

Constraint-based Modeling: Part II

LP, Lessons Learned, and the Growing Field of CBM

Tuesday, May 4, 2004

Timothy E. Allen / Bernhard Ø. Palsson

BE 203 Lecture

Outline

- Constraints in biology
- Reconstructions and applying constraints
- Constraint-based modeling (CBM):
philosophy and overview
- Basics of flux balance analysis (FBA)
- **Lessons learned**
- **CBM: an expanding field**

Lessons Learned:
Applications to Genome-scale in silico
Reconstructions

Lessons Learned:

Applications to Genome-scale in silico Reconstructions

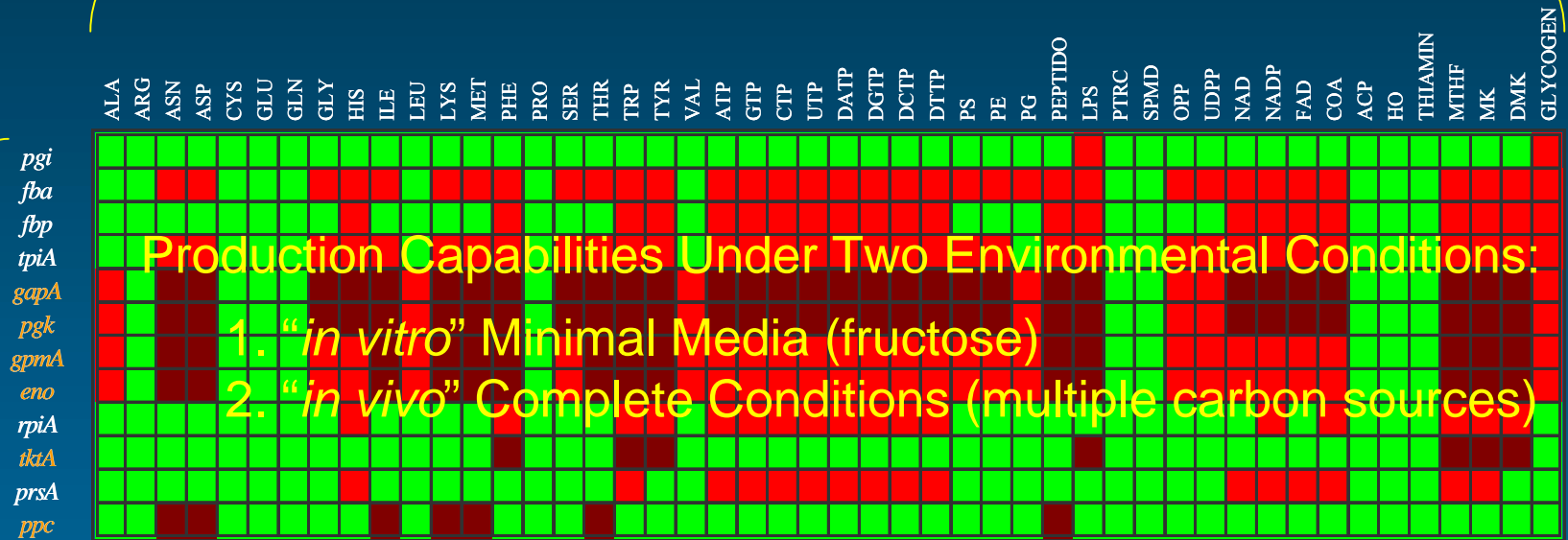
1. Deletion studies (*H. influenzae*)
2. Essential amino acids (*H. pylori*)
3. Reaction subsets / operons (*E. coli*)
4. Gap analysis (*E. coli*)
5. Optimal growth predictions / adaptive evolution (*E. coli*)
6. Iterative hypothesis generation (*E. coli*)
7. Integration of heterogeneous datasets (*E. coli*)

Example #1: Gene Deletions & Production Deficiencies

H. Influenzae Central Metabolism

50 Biomass Requirements

Genes of Central Metabolism

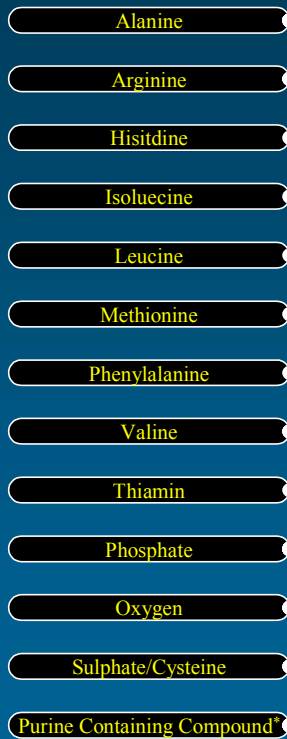


Minimal Substrate
Conditions
(fructose)

