A theory of optimal differential gene expression

Wolfram Liebermeister lieberme@molgen.mpg.de

Max Planck Institute for Molecular Genetics, Ihnestr. 73, 14195 Berlin. www.molgen.mpg.de/∼ag_klipp

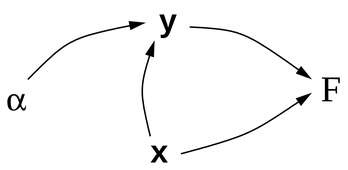
BCB - Berlin Center for Genome Based Bioinformatics

Kinetic Modeling Group

Introduction

The large-scale structure of gene regulation has been intensely studied by measuring gene expression on genomic scale. Cluster analyses and linear models of gene expression data have revealed groups of coregulated genes that often share biological functions. Conversely, expression data have been used for annotating genes and reconstructing metabolic pathways [2], but in a purely heuristical way. Here we derive a relation between expression and function 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. As a consequence, optimal gene expression profiles and regulatory networks may portray the topology of the metabolic network.

The model

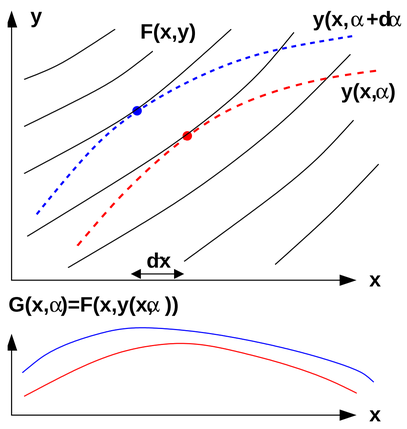


Model quantities

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

The regulators x always adapt themselves to the perturbation in order to maximize a fitness function F.

Optimal response to perturbations



Local description by derivatives:

- Response coefficients: $R^y_x = \partial y/\partial x$, $R^y_\alpha = \partial y/\partial \alpha$
- Fitness derivatives: $F_x = \partial F/\partial x$, $F_y = \partial F/\partial y$, $F_{xx} = \partial^2 F/\partial x^2$, $F_{yy} = \partial^2 F/\partial y^2$

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

Different kinds of perturbations

Achieving a fixed change $\mathrm{d}y = R^y_x\ \mathrm{d}\bar{x}$ $\quad\quad$ $\mathrm{d}\bar{x} = F^{-1}_{xx}\ (R^y_x)^{\mathrm{T}}\ (R^y_x\ F^{-1}_{xx}\ (R^y_x)^{\mathrm{T}})^{-1}\ \mathrm{d}y$

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

Single value x_i perturbed: A single component x_i becomes constrained to a fixed value $x_i + \mathrm{d}\hat{x}_i$ $\quad\quad$ $d\bar{x} = \frac{1}{(G^{-1}_{xx})_{ii}}\ G^{-1}_{xx}\ \mathrm{d}\hat{x}$

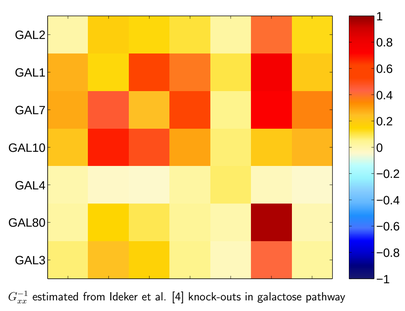
Perturbations $\mathrm{d}\hat{\alpha}$ of y $\quad\quad$ $\mathrm{d}\bar{x} = -G^{-1}_{xx}\ ((R^y_x)^{\mathrm{T}}\ F_{yy}\ \mathrm{d}\hat{y} + (\mathrm{d}\hat{R}^y_x)^{\mathrm{T}} F_y)$

Superposed response to the perturbation of variables and response coefficients.

Reciprocal response in knock-out experiments

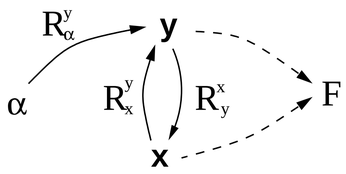
- Knock-out experiment → expression data matrix M (columns: genes knocked out, rows: the same genes, measured, log-values)
- Model prediction: $M = D\ G^{-1}_{xx}$ where D is diagonal and G^{-1}_{xx} is symmetric.

  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.



G^{-1}_{xx} estimated from Ideker et al. [4] knock-outs in galactose pathway

Optimal linear feedback



The optimality postulate for perturbations $\mathrm{d}\alpha$ can be implemented by a linear feedback.

$$R^x_y = -F^{-1}_{xx} (R^y_x)^{\mathrm{T}} F_{yy}$$

- The resulting reaction $\mathrm{d}x = (1 - R^x_y\ R^y_x)^{-1} R^x_y\ R^y_\alpha\ \mathrm{d}\alpha$ is optimal
- The feedback connections are related to the response coefficients (≈ functions of a regulators)
- Nonlinear systems (signalling pathways etc.) may locally implement the linear response.

Metabolic systems

A metabolic system is characterized by

- Stoichiometric matrix N and the kernel matrix K of stationary fluxes, fulfilling $NK = 0$
- Elasticities ϵL: linear influences of independent metabolites on isolated reactions

Metabolic response coefficients describe the (linearised) systemic response to parameter perturbations.

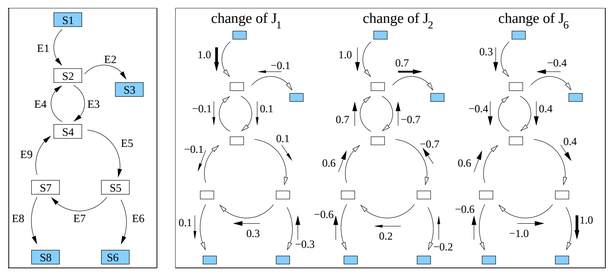
- The control coefficients C^J_{ik}, C^S_{ik} describe the linear influence of flux change of reaction k on global stationary flux J_i or concentration S_i
- Theorems of metabolic control theory [3] → constraints on control coefficients C^S, C^J

The summation and connectivity theorems $C^J \epsilon L = 0$ and $C^S K = 0$ yield general properties of optimal regulation patterns $\mathrm{d}\bar{x} = -F^{-1}_{xx}\ (\epsilon^{\mathrm{p}})^{\mathrm{T}} (C^Y)^T M$

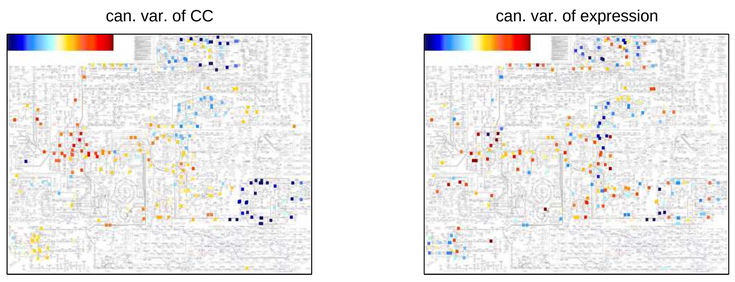
Assume that fitness curvature matrix F_{xx} and parameter-elasticities ϵ^{p} are diagonal:

- If fitness depends only on Fluxes: Theorem $\mathrm{d}\tilde{x}^T \epsilon L = 0$ → Regulation profiles for n adjacent reactions (sharing a metabolite) are confined to a $(n-1)$-dimensional subspace.
- If fitness depends only on Concentrations: Theorem $\mathrm{d}\tilde{x}^T K = 0$ → regulation values over any stationary flux mode sum to zero.

Example: a simple metabolic network



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

General predictions

- Expression patterns reflect the metabolic response coefficients
- Reciprocal behaviour for small perturbations in deletion or RNAi experiments.
- Relation between differential expression and fitness loss after deletions.

Predictions for metabolic systems assuming expression ∝ enzymatic activity

- If the fitness depends only on fluxes, and elasticities ϵ represent only stoichiometry: 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 "training conditions" during evolution)
- The approach is general:
  Only requirement: Metabolic response coefficients must be defined
- Time-dependent perturbations of a stationary state can be treated analogously.
- Quantitative tests are difficult, because
  - Relatively few response coefficients can be measured (but some properties are known)
  - Fitness function is not known
- Results suggest to use sparse linear models (e.g., ICA) for microarray data analysis.

Acknowledgement: This work was funded by the European commission, grant No. 503269

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] D. Hanisch, A. Zien, R. Zimmer, and T. Lengauer. Co-clustering of biological networks and gene expression data. *Bioinformatics*, 18(90001):145S–154, 2002.

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

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

[5] W. Liebermeister, E. Klipp, S. Schuster, and R. Heinrich. A theory of optimal differential gene expression. *BioSystems*, 76:261–278, 2004.

Berlin Center for Genome Based Bioinformatics

SPONSORED BY THE Federal Ministry of Education and Research

NGFN Nationales Genomforschungsnetz