Modelling of optimal differential gene expression

Wolfram Liebermeister and Edda Klipp

{lieberme|klipp}@molgen.mpg.de

SPONSORED BY THE Federal Ministry of Education and Research

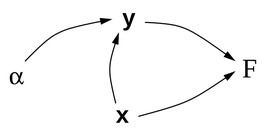
Abstract

An optimality-based model of regulatory systems (e.g. differential gene expression) is studied. The model predicts that the behaviour of regulators is related to their function, i.e. their influence on variables that are relevant for the organism's fitness. As a consequence, optimal gene expression profiles may portray the topology of the metabolic network. Optimal regulation can be realized by linear feedback signals, which, again, are related to gene function.

Introduction

During the last years, the large-scale structure of gene regulation has been studied intensely by measuring gene expression on the genomic scale. Cluster analyses and linear models of gene expression data reveal groups of coregulated genes sharing biological functions. Expression data have been used for annotating genes and reconstructing metabolic pathways, but in a purely heuristical way: a relation between expression and function can be derived from a principle of optimal regulation [5]. The proposed model predicts a relation between optimal differential expression (after external perturbations), function (quantified by response coefficients), and the regulatory mechanism, as it is found in operons.

The model



Model quantities

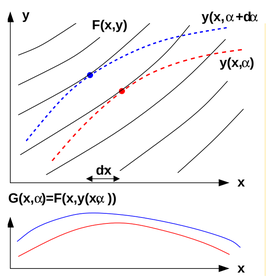
- x: regulatory variables (gene expression,..)
- $y(x, \alpha)$: "output" variables (metabolites, fluxes..)
- $F(x, y, \beta)$: fitness function (reproduction rate,..)
- α, β: perturbation parameters (nutrients,..)

The regulators x always adapt themselves such as to maximize a fitness function F.

Local description by derivatives

- Response coefficients: derivatives $R^y_x = \partial y/\partial x$, $R^y_\alpha = \partial y/\partial \alpha$
- "Marginal fitness": derivatives $F_x = \partial F/\partial x$, $F_y = \partial F/\partial y$.
- "Curvatures": second derivatives $F_{xx} = \partial^2 F/\partial x^2$, $F_{yy} = \partial^2 F/\partial y^2$

Optimal response to perturbations



- Define regulatory fitness $G(x, \alpha, \beta) = F(x, y(x, \alpha), \beta)$
- Initially, the system is in a locally optimal state where $G_x = F_x + R^{y^T}_x F_y = 0$
  $G_{xx} = F_{xx} + R^{y^T}_x F_{yy} R^y_x$ has negative eigenvalues
- Small perturbation (of y, x, F_y, ...)
- Find response $d\bar{x}$ to reach a new optimal state
- Condition: $G_x = 0$ before and after perturbation

Different kinds of perturbations

Achieving a fixed change $dy = R^y_x \, d\bar{x}$

$$d\bar{x} = F^{-1}_{xx} \, R^{y^T}_x \, (R^y_x \, F^{-1}_{xx} \, R^{y^T}_x)^{-1} \, dy$$

The scaled expression profile $F_{xx} \, d\bar{x}$ is a linear combination of regulatory profiles.

Single value x_i perturbed

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

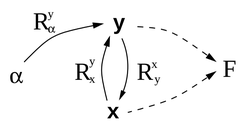
$$d\bar{x} = \frac{1}{(G^{-1}_{xx})_{ii}} \, G^{-1}_{xx} \, d\hat{x}$$

Perturbations $d\hat{\alpha}$ of y or $d\hat{\beta}$ of F

$$d\bar{x} = -G^{-1}_{xx} \, (R^{y^T}_x \, F_{yy} \, \mathrm{d}\hat{y} + \mathrm{d}\hat{R}^{y^T}_x F_y)$$

Superposed responses to a perturbation of variables and response coefficients.

Optimal linear feedback



The optimality postulate for perturbations $d\alpha$, $d\beta$ can be implemented by linear feedbacks.

$$R^x_y = -F^{-1}_{xx} R^{y^T}_x F_{yy}$$

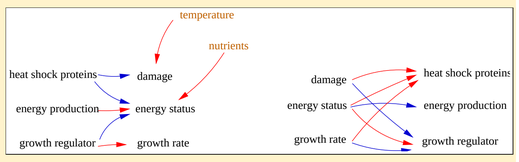
- The resulting reaction $dx = (1 - R^x_y \, R^y_x)^{-1} R^x_y \, R^y_\alpha \, d\alpha$ is optimal
- The feedback connections are related to the response coefficients, and therefore to the function of a regulator
- Nonlinear systems (signalling pathways etc.) may locally implement the linear response.

Example: balancing growth and repair

Regulatory network R^y_x, R^y_α and optimal feedback network R^x_y, R^x_β

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

$$F_{yy} = -\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$



Metabolic systems

According to the model, genes with similar impact on important cell variables show correlated expression. Thus, we can expect coexpression for protein complexes, functional modules, and functional gene classes.

A metabolic 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

The systemic response to perturbations is described by

- **Response coefficients** C^J_{ik}, C^S_{ik}
  Linear influence of **flux change** of reaction k on global steady state **flux** J_i or **concentration** S_i
- Theorems of metabolic control theory [2] → constraints on control coefficients C^S, C^J

Consider a perturbation that would change steady-state fluxes, concentrations, etc..
An additional regulation $\mathrm{d}\bar{E}$ by enzymes leads to a **better** steady state.

The **summation and connectivity theorems**, in particular,

$$C^J \epsilon L = 0 \text{ and } C^S K = 0$$

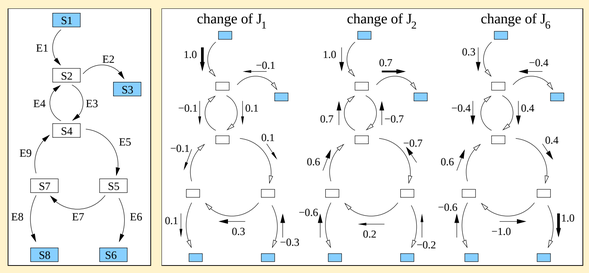
yield general properties of optimal regulation patterns $d\bar{E} = -F^{-1}_{EE} \, C^{Y^T} M$

If F_{EE} is diagonal, and fitness depends only on

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

Example: a simple metabolic network

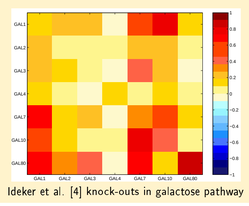
- 8 metabolites, (4 external), 9 reactions (4 external)
- Enzymes are regulators x, fluxes J_1, J_2, J_6 are relevant for the fitness
- Simple fitness curvatures $F_{xx} = -1$, $F_{yy} = -1$



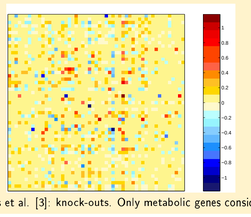
Reciprocal response in knock-out experiments

- Knock-out experiment → expression data matrix M
  (rows: genes knocked out, columns: same genes, measured, log-values)
- Model prediction: $M = D \, G^{-1}_{xx}$ where D is diagonal and G^{-1}_{xx} is symmetric.
  If perturbing gene i affects the expression of gene j, the opposite should also hold.
- The predicted compensation should also appear in phylogenetic gene profiles.

Experimental data: estimated G^{-1}_{xx}

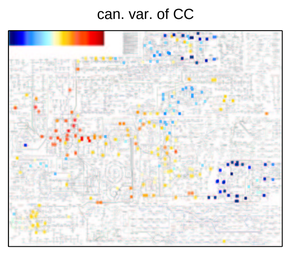


Ideker et al. [4] knock-outs in galactose pathway



Hughes et al. [3]: knock-outs. Only metabolic genes considered

Relating expression to control coefficients



Comparing simulated flux control coefficients to gene expression data (Gasch et al. [1]) by canonical analysis. The first components found (shown above) are significantly similar.

Model predictions

Expression patterns reflect the **response coefficients** on the relevant variables
The response coefficients are in general unknown, but one can derive

General results

- Reciprocal behaviour for small perturbations in deletion or RNAi experiments.
- Relation between differential expression and fitness loss after deletions.

Predictions for metabolic systems (perturbations $d\alpha, d\beta$)
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 set of m reactions controls $n < m$ independent fluxes:
  its expression pattern should be confined to a n-dimensional subspace.

Discussion

- The approach is **limited** to
  - Small perturbations
  - Physiological conditions (optimal behaviour is based on evolutionary "training set")
  - Homogeneous cell populations
- Time-dependent perturbations of a stationary state can be treated in the same manner.
- The use of sparse linear models for data analysis is justified.
- **Quantitative tests are difficult,** because
  - Relatively few response coefficients can be measured (but some properties are known)
  - Fitness function is not known
- In physics, extremal principles are an alternative way to state the laws of nature, and often equivalent to a causal formulation. In biology, the teleological approach (explaining facts as outcome of evolution) supplements the mechanistical explanation (by a "causa efficiens") with a view on the objectives ("causa finalis") of regulatory networks.

References

[1] A.P. Gasch et al. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*, 11:4241–4257, 2000.

[2] R. Heinrich and S. Schuster. *The regulation of cellular systems*. Chapman & Hall, 1996.

[3] T.R. Hughes et al. Functional discovery via a compendium of expression profiles. *Cell*, 102:109–126, 2000.

[4] T. Ideker et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292:929–934, 2001.

[5] Wolfram Liebermeister et al. A theory of optimal differential gene expression. *submitted to BioSystems*, 2003.