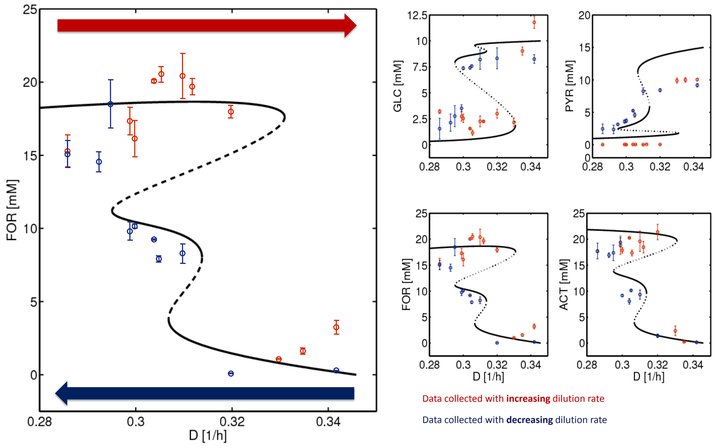
INVESTIGATION OF METABOLIC PHENOMENA USING INFORMATION THEORY

FRANK DEVILBISS[1] PABLO ROBLES-GRANDA[2], AND DORAISWAMI RAMKRISHNA[1]
[1]SCHOOL OF CHEMICAL ENGINEERING, PURDUE UNIVERSITY
[2]DEPARTMENT OF COMPUTER SCIENCE, PURDUE UNIVERSITY

INTRODUCTION

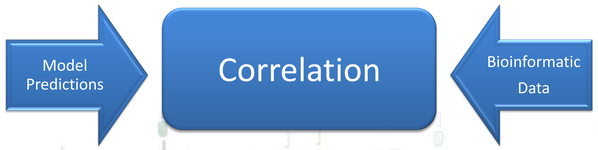
Metabolism embodies all chemical reactions that take place in order to maintain life. Intuitively, an organism performs various metabolic reactions with the goal of survival, but the control goal driving metabolism, programmed into said organism's DNA, is not fully understood at this juncture. Modeling metabolism has relied on steady state concepts without a full dynamic description of system-wide behavior. The predictive power of such approaches is limited. However, a cybernetic mathematical theory, based on viewing the regulation of metabolism as motivated by a survival goal, has been recently shown to successfully anticipate many metabolic phenomena. These include diverse uptake patterns of nutrients as well as a multiplicity of metabolic states at particular growth rates. These results show that metabolism's control objective may be much different from other constraint-based models that are widely used. The cybernetic models have revealed that metabolism may function by dynamic optimal goals such as maximizing the carbon uptake rate at each instant. Understanding what control goal is programmed into DNA is essential to effective metabolic engineering efforts and, more fundamentally, to the study of biology itself.



GLC=10mM, PYR=15mM
(Kim et al., Unpublished)

MATERIALS AND METHODS

In essence, the goal is to validate a theory of metabolism qualitatively through extracting relevant information revealed in volumes of "omic" data. Given the development of many high throughput bioinformatic technologies, the ability to take thousands of simultaneous measurements on the genome scale has come of age. It is our goal to analyze this data and compare this with our model predictions.

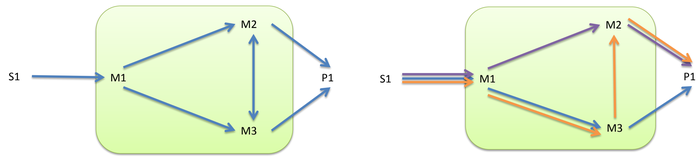


DISCUSSION

MODEL PREDICTIONS

Using cybernetic models of metabolism, model predictions are generated describing dynamic growth patterns for a system specified by published data. With the metabolic models are two integral features that lend to their accurate approximation of single cellular organisms: cybernetic controls and elementary mode reduction.

Elementary modes are a breakdown of a metabolic reaction network's topology into simple reactions pertaining to cellular input and cellular output.[4] This network simplification still retains the detail of the various metabolic pathways.



Network Reactions

Network Elementary Modes

MODEL PREDICTIONS CONTINUED

Cybernetic controls within the model represent the cell's modulation of metabolic activity towards a certain goal.[5] In this case, the cell is trying to eat the fastest as it competes for resources with other cells. At the same time, there are resource constraints that prevent the cell from digesting all sugars available in a culture due to the fact that the cell must invest resources in the production of enzymes.

$$u_i = \frac{r_i}{\sum_j r_j} \qquad v_i = \frac{r_i}{\max_j r_j}$$

In these formulations above, each rate of uptake through a specific pathway is modulated by these cybernetic variables in which modes with higher rates have higher uptake rates and larger enzyme production than modes with lower rates.

BIOINFORMATIC DATA

There are a variety of high throughput experimental techniques that provide transcriptomic data. While RNA-seq and SAGE provide the best quantification of mRNA, relevant data is not as accessible as it is for cDNA microarrays due to cost constraints and ease of access to these technologies.[1]

Typically, microarray data is analyzed in terms of differential expression between different conditions (e.g. growth on glucose vs. growth on pyruvate). Towards the consideration of error present in data collection as well as biological errors, a Bayesian hierarchical error model isolates differentially expressed genes:[2]

Error Model:

$$y_{i,j,k,l}|\{x_{i,j,k},\sigma_e^2\} = x_{i,j,k} + e_{i,j,k,l} \qquad x_{i,j,k}|\left\{\mu, g_i, c_j, r_{i,j}, \sigma_{b_{i,j}}^2\right\} = \mu + g_i + c_j + r_{i,j} + b_{i,j,k}$$

Computation:

$$\Pr(\boldsymbol{y},\boldsymbol{x};\boldsymbol{\theta}) = \prod_{i,j,k,l}\phi\left(\frac{y_{i,j,k,l}-x_{i,j,k}}{\sigma_e}\right)\times\prod_{i,j,k}\phi\left(\frac{x_{i,j,k}-\mu-g_i-c_j-r_{i,j}}{\sigma_{b_{i,j}}}\right)$$

$$\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) \propto \Pr(\boldsymbol{y},\boldsymbol{x};\boldsymbol{\theta})\times\prod_i\phi\left(\frac{g_i}{\sigma_g}\right)\times\prod_j\phi\left(\frac{c_j}{\sigma_c}\right)\times\prod_{i,j}\phi\left(\frac{r_{i,j}}{\sigma_r}\right) \times\prod_{i,j}\Gamma\left(\sigma_{b_{i,j}}^{-2};\alpha_b,\beta_b\right)\times\Gamma(\sigma_e^{-2};\alpha_e\beta_e)$$

DISCUSSION CONTINUED

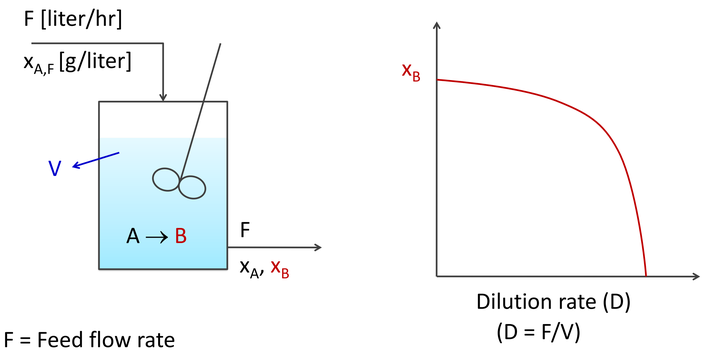
Gibb's sampling or Markov Chain Monte Carlo can be used to estimate these parameters given that direct estimation is limited by the fact that there are so many parameters. From here, an F-Statistic can isolate differentially expressed genes based off of parameter estimation.

$$F_i = \sum_{j=1}^{C}\frac{m_{i,j}\left(\bar{\mu}_{i,j}-\bar{\bar{\mu}}_i\right)^2}{M_i\left(\bar{\sigma}_{b_{i,j}}^2+\bar{\sigma}_e^2\right)} \qquad \bar{\bar{\mu}}_i = \sum_{j=1}^{C}\frac{\bar{\mu}_{i,j}}{C}\ M_i = \sum_{j=1}^{C} m_{i,j}$$

Given statistical agreement between model predictions and gene expression data, steps towards the validation of a metabolic theory will have been made. The data set that has been elected for the model formulation and differential gene expression analysis is from Ishii et. al.

LITERATURE CITED

1. Baggerly, K. A., Coombes, K. R., & Morris, J. S. (2006). An Introduction to High-Throughput Bioinformatics Data. In *Bayesian Inference for Gene Expression and Proteomics* (pp. 1-39). Cambridge University Press.
2. Cho, H., & Lee, J. K. (2004). Bayesian hierarchical error model for analysis of gene expression data. *Bioinformatics* , 2016-2025.
3. Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., et al. (2007). Multiple High-Throughput Analyses Monitor the Response of E. coli to Perturbations. *Science* , 593-597.
4. Kim, J. I., Varner, J. D., & Ramkrishna, D. (2008). A Hybrid Model of Anaerobic E. coli GJT0001: Combination of Elementary Flux Modes and Cybernetic Variables. *Biotechnology Progress* , 993-1006.
5. Kompala, D. S., Ramkrishna, D., Jansen, N. B., & Tsao, G. T. (1986). Investigation of Bacterial Growth on Mixed Substrates: Experimental Evaluation of Cybernetic Models. *Biotechnology and Bioengineering* , 1044-1055.



F = Feed flow rate
$x_{A,F}$ = Concentration of A in the feed
x_A = Concentration of A at the outlet
x_B = Concentration of B at the outlet

CENTER FOR SCIENCE OF INFORMATION
A National Science Foundation Science and Technology Center
soihub.org