Carbon-supplemented
Conditions
(fructose, glucose, glycerol, galactose,
fucose, ribose, and sialic acid)

Example #2: *H. Pylori* Minimal Requirements

Helicobacter Pylori Minimal Substrate Requirements



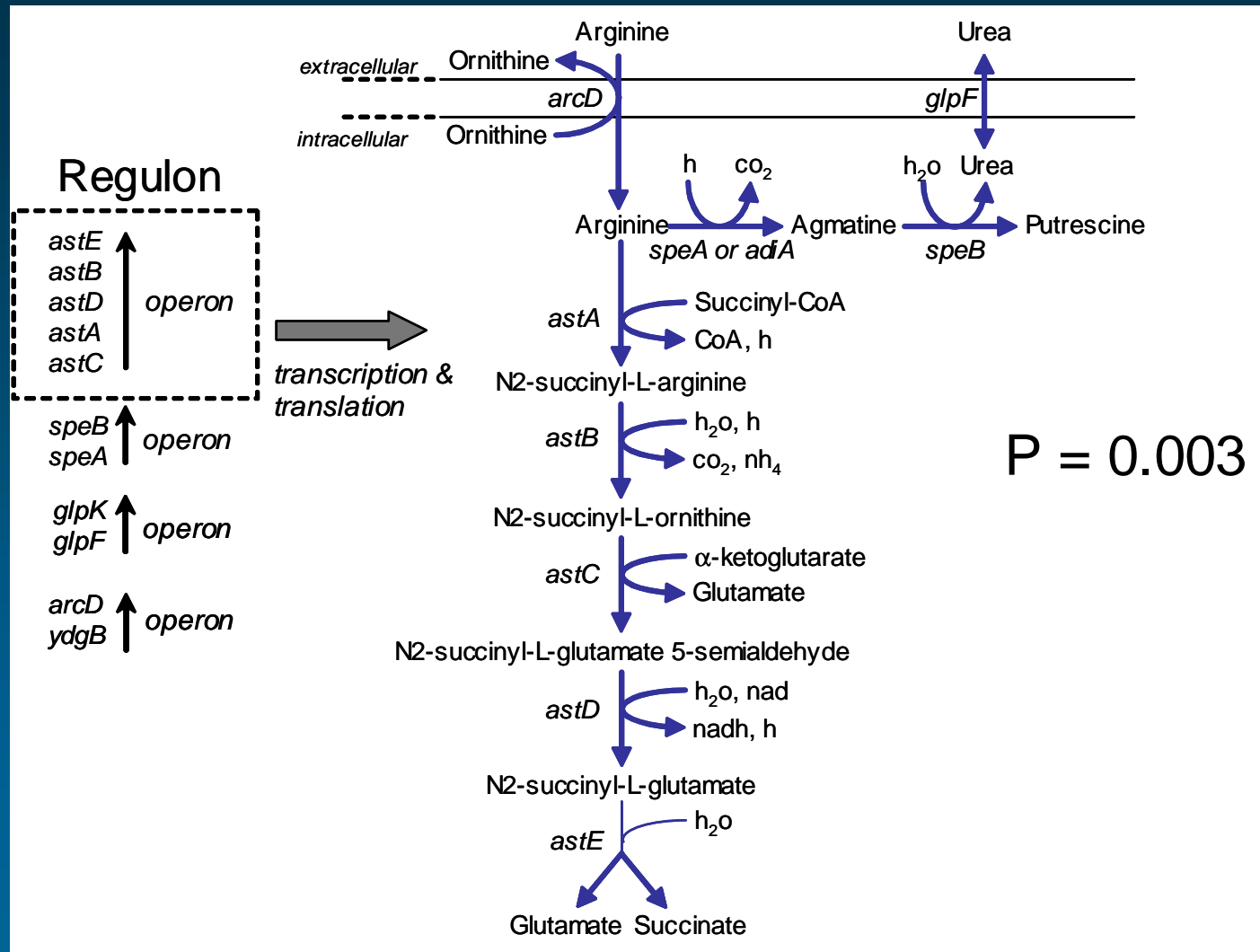
Biomass Requirements



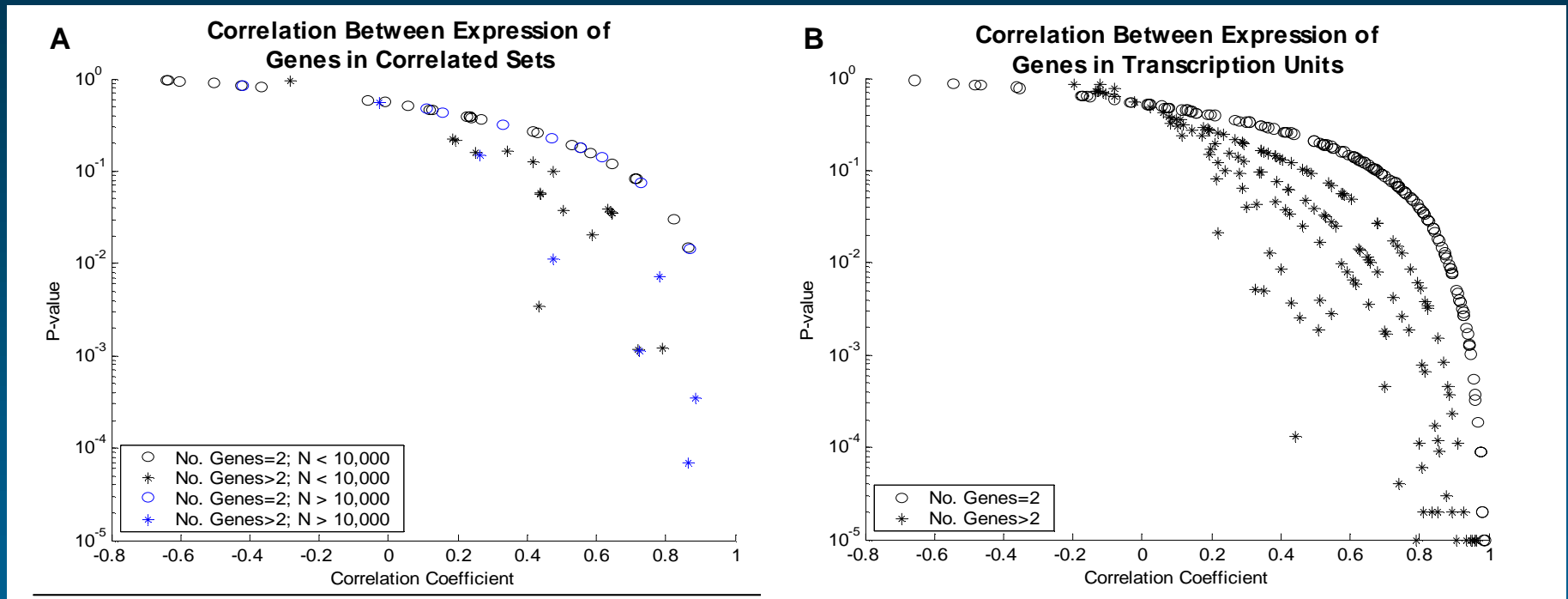
* Adenine, Adenosine, Guanine,
Guanosine, Hypoxanthine

- 8 amino acids required
- purine sources
 - Adenine
 - Adenosine
 - Guanine
 - Guanosine
 - Hypoxanthine
- sulfur source
 - Cysteine
 - Sulphate
- oxygen
 - no substrate level
 - phosphorylation (lacks PYK)
- glutamate
 - Requires alanine or arginine,
only component not
dependent on one substrate

Example #3: Reaction Subsets (*E. coli*)

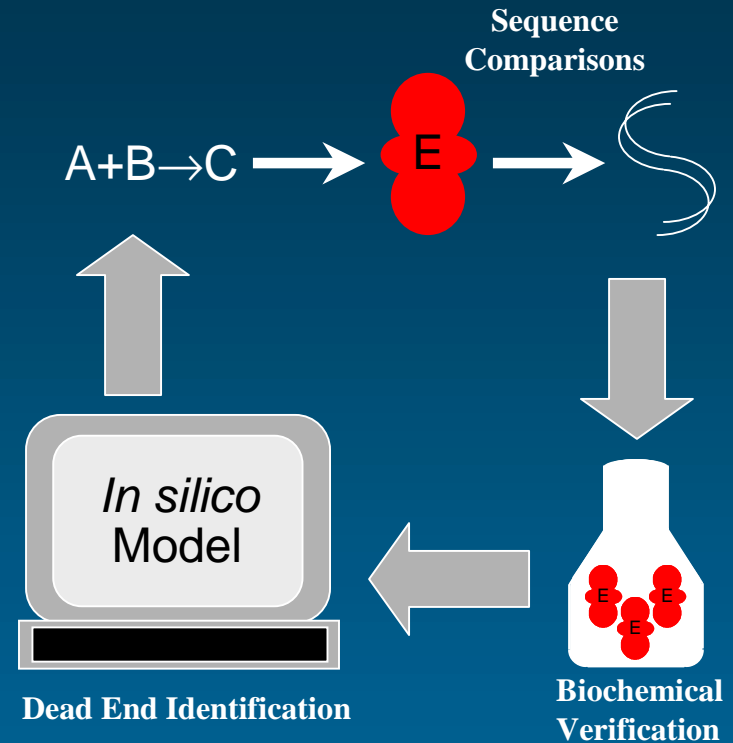
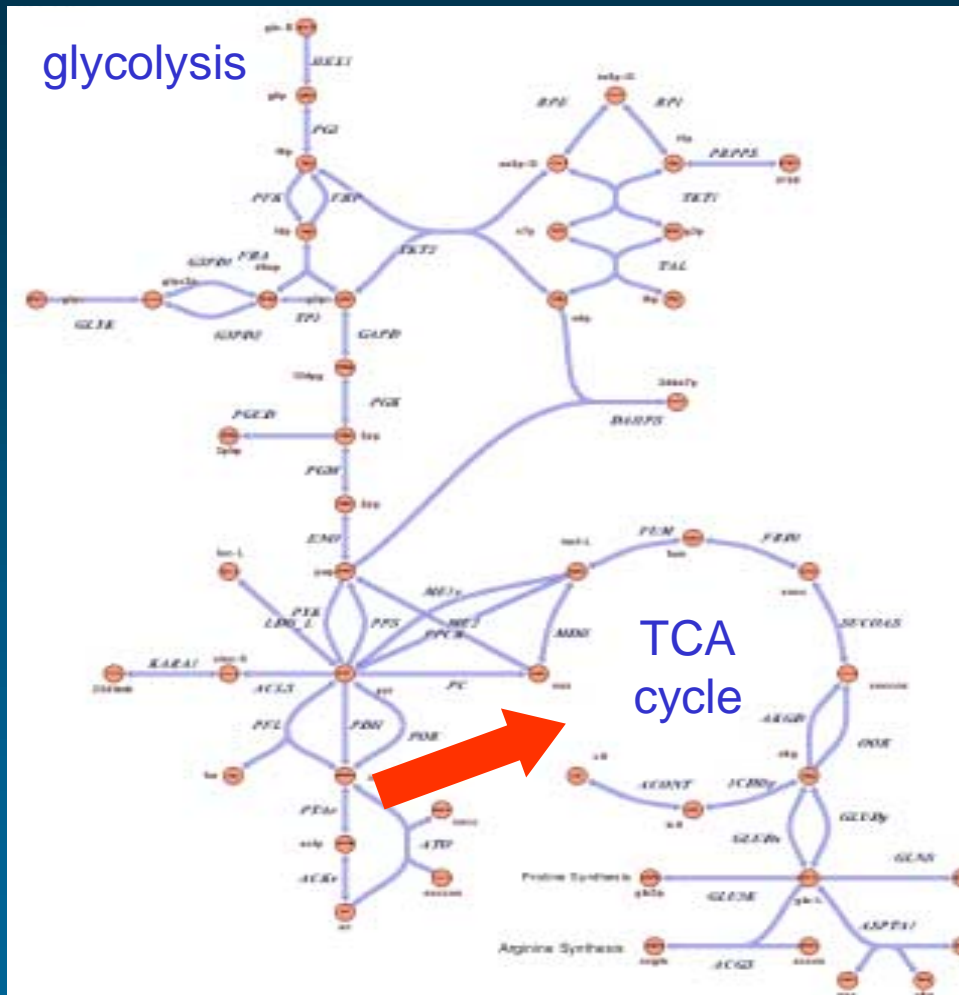


Correlated Sets / Operons (*E. coli*)



Correlation of Genes in Correlated Sets and Operons Using Expression Data for *E. coli*

Example #4: Network Gap Analysis



Models can be used to guide biological discovery.

55 Putative Annotations

Bnum	EC number	Published Annotation [Serres et al.]	Suggested Annotation
b3718	3.1.1.17	putative isomerase	gluconolactonase
b2160	2.7.1.13	putative sugar kinase	dehydrogluconokinase
b2166	2.7.1.14	putative sugar kinase	sedoheptulokinase
b2661	1.2.1.19; 1.2.1.24	succinate-semialdehyde dehydrogenase I, NADP-dependent	aminobutyraldehyde dehydrogenase; succinate-semialdehyde
b4266	1.1.1.6	5-keto-D-gluconate-5-reductase	glycerol dehydrogenase.
b3003	1.1.1.6	putative oxidoreductase, NAD(P)-binding	glycerol dehydrogenase. acetoin dehydrogