six lectures on systems biology

jeremy gunawardena
department of systems biology
harvard medical school

lecture 3
5 april 2011

part 2 seminar room, department of genetics

a rather provisional syllabus

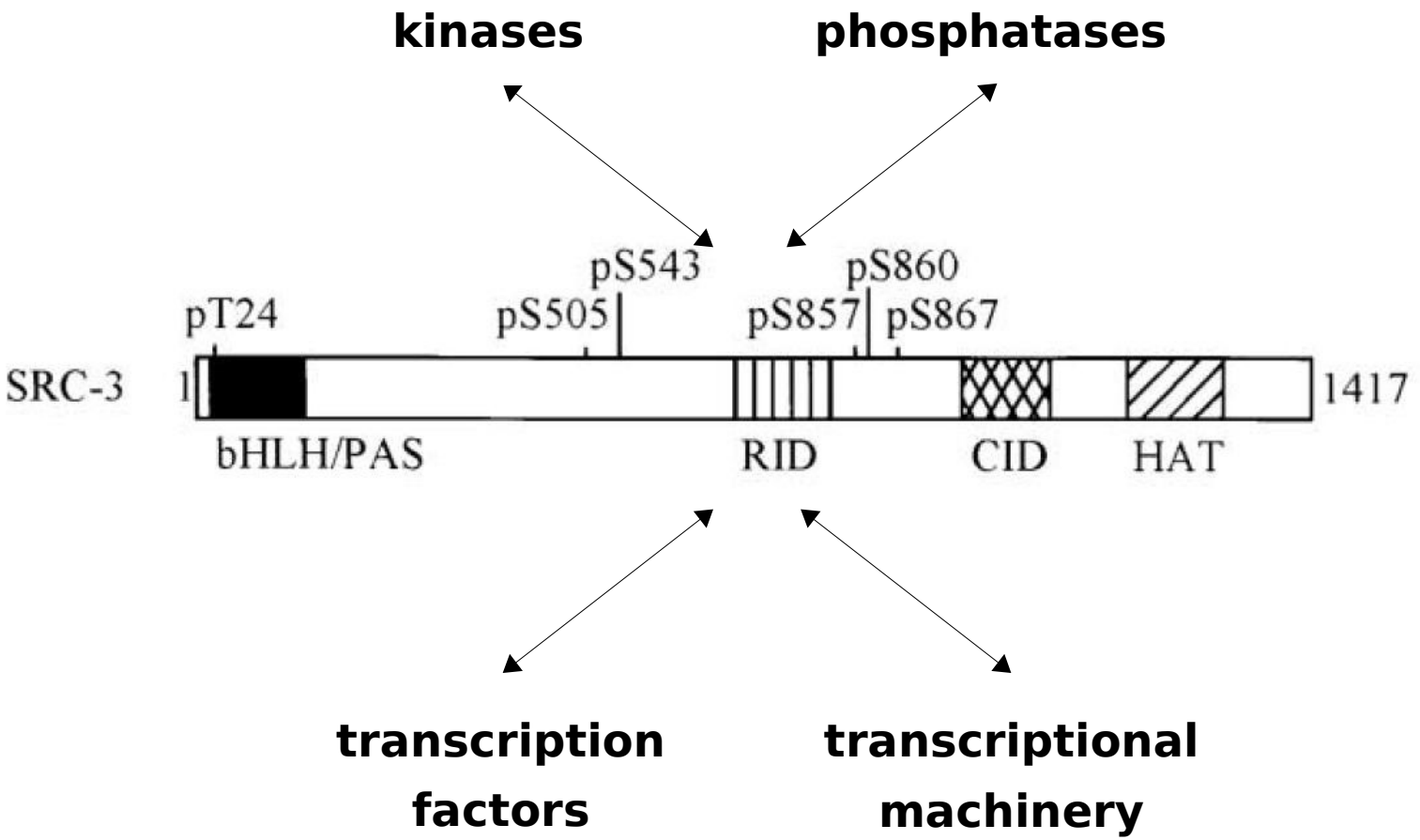
0. why mathematical models?
1. post-translational modification of proteins
2. microscopic cybernetics
3. development and evolution

1. post-translational modification (PTM)

how can we measure mod-form distributions?

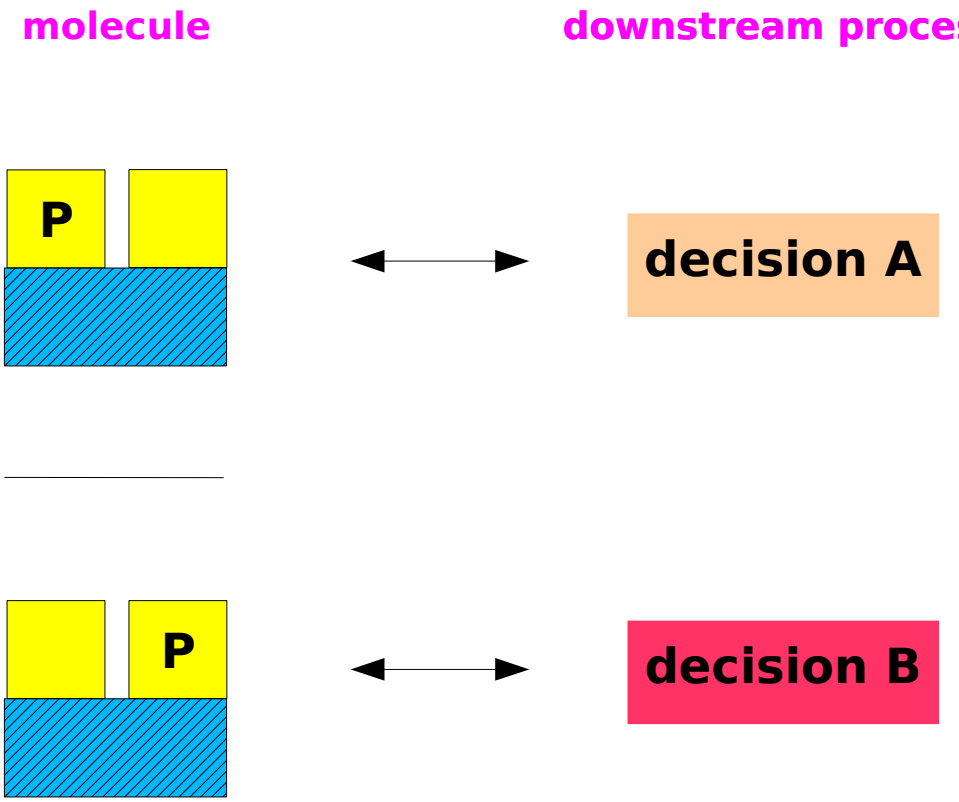
how do PTM networks regulate the distributions?

information processing by PTMs



the cartoon fallacy about PTMs

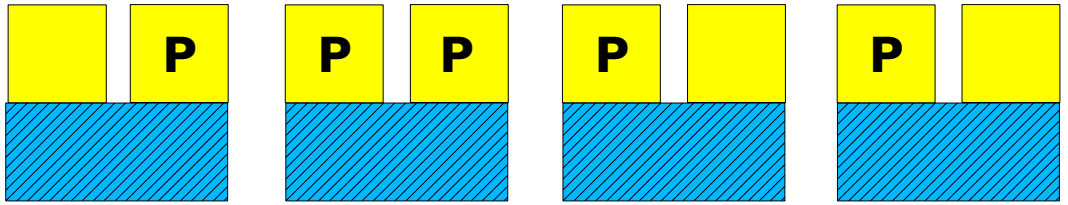
“a protein with more sites of modification carries more information”



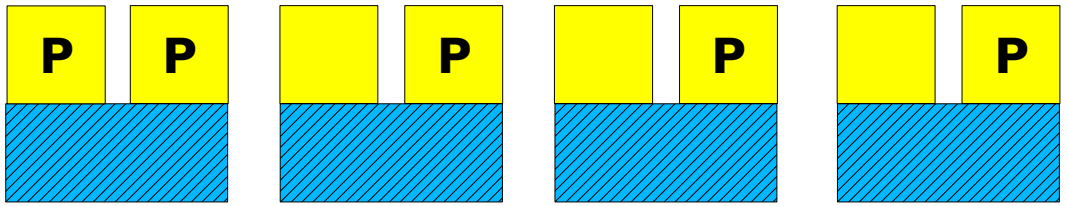
but under the hood

distribution

downstream process



decision A

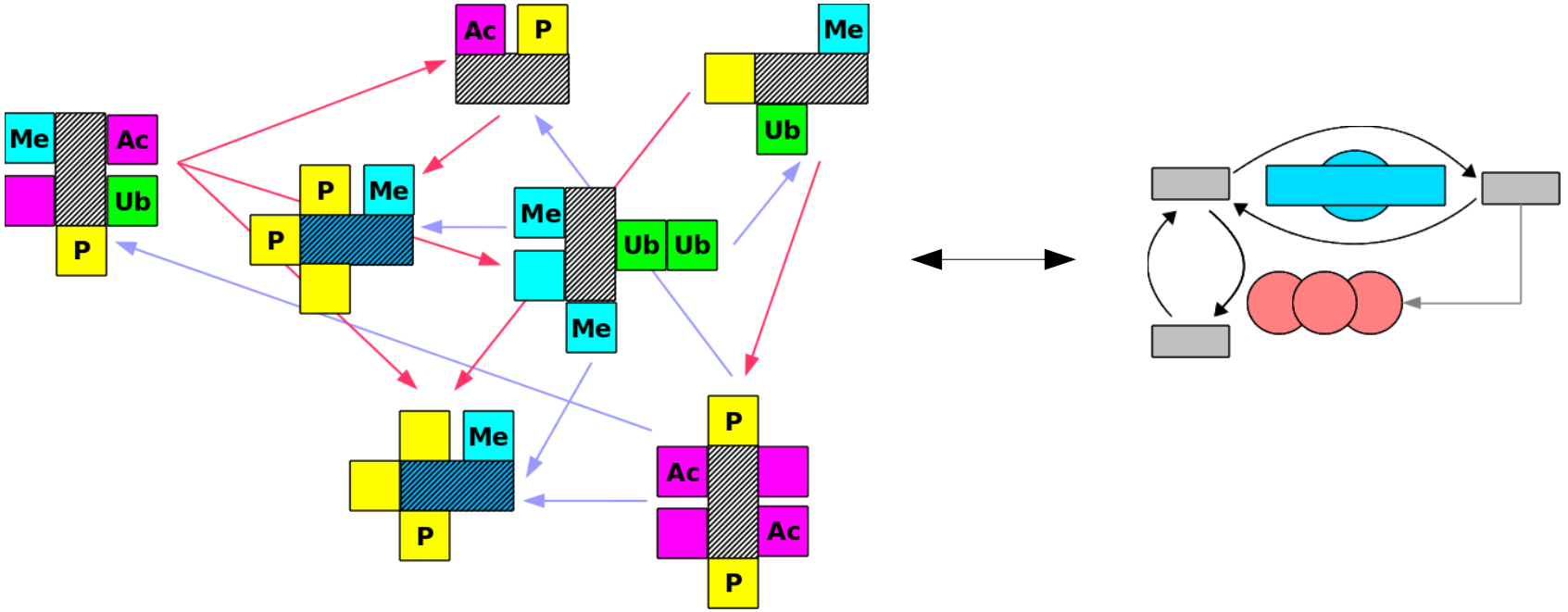


decision B

separation of time scales

network of enzymes and substrates

downstream process



fast sub-system

slow sub-system

the fast sub-system comes to steady state quickly and the slow sub-system “reads” that steady state

polynomial dynamical systems

reaction

monomial

$2A + 3B \xrightarrow{k} C$ mass action $\frac{dC}{dt} = kA^2B^3$

polynomials

reaction network $\frac{dx}{dt} = f(x;k)$

$x = (x_1, \cdots, x_n)$ species concentrations

$k = (k_1, \cdots, k_p)$ rate constants (parameters)

seeing the wood for the trees

simulation requires the numerical values of all parameters to be known beforehand

for PTM systems, the numbers of sites of modification need to be known and the combinatorial complexity increases exponentially ...

however, at steady state, $\frac{dx}{dt} = 0$

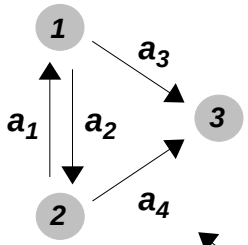
solutions of a system of polynomial algebraic equations $f(x;k) = 0$

and this can be analysed with the parameters treated as symbols

Manrai & Gunawardena, *"The geometry of multisite phosphorylation"*, Biophys J **95**:5543-33 2008

linear chemistry

directed graph



mass action

Laplacian matrix

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

symbolic label

each edge is a unimolecular reaction, with the label as rate constant

Gunawardena, *"A linear elimination framework for nonlinear biochemical systems"*, submitted, 2011.

Gustav Kirchhoff, *"Uber die Auflosung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Strome gefuhrt wird"*, Ann Phys Chem, **72**:497-508 1847

with nonlinear applications

labelled directed graph → **Laplacian dynamics**

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

the labels in G may be complicated algebraic expressions

$$a_1 = \left(\frac{k_1 + k_2}{k_3}\right)(L_1)^2 L_2$$

provided no concentration – L_1 or L_2 – is that of a node in G

(although this constraint can be weakened ...)

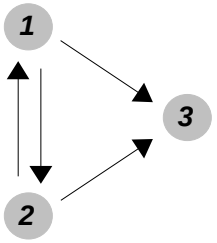
steady states in the linear framework

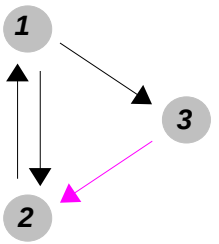
$$\mathcal{L}(G).x = 0 \qquad x \in \ker \mathcal{L}(G)$$

if G is **strongly connected** with positive labels, then there is a unique steady state, up to a constant factor

$$\ker \mathcal{L}(G) = \langle \rho \rangle$$

strongly connected – *any two distinct nodes are connected by a contiguous path of edges pointing in the same direction*





strongly connected

the matrix-tree theorem

$$\ker \mathcal{L}(G) = \langle \rho \rangle$$

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

polynomial in the symbolic labels

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

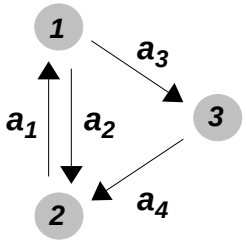
W T Tutte, *"The dissection of equilateral triangles into equilateral triangles"*, Proc Camb Phil Soc **44**:463-82 1948

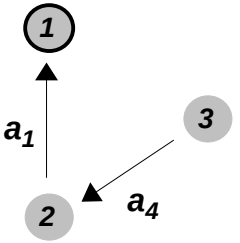
spanning trees

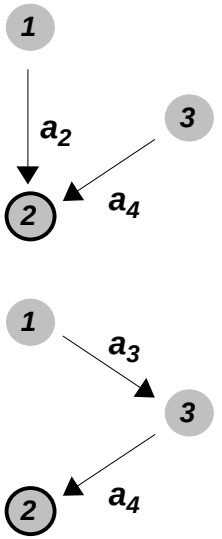
rooted spanning tree – a sub-graph T of G that

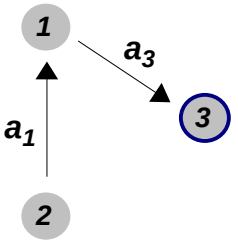
1. spans G – every node of G is also a node of T
2. is a TREE – T has no cycles, ignoring edge directions
3. is rooted at *i* – *i* is the only node of T with no outgoing edges

spanning tree formula









$a_1 a_4$ $\quad a_2 a_4 + a_3 a_4$ $\quad a_1 a_3$

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

Laplacian $\qquad$ **ρ**

elimination of internal complexity

if there is a steady state $\frac{dx}{dt} = 0 \qquad \mathcal{L}(G).x = 0$

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

then each x_i can be eliminated

reference node

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

rational expression in the symbolic labels

(non-strongly connected graphs)

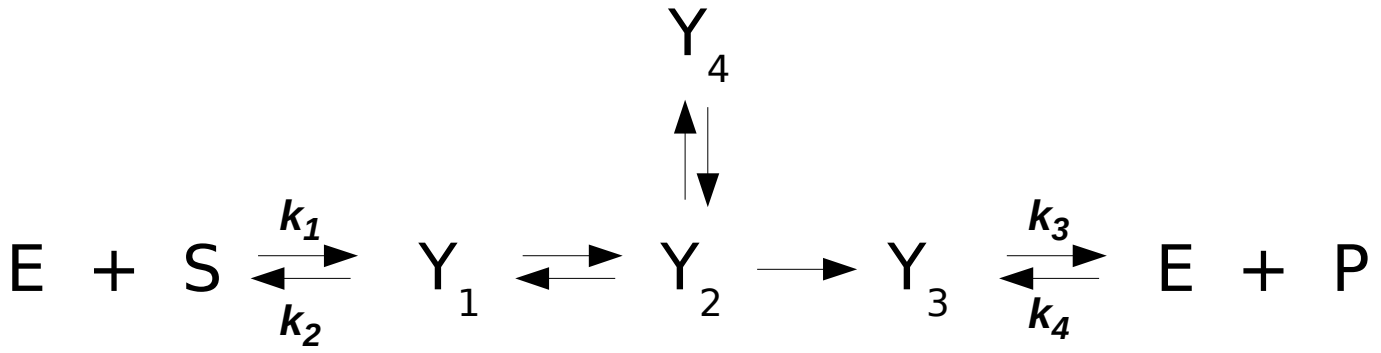
if G is not strongly connected, there may be multiple independent steady states

they can be completely described but this is not needed here

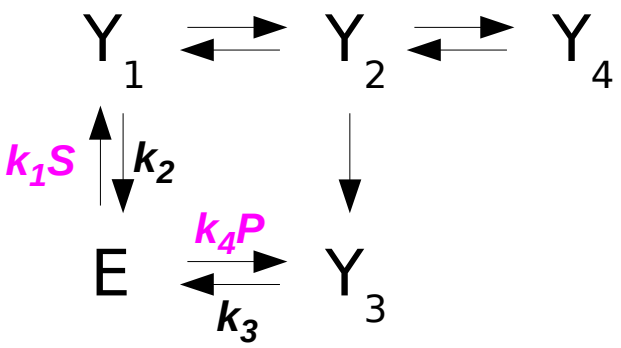
see below for more details

Gunawardena, *"A linear elimination framework for nonlinear biochemical systems"*, submitted, 2011

applications – enzyme kinetics



labelled directed graph



MTT

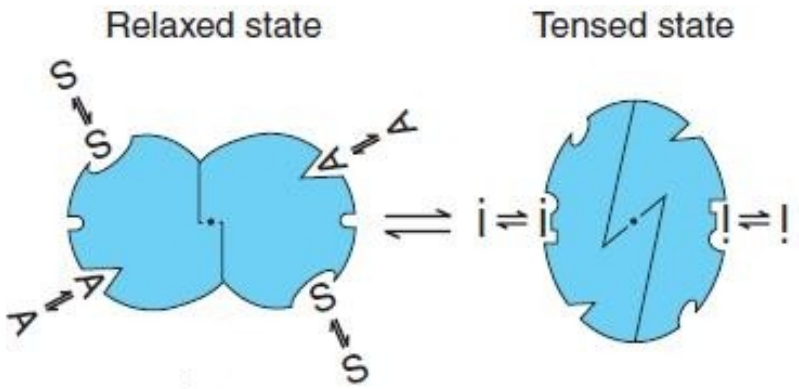
$$Y_3 = \left(\frac{\rho_{Y_3}}{\rho_E + \rho_{Y_1} + \rho_{Y_2} + \rho_{Y_3}} \right) E_{tot}$$

$$\frac{dP}{dt} = k_3 Y_3$$

King & Altman, *"A schematic method of deriving the rate laws for enzyme-catalyzed reactions"*, J Phys Chem **60**:1375-8 1956

applications – protein allostery

Monod-Wyman-Changeux formula



fractional saturation

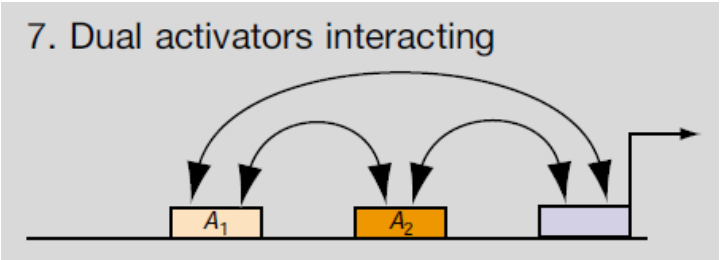
$$Y = \frac{(1+x)^{n-1}x + cL'(1+cx)^{n-1}x}{(1+x)^n + L'(1+cx)^n}$$

$$\frac{dP}{dt} = V_{max}Y$$

Monod, Wyman, Changeux, *"On the nature of allosteric transitions: a plausible model"*, J Mol Biol **12**:88-118 1965

Changeux & Edelstein, *"Allosteric mechanisms of signal transduction"*, Science **308**:1424-8 2005

applications – transcriptional regulation



rate of gene expression

$$\frac{1+\frac{[A_1]}{K_{A_1}}f_1+\frac{[A_2]}{K_{A_2}}f_2+\frac{[A_1]}{K_{A_1}}\frac{[A_2]}{K_{A_2}}f_1f_2\omega}{1+\frac{[A_1]}{K_{A_1}}+\frac{[A_2]}{K_{A_2}}+\frac{[A_1]}{K_{A_1}}\frac{[A_2]}{K_{A_2}}\omega}$$

Bintu, Buchler, Garcia, Gerland, Hwa, Kondev, Phillips, *"Transcriptional regulation by the numbers"* Curr Opin Gen Dev **15**:116-24 2005 – see also **15**:125-35 2005

Ackers, Johnson, Shea, *"Quantitative model for gene regulation in lambda phage repressor"*, PNAS **79**:1129-33 1982

applications – post-translational modification

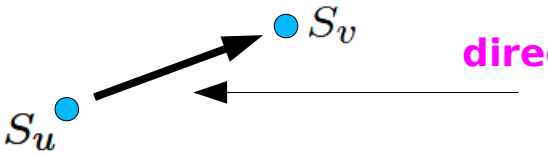
substrate with multiple types of modifications on multiple sites

$$S_1, \cdots, S_N$$ **modforms**

multiple forward and reverse enzymes

$$E_1, \cdots, E_L$$

directed “system graph” on the modforms

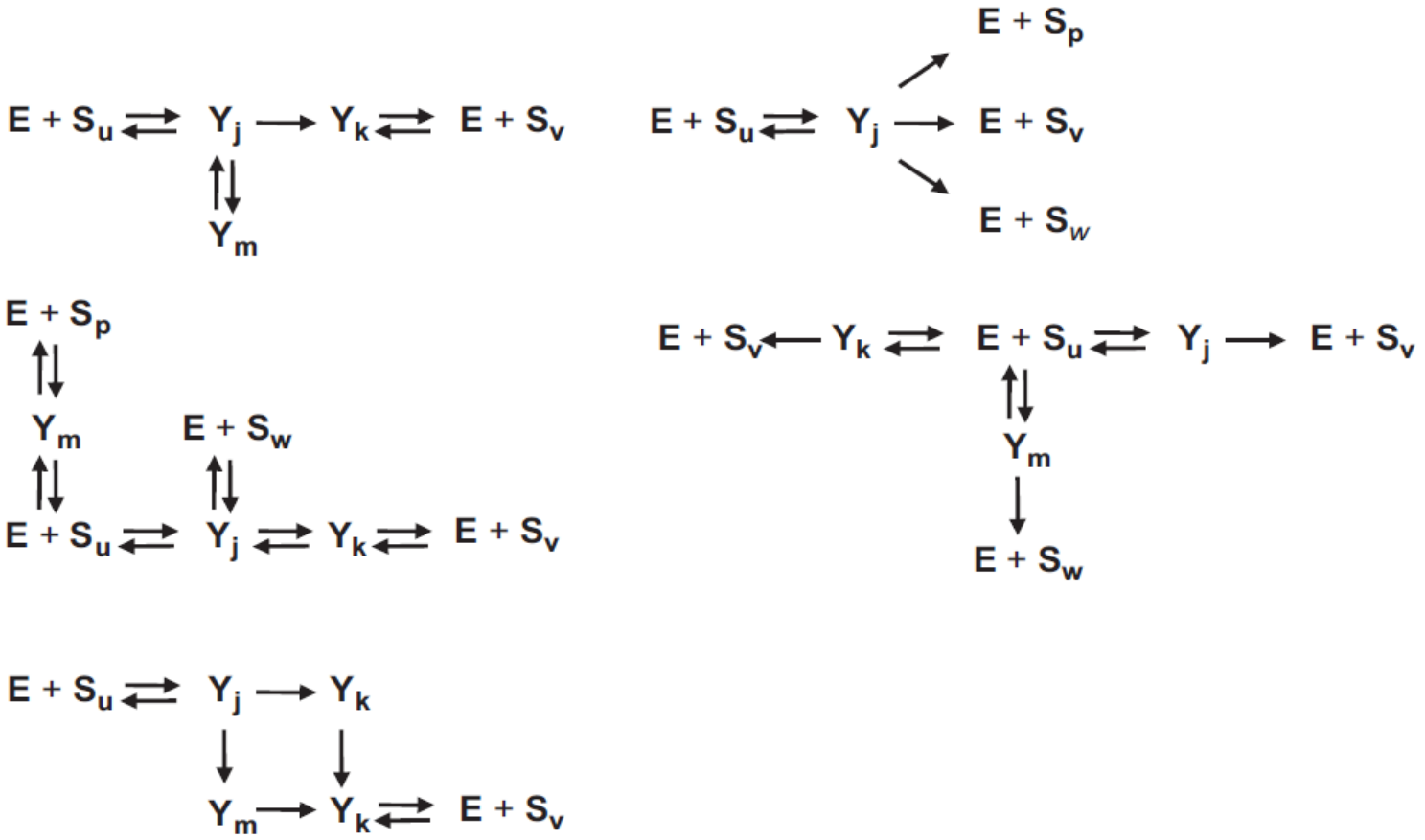


directed edge if some enzyme catalyses the change

naturally strongly connected

Thomson, Gunawardena, *“The rational parameterisation theorem for multisite post-translational modification systems”,* J Theor Biol **261**:626-36 2009

biochemistry of modification



donor “charge” maintained constant by external mechanisms

no ubiquitin-like modifications

hierarchical elimination

eliminate the intermediate complexes in favour of substrates and enzymes

this gives a label for the system graph

provided no enzyme is also a substrate, the substrates can then be eliminated in favour of the enzymes

$$S_u = r_u(E_1, \cdots, E_L)S_0$$

rational expression

$$\begin{array}{ccc} E_{1,tot} & = & \Phi_1(E_1, \cdots, E_L) \\ \vdots & \vdots & \vdots \\ E_{L,tot} & = & \Phi_L(E_1, \ldots, E_L) \end{array}$$

complete analytical description, with an exponential reduction in the number of independent variables

information capacity of PTM systems

the modform distribution of a single substrate with n sites has a maximal information storage capacity of at least

$$\log_2 \left[\frac{n+2}{2} \right] \text{ bits}$$

Thomson, Gunawardena *"Unlimited multistability in multisite phosphorylation systems"*, Nature **460**:274-7 2009

summing up

1. separation of time scales allows internal complexity to be eliminated

2. this can be done using a linear, graph-theoretic elimination framework

3. which underlies the analysis of enzyme kinetics, gene regulation, protein allostery and post-translational modification

4. PTM systems can be completely analysed at steady state, symbolically in their parameters, with an exponential reduction in algebraic complexity

5. information can be carried in PTM distributions and the information capacity increases logarithmically with the number of sites