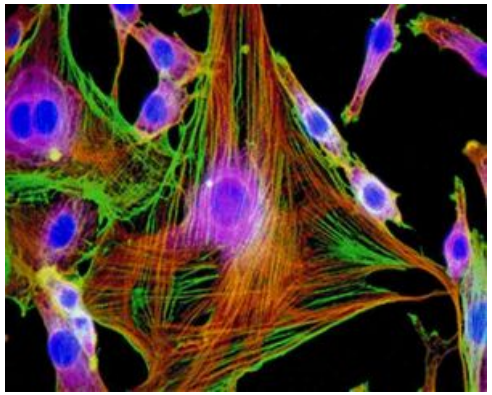
lecture 5
timescale separation and the linear framework

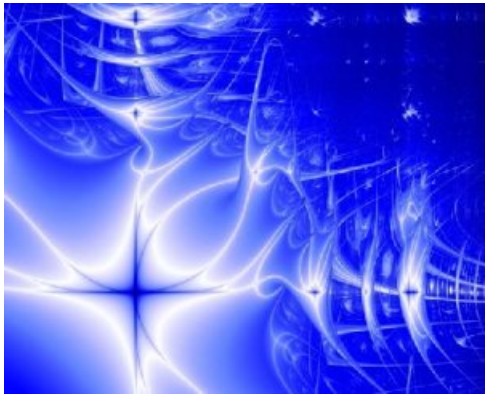
jeremy gunawardena



department of systems biology
harvard medical school
200 longwood avenue
boston, ma 02115

jeremy@hms.harvard.edu

http://vcp.med.harvard.edu/



“A Systems Approach to Biology”, UBA Buenos Aires, 11-22 June 2018

syllabus

1. the role of mathematics in biology

2. homeostasis of the organism

3. the complexity of evolution

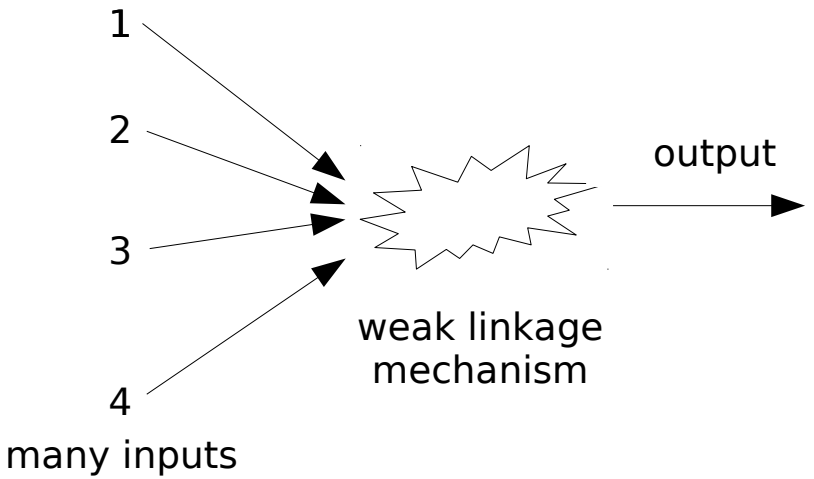
4. weak linkage and learning

5. timescale separation and the linear framework

0. preamble

weak linkage yields combinatorial complexity

by undertaking scalable integration, weak linkage mechanisms acquire substantial internal state through combinatorial complexity



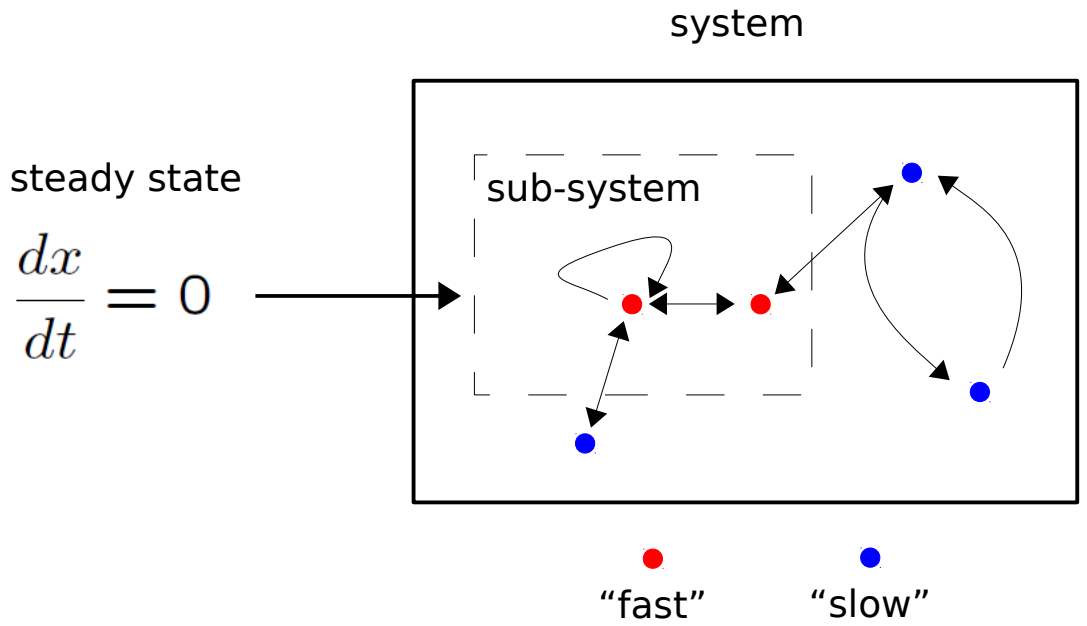
- allostery – multiple conformations and patterns of binding
- PTMs – multiple patterns of modification (“modforms”)
- genes – multiple patterns of transcription factor (TF) binding

the “**linear framework**” is a mathematical method for calculating the input-output responses of such mechanisms, using **timescale separation** to eliminate the combinatorial complexity

1. timescale separation

timescale separation

mathematically eliminating the components of a sub-system by assuming the sub-system is at steady state

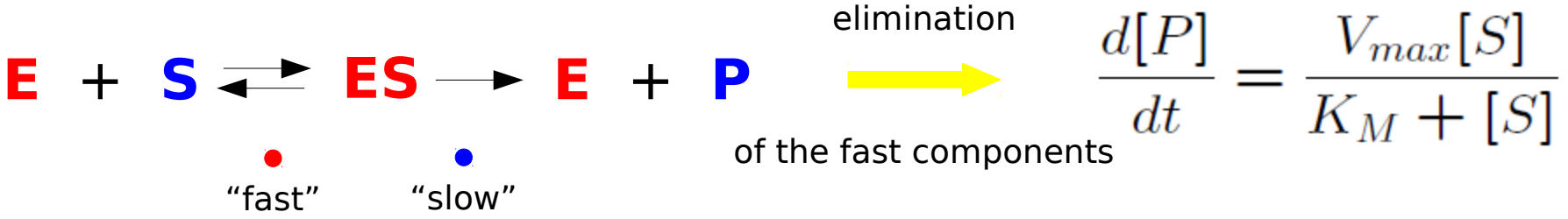


timescale separation is often used to simplify the mathematics even when the steady-state assumption is difficult to justify – often with surprisingly good results

michaelis & menten

under the in-vitro conditions which they used, substrate was in substantial excess over enzyme, so it seemed reasonable to assume that they (hypothetical) enzyme-substrate complex was a “fast” component

1879-1960 *1875-1949*



Michaelis & Menten, *“Die kinetik der Invertinwirkung”*, Biochem Z, **49**:333-69, 1913

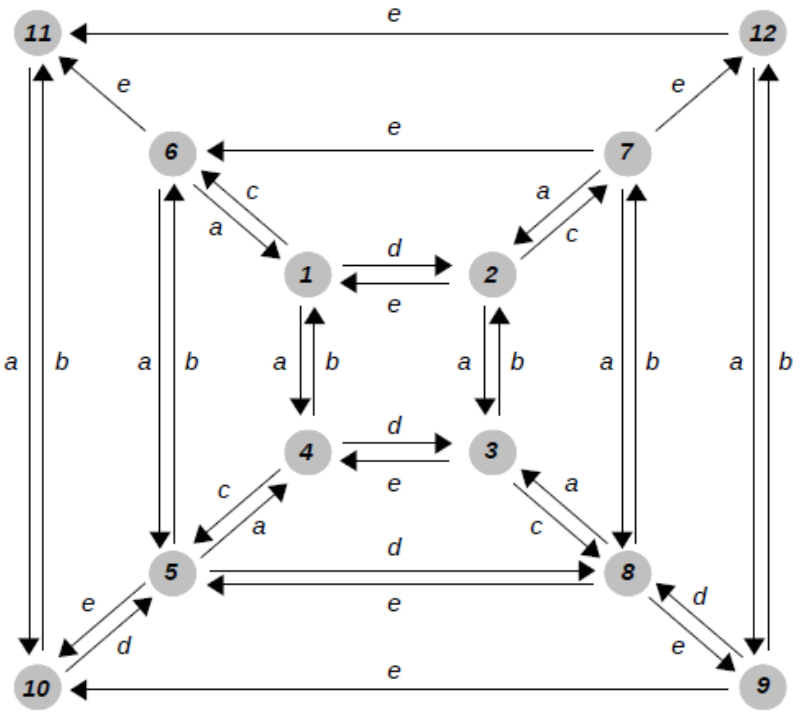
2. graphs and the linear framework

labelled, directed graphs

a **graph** consists of **vertices** (or **nodes**), with at most one **edge** between any two distinct vertices

the graph is **directed** – each edge has a specified direction, denoted by an arrow at one end

the graph is **labelled** on each edge. labels have **units of (time)$^{-1}$**

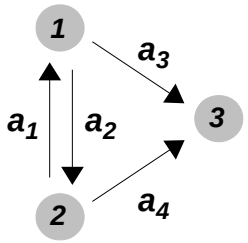


we shall work with graphs which are **connected** (in one piece, forgetting edge directions) and which have **no self-loops**

such graphs usually represent the sub-system that is at steady state. the vertices represent “fast” components; the edges represent reactions; and the labels represent the influence of the “slow” components

the linear framework

the dynamics on graphs is “one-dimensional” chemistry – each edge is considered as a chemical reaction under mass-action kinetics, with the label as the rate constant



Laplacian matrix

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -(a_2+a_3) & a_1 & 0 \\ a_2 & -(a_1+a_4) & 0 \\ a_3 & a_4 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$G \quad \Rightarrow \quad \frac{dx}{dt} = \mathcal{L}(G).x$$

system of linear ODEs
discretised diffusion equation

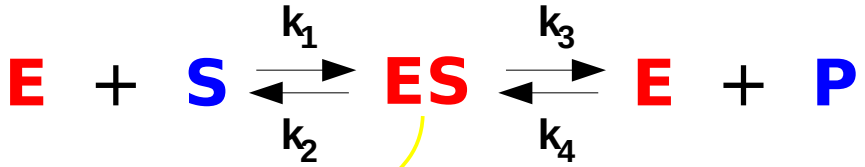
conservation law:

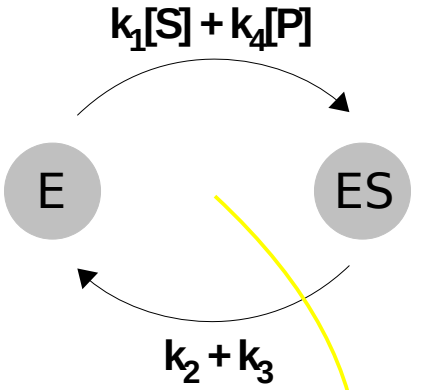
$$x_1(t) + x_2(t) + \cdots + x_n(t) = x_{tot} \qquad 1.\mathcal{L}(G) = 0$$

Gunawardena, *“A linear framework for time-scale separation in nonlinear biochemical systems”*, PLoS ONE **7**:e36321 2012; Mirzaev & Gunawardena, “*Laplacian dynamics on general graphs*”, Bull Math Biol **75**:2118-49 2013; Gunawardena, *“Time-scale separation: Michaelis and Menten's old idea, still bearing fruit”*, FEBS J **281**:473-88 2014.

the nonlinearity is rewritten using the labels

reversible Michaelis-Menten scheme





nonlinear dynamics

$$\frac{d[E]}{dt} = (k_2 + k_3)[ES] - k_1[E][S] - k_4[E][P]$$

$$\frac{d[ES]}{dt} = -(k_2 + k_3)[ES] + k_1[E][S] + k_4[E][P]$$

linear Laplacian dynamics

$$\frac{d}{dt}\begin{pmatrix} [E] \\ [ES] \end{pmatrix} = \begin{pmatrix} -(k_1[S] + k_4[P]) & (k_2 + k_3) \\ (k_1[S] + k_4[P]) & -(k_2 + k_3) \end{pmatrix}\begin{pmatrix} [E] \\ [ES] \end{pmatrix}$$

uncoupling and the labels

uncoupling condition – the labels cannot have concentrations of components which are also vertices in the graph

the nonlinearity in the labels is dealt with in different ways, depending on the application

- approximation $[E] \approx E_{tot}$
- conservation law $E_{tot} = [E] + [ES^{(0)}] + [ES^{(1)}] + \cdots + [ES^{(K)}]$
- singular perturbation $\frac{d[P]}{dt} = \alpha [ES^{(j)}]$
- hierarchical graphs – labels are given by components in a different graph

microscopic interpretation

let $X(t)$ be a time-homogeneous **Markov process** on the states $1, \cdots, n$
for which **infinitesimal transition rates** exist –

$$\lim_{\Delta t \to 0} \frac{\Pr(X(t+\Delta t) = i \mid X(t) = j)}{\Delta t} = a_{ij}$$

define the graph, G_X , with vertices $1, \cdots, n$ and an edge $j \to i$ iff $a_{ij} \neq 0$
give this edge the label a_{ij}

the **master equation** (Kolmogorov forward equation), for the probability of $X(t)$
being in state i at time t, is identical to Laplacian dynamics on G_X

$$x_i(t) = \Pr(X(t) = i) \qquad \boxed{\frac{dx}{dt} = \mathcal{L}(G_X).x}$$

Mirzaev & Gunawardena, "*Laplacian dynamics on general graphs*", Bull Math Biol **75**:2118-49 2013

3. calculating steady states

uniqueness of steady states

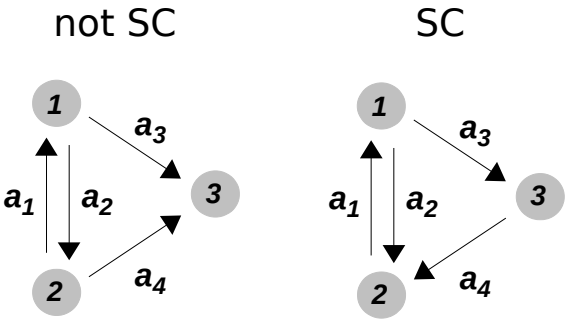
for any graph, G, Laplacian dynamics always tends to a steady state

$$x(t) \to x^* \qquad \left.\frac{dx}{dt}\right|_{x=x^*} = 0 \qquad x^* \in \ker \mathcal{L}(G)$$

if G is **strongly connected**, the steady state is unique up to a scalar multiple

$$\dim \ker \mathcal{L}(G) = 1$$

strongly connected – there is a directed path between any two distinct vertices



Mirzaev, Gunawardena, Bull Math Biol **75**:2118-49 2013

the matrix-tree theorem gives a canonical s.s

Matrix-Tree Theorem (MTT): whenever G is strongly connected

$$\ker \mathcal{L}(G) = \langle \rho \rangle \qquad \rho_i = \sum_{T \in \Theta_i(G)} \left(\prod_{j \xrightarrow{a} k \in T} a \right)$$

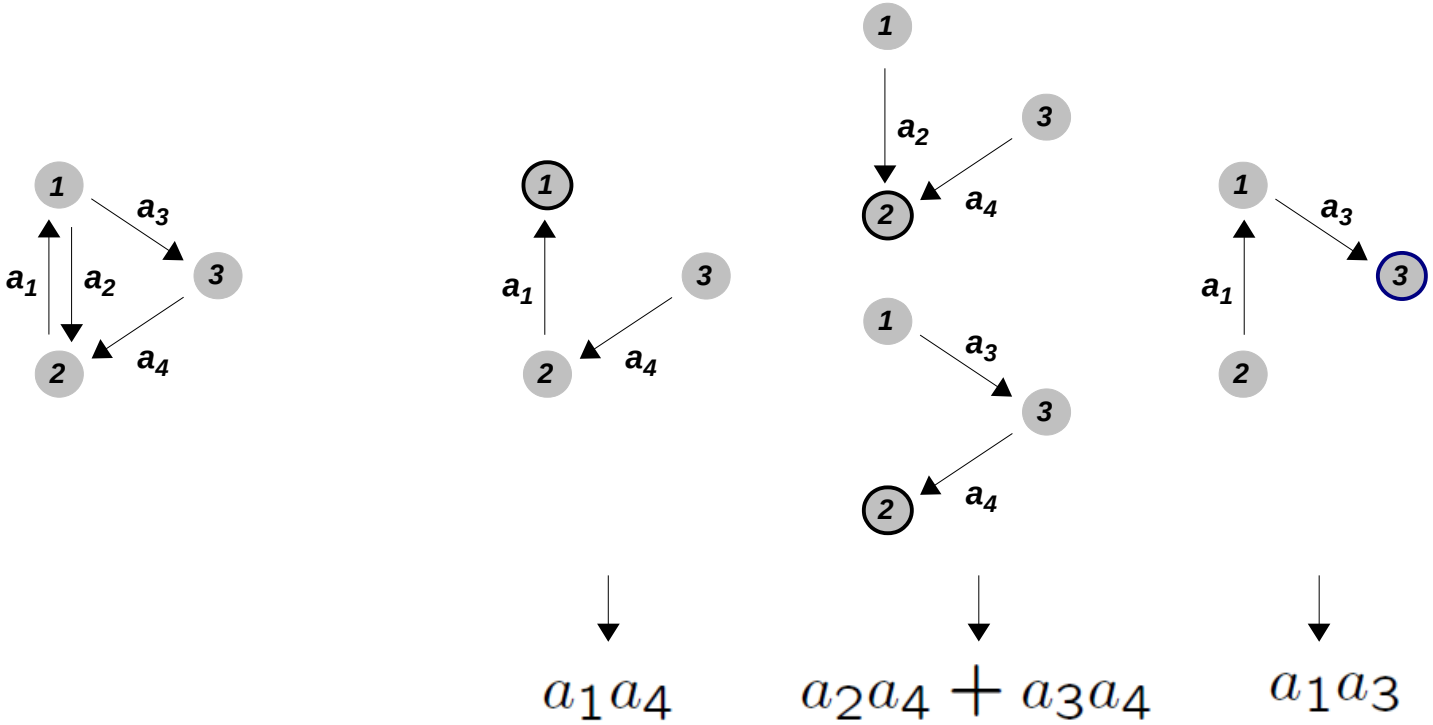
$\Theta_i(G) =$ set of spanning trees rooted at *i*

rooted spanning tree – a sub-graph T of G which

- **SPANS** G – every node of G is also a node of T
- is a **TREE** – T has no cycles, ignoring edge directions
- is **ROOTED** at *i* – *i* is the only node of T with no outgoing edges

Bill Tutte, *"The dissection of equilateral triangles into equilateral triangles"*, Proc Camb Phil Soc **44**:463-82 1948; Mirzaev & Gunawardena, Bull Math Biol **75**:2118-49 2013 – Appendix gives a proof of the MTT

spanning trees and the MTT



$$\begin{pmatrix} -(a_2+a_3) & a_1 & 0 \\ a_2 & -a_1 & a_4 \\ a_3 & 0 & -a_4 \end{pmatrix} \begin{pmatrix} a_1 a_4 \\ (a_2+a_3)a_4 \\ a_1 a_3 \end{pmatrix} = 0$$

Laplacian **ρ**

how elimination works

when G is strongly connected, so that $\ker \mathcal{L}(G) = \langle \rho \rangle$

if there is a steady state $x^* \in \ker \mathcal{L}(G)$

$$x^* = \lambda \rho \qquad \begin{pmatrix} x_1^* \\ \vdots \\ x_n^* \end{pmatrix} = \lambda \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_n \end{pmatrix}$$

then each of the x_i^* can be **eliminated** in favour of the ρ_i

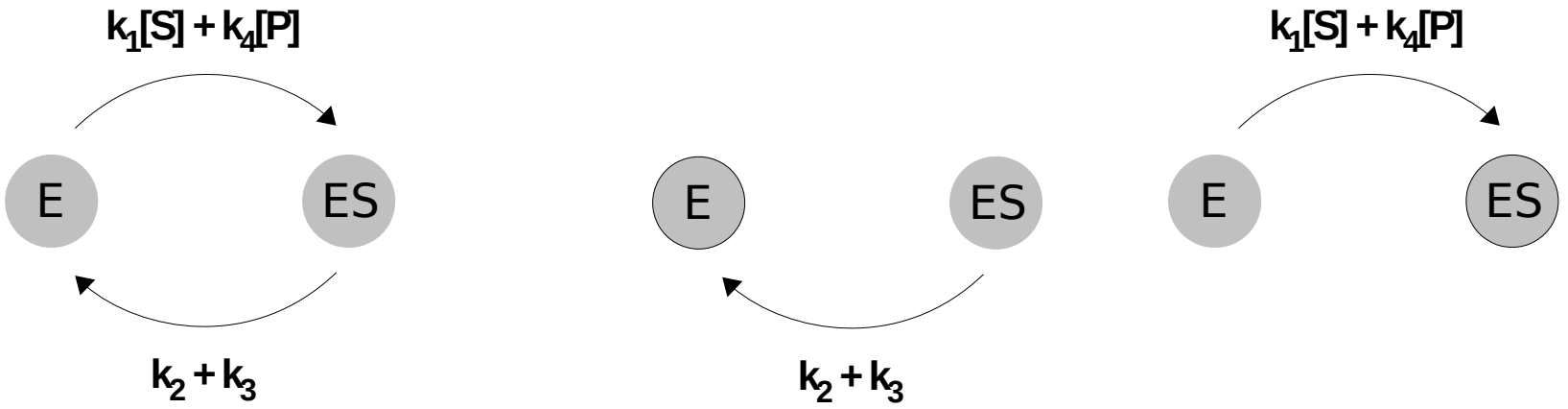
reference node

$$x_i^* = \left(\frac{\rho_i}{\rho_1 + \cdots + \rho_n} \right) x_{tot} \qquad\qquad x_i^* = \frac{\rho_i}{\rho_1} x_1^*$$

and the ρ_i are given in terms of the edge labels by the MTT

example – reversible michaelis-menten II

enumeration of spanning trees

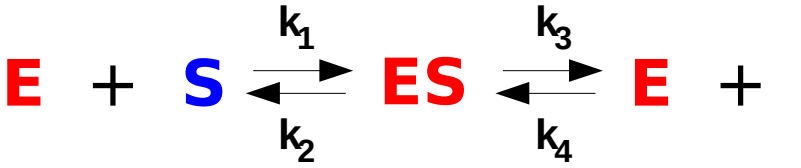


$$\rho = \begin{pmatrix} k_2 + k_3 \\ k_1[S] + k_4[P] \end{pmatrix}$$

elimination

$$[ES] = \left(\frac{k_1[S] + k_4[P]}{k_2 + k_3 + k_1[S] + k_4[P]} \right) E_{tot} \quad [E] = \left(\frac{k_2 + k_3}{k_2 + k_3 + k_1[S] + k_4[P]} \right) E_{tot}$$

example – reversible michaelis-menten II



$$\frac{d[P]}{dt} = k_3[ES] - k_4[E][P]$$

substitute steady-state values of "fast" components from MTT

$$\frac{d[P]}{dt} = \left(\frac{V_f[S]/K_f - V_r[P]/K_r}{1 + [S]/K_f + [P]/K_r} \right)$$

$$V_f = k_3 E_{tot} \quad V_r = k_2 E_{tot} \qquad K_f = \frac{k_2 + k_3}{k_1} \quad K_r = \frac{k_2 + k_3}{k_4}$$

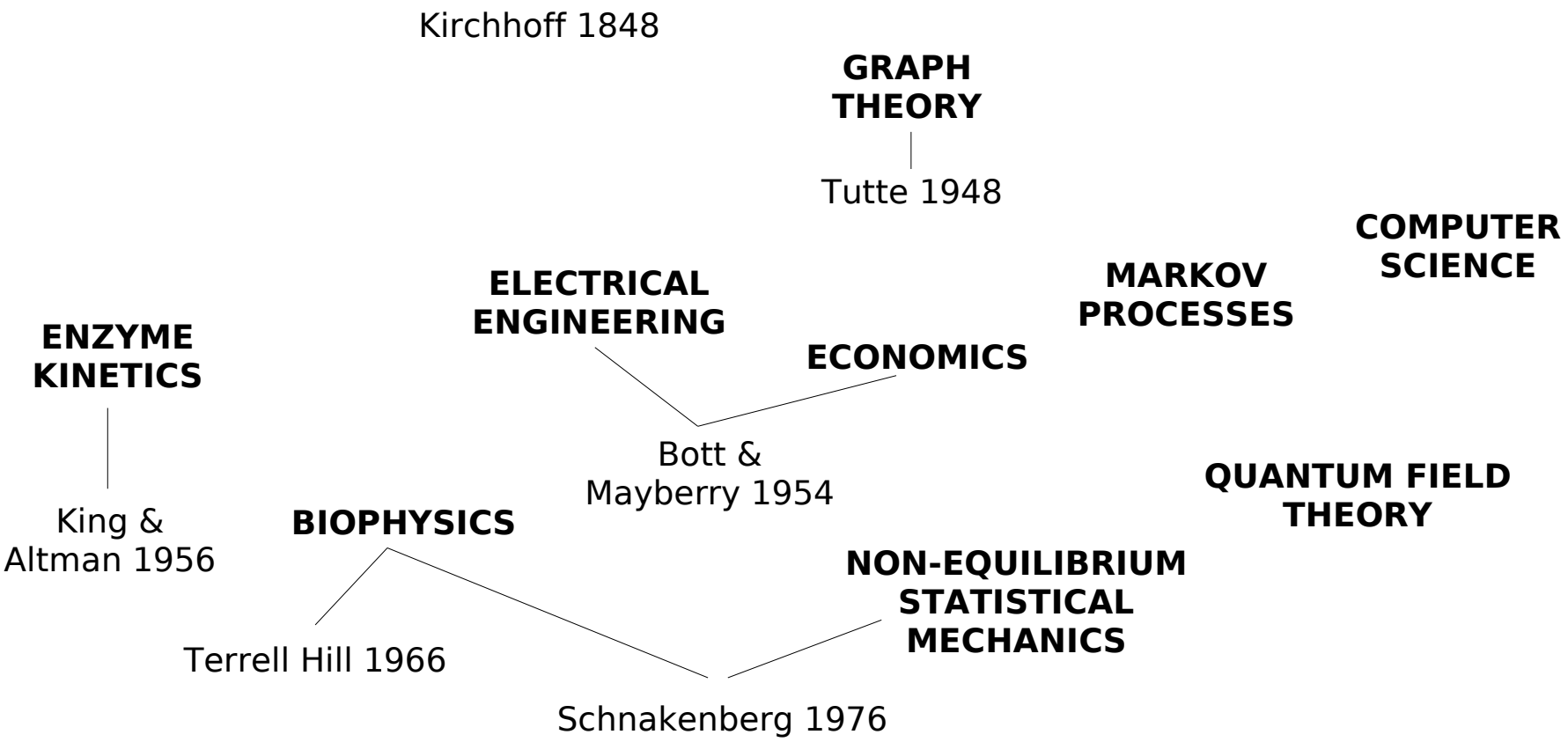
forward & reverse maximal rates

forward & reverse Michaelis-Menten constants

Athel Cornish-Bowden, **Fundamentals of Enzyme Kinetics**, 2nd edition, Portland Press, 2001

e pluribus unum

independent discoveries of the MTT



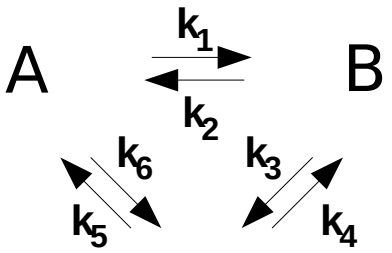
Gunawardena, PLoS ONE **7**:e36321 2012; Mirzaev, Gunawardena; Bull Math Biol **75**:2118-49 2013; Gunawardena, FEBS J **281**:473-88 2014

4. equilibrium and energy

(thermodynamic) equilibrium is a very special s.s.

principle of detailed balance: at thermodynamic equilibrium, every reaction is reversible and each pair of reversible reactions is separately at equilibrium, irrespective of any other reactions in which the components participate

$$k_1[A] = k_2[B]$$



$$k_5[C] = k_6[A]$$

$$k_3[B] = k_4[C]$$

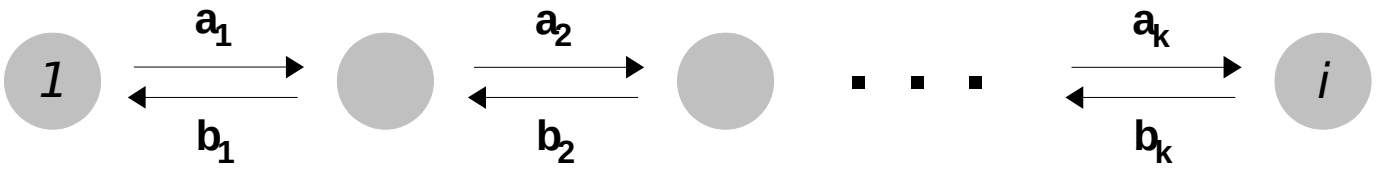
$$k_1 k_3 k_5 = k_2 k_4 k_6$$

cycle condition

Gilbert Lewis, *"A new principle of equilibrium"*, PNAS **11**:179-83 1925; Mahan, "*Microscopic reversibility and detailed balance; an analysis*", J Chem Edu **52**:299-302 1975

the MTT simplifies at equilibrium

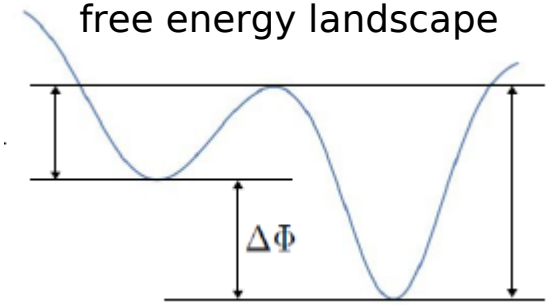
if the steady-state is one of thermodynamic equilibrium, then it is not necessary to enumerate spanning trees -



detailed balance

$$\rho_i = \left(\frac{a_1}{b_1}\right)\left(\frac{a_2}{b_2}\right)\cdots\left(\frac{a_k}{b_k}\right)\mu$$

free energy landscape



$$\left(\frac{a_1}{b_1}\right) = \exp\left(\frac{-\Delta\Phi}{k_B T}\right)$$

steady-state calculations become equivalent to equilibrium statistical mechanics

BUT the MTT remains valid away from equilibrium and thereby gives a restricted form of non-equilibrium statistical mechanics

the hopfield barrier

Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity

(protein synthesis/DNA replication/amino-acid recognition)

J. J. HOPFIELD

Proc. Nat. Acad. Sci. USA
Vol. 71, No. 10, pp. 4135–4139, October 1974

"THE HOPFIELD BARRIER"

thermodynamic equilibrium sets an upper bound to how well information processing tasks can be undertaken by a biochemical system.

the only way to exceed this barrier is to dissipate energy and maintain the system away from equilibrium

Estrada, Wong, DePace, Gunawardena, "*Information integration and energy expenditure in gene regulation*", Cell **166**:234-44 2016

the problem of path-dependence

at thermodynamic equilibrium, the MTT simplifies – it is only necessary to use a single path in the graph to calculate steady-state probabilities

away from thermodynamic equilibrium, it is necessary to enumerate all rooted spanning trees in the graph – each path in the graph contributes

$$\rho_i = \sum_{T \in \Theta_i(G)} \left(\prod_{j \xrightarrow{a} k \in T} a \right)$$

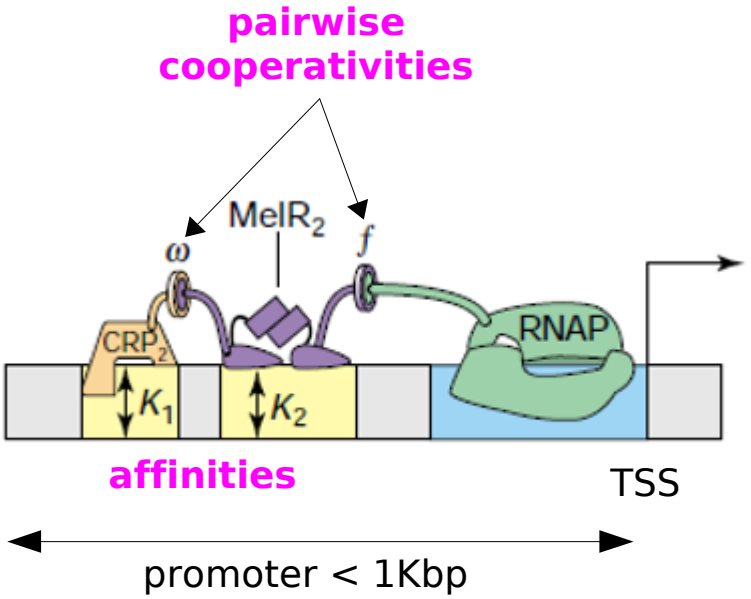
the number of spanning trees increases **super-exponentially** in the size of the graph

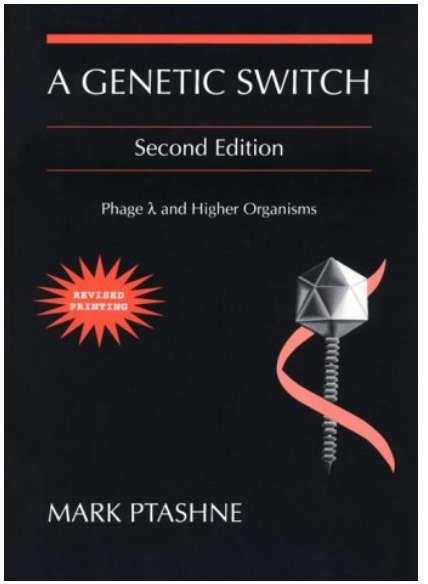
we need new mathematical ideas to address this problem

Ahsendorf, Wong, Eils, Gunawardena, BMC Biol **12**:102 2014; Estrada, Wong, DePace, Gunawardena, Cell **166**:234-44 2016

4. gene regulation

bacterial gene regulation

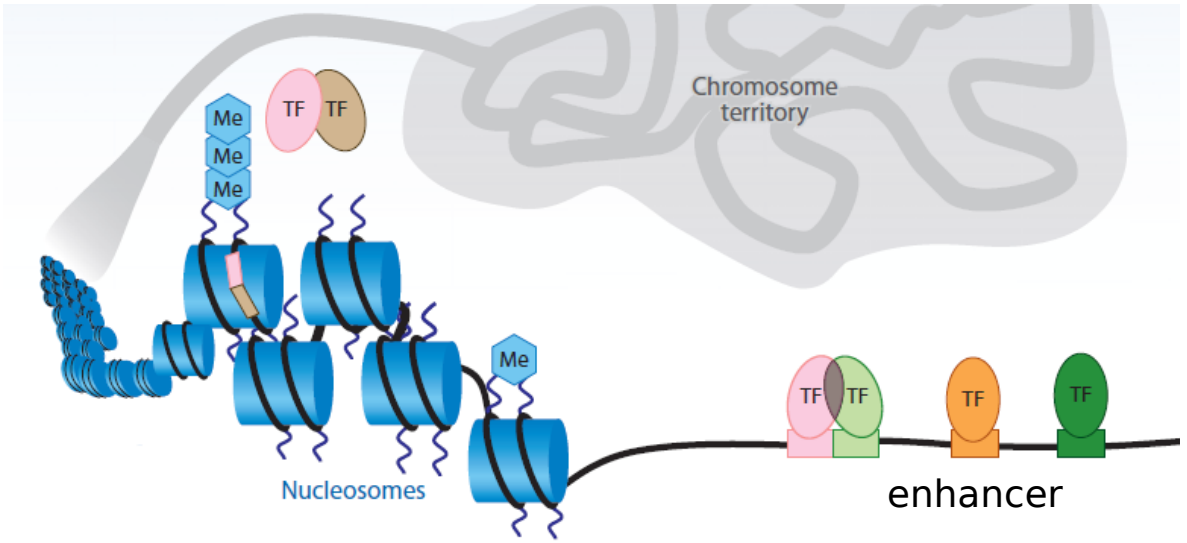


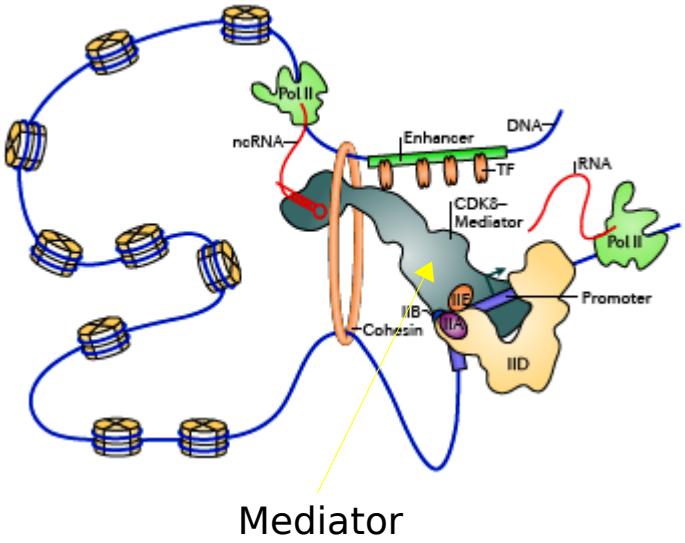


- specificity comes through transcription factors (TFs)
- long binding motifs, ~16bp on average
- information is conveyed over short distances
  through pairwise cooperative interactions between TF-TF, TF-RNAP
- regulation takes place without energy expenditure

Bintu, Buchler, Garcia, Gerland, Hwa, Kondev, Kuhlman, Phillips, "*Transcriptional regulation by the numbers I & II*", Curr Opin Gen Dev **15**:116-24 & 125-35 2005

eukaryotic gene regulation



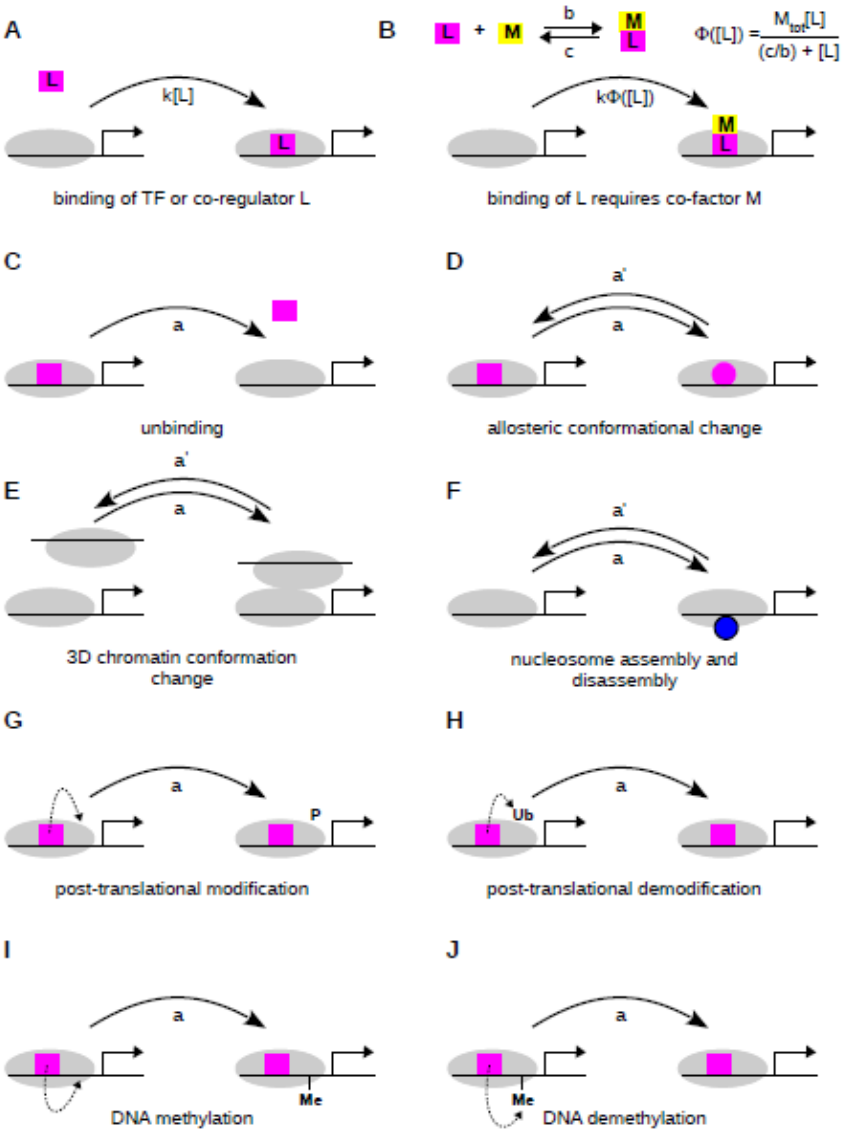


- hierarchical spatial organisation
- **information integration** over long distances
- co-regulatory complexes linking information sources
- short TF binding motifs, ~8bp on average
- many forms of energy expenditure –
    - chromatin reorganisation
    - nucleosome remodelling
    - PTM of histones, TFs, co-regulators, RNAP, ...

Lelli, Slattery, Mann, Annu Rev Genet **46**:43-68 2012; Allen, Taatjes, Nature Rev Mol Cell Biol **16**:155-66 2015; (*) Ong, Corces, Nature Rev Genet **12**:283-93 2011

linear framework graphs

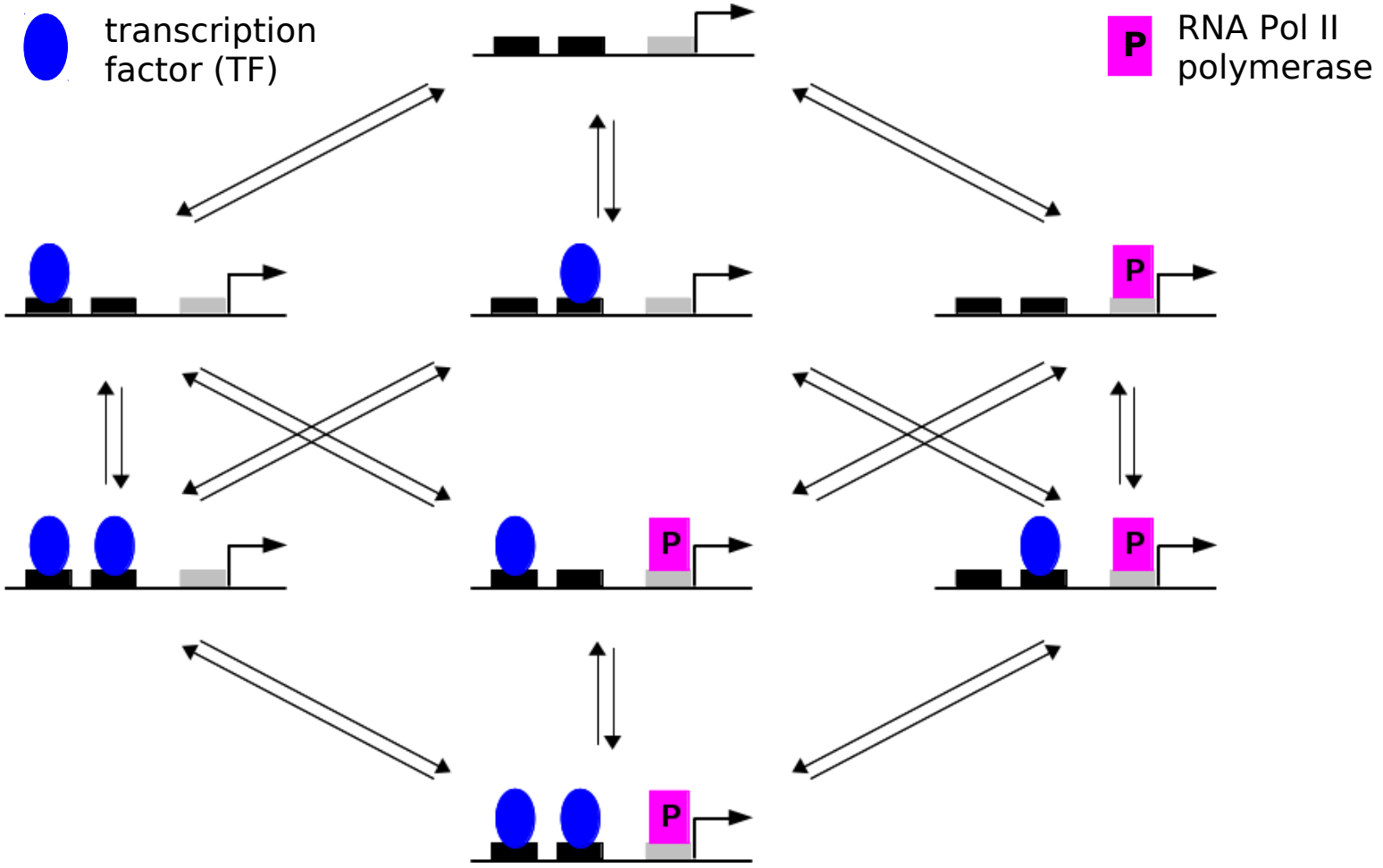
vertices are "snapshots" of DNA context at an appropriate level of granularity



Ahsendorf, Wong, Eils, Gunawardena, "*A framework for modelling gene regulation which accommodates non-equilibrium mechanisms*", BMC Biol **12**:102 2014.

graph for studying information integration

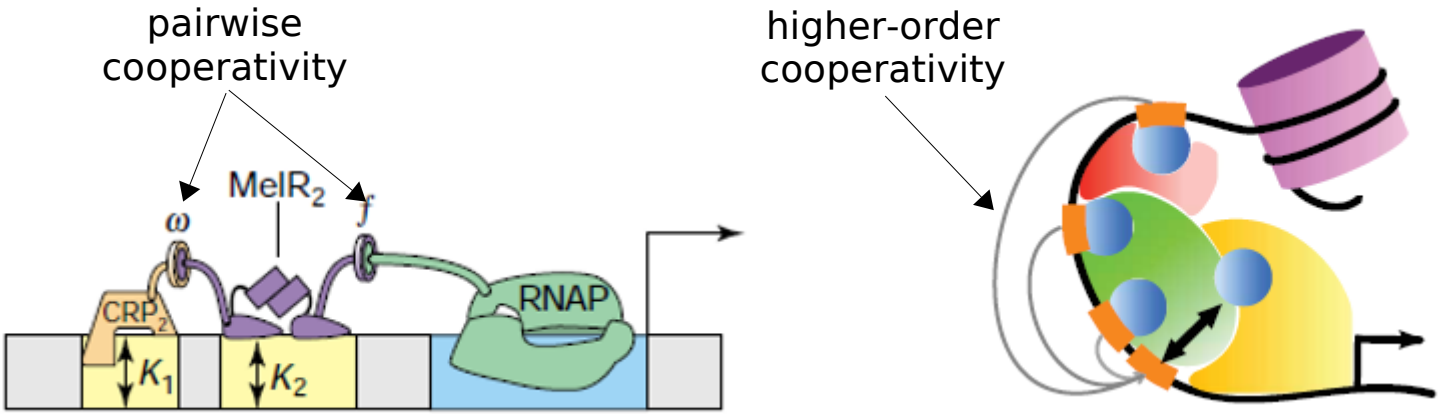
single transcriptional activator binding at n sites (n = 2 is shown below)



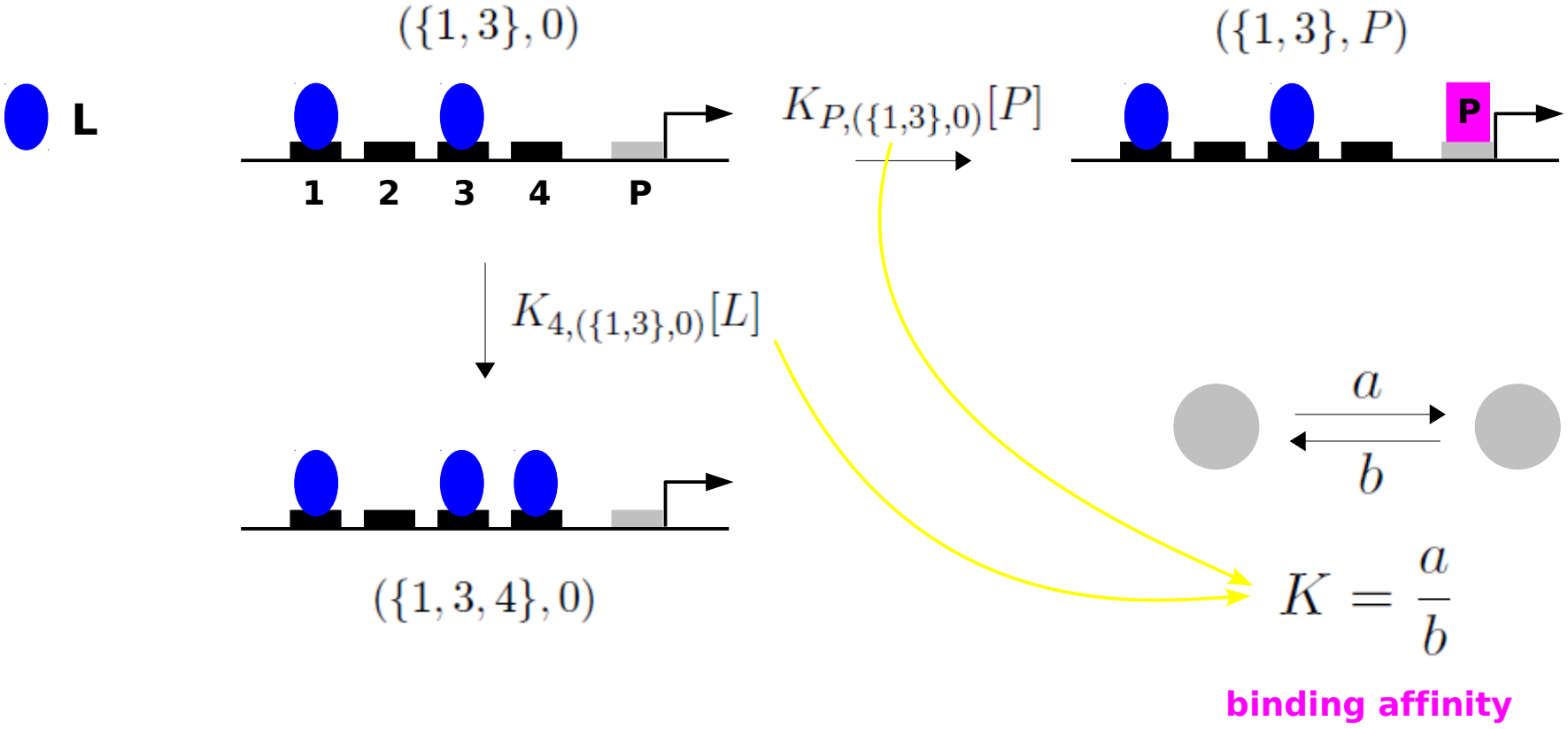
updated from Estrada, Wong, DePace, Gunawardena, Cell **166**:234-44 2016.

underlying assumptions

- the graph is at thermodynamic equilibrium
- expression is averaged – the rate of mRNA expression is proportional to the steady-state probability of RNA polymerase being present
- molecular complexity arising from chromatin, co-regulators, etc is not explicitly represented but is assumed to influence the edge labels
- in particular, this allows for “higher-order” cooperativity



binding and unbinding at equilibrium



higher-order cooperativity

two kinds of “higher-order” cooperativity

$$\omega_{i,(S,*)} = \frac{K_{i,(S,*)}}{K_{i,(\emptyset,*)}} \qquad \omega_{P,(S,0)} = \frac{K_{P,(S,0)}}{K_{P,(\emptyset,0)}}$$

TF-TF TF-RNAP

higher-order cooperativities

the “order” of cooperativity is $\#S$; pairwise cooperativity is $\#S = 1$

detailed balance must be considered – the parameters are not independent!

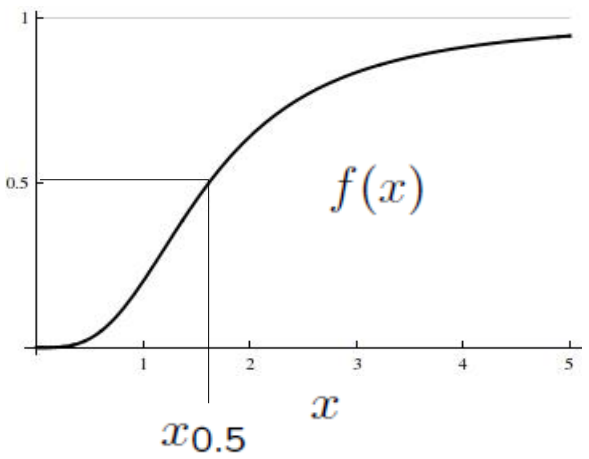
$$\omega_{i,(S\cup\{j\},*)}\omega_{j,(S,*)} = \omega_{j,(S\cup\{i\},*)}\omega_{i,(S,*)}$$

independent generators

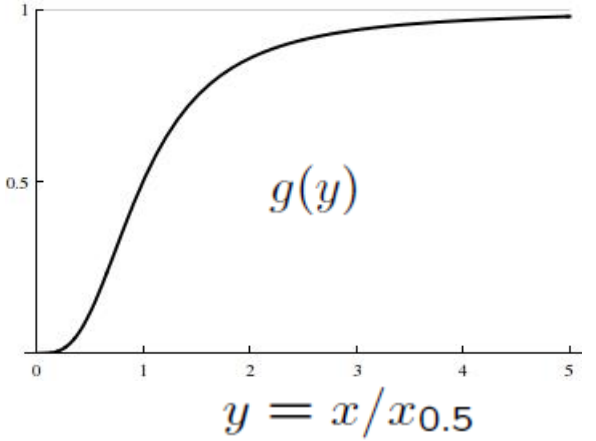
$$\omega_{i,(S,*)} \quad i < S$$

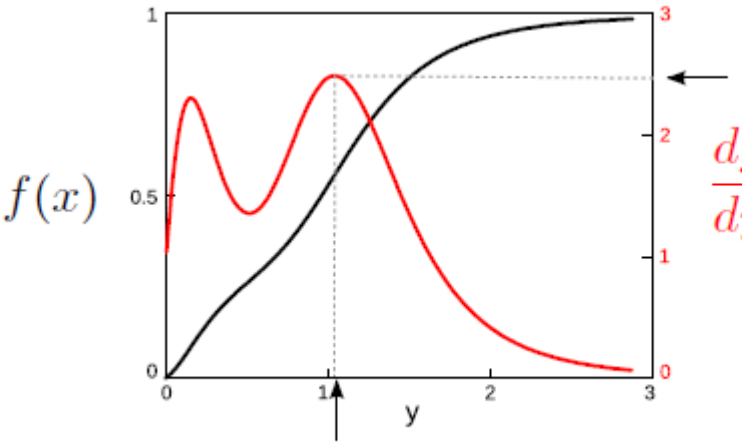
sharpness in gene regulation

gene regulation function – $f(x) = \Pr(\text{RNAP is bound})(x) \quad x = [L]$



normalisation



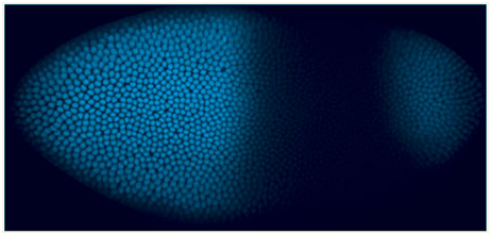


“steepness”

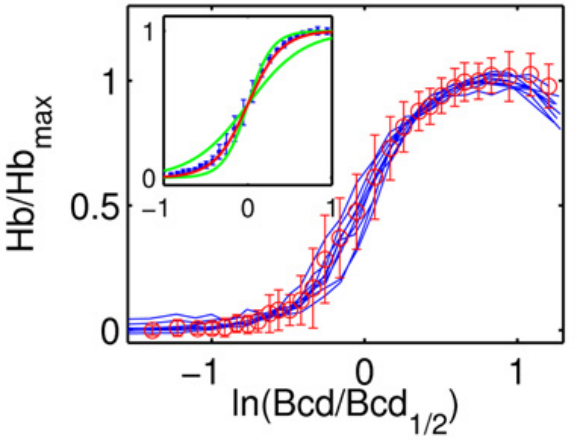
“position”

sharpness in Drosophila embryo patterning

Hunchback is sharply expressed in response to maternal Bicoid



Hb expression



Hill function

$$\frac{[\mathrm{Hb}]}{[\mathrm{Hb}]_{max}} \approx \frac{x^5}{1+x^5} \qquad x = [\mathrm{Bcd}]/[\mathrm{Bcd}]_{0.5}$$

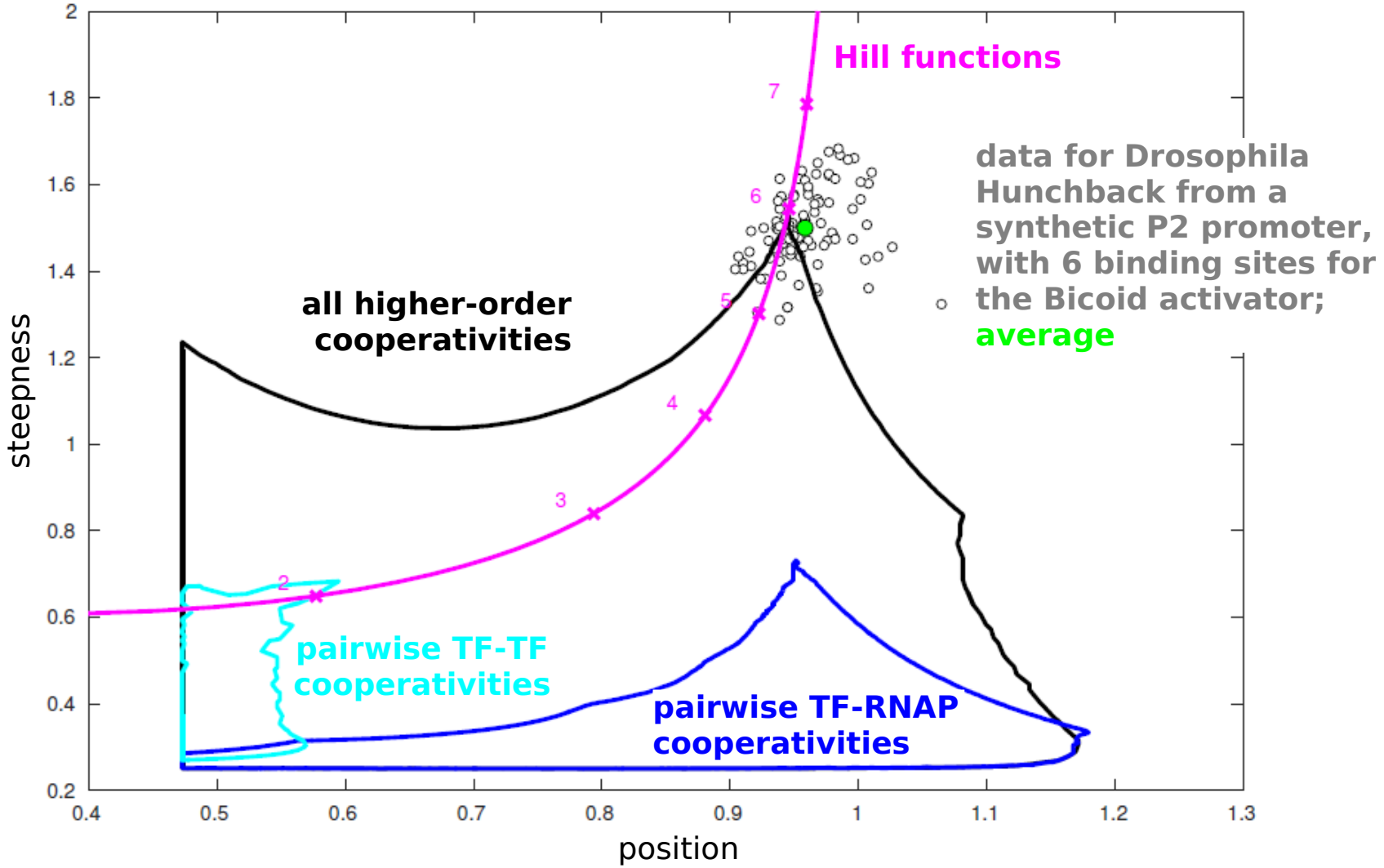
pairwise cooperativity

"consistent with the idea that Hb transcription is activated by cooperative binding of effectively five Bcd molecules"

Gregor, Tank, Wieschaus, Bialek, *"Probing the limits to positional information"*, Cell **130**:153-64 2007

higher-order cooperativity is essential

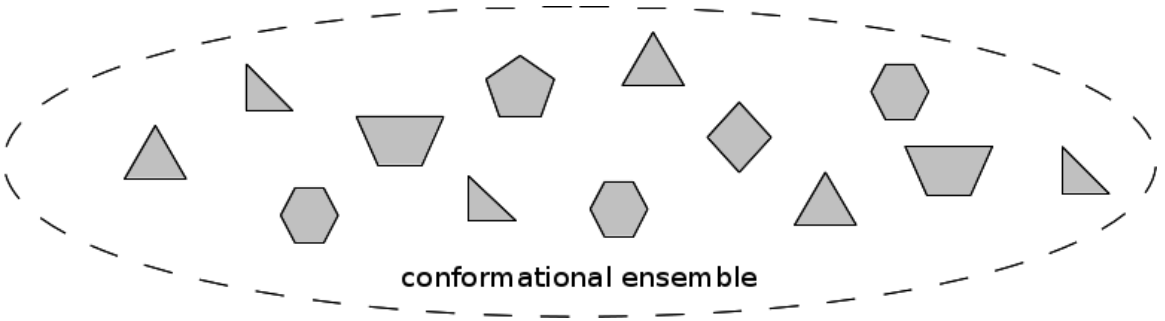
transcriptional activator on 6 sites,
$\log_{10}(\omega_{*,(S,*)})$ sampled uniformly in [-3,3]



updated from Estrada, Wong, DePace, Gunawardena, Cell **166**:234-44 2016.

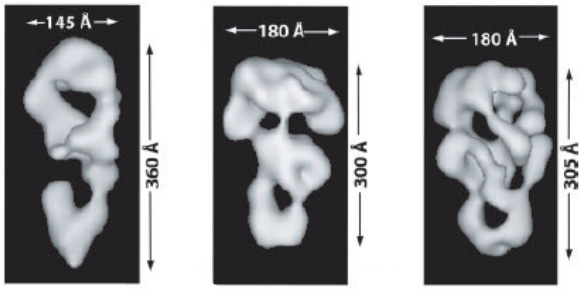
how does higher-order cooperativity arise?

conformational ensembles can yield arbitrary higher-order cooperativities



provided the ensemble is sufficiently complex

which seems to happen in gene regulation



Mediator conformations from cryoEM

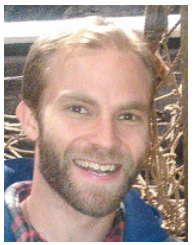
Biddle, Martinez-Corral, Wong, Gunawardena,"*Allosteric conformational ensembles integrate information through higher-order cooperativity*", in preparation, 2018; Taatjes et al, Science **295**:1058-62 2002

those who did the work

Angela
DePace

Javier
Estrada

Jeehae
Park



John
Biddle

Rosa
Martinez-Corral

Felix
Wong

Kate
Shulgina

summing up and some questions

- systems biology – how do we get from dead molecules to living organisms?
- are there levels of representation for cells and organisms, analogous to those found in neuroscience?
- what is a minimal mathematical model for homeostasis that accounts for known physiology and individual variation?
- scalable integration through weak linkage can reconcile population genetics, developmental biology and the evolution of complexity
- does learning by weak linkage take place within the organism as it develops, thereby allowing for “developmental selection”?
- steady-state input-output response characteristics of weak linkage mechanisms can be mathematically analysed using the linear framework
- how do we solve the problem of path dependence away from equilibrium?
- what are the hopfield barriers for different information processing tasks?
- how do we analyse time-varying input-output responses?