

Optimality constraints on gene expression patterns



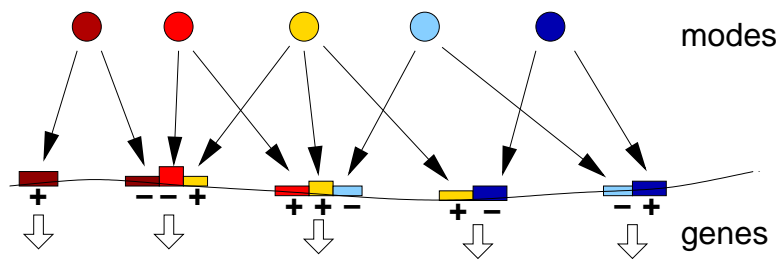
Wolfram Liebermeister
MPI für molekulare Genetik
Theoretische Biophysik, HU Berlin



- Simple mathematical model of steady-state regulation
- Teleological assumption \rightarrow optimal linear response to perturbation
- Relation between regulation and control coefficients (function)

Linear models of gene expression represent correlations

Unobserved (“latent”) variables exert a linear influence on genes



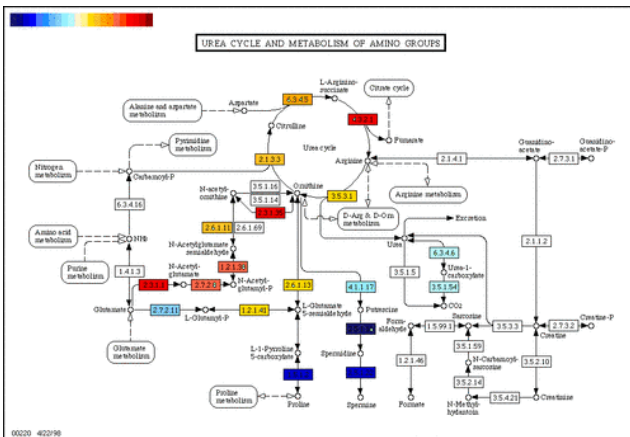
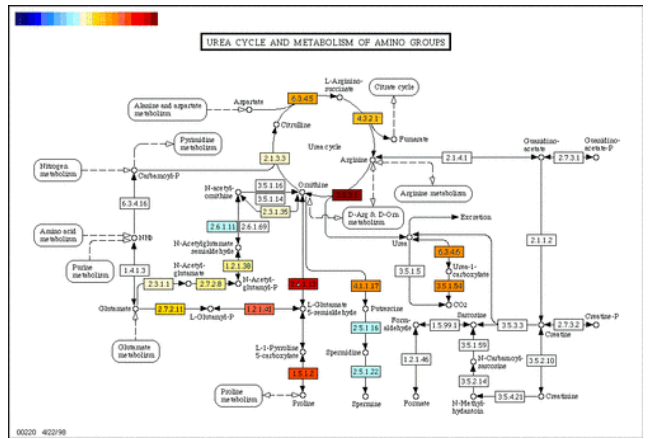
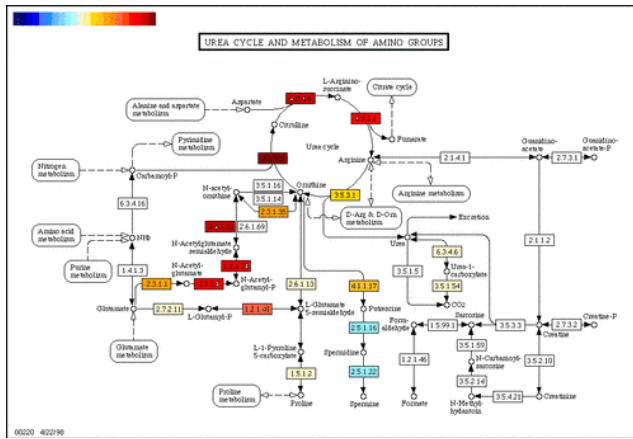
Identifying the parameters from data requires additional assumptions

- factor analysis (different rotation criteria...)
- independent components (minimal mutual information,...)

Interpret modes as

- regulators (signalling pathways, transcription factors etc.) ?
- biological functions (common metabolic pathways, annotations) ?

Example: urea cycle in yeast



Regulation of urea cycle, 3 linear modes
(factor analysis with varimax criterion)

Data from:
Hughes et al., Cell 102 (2000), 109

Optimal regulation, related to function?

Often, regulation seems to make sense, e.g.

- cooperating genes are coregulated (e.g. metabolic pathways, components of complexes)
- a perturbation activates genes that “buffer” the perturbation (e.g. starvation, shock)
- genes that are not needed are downregulated (e.g. during cell cycle)
- important variables feed back to their regulators

Empirical view (*causa efficiens*)

Explain regulation by signalling cascades, feedback from metabolism etc.

Teleological view (*causa finalis*)

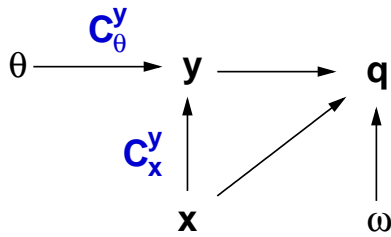
The cell aims at achieving the most with the least effort.

A gene's regulation depends on its ability to change relevant variables.

two sides of a coin...?

The regulatory **machinery** may be optimized when reflecting the functions of genes (e.g. shock response elements, nutrient sensing,..)

The teleological model



Model quantities

- x : controlling variables to be chosen (gene expression,..)
- $y(x, \theta)$: controlled quantities (metabolites, fluxes..)
- $q(x, y, \omega)$: fitness function (reproduction rate,..)
- θ, ω : perturbation parameters (nutrients,..)

Regulatory system \mathcal{X} controls "relevant" system \mathcal{Y} .

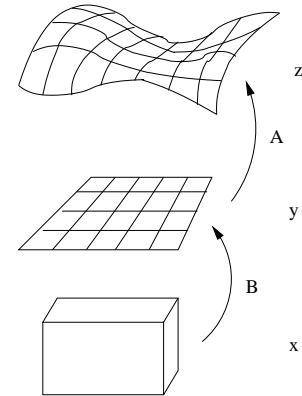
\mathcal{X} behaves such as to maximize a fitness function $q(x, y(x, \theta), \omega)$.

Second-order expansion

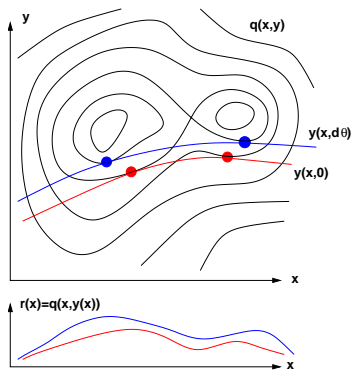
- Control coefficients: $C_x^y = \frac{\partial y}{\partial x}$, $C_\theta^y = \frac{\partial y}{\partial \theta}$ (also $C_{xx}^y, C_{x\theta}^y, C_{\theta\theta}^y$)
- "Prizes": $q_x = \frac{\partial q}{\partial x}$, $q_y = \frac{\partial q}{\partial y}$.
- "Prize stiffness": $q_{xx} = \frac{\partial^2 q}{\partial x^2}$, $q_{yy} = \frac{\partial^2 q}{\partial y^2}$

Assumptions

- q_{xx} and q_{yy} are invertible (no constant prizes)
- C_x^y has full row rank (the y_i can be regulated independently)



Optimal response to perturbations



- Define effective fitness $r(x, \theta, \omega) = q(x, y(x, \theta), \omega)$
- **Initially**, the system is in a locally optimal state where $r_x = 0$
 r_{xx} has negative eigenvalues
- Small perturbation (of y , x , q_y , ...)
- Find response $d\bar{x}$ to reach a **new** optimal state
- Condition: $r_x = 0$ before and after perturbation

Interpretation in the case of expression data

- Log expression values are considered as regulatory system \mathcal{X}
- Assume small experimental perturbations of the cells

Optimal response

Second-order expansion of $r(x, \theta, \omega)$ yields

$$\begin{pmatrix} r_{xx} & r_{x\theta} & r_{x\omega} \\ r_{\theta x} & r_{\theta\theta} & r_{\theta\omega} \\ r_{\omega x} & r_{\omega\theta} & r_{\omega\omega} \end{pmatrix} = \begin{pmatrix} q_{xx} + C_x^{yT} q_{yy} C_x^y + D_{xx} & C_x^{yT} q_{yy} C_\theta^y + D_{x\theta} & C_x^{yT} q_{y\omega} \\ C_\theta^{yT} q_{yy} C_x^y + D_{\theta x} & C_\theta^{yT} q_{yy} C_\theta^y + D_{\theta\theta} & q_{\omega y} C_\theta^y \\ q_{\omega y} C_x^y & C_\theta^{yT} q_{y\omega} & q_{\omega\omega} \end{pmatrix}$$

where $D_{xx} = (q_y C_{xx}^y)$, $D_{x\theta} = (q_y C_{x\theta}^y)$, and $D_{\theta\theta} = (q_y C_{\theta\theta}^y)$

The total differential of r_x reads

$$dr_x = [q_{xx} + C_x^{yT} q_{yy} C_x^y + D_{xx}] dx + [C_x^{yT} q_{yy} C_\theta^y + D_{x\theta}] d\theta + [C_x^{yT} q_{y\omega}] d\omega$$

An optimal initial state with $r_x(x, \theta, \omega) = 0$ becomes perturbed by $d\hat{\theta}$ and $d\hat{\omega}$.

The optimal reaction $d\bar{x}$ must assure

$$dr_x = r_x(x + d\bar{x}, \theta + d\hat{\theta}, \omega + d\hat{\omega}) - r_x(x, \theta, \omega) = 0$$

Therefore

$$d\bar{x} = - [q_{xx} + C_x^{yT} q_{yy} C_x^y + D_{xx}]^{-1} [(C_x^{yT} q_{yy} C_\theta^y + D_{x\theta}) d\hat{\theta} + C_x^{yT} q_{y\omega} d\hat{\omega}]$$

if the inverse exists. Using $d\hat{q}_y = q_{yy} C_\theta^y d\hat{\theta} + q_{y\omega} d\hat{\omega}$, we get

$$d\bar{x} = - [q_{xx} + D_{xx} + C_x^{yT} q_{yy} C_x^y]^{-1} [C_x^{yT} d\hat{q}_y + D_{x\theta} d\hat{\theta}]$$

Different kinds of perturbations

1. Achieving a fixed change $dy = C_x^y d\bar{x}$

$$d\bar{x} = q_{xx}^{-1} C_x^{yT} (C_x^y q_{xx}^{-1} C_x^{yT})^{-1} dy$$

2. Single value x_i perturbed

One component x_i becomes constrained to a fixed value $x_i + d\hat{x}_i$

$$d\bar{x} = \frac{1}{(r_{xx}^{-1})_{ii}} r_{xx}^{-1} d\hat{x}_i$$

3. Perturbations $d\hat{\theta}$ of y or $d\hat{\omega}$ of q

$$d\bar{x} = -r_{xx}^{-1} C_x^{yT} d\hat{q}_y = -q_{xx}^{-1} C_x^{yT} d\bar{q}_y$$

if $y(x, \theta)$ is expanded to first order.

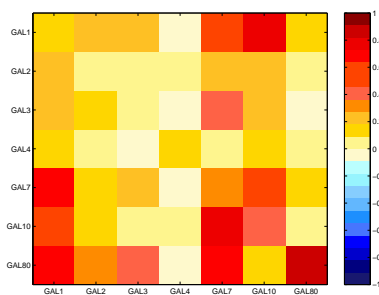
with "virtual" prize change $d\hat{q}_y = q_{yy} C_\theta^y d\hat{\theta} + q_{y\omega} d\hat{\omega}$

Reciprocal response in knock-out experiments

- Knock-out experiment \rightarrow expression data matrix M
(rows: genes knocked out, columns: same genes, measured)
- log mRNA concentrations are considered as regulators x_i
- Model prediction: $M = D r_{xx}^{-1}$ where D is diagonal and r_{xx}^{-1} is symmetric.
- If perturbing gene i affects the expression of gene j , the opposite should also hold.

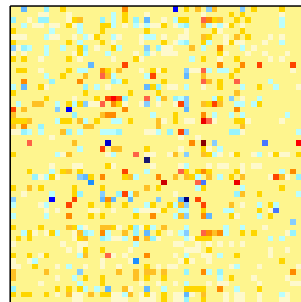
Experimental data: estimated r_{xx}^{-1}

Quantify symmetry of matrix A by symmetry ratio $s = \frac{\sum (A_{ik} + A_{ki})^2}{\sum (A_{ik} - A_{ki})^2}$.



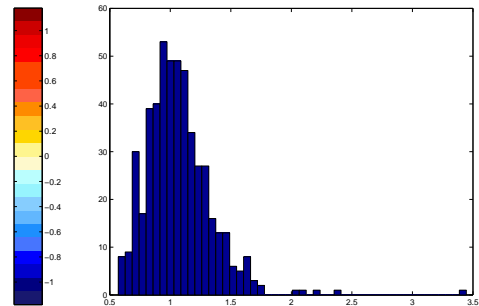
Ideker et al. [?]

knock-outs in galactose pathway



Hughes et al. [?]: knock-outs,

only metabolic genes considered



Hughes data: $s = 1.85$ is significant at $\alpha = 0.01$,
estimated using data with permuted rows

Metabolic control theory

Theorems of metabolic control theory → constraints on control coefficients C_x^y

Reaction system is characterized by

- **Stoichiometric matrix** N : The kernel matrix K of steady state fluxes fulfills $NK = 0$
- **Elasticities** ϵL : linear influences of independent metabolites on isolated reactions
- **Control coefficients** C_{ik}^J, C_{ik}^S linear influence of **parameter change** in enzyme k on global steady state **flux** J_i or **concentration** S_i

Consider a perturbation that would change steady-state fluxes, concentrations, etc..

An additional regulation by enzymes leads to a **better** steady state.

Summation and connectivity theorems, in particular, $C^J \epsilon L = 0$ and $\epsilon C^S K = 0$ yield general properties of optimal regulation patterns $d\tilde{x} = -q_{xx}^{-1} C_x^{yT} d\tilde{q}_y$:

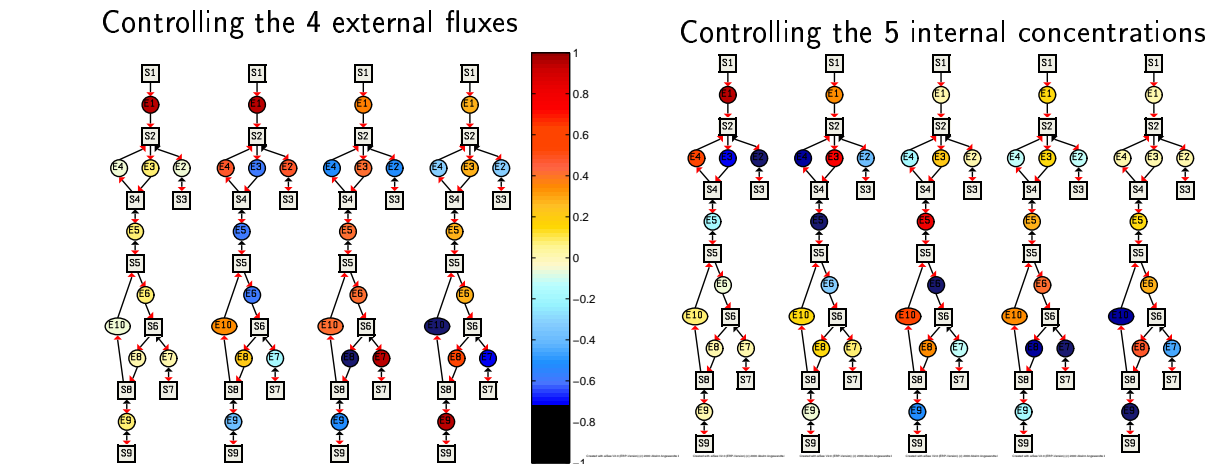
If q_{xx} is uniform, and if fitness depends only on

- **Fluxes**: $d\tilde{x}^T \epsilon L = 0$
→ Regulation profiles for n adjacent reactions (sharing a metabolite) are confined to a $(n - 1)$ -dimensional subspace.
- **Concentrations**: $d\tilde{x}^T K = 0$
→ sum of regulation values over any steady-state flux mode vanishes.

Example: a simple metabolic network

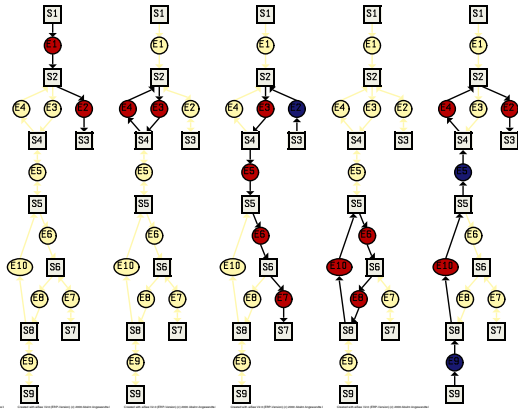
- 9 metabolites, (4 external), 10 reactions (4 external), 11 elementary flux modes (see [?])
- reactions are regulators x
- simple fitness structure $q_{xx} = -1$, $q_{yy} = -1$

Reaction to perturbation $d\theta < 0$

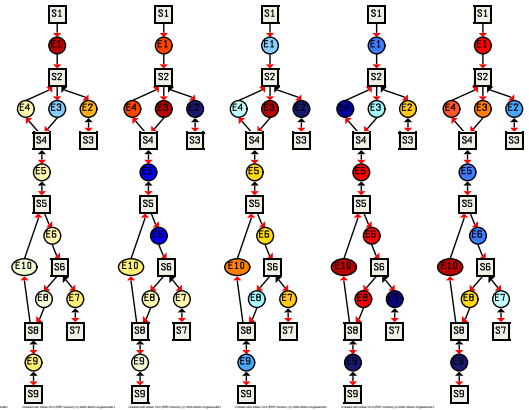


Controlling the coefficients of 5 independent elementary modes

Elementary flux modes



Reaction to perturbation $d\theta < 0$



Model predictions I

- Linear response to small perturbations (scalable and additive)
- no memory effects (fitness is a potential function)
- Perturbations are distributed by a cascade of reactions.

Expression patterns reflect the **control coefficients** on the relevant variables

The control coefficients are in general unknown, but one can derive **general results**

- Reciprocal behaviour for small perturbations in knock-out or RNAi experiments,
- Genes that have no effect on the relevant variables remain unchanged.
- Cooperating genes show correlated expression (e.g. protein clusters)
- The reaction contributes to homeostasis, variables and their prizes are buffered.
- If $q_{xx} \gg C_x^{yT} q_{yy}$, r_{xx}^{-1} consists of a cascade $\sum_{n=0}^{\infty} (-q_{xx}^{-1} C_x^{yT} q_{yy} C_x^y)^n q_{xx}^{-1}$ of effects
Due to compromises, perturbations become distributed
- Reaction of subsystems depends on the actual prize change of the variables affected by the subsystem
- The response does not depend on the (balanced) prizes, only on their derivatives
- A second-order expansion of y yields a correction of q_{xx} and the reaction to an additional perturbation dC_x^y

Model predictions II

Modules

Large disjoint modules of regulators that control a few variables each:

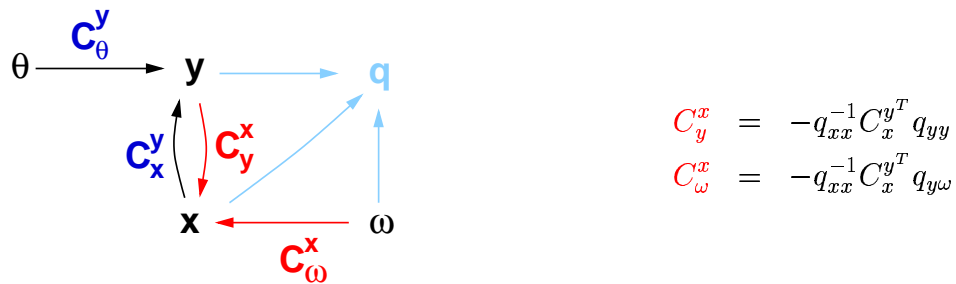
- Low-dimensional expression patterns (\approx factor analysis model)
- Sparse (ICA-like) model for whole-genome patterns

Predictions for metabolic systems (perturbations $d\theta, d\omega$)
assuming expression \propto enzymatic activity

- If the fitness depends only on fluxes, and elasticities ϵ represent only stoichiometric influences: correlated expression of neighbour enzymes
- If the fitness depends only on concentrations:
the expression profile, summed over any stationary flux, vanishes.
- If a module of m reactions controls $n < m$ independent fluxes:
its expression pattern should be confined to a n -dimensional subspace.
- If fitness depends on fluxes and flux is on the boundary of allowed cone
 \rightarrow reaction stays on the boundary

Optimal linear feedback

- The optimality postulate for perturbations $d\theta$, $d\omega$ can be implemented by linear feedbacks.



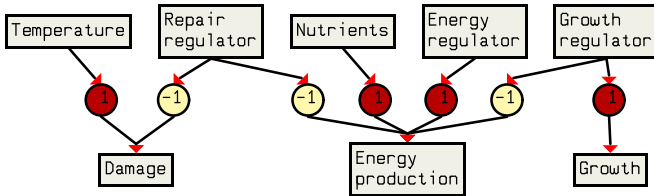
- The resulting reaction $dx = (1 - C_y^x C_x^y)^{-1} (C_y^x C_\theta^y d\theta + C_\omega^x d\omega)$ is optimal
- If q_{xx} is uniform, the system is also optimal for knock-out perturbations
- The feedback connections are related to the control coefficients, and therefore to the function of a regulator
- Nonlinear systems (signalling pathways etc.) may locally implement the linear response.

Optimal feedback example: balancing growth, repair, and energy

- variables y (1) growth, (2) energy production, (3) damage
- regulators x (1) growth control, (2) energy control, (3) repair
- perturbations θ (1) temperature, (2) nutrients
- perturbations ω (1) competition in the population

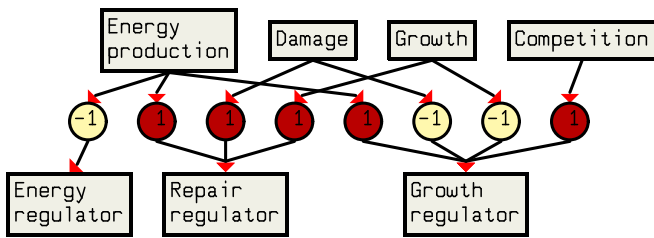
$$q_{xx} = - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Regulatory network C_x^y, C_θ^y



$$q_{yy} = - \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad q_{y\omega} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Optimal feedback network C_y^x, C_ω^x



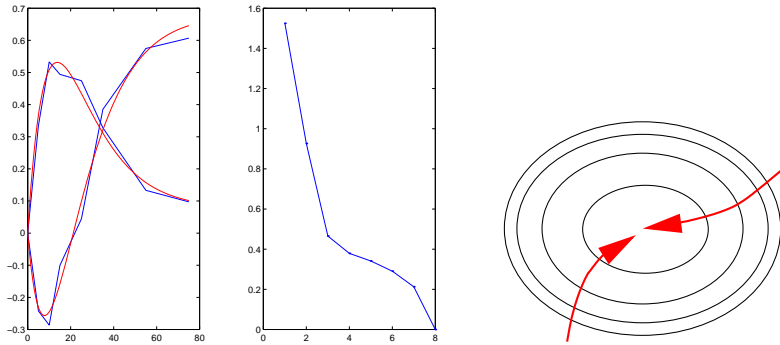
Not yet done...

Test predictions for metabolic systems

- need for control coefficients
- map expression values to reactions (problem: ORF \rightarrow EC \leftarrow reaction)

Time-dependent regulation

- Switching processes



Heat shock expression data, 2 principal components,
fitted with exponential relaxation

- Frequency-dependent regulation of stochastic perturbations
Systems on different timescales (kinetics, expression, evolution,...)
Basic idea: slow system adapts operating point for fast system

Acknowledgements

- Reinhart Heinrich
- Stefan Schuster
- Edda Klipp