Canonical Modeling as a Tool in Metabolic Engineering

Eberhard O.Voit

Integrative BioSystems Institute and
Department of Biomedical Engineering
Georgia Institute of Technology and Emory University
Atlanta, Georgia

GT School of Chemical and Biomolecular Engineering
November 12, 2008

Preamble

In College I was told that Biology was too complicated to use Math.

I have learned by now that Biology is too complicated *not* to use Math.

Overview

Metabolic Engineering

Challenges and Need for Models

Canonical Modeling; Biochemical Systems Theory

Advantages, Examples

Issues of Optimization

IBSI: An Overview

Metabolic Engineering

Metabolic Engineering is a new approach to understanding and using metabolic processes. As the name implies, ME is the targeted and purposeful alteration of metabolic pathways found in an organism in order to better understand and use cellular pathways for chemical transformation, energy transduction, and supramolecular assembly.

Metabolic Engineering Working Group

http://www.metabolicengineering.gov/

Metabolic Engineering

In the past:
Strain and productivity improvement through random mutagenesis and selection (strain and medium); maybe adaptive evolution

Q: Isn't there a more rational way of doing this?

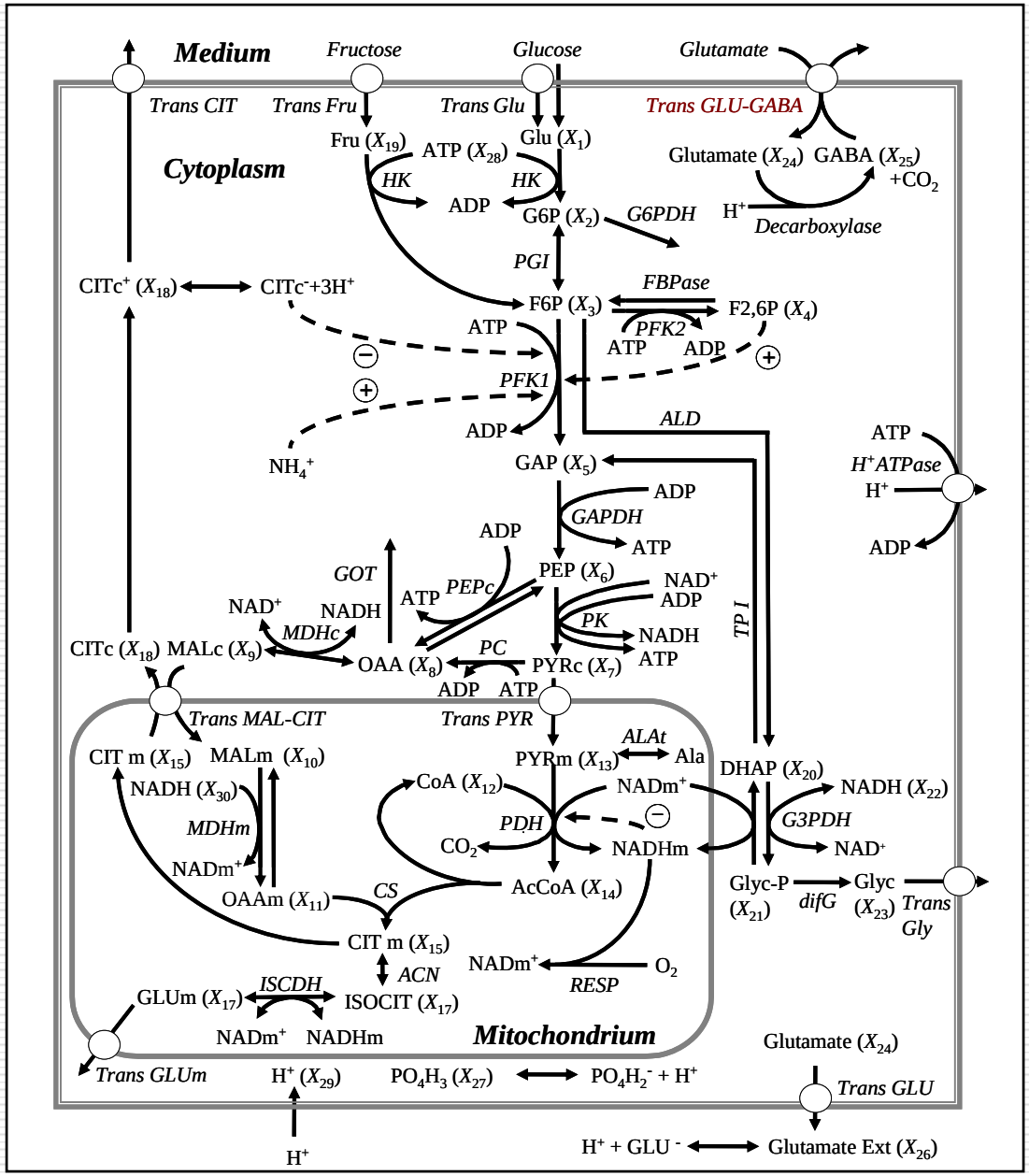
Now and Future:
Develop mathematical models of organisms, manipulate models, implement successful implementation in actual organisms

Example: Citric Acid (*A. niger*)

Currie started in 1915

Our Task:
Compute how to reroute flux in an optimal fashion; *e.g.*, maximize citric acid output

Big challenge:
Complexity of biological systems



Features of Organizational Complex Systems

Large numbers of components

Large number of processes

Processes are nonlinear

Quantitative changes in parameters cause qualitative changes in response

Large Numbers of Components and Processes

6,000 genes in *S. cerevisiae*

Thousands of proteins in *E. coli*

100 billion neuronal components in human brain

100s of trillions of interconnections between neurons

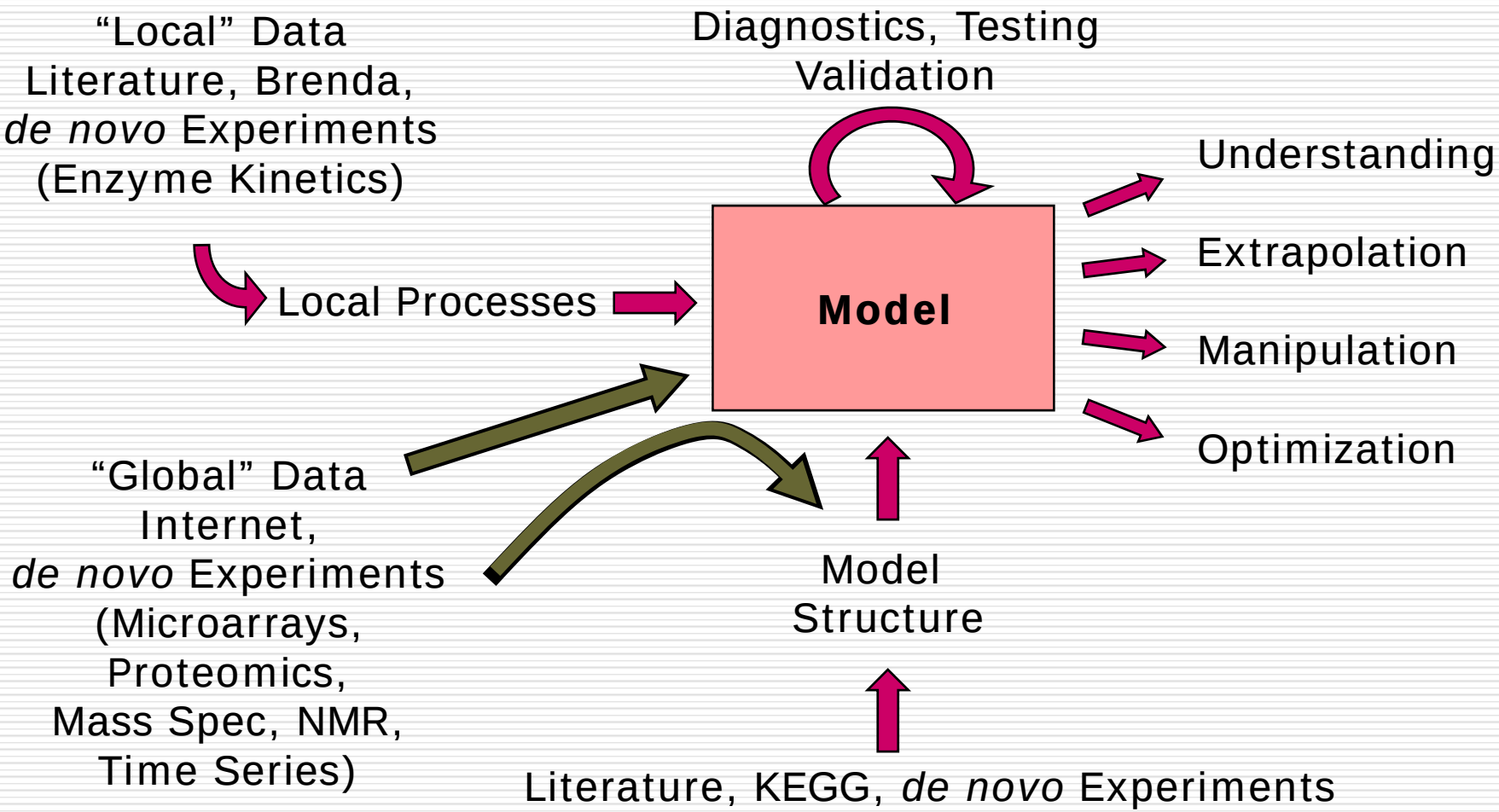
5 octillion atoms in human body:
5,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000

Miller (1956): Human brain can manage 7±2 items at once

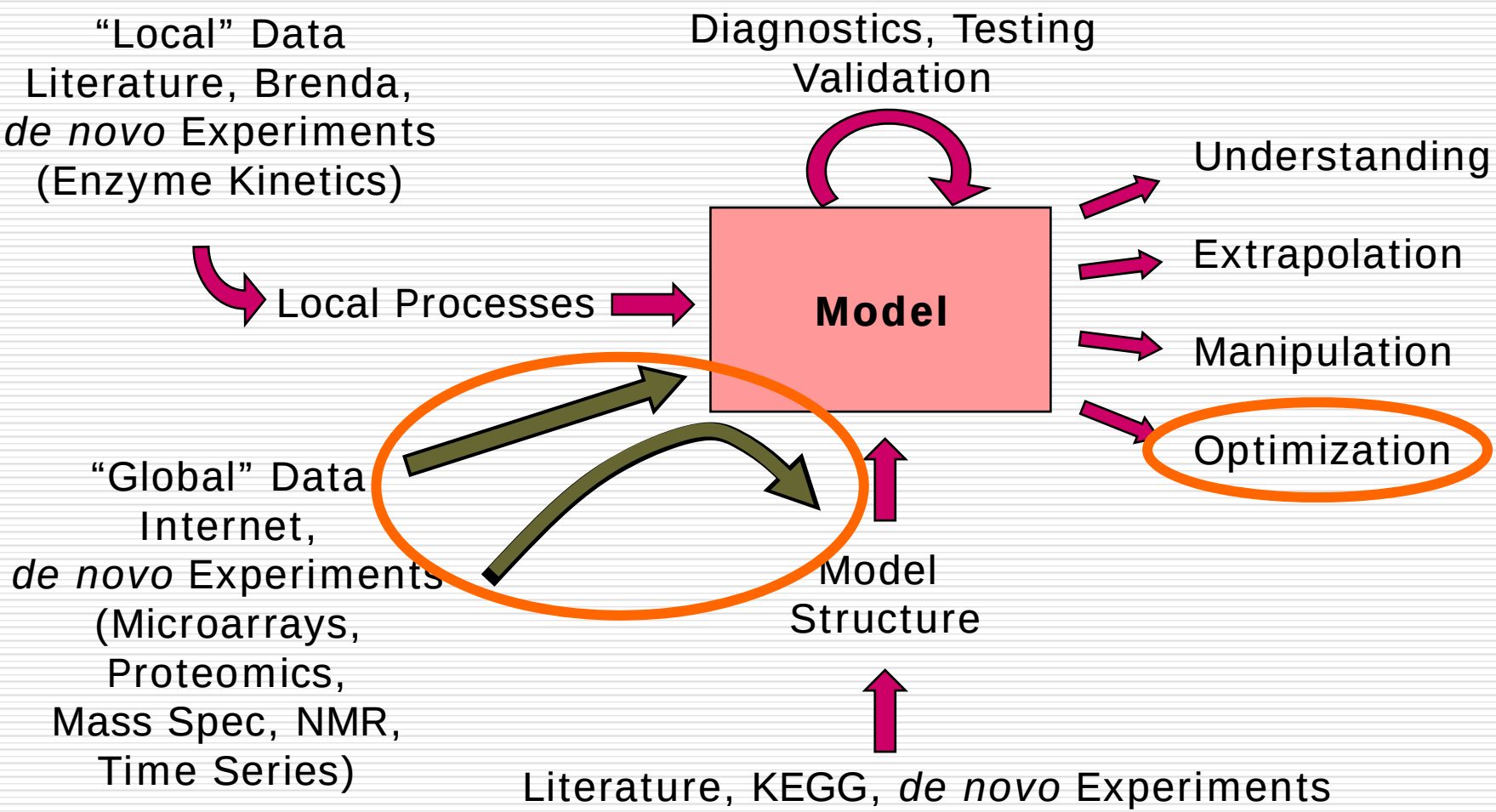
Biological Systems are Scary !

ystems Analysis to the Rescue!

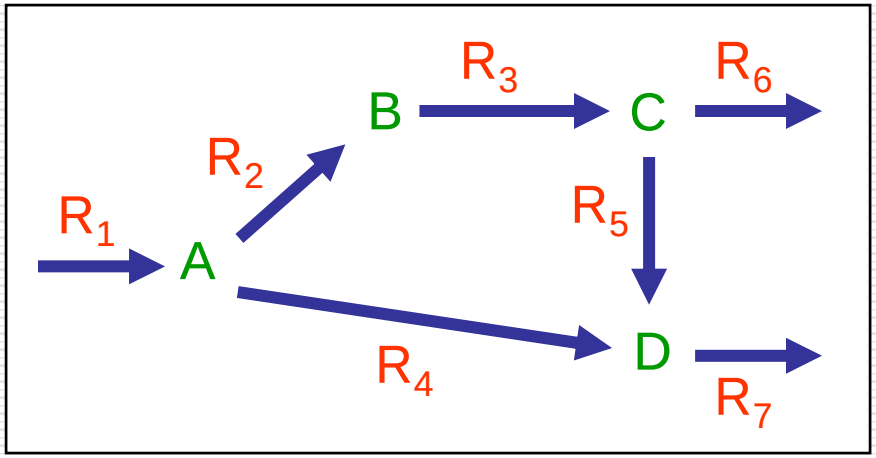
Generic Modeling of Pathway Systems



Generic Modeling of Pathway Systems



Stoichiometric Systems



$$\mathbf{N}= \begin{array}{c} \begin{array}{ccccccc} R_1 & R_2 & R_3 & R_4 & R_5 & R_6 & R_7 \end{array} \\ \left(\begin{array}{ccccccc} 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \end{array}\right) \end{array} \begin{array}{c} \\ A \\ B \\ C \\ D \end{array}$$

Change in substrate concentrations (**S**) is function of fluxes (**V**) and stoichiometric matrix **N**: $d\mathbf{S}/dt = \mathbf{N}\cdot\mathbf{V} = 0$

Flux-Balance Analysis (Bernhard Palsson group):
Reduce solution space with physico-chemical constraints

Stoichiometric Systems

Advantages:

No kinetic details needed, just topology and fluxes
Linear system (no real size limitation)
Straightforward optimization
Steady-state solution space given by kernel

Limitations:

Kinetic information cannot be used
No nonlinearities
No true regulatory signals possible
Optimal strategies of flux alteration
affected by signals

Wanted

A nonlinear formalism that:

captures the essence of biological systems

accounts for physiological and pathological processes, regulation, signal transduction…

yields insight

lends itself to (manageable) analysis

lends itself to (manageable) simulation

lends itself to optimization

is computationally efficient

Toward a Nonlinear Modeling Framework

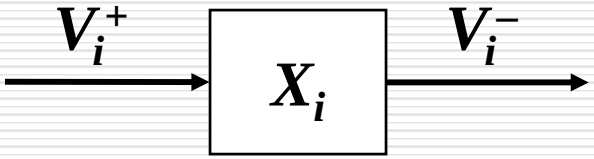
Needs to be Based on Integrated Systems Analysis:

Each component of the system may potentially depend on all other components and outside factors.

To “understand” the system, we need to know how every component changes over time.

Dynamic changes in a system component are driven by inputs and outputs.

Formulation of a Dynamical Systems Model



$$\dot{X}_i = \frac{dX_i}{dt} = V_i^+ - V_i^-$$

$$V_i^+ = V_i^+(\underbrace{X_1, X_2, \ldots, X_n}_{\textit{inside}}, \underbrace{X_{n+1}, \ldots, X_{n+m}}_{\textit{outside}})$$

complicated

Big Problem: Where do we get functions from?

Sources of Functions for Complex Systems Models

Physics: Functions come from theory

Biology: No theory available

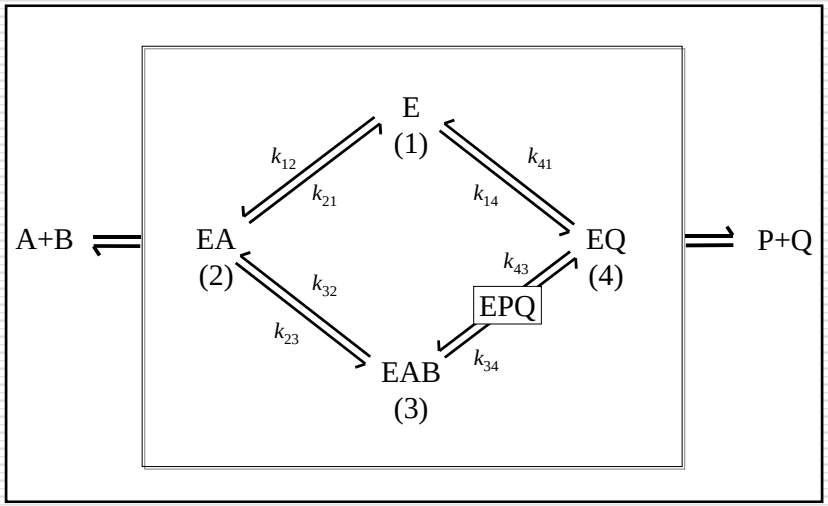
Solution 1: Educated guesses: growth functions

Solution 2: “Partial” theory: Enzyme kinetics

Solution 3: Generic approximation

Why not Use "True" Functions?

$$A+B \rightleftharpoons P+Q$$

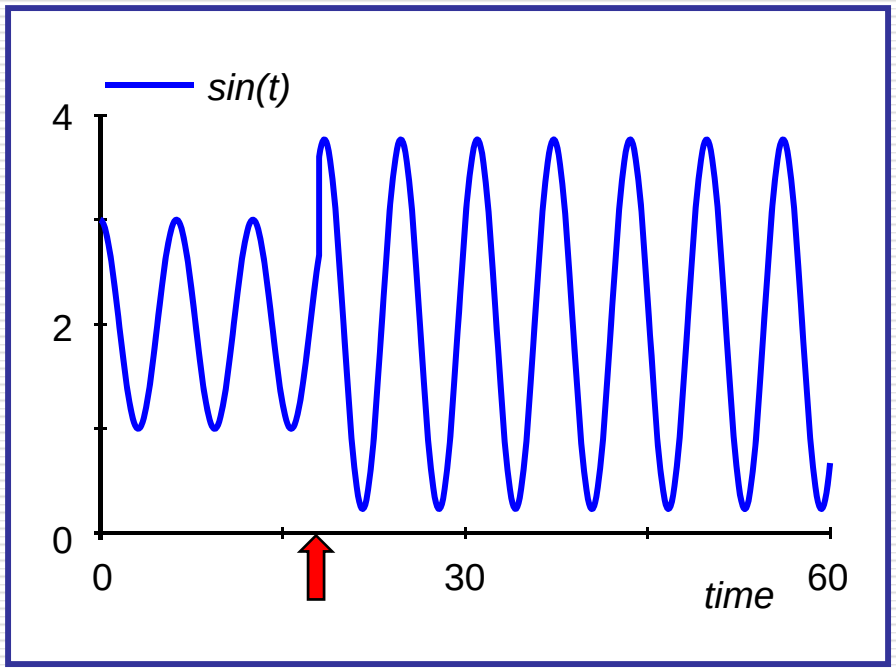


$$v = \frac{\left(\frac{\text{num.1}}{\text{coef. AB}}\right)(A)(B) - \left(\frac{\text{num.1}}{\text{coef. AB}} \times \frac{\text{num.2}}{\text{num.1}}\right)(P)(Q)}{\left(\frac{\text{constant}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right) + \left(\frac{\text{coef. A}}{\text{coef. AB}}\right)(A) + \left(\frac{\text{coef. B}}{\text{coef. AB}}\right)(B)}$$
$$+ \left(\frac{\text{coef. AB}}{\text{coef. AB}}\right)(A)(B) + \left(\frac{\text{coef.P}}{\text{coef. AP}} \times \frac{\text{coef. AP}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right)(P)$$
$$+ \left(\frac{\text{coef.Q}}{\text{constant}} \times \frac{\text{constant}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right)(Q)$$
$$+ \left(\frac{\text{coef.AP}}{\text{coef. A}} \times \frac{\text{coef.A}}{\text{coef. AB}}\right)(A)(P) + \left(\frac{\text{coef.BQ}}{\text{coef. B}} \times \frac{\text{coef.B}}{\text{coef. AB}}\right)(B)(Q)$$
$$+ \left(\frac{\text{coef.PQ}}{\text{coef.Q}} \times \frac{\text{coef.Q}}{\text{constant}} \times \frac{\text{constant}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right)(P)(Q)$$
$$+ \left(\frac{\text{coef.ABP}}{\text{coef. AB}}\right)(A)(B)(P)$$
$$+ \left(\frac{\text{coef.BPQ}}{\text{coef. BQ}} \times \frac{\text{coef.BQ}}{\text{coef. B}} \times \frac{\text{coef.B}}{\text{coef. AB}}\right)(B)(P)(Q)$$

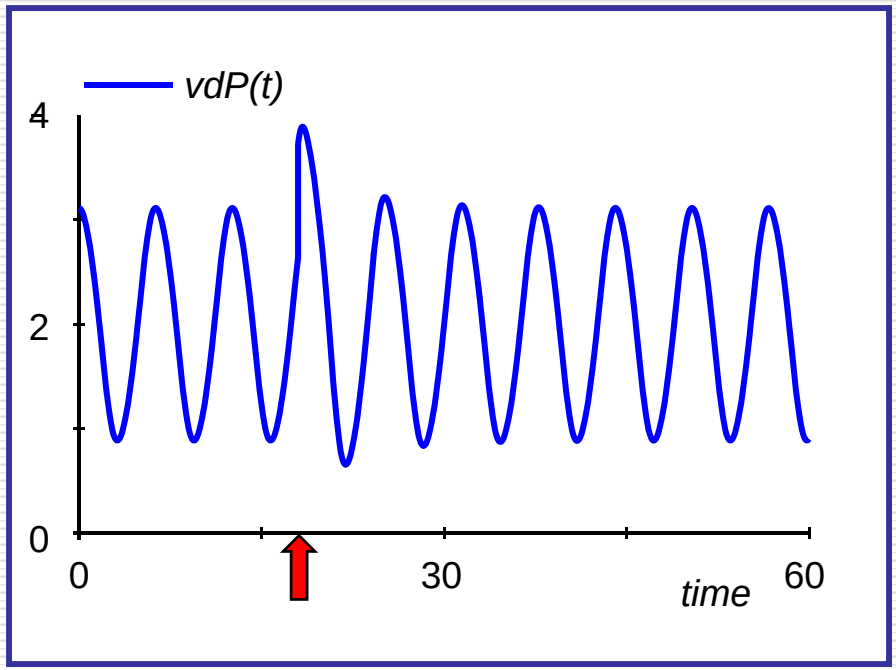
from Schultz (1994)

Why not Use Linear Functions?

Example: Heartbeat modeled as stable limit cycle



System of linear differential equations



System of non-linear differential equations

Formulation of a Model for Complex Systems

Challenge:

Linear approximation unsuited

Infinitely many nonlinear functions

Solution with Potential: $$\dot{X}_i = \frac{dX_i}{dt} = V_i^+ - V_i^-$$

Savageau (1969): Approximate V_i^+ and V_i^- in a logarithmic coordinate system, using Taylor theory.

Result: *Canonical Modeling; Biochemical Systems Theory*.

Result: S-system

$$\dot{X}_i = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} \ldots X_{n+m}^{g_{i,n+m}} - \beta_i X_1^{h_{i1}} X_2^{h_{i2}} \ldots X_{n+m}^{h_{i,n+m}}$$

Each term is represented as a product of power-functions.

α's and β's are *rate constants*, *g*'s and *h*'s *kinetic orders*.

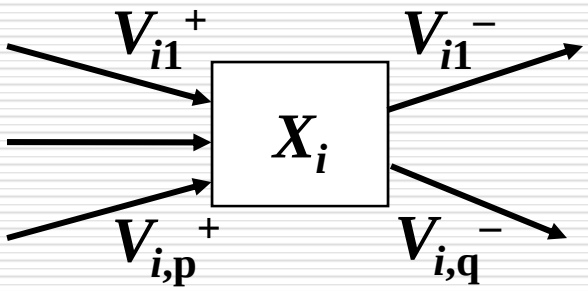
Important:

Each term contains and only those variables that have a direct effect; others have exponents of 0 and drop out. Automated model design possible!

Alternative Formulations within BST

S-system Form:

$$\dot{X}_i = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} \ldots X_{n+m}^{g_{i,n+m}} - \beta_i X_1^{h_{i1}} X_2^{h_{i2}} \ldots X_{n+m}^{h_{i,n+m}}$$

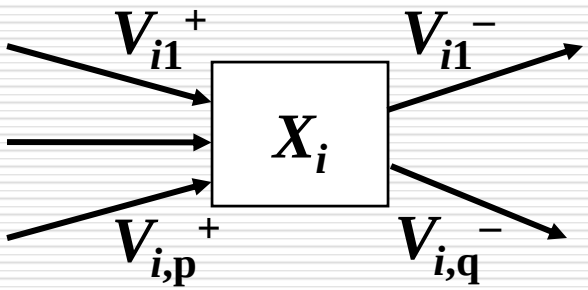


$$\dot{X}_i = \frac{dX_i}{dt} = \sum V_{ij}^+ - \sum V_{ij}^-$$

Alternative Formulations

S-system Form:

$$\dot{X}_i = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} \cdots X_{n+m}^{g_{i,n+m}} - \beta_i X_1^{h_{i1}} X_2^{h_{i2}} \cdots X_{n+m}^{h_{i,n+m}}$$



$$\dot{X}_i = \frac{dX_i}{dt} = \sum V_{ij}^{+} - \sum V_{ij}^{-}$$

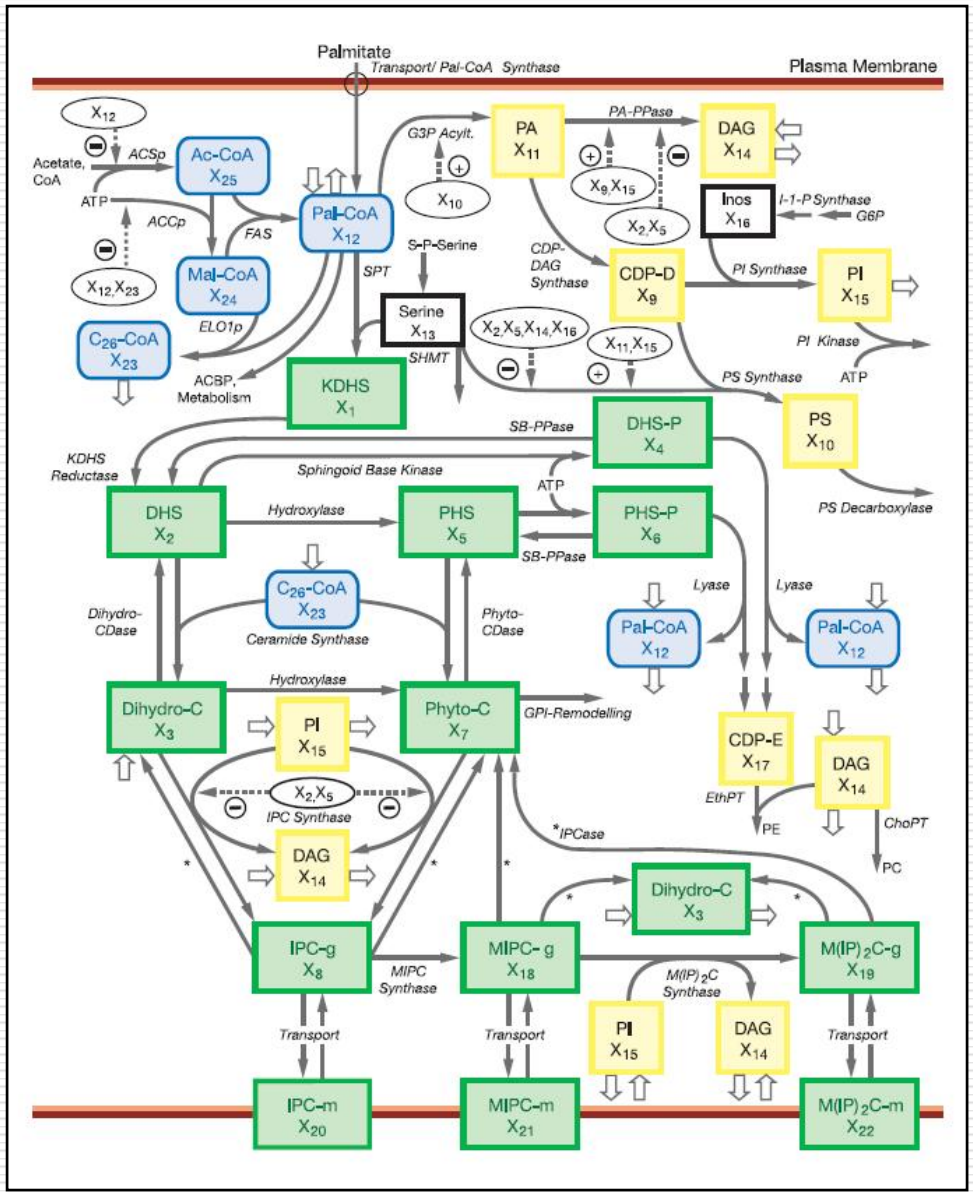
Generalized Mass Action Form:

$$\dot{X}_i = \sum \pm \gamma_{ik} \prod X_j^{f_{ijk}}$$

Doable Size

Sphingolipid pathway (purely metabolic)

1. ~25 metabolites
2. ~ 30 enzyme steps
3. Many parameters
4. Values from literature; some *in vivo* experiments



Alvarez, Sims , Hannun, Voit
JTB, 2004; *Nature,* 2005

Applications of BST

Pathways: purines, glycolysis, citric acid, TCA, red blood cell, trehalose, sphingolipids, dopamine, lignin synthesis, ...

Genes: circuitry, regulation,...

Genome: explain expression patterns upon stimulus

Metabolic engineering, maximization of yield,...

Growth, immunology, pharmaceutical science, forestry, ...

Math: recasting, function classification, bifurcation analysis,...

Statistics: S-system representation, S-distribution, trends; applied to seafood safety, marine mammals, health economics

Steady-State Equations Linear

$$\dot{X}_i = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} \ldots X_{n+m}^{g_{i,n+m}} - \beta_i X_1^{h_{i1}} X_2^{h_{i2}} \ldots X_{n+m}^{h_{i,n+m}} = 0$$

Define $Y_i = \log(X_i)$:

$$\log\alpha_i + g_{i1}Y_1 + g_{i2}Y_2 + g_{i,n+m}Y_{n+m}$$
$$= \log\beta_i + h_{i1}Y_1 + h_{i2}Y_2 + h_{i,n+m}Y_{n+m}$$

$$\mathbf{Y}_D = \mathbf{A}_D^{-1} \cdot \mathbf{b} - \mathbf{A}_D^{-1} \cdot \mathbf{A}_I \cdot \mathbf{Y}_I$$

S-system highly nonlinear, but steady-state equations linear.

Pathway Optimization with S-systems

Optimization under steady-state (batch) conditions becomes

Linear Program

even though (nonlinear) kinetics is taken into account:

maximize log(flux) [or log(variable)]

subject to:

Steady-state conditions in log(variables)

Constraints on log(variables)

Constraints on log(fluxes)

Pathway Optimization (cont'd)

***Great** Advantage:*

Methods of *Operations Research* applicable

- very well understood
- applicable for over 1,000 simultaneous variables
- robust and efficient
- incomparably faster than nonlinear methods

Torres, Alvarez, Voit, …: Applications (*e.g.*, citric acid, ethanol, glycerol, L-carnitine)

Hatzimanikatis, Bailey, Floudas, 1996: Use these features for optimization of pathway structure

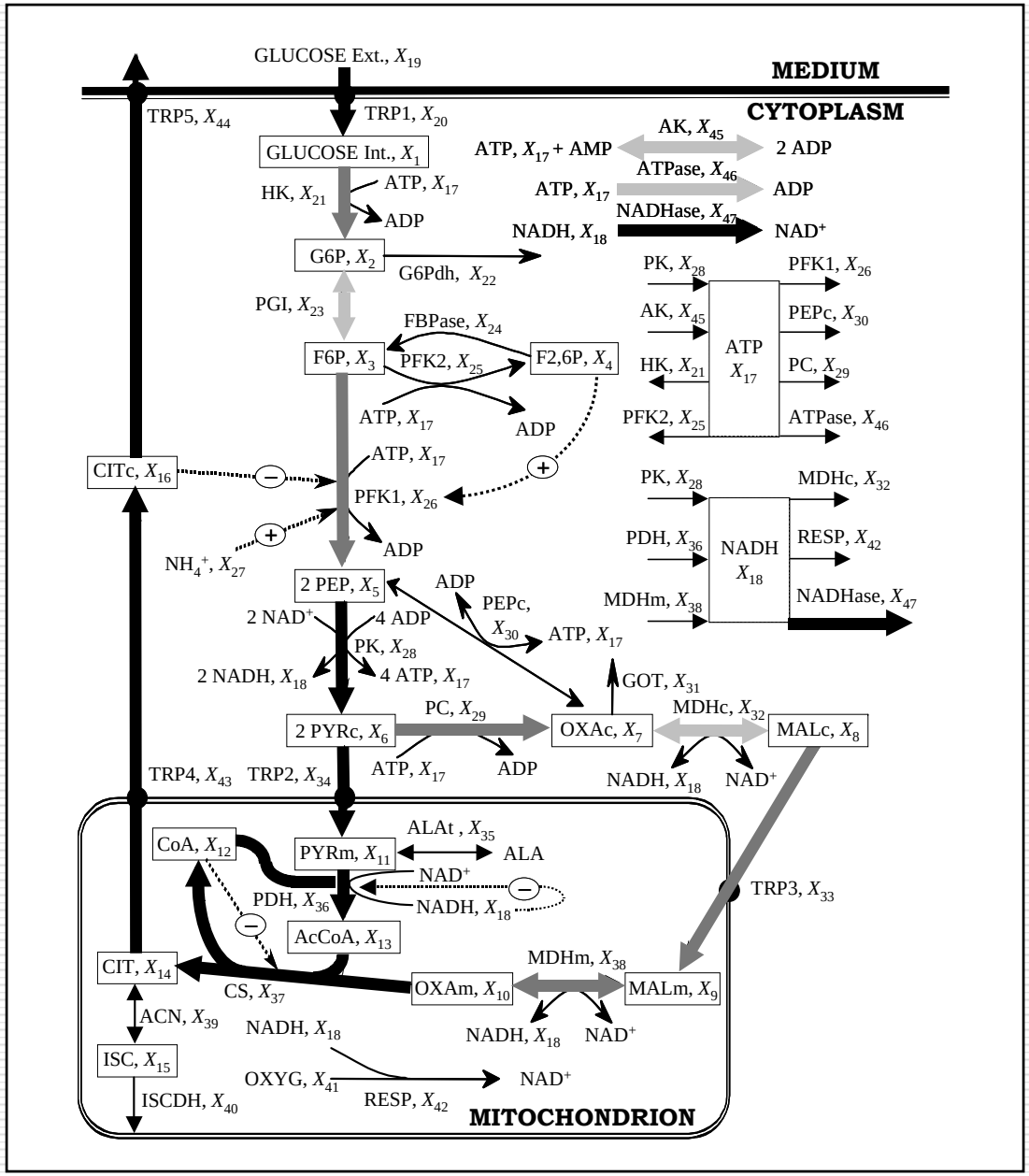
Marin-Sanguino, Torres, Polisetty, Gatzke, Voit, …:
Extension to GMA models via
iterative methods, branch-and-reduce methods, geometric programming

Example

Citric acid yield:

Optimization prescribes enzyme activity levels that lead to maximal citric acid production while satisfying constraints on metabolites and fluxes.

Maximal increase:
~ 12 fold



Notable Results

Citric acid system contains ~ 20 accessible enzymes / genes

Optimize by allowing changes in all enzymes:
Yield increased ~ 12 fold

Q: If only a single enzyme may be changed, which one should it be? How much could yield be increased?

A: No matter which enzyme is changed, yield does not really increase!

Q: Change 2, 3, … enzymes. Yield improvement?

A: 2: none; 3: none, 4-6: almost none; 7 needed for ~3 fold yield!

Interpretation: Standard techniques have found the easy solutions!

‘The Other Optimization’: Parameter Estimation from Time Series Data

- According to computer scientists: trivial, solved.
- Many methods
- Most work sometimes
- None works always
- Estimation remains to be a challenging topic!
- Example: Kikuchi *et al.* 2003
- Over 100 papers on BST estimation since 2000

Problems with Traditional Methods

Time to (global) convergence

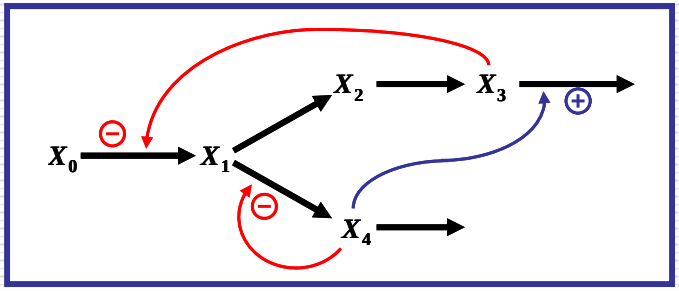
Problems with collinear and or noisy data

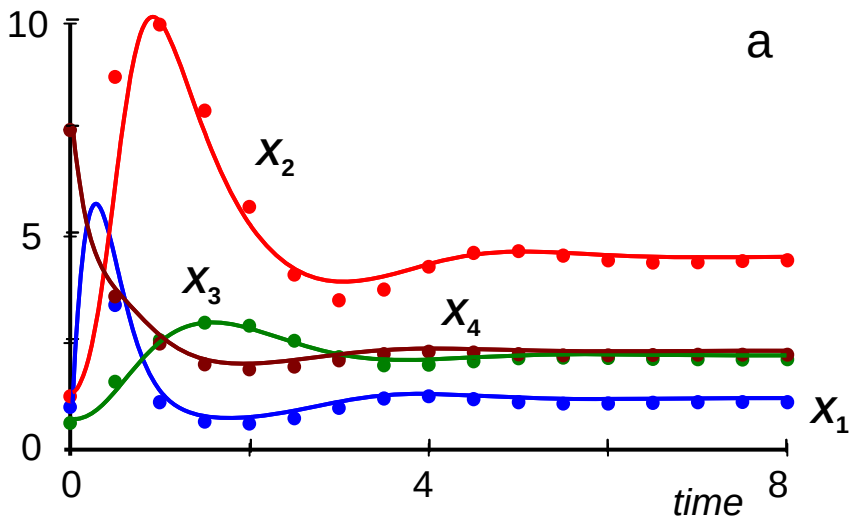
Problems with models permitting redundancies

Problems with compensation of error among terms

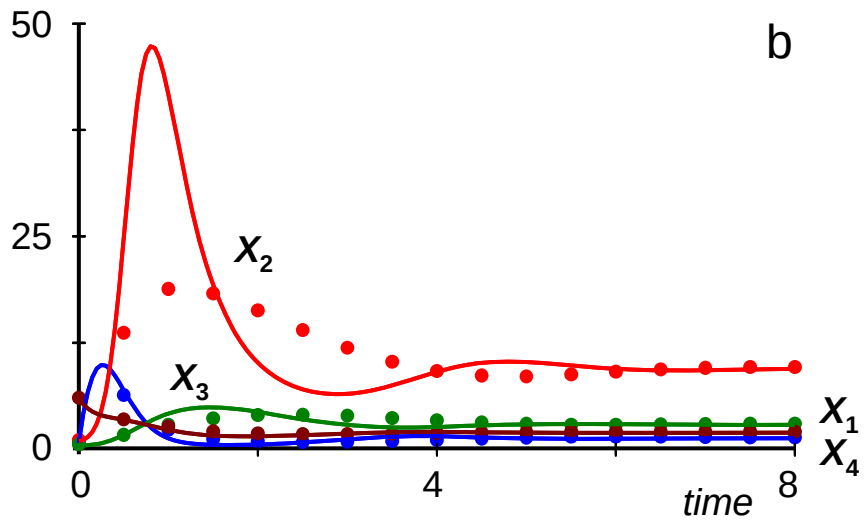
Problems with Traditional Methods: Extrapolation

Generic model in GMA form





Bad parameters, but good fits because of error compensation



Problem with the "misestimated" system during extrapolation

Dynamic Flux Estimation (DFE)

Inspired by Stoichiometric and Flux Balance Analysis

Extended to dynamic time courses

Study flux balance at each time point

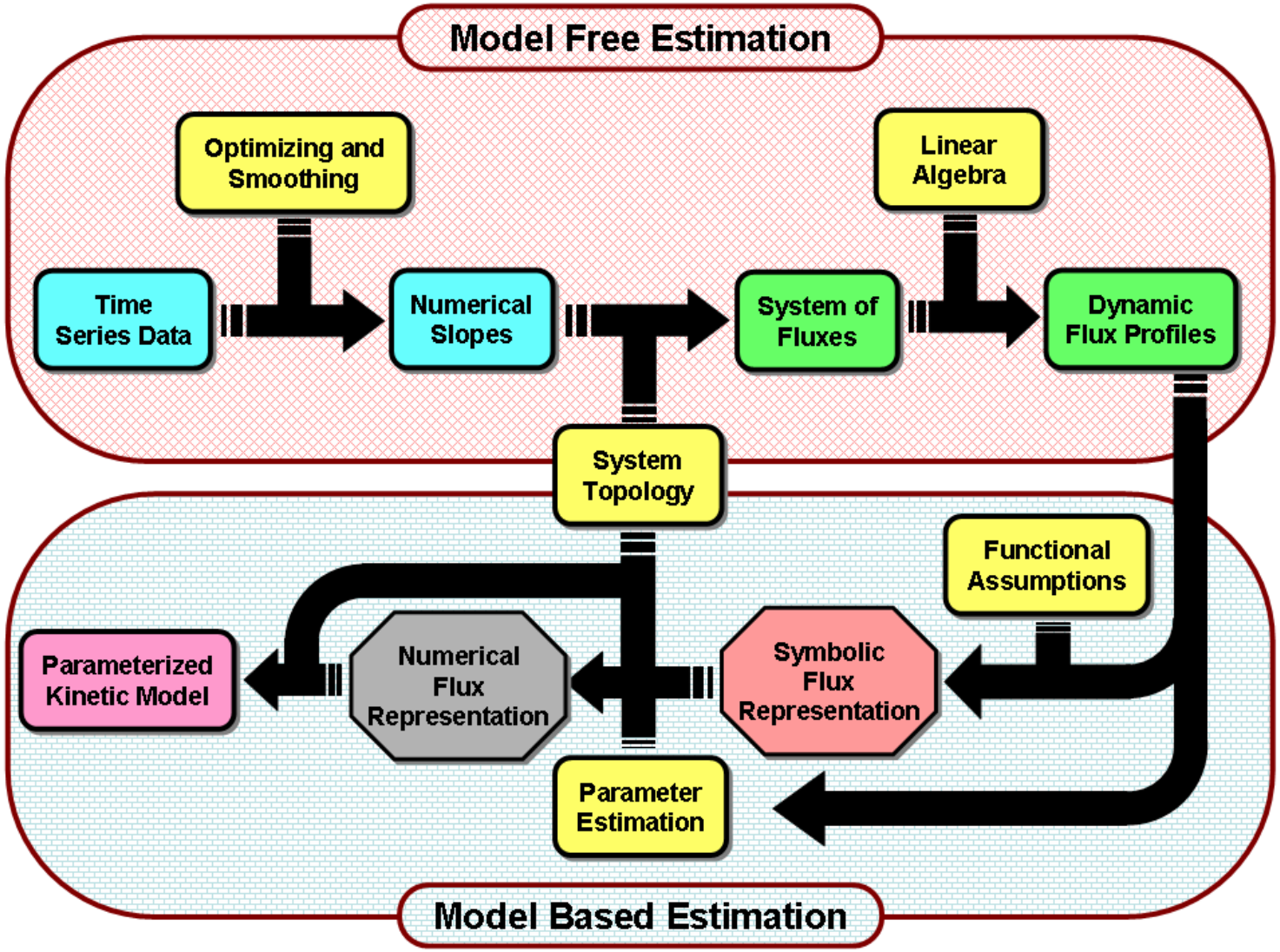
Change in variable @ t = all influxes @ t – all effluxes @ t

Linear system; solve as far as possible

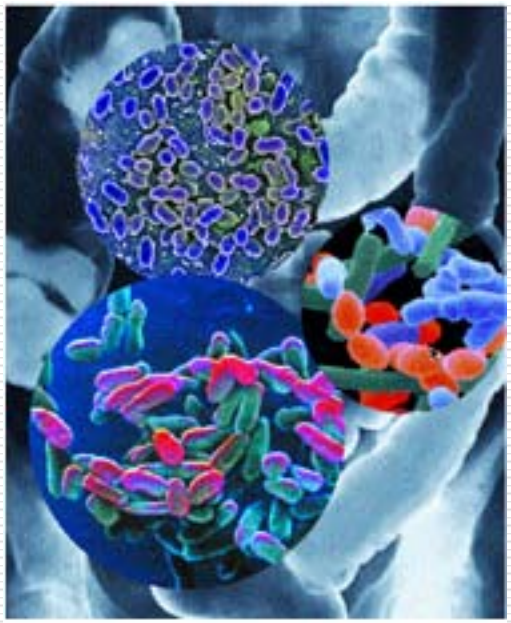
Result: values of each flux @ all *t*

Represent fluxes with appropriate models

Dynamic Flux Estimation (DFE)



Test Bed: Regulation of Glycolysis in *Lactococcus lactis*



Bacteria found in yogurt and cheese: *Lactococcus lactis* (top), *Lactobacillus bulgaricus* (blue), *Streptococcus thermophilus* (orange), *Bifidobacterium* spec (magenta).

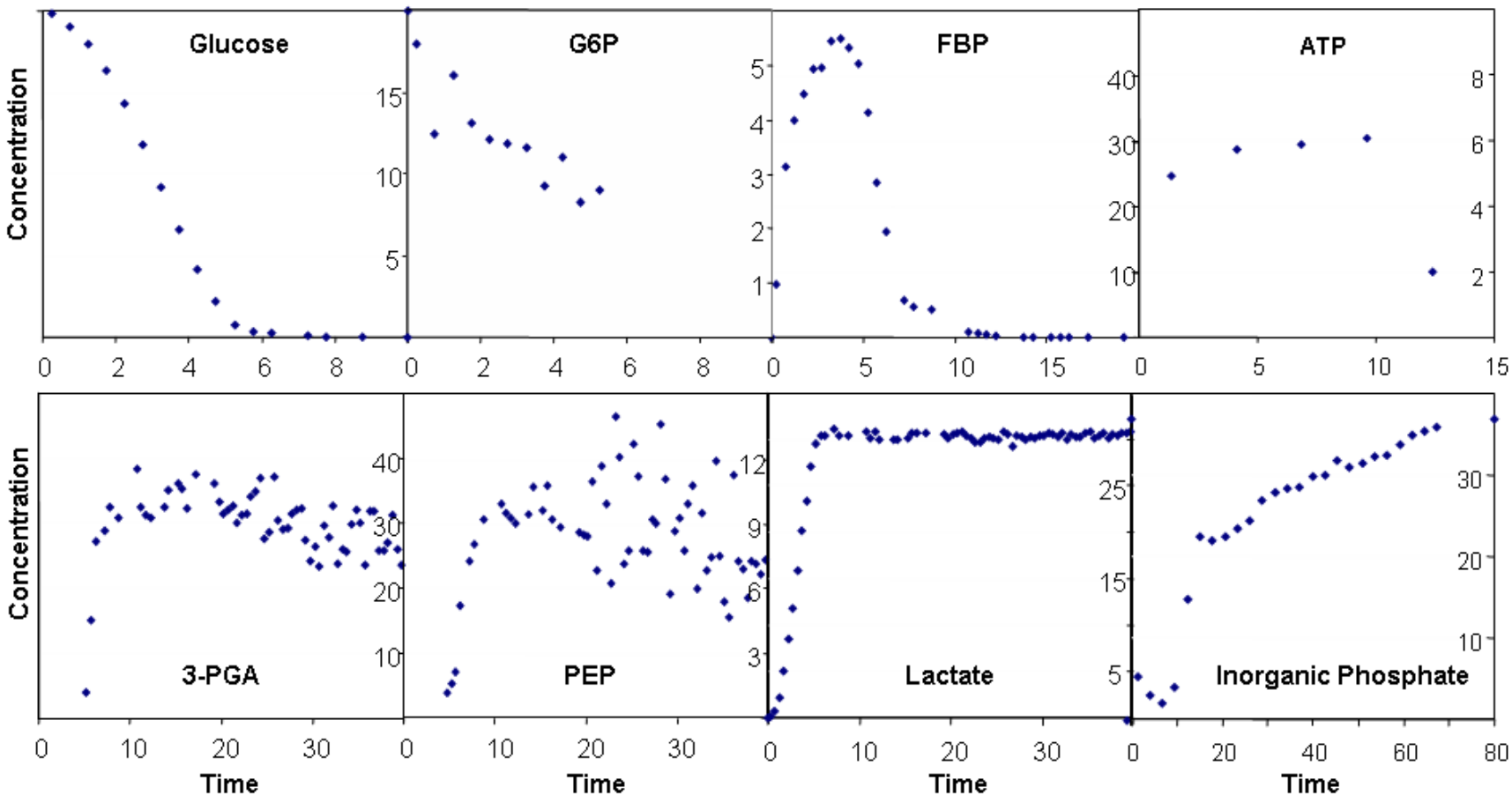
www.hhmi.org/bulletin/winter2005/images/bacteria5.jpg

Bacterium involved in dairy, wine, bread, pickle production.
Relatively simple organization.
Here: study regulation of glycolysis.

Goals of Modeling

- Understand pathway; design, operation
- Allow extrapolation to new situations
- Allow prediction for manipulation
- Maximize yield of main product
- Optimize yield of secondary products
- Eventually develop a cell-wide model

Experimental Time Series Data



E.O. Voit, J.S. Almeida, S. Marino, R. Lall, G. Goel, A.R. Neves, and H. Santos: *IEE Proc. Systems Biol. 2006*

Lactococcus Data

Had modeled (fitted) these data before

First, difficult to find any solutions

Combination of methods led to good fit

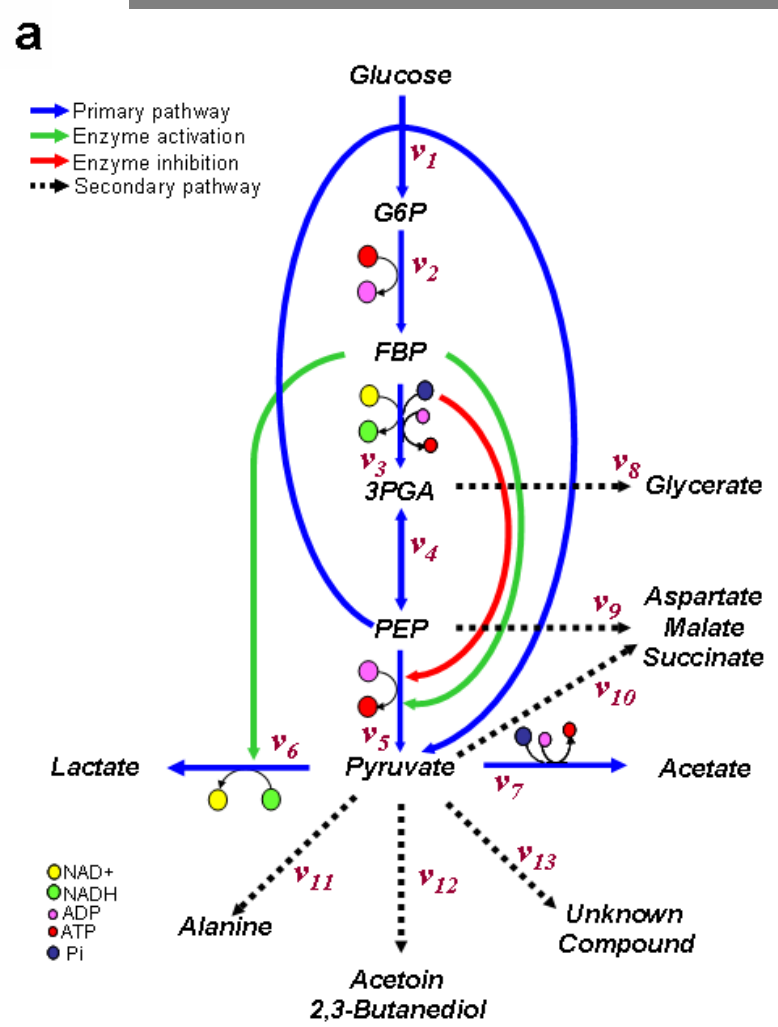
Later, many distinctly different solutions

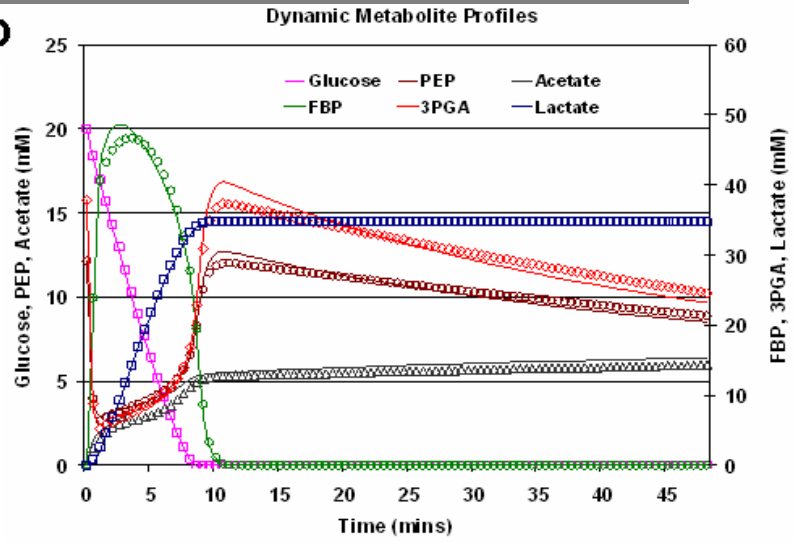
Question: Is any of these solutions optimal?

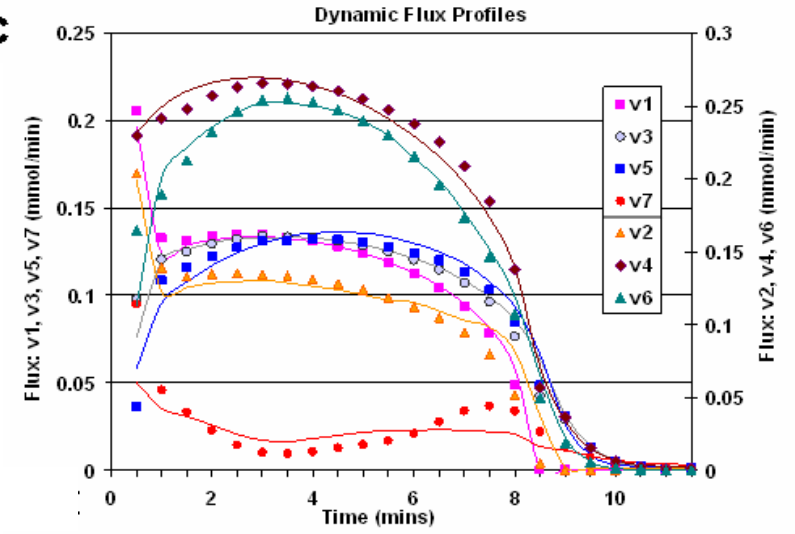
Question: Is the BST model appropriate?

Problems with extrapolation

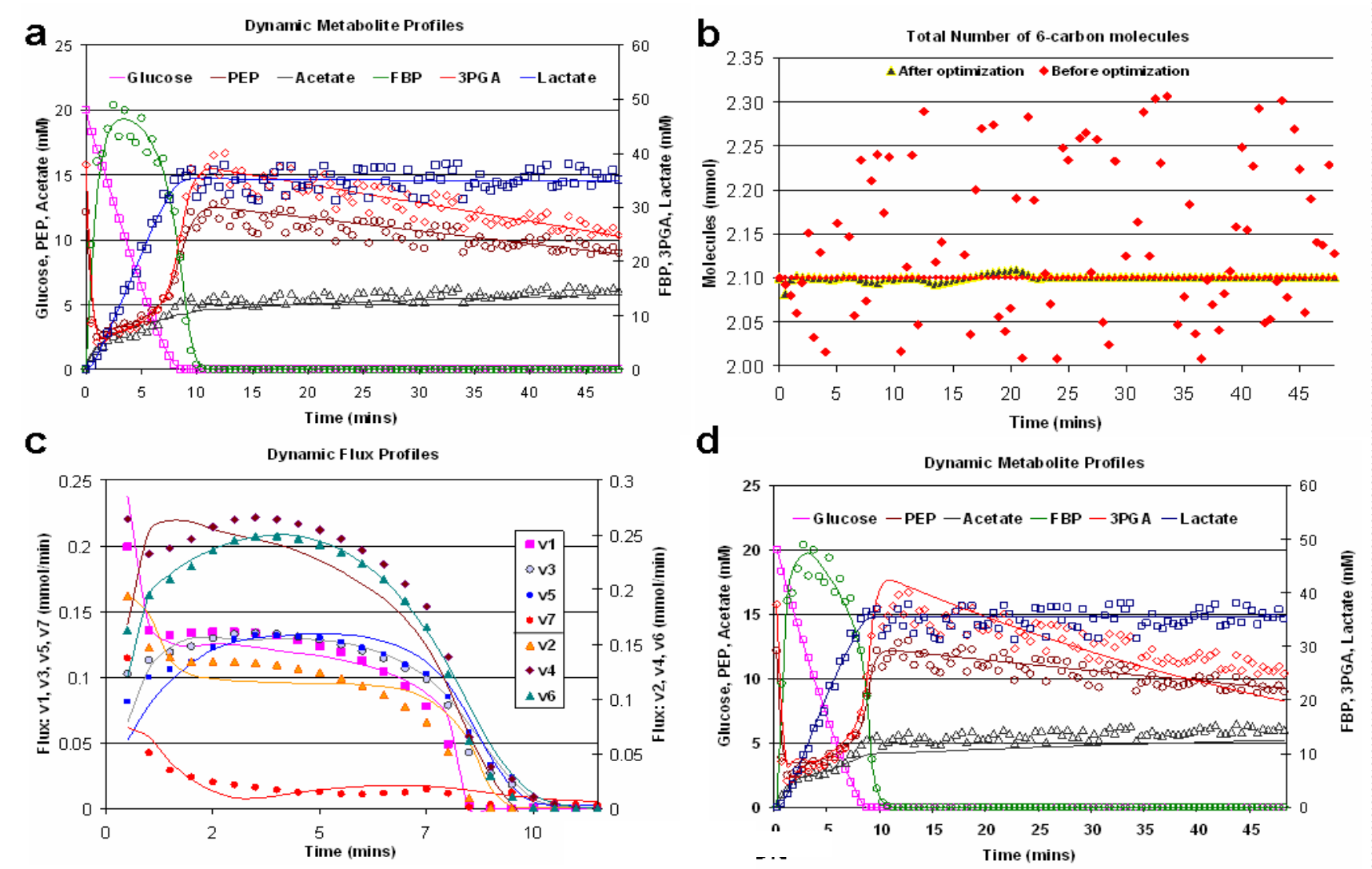
Dynamic Flux Estimation (DFE)



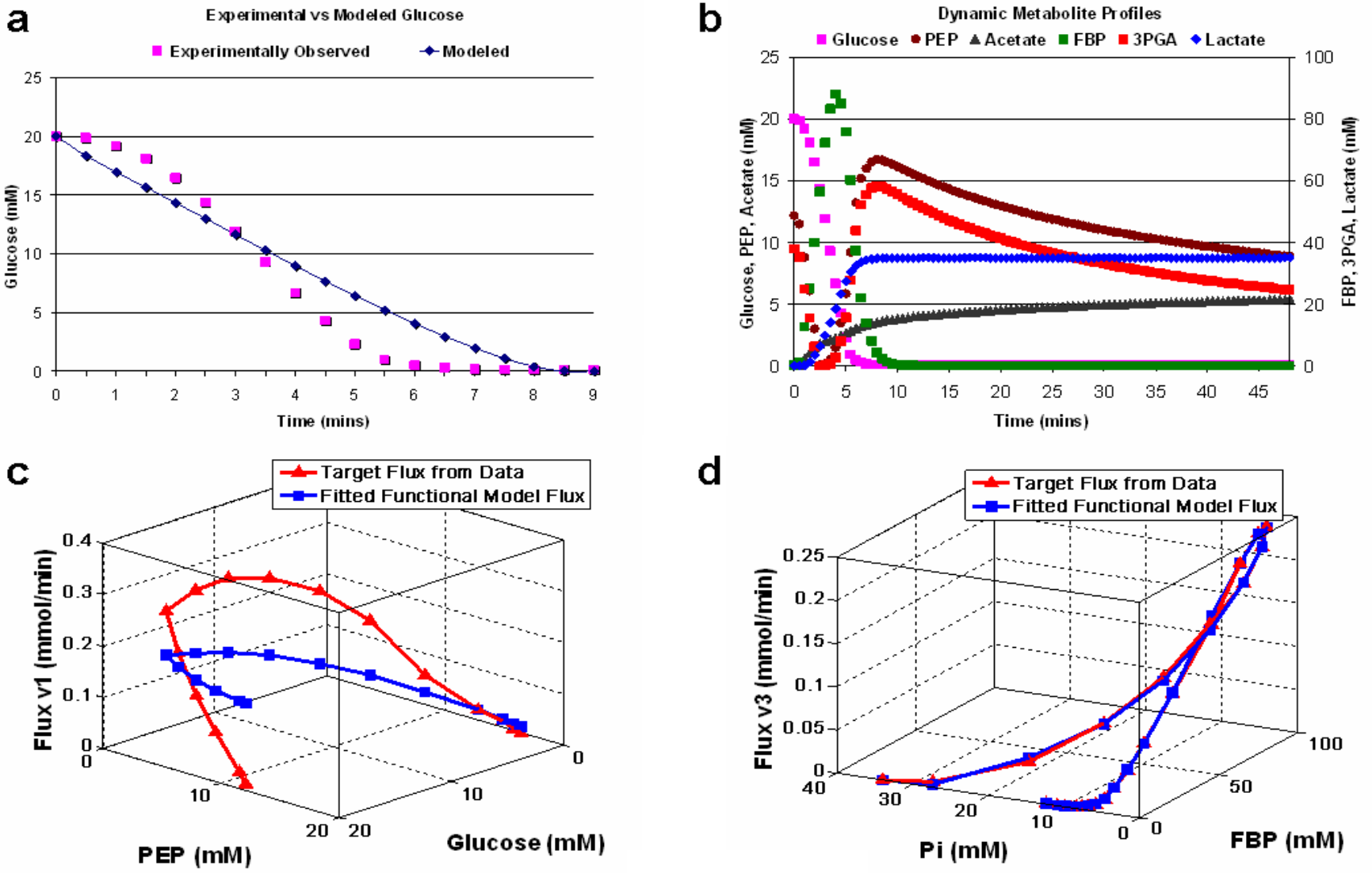




DFE Rather Tolerant to Noise



DFE Spots "Wrong" Assumptions



Open Problems

Technical Issues:

Convergence to slow and not always global

Smoothing and Mass Conservation:

Noise in the data leads to loss or gain of mass

Underdetermined Flux Systems:

Linear system of flux often not of full rank

Augment DFE with other methods

(e.g., regression or bottom-up estimation)

Characterization of Redundancies:

Data collinear or non-informative (pooling?)

