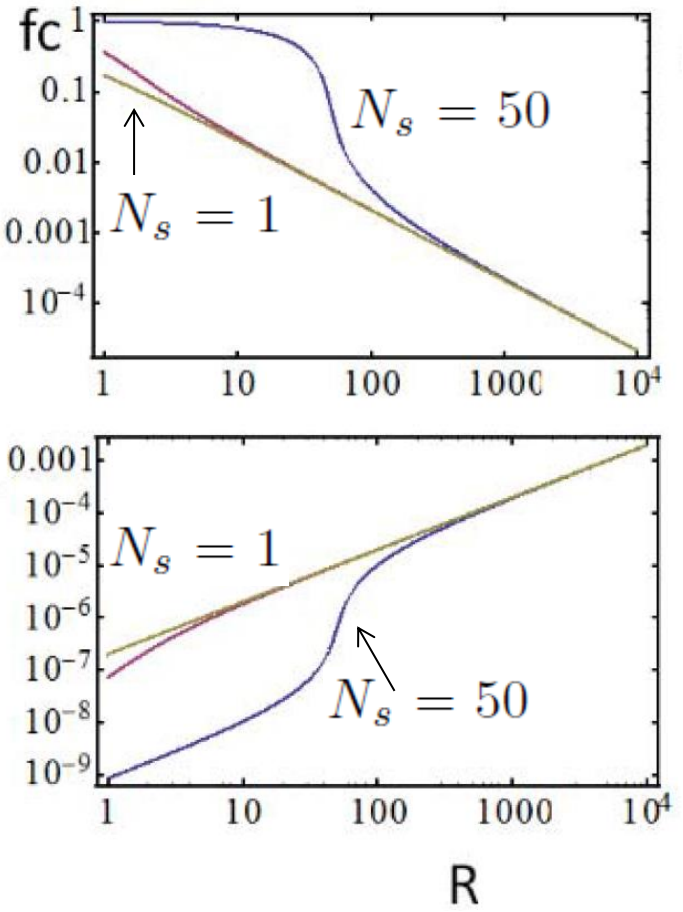
$$\text{Fold-change} = \frac{1}{1 + \lambda_r e^{-\beta \epsilon_s}}$$

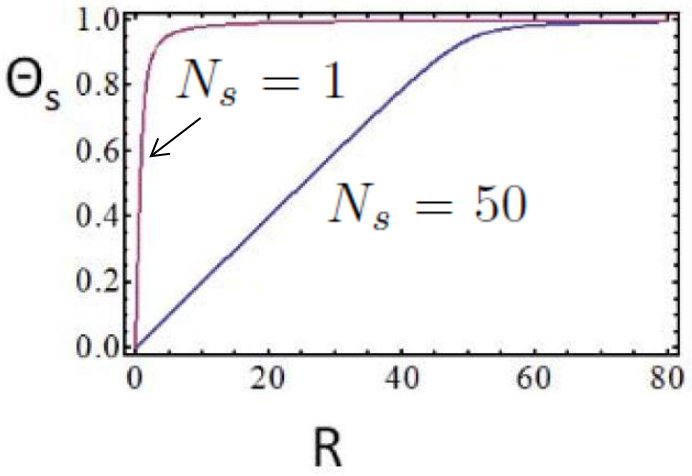




Weinert, F. M., Brewster, R. C., Rydenfelt, M., Phillips, R., & Kegel, W. K. (2014) *Physical review letters*, 113(25), 258101.

Fold-change with 1 & 50 promoters (no competitors)





$\theta_s =< R_s > /N_s$

$N_{NS} = 5 \times 10^6$

$\epsilon = -17kT$

green line is the canonical result for $N_s = 1$

Ensemble equivalence

Take single gene (N_s = 1) + non-specific reservoirs (N_{ns}). Enumerate configurations taking R = constant (and weak polymerase limit)
-> canonical result:

$$\text{Fold-change} = \frac{1}{1+(R/N_{ns})e^{-\beta\varepsilon_s}}$$

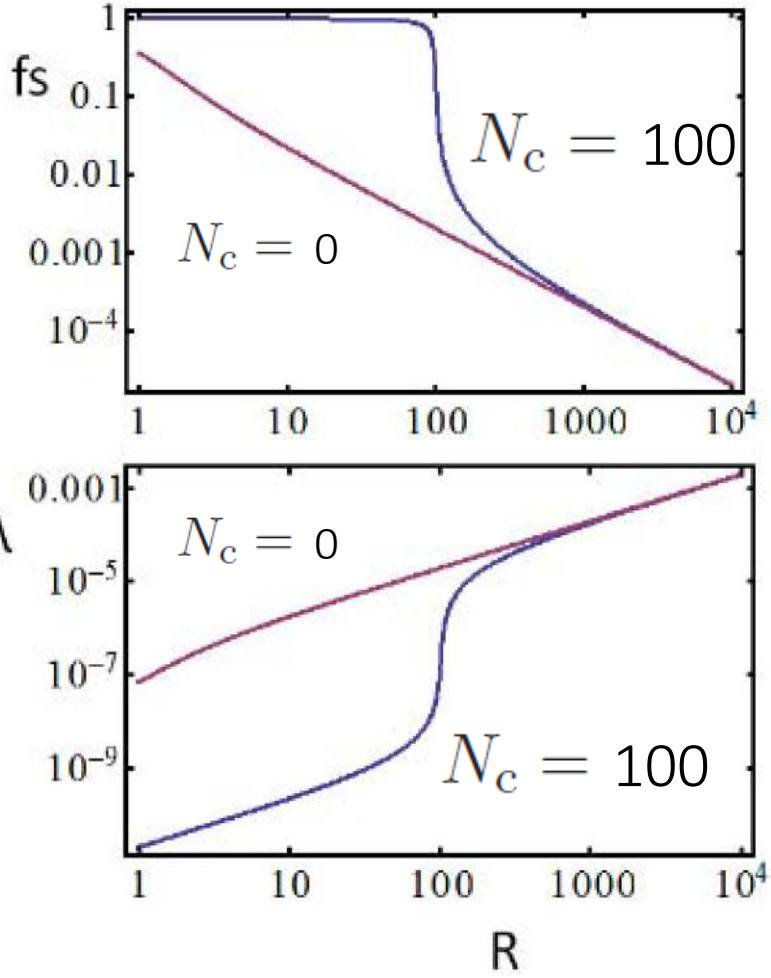
Compare with grand canonical expression, using that $\lambda_r \approx \frac{\langle R_{ns} \rangle}{N_{ns}}$

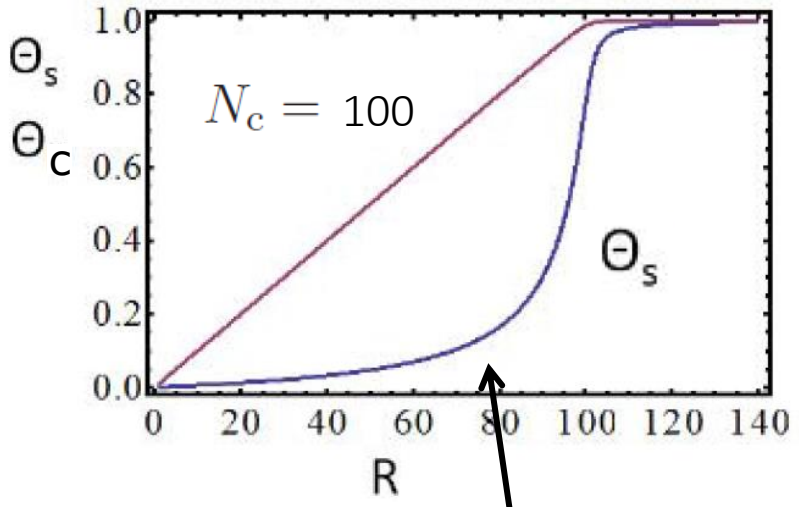
$$\text{Fold-change} = \frac{1}{1+(\langle R_{ns} \rangle / N_{ns})e^{-\beta\varepsilon_s}}$$

Asymptotically equal for large R – so, NOT if R=O(1).

In GC, at small R, there is a finite probability that $R \neq \langle R_s \rangle + \langle R_{ns} \rangle$

Add competitor sites (decoys)





$N_{NS} = 5 \times 10^6$

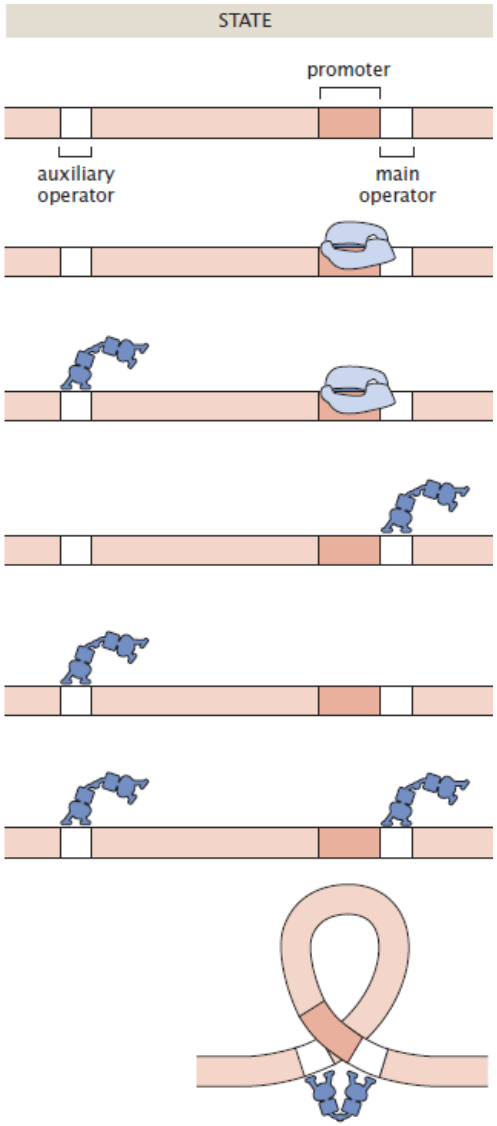
$N_s = 1$

$\epsilon_s = -17kT$

$\epsilon_c = -20\, k_B T$

Looks like ‘collective’ or ‘allosteric’ mechanism – yet ‘simple’ adsorption

More complex regulatory scenario's: DNA looping



$$x_p = e^{-\beta \varepsilon_p}$$

$$x_a = e^{-\beta \varepsilon_a}$$

$$x_m = e^{-\beta \varepsilon_m}$$

$$x_L = e^{-\beta F_L}$$

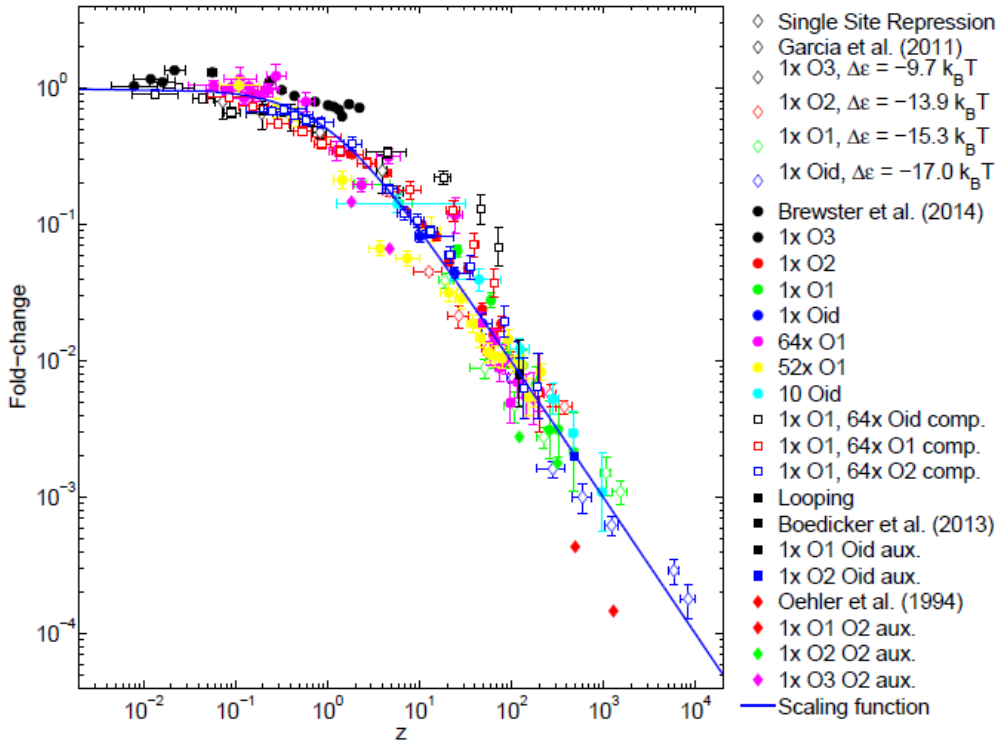
$$\Xi = \sum_{p=0}^{1} \sum_{r=0}^{2} \lambda_p^p \lambda_r^r Z(p,r) = 1 + \lambda_p x_p + \lambda_r \left(x_a + x_m + x_a x_m x_L \right) + \lambda_p \lambda_r x_a x_p + \lambda_r^2 x_a x_m$$

Complex regulatory scenario's can be expressed into universal scaling form:

$$\text{fold-change} = \begin{cases} \dfrac{1}{1+z_L} & \text{(looping)} \\ \dfrac{1}{1+z_{EL}}, & \text{exclusive looping} \end{cases}$$

$$z_L = \frac{\lambda_r(x_m + x_a x_m x_L) + \lambda_r^2 x_a x_m}{1 + \lambda_r x_a}.$$

$$z_{EL} = \frac{\lambda x_a x_m x_L}{1 + \lambda(x_a + x_m) + \lambda^2 x_a x_m}$$



Data collapse of several regulatory scenario's

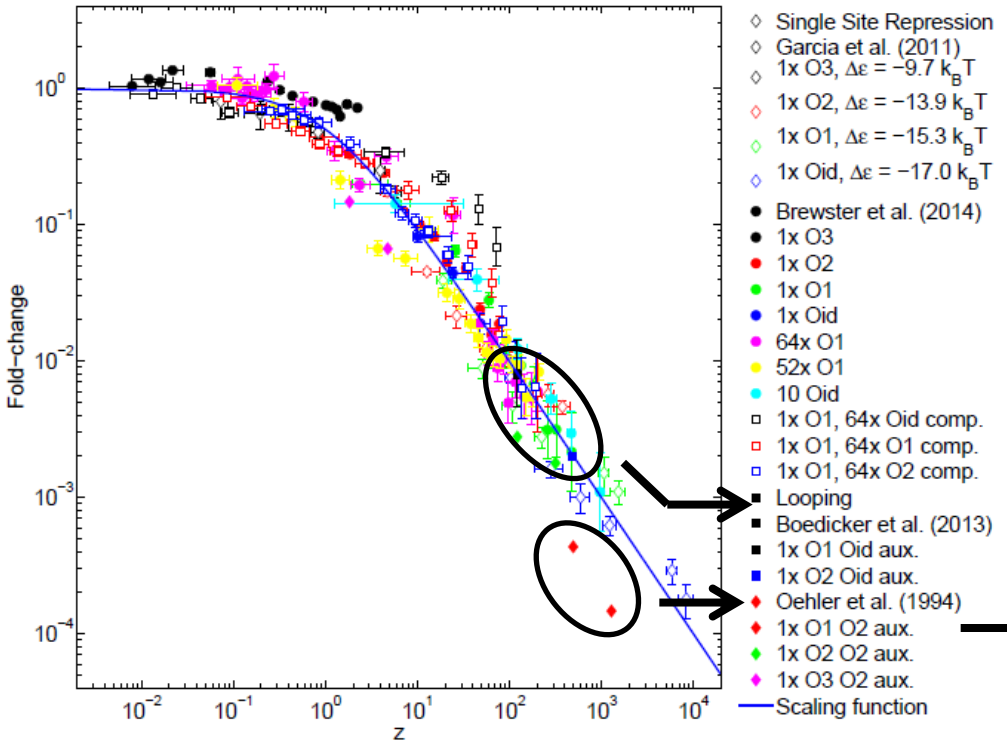
Simple repression: $z=\lambda_r e^{-\beta\varepsilon_s}$

Complex regulatory scenario's can be expressed into universal scaling form:

$$\text{fold-change} = \begin{cases} \dfrac{1}{1+z_L} & \text{(looping)} \\ \dfrac{1}{1+z_{EL}}, & \text{exclusive looping} \end{cases}$$

$$z_L = \frac{\lambda_r(x_m + x_a x_m x_L) + \lambda_r^2 x_a x_m}{1 + \lambda_r x_a}.$$

$$z_{EL} = \frac{\lambda x_a x_m x_L}{1 + \lambda(x_a + x_m) + \lambda^2 x_a x_m}$$



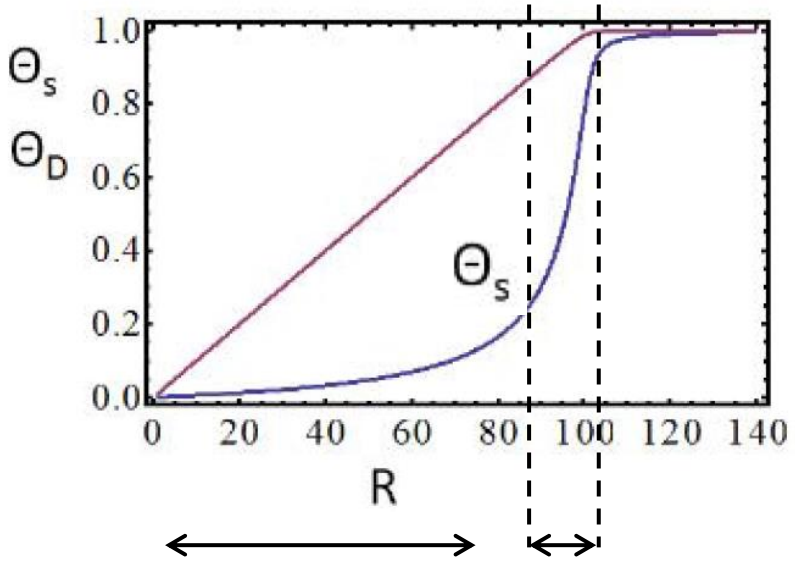
→ 'spot on!'

→ Deviation likely due to variation in R

Prediction: **A bacterial transistor**

Strong non-linear Fold-change response **in the presence of decoys**

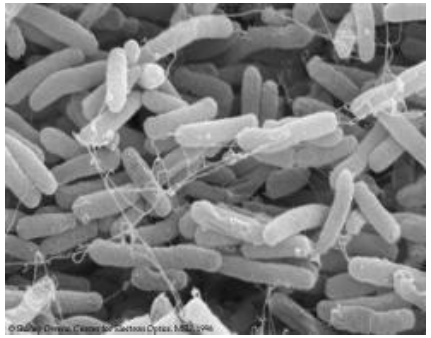
-> Effective value of R is often regulated in cells. R can be active or inactive.



Weak response Strong response

Input R(t)

(via additional regulation factor)



Output Fc(t)

Input sinusoidal variation

$$R(t) = R_0 + \Delta R \sin(\omega t)$$

$N_{NS} = 5 \times 10^6$

$N_s = 1$

$N_c = 100$

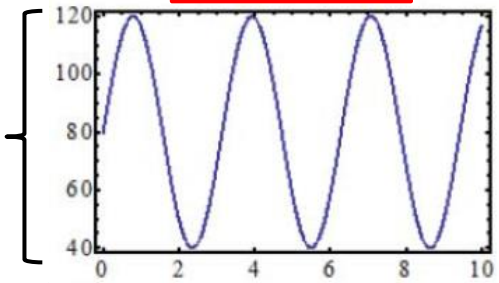
$\epsilon_s = -17kT$

$\epsilon_c = -20\, k_B T$

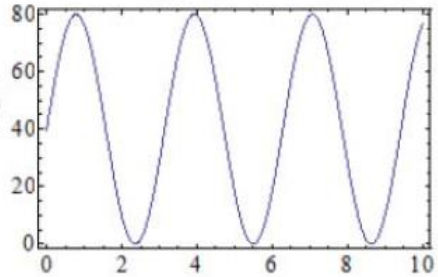
$\Delta R = 40$

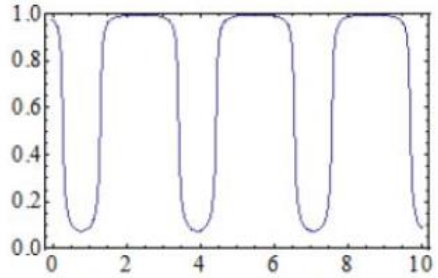
$R_0 = 80$

$R_0 = 40$

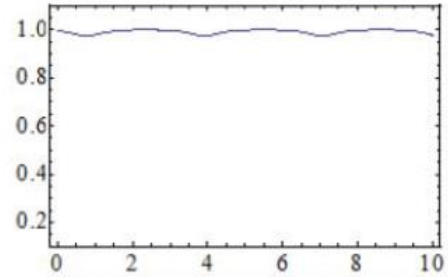


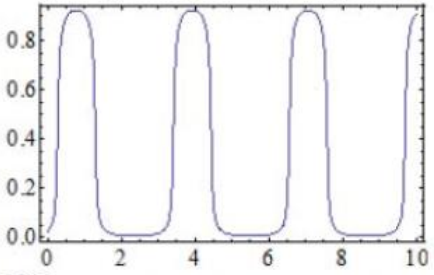
$R_0 + \Delta R \sin(\omega t)$



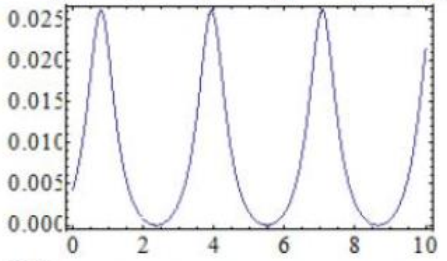


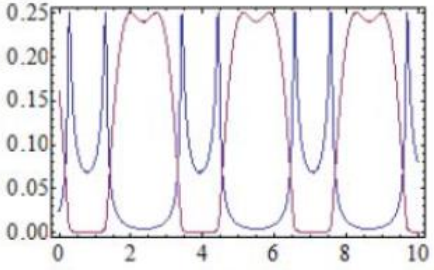
fc





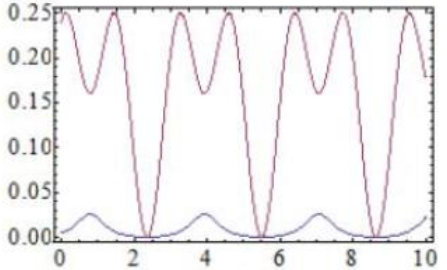
Θ_s





σ_s^2 / N_s (blue)

σ_D^2 / N_D



Response

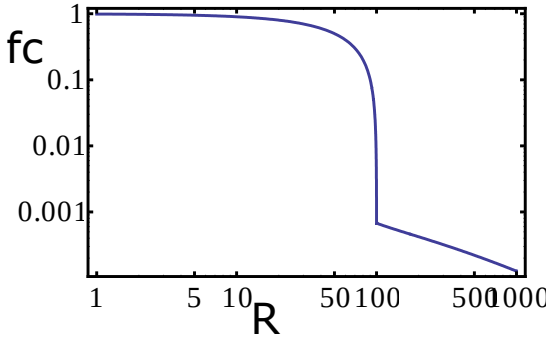
ωt

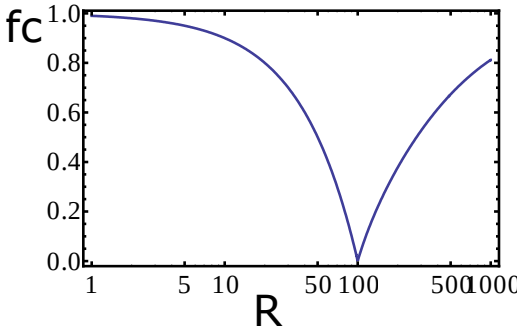
ωt

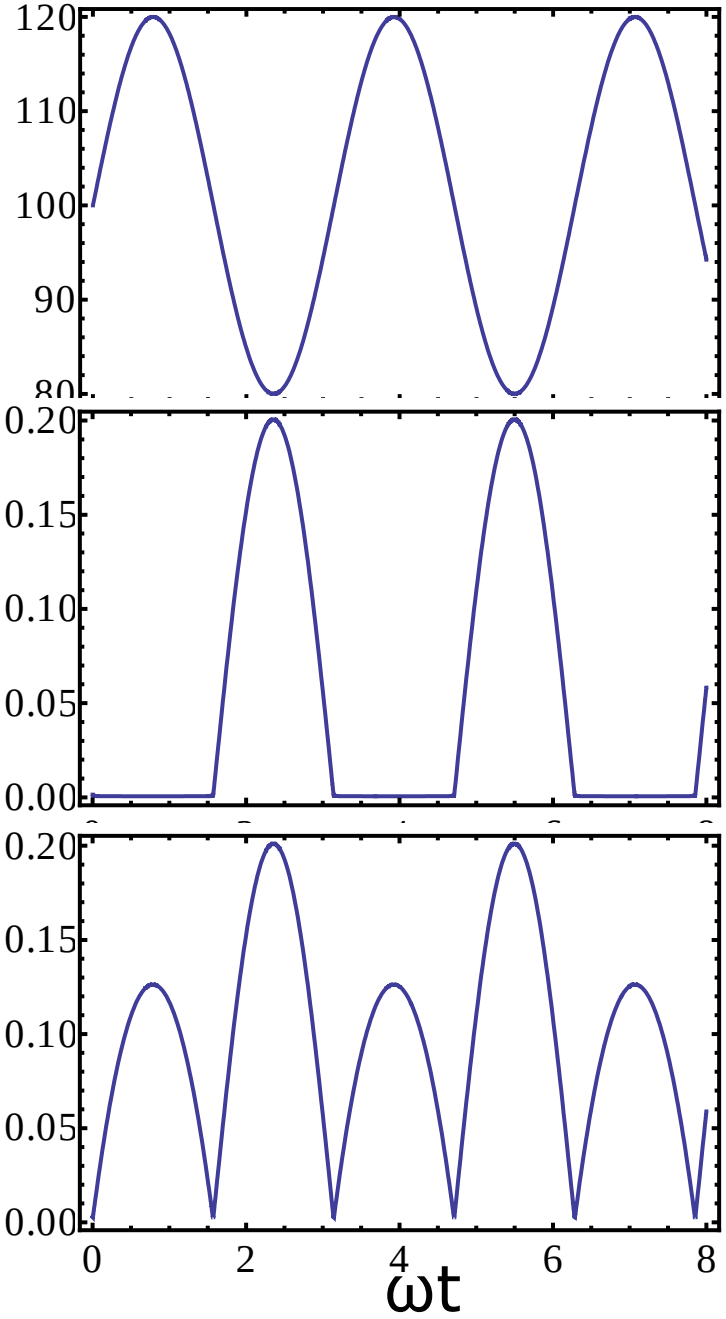
In case of looping / exclusive looping:

$$R(t) = R_0 + \Delta R \sin(\omega t)$$

genes = 100
R_0=100
ΔR = 20
ε_m=-17 kT
ε_a=-17 kT
F_L = 10 kT







Input

Looping

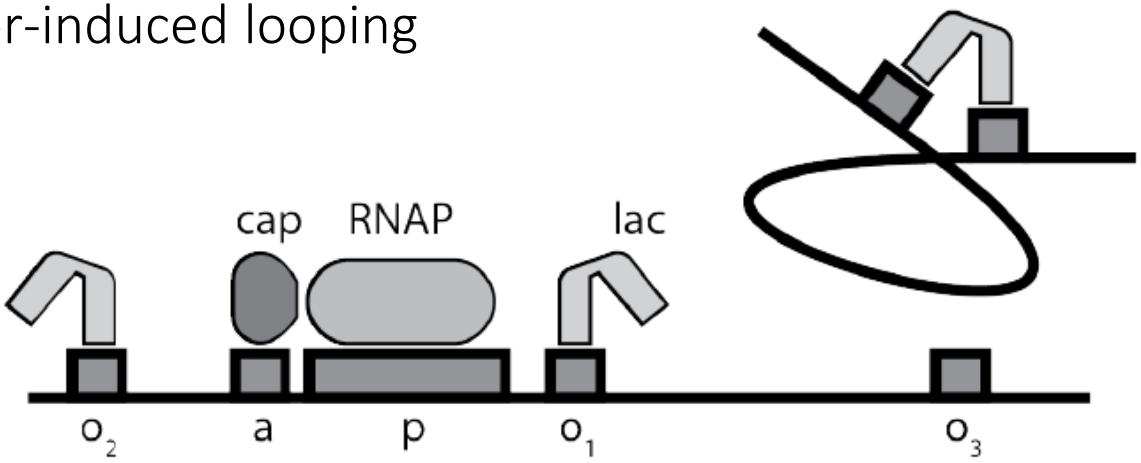
Exclusive looping

↓

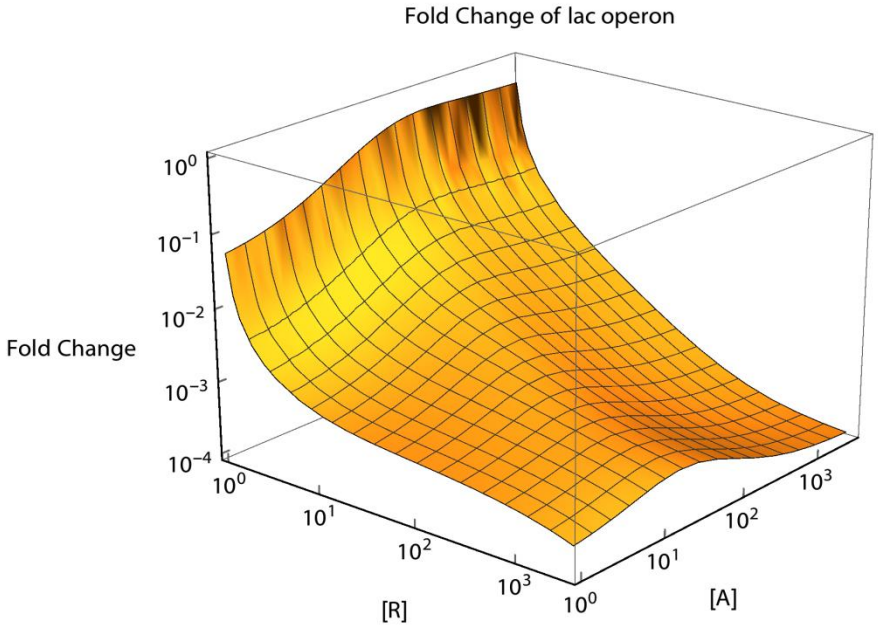
Period doubling

IN PROGRESS:

The full Lac-operon: 1 main operator, 2 auxiliary operators, 1 activator, activator-induced looping

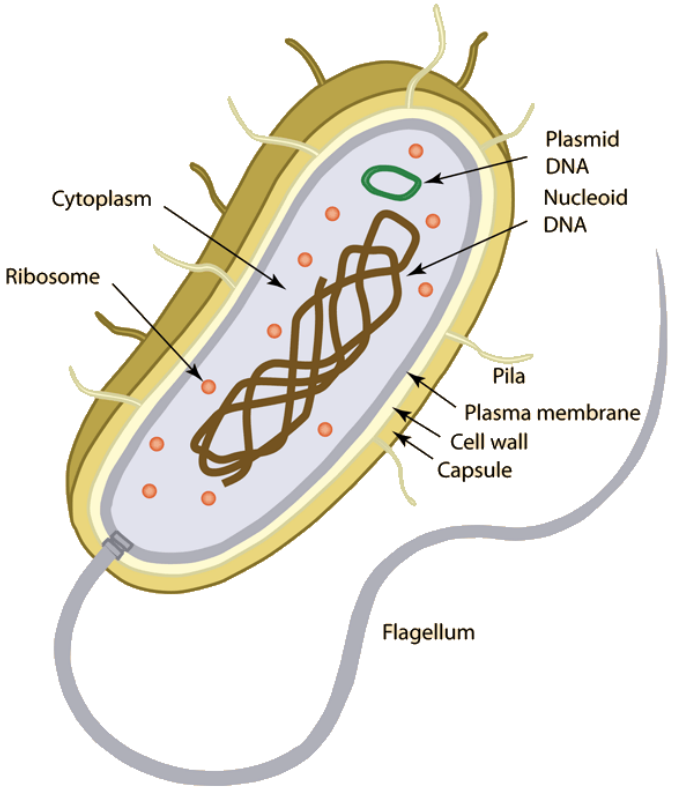


with Jasper Landman



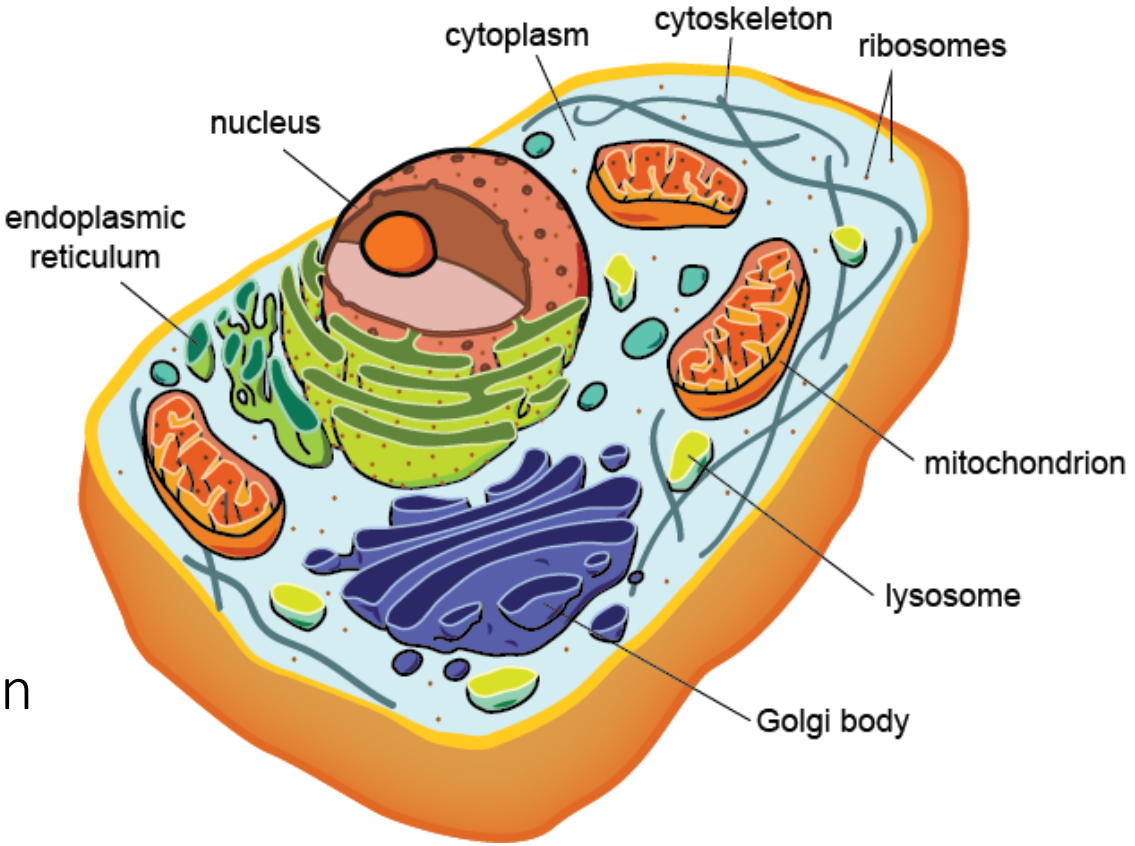
Repressor – operator energies
as in Boedicker ea PRL 2013 ;
Activator properties as in Phillips ea PBoC 2012
Looping energies O1-O2 = O1-O3 = 9.1 kT
Ignore O2 – O3.

Outlook: Does this work in cells of a eukaryote?



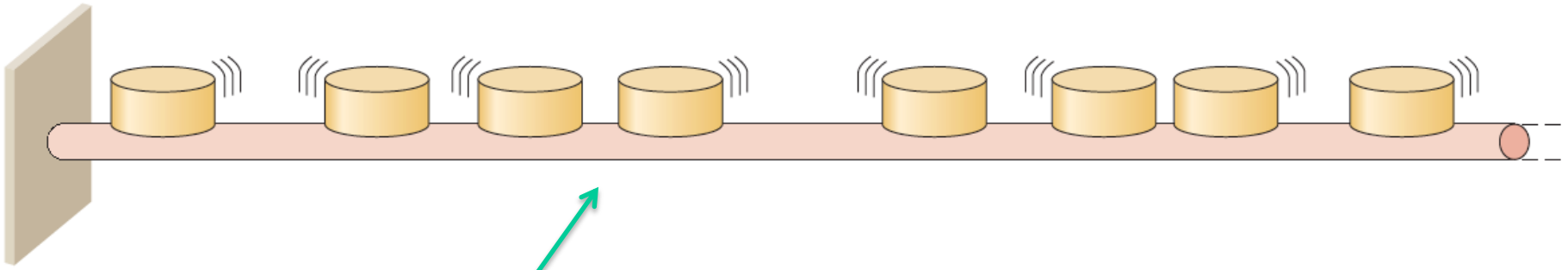
Bacterial

Eukaryote

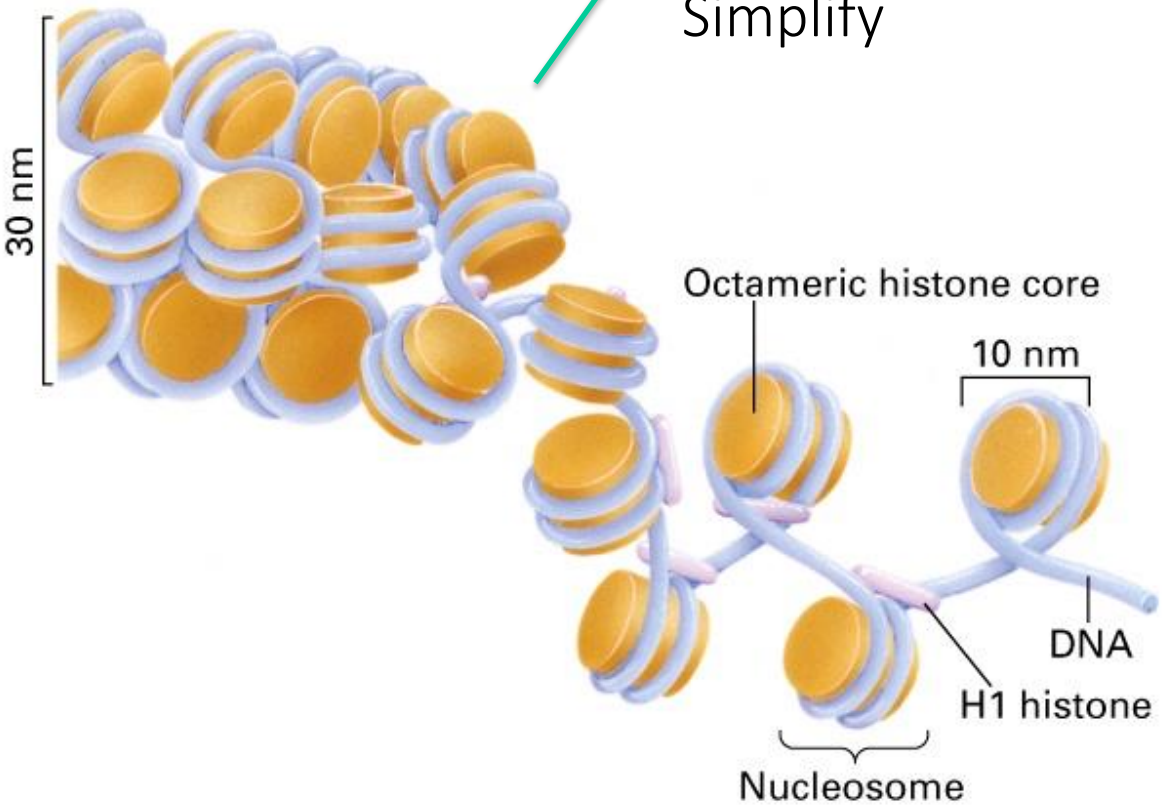


- Nucleosomes
- Compartimentalisation
- Different time scales?

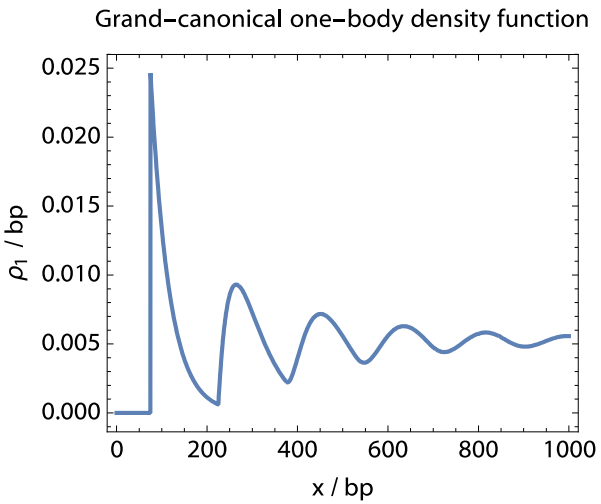
Outlook - Nucleosomes



Simplify



Analytical solution to 1D gas of hard rods exists



Conclusions & further work

- Scaling of all the relevant available data implies: linear combination of effective concentration (λ) times Boltzmann factor fully determines the genetic activity. Mechanism == binding / unbinding of regulatory proteins.

- More reservoirs / lattices can be added – leads to higher order polynomials in λ

- Fake-genes & stat mech make quantitative predictions and falsification possible in biology – theory provides PREJUDICE

- 'Coarse grain' multiple regulatory mechanisms into equivalent circuits with well-defined I/O's -> extendable to dynamics?

- Could this work in principle in eukariotic cells? How to treat 'compartments', chromatin 'action'?

Thank You

Rob Phillips

and

Mattias Rydenfelt

Franz Weinert

Jasper Landman

Rob Brewster

Quantitatively measuring gene expression