Model allows transformation groups (Lie analysis?)

Summary of Technical Talk

Biology has become too complicated not to use math

At the center of computational systems biology is a model

Nature does not provide guidelines for model design

Infinite choices; canonical models provide good default

Advantages of canonical models

- Interpretation of parameters and other features
- Computations associated with steady state
- Optimization facilitated by model structure
  - (yield, parameters)

Acknowledgements

The Current Crew:

Funding: NIH, NSF, DOE, Woodruff Foundation, Georgia Research Alliance, University System of Georgia

Information: www.bst.bme.gatech.edu

Integrative
BioSystems
Institute

IBSI: An Overview

www.ibsi.gatech.edu


CAMPUS MAP
DIRECTORIES
SITE MAP
SEARCH:
GO
Georgia Tech
INTEGRATIVE BIOSYSTEMS INSTITUTE
IBSI
Home
People
Events
Conference
Programs
Facilities
IBSI in the News
Contact
International Launch Conference
October 18-21, 2008
Frontiers in Multi-Scale Systems Biology
Welcome to IBSI
The Integrative BioSystems Institute (IBSI) at Georgia Tech is a forum for multi-scale, multi-disciplinary systems approaches toward solutions of grand-challenge problems in biology. IBSI's main focus is the development and application of enabling technologies that are needed to solve some of the grand-challenge questions of biology and medicine of the 21st century. Our initial target applications are developmental processes leading to cancer and the interactions between humans and microbial systems in the environment. Accompanying and supporting these threads are initiatives in high-performance computing, computational modeling, and the creation of macro-, micro- and nano-devices for biosystems research.
Executive Committee:
Eberhard O. Voit, Director
Jeffrey Skolnick, Associate Director
Richard Fujimoto
Integrative Systems Biology at Georgia Tech
Rationale and Vision
IBSI Spotlight
Center for the Study of Systems Biology
Laboratory for Biological Systems Analysis
Center for Bioinformatics and Computational Genomics
Mass Spectrometry Tissue Imaging Center
Environmental Systems Microbiology
Computational Science and Engineering Division
High Performance Computing
ACCOUNTABILITY | CONTACT US | LEGAL & PRIVACY INFORMATION
©2008 Georgia Institute of Technology :: Atlanta, Georgia 30332


Horizons
Research
A Publication of the Georgia Institute of Technology
Summer 2008
SYSTEMS BIOLOGY:
SOLVING GRAND CHALLENGES
Georgia Institute of Technology
Tongue Drive
Fuel Efficiency
Reducing Noise
Worm Sorter
Fate of Nanomaterials

History

Task Force 2004: Study Feasibility of an Entity like “The Institute for Computational and Systems Biology at Georgia Tech”

(Mark Borodovsky, Steve Harvey, John McDonald, Larry McIntire, Konstantin Mischaikow, Francois Sainfort, Anderson Smith, Eberhard Voit, Loren Williams)

Planning Committee 2005: What should such an entity look like?

(Steve Harvey, Richard Lipton, Eberhard Voit)

Implementation Committee 2006: Hammer out the specifics

(Richard Fujimoto, Jeff Skolnick, Eberhard Voit)

Integrative BioSystems Institute

Phase 1 Activities

- Send surveys to faculty for input
- Had several meetings
- Talked with key individuals
- Explored existing strengths
- Identified areas of need
- Explored models for integration and leadership
- Summarized thoughts in preliminary vision statement
- Discussed vision statement with key individuals
- Discussed implementation ideas with Deans, Science Chairs
- Initiated chalk-talk series (open to all interested parties)
- Submitted white paper to Deans (April 2006)

Integrative BioSystems Institute

Vision

Georgia Tech will be a nationally and internationally recognized, leading institution for research and study of integrative biological systems.

This vision will require and be accomplished by:

- institution-wide participation and support
- leveraging existing strengths in biology, engineering, computation, the natural sciences and mathematics
- creation of a physical (not only virtual) institute supporting systems research in biology
- strategic hiring in relevant areas

Integrative BioSystems Institute

Goals

1. **Create** a nationally and internationally recognized **research environment** dedicated to the investigation of integrative biological systems.

2. **Leverage and consolidate existing** biological, analytical, computational and engineering **strengths at Georgia Tech** to form a synergistic, multidisciplinary task force for the analysis and manipulation of complex biomedical phenomena.

3. **Develop graduate programs** for the study of integrative biological systems.

Integrative BioSystems Institute

Rationale: Why? Why Now?

Motive: Molecular biology needs and facilitates systems-level *reconstructionist* approaches complementing traditional *reductionist* research.

Opportunity: Biology, engineering and computer science have developed technologies and methods with unprecedented potential, including:

- high-throughput genome, proteome, metabolic analysis
- data mining
- large scale simulation
- optimization techniques
- pattern recognition
- visualization
- *in vivo* sensing and imaging
- miniaturization, nanotechnology
- microfluidics, labs on chips
- synthetic biology

These are areas where GT is really strong!

Integrative BioSystems Institute

Approaches within Biology

Data Collection: Observe, measure, create data:
hypothesis driven (*e.g.*, enzyme kinetics)
high-density (*e.g.*, microarrays)
From observation to quantitative data

Bioinformatics: Organize high-throughput data
"Data mining"
Extract patterns
From quantitative data to information

Biological Systems Analysis:
Make sense of data and patterns
Integrate information into predictive models
From information to understanding

Synthetic Biology: Reconstruct biology from bottom up
From understanding to innovation

Integrative Systems Biology

From observation to quantitative data

From quantitative data to information

From information to understanding

From understanding to innovation

Integrative BioSystems Institute

Question of "Value Added" Important:

What's in it for GT?
What's in it for each stakeholder?

Prospect of solving grand-challenge problems in biology and medicine

Solutions require complex, flexible teams of multi- and interdisciplinary scientists.

Integrative BioSystems Institute

Guiding Themes:

Multi-level assessments

Development of molecular inventories

Development of enabling technologies for systemic approaches

Understand ➡ manipulate ➡ engineer systems ➡ create synthetic pathway systems

Integrative BioSystems Institute

Begin with Three Streams:

Development of normal cells toward cancer

Health related microbial systems in the environment

High-performance computing in biology and medicine

Integrative BioSystems Institute

Initial Themes within IBSI

Multi-Level Assessments of Carcinogenesis

Multi-Level Assessments of Health-Related Environmental Systems

Computational Systems Biology

Modeling & Simulation
Mathematical Analysis
HPC

Experimental Systems Biology

Molecular Inventories
Genomics
Proteomics

Technological Systems Biology

Devices
Sensing & Probing
Miniaturization

Enabling Technologies

Integrative BioSystems Institute

Implementation has Begun!

Personnel:

Executive Committee Named (Voit, Skolnick, Fujimoto)

Advisory Board Named (high-ranking GT faculty)

Rotating Directorship (Voit: 2008-2009; Skolnick 2010-2011)

Recruitment and Hiring: Through Schools; but other entities part of recruitment process

Invitation to suggest potential senior hires to provost

Integrative BioSystems Institute

Implementation has Begun!

Research: -

Funds for seed projects (at least two units to be involved)

Funds for shared graduate students (at least two faculty from different units to be involved)

Funds for distinguished seminar series

Integrative BioSystems Institute

Implementation has Begun!

Space:

New BioSystems Building

Building is top priority in GT's capital campaign and planning

Swing space being negotiated

Integrative BioSystems Institute

Implementation has Begun!

Activities:

Chalk-talk series

Entered capital campaign as line item

Committees formed:

- Advisory Board
- Shared graduate students
- Conference organizing committee

Website: www.ibsi.gatech.edu

Integrative BioSystems Institute

Implementation has Begun!

Events:

Brainstorming Workshop at Chateau Élan (October 16-17, 2006)

Poster Workshop (March 14, 2007) (featuring, *e.g.*, Hang Lu)

Announcement Celebration: February 1, 2008

International Launch Conference: October 18-21, 2008

- Roster of high-powered speakers
- ~ 150 attendees
- Poster session
- Venue: Georgian Terrace Hotel

Distinguished Seminar Series

Integrative BioSystems Institute

Implementation has Begun!

Work in Progress

Develop graduate and postdoc programs

Training grants

Planning for IBSI building

Develop strategic partnerships
- Emory, UGA, ISB, …
- Industry

Eminent Visiting Scholar program (partially GRA funded)

Work toward self-sustainability

Integrative BioSystems Institute

Wish List

Funds for:

-

IBSI building

Named directorship, faculty chairs, early career professorships

Research development awards

Named postdoc and graduate student awards

Seed grant program

Visiting scholar program

Institutional membership program

Integrative BioSystems Institute

This is a great opportunity and there is a lot of buzz

… all the way from the President and Provost

… to individual faculty all across campus

Integrative
BioSystems
Institute

Together, let's integrate the buzz!

$\int$

Integrative BioSystems Institute

www.ibsi.gatech.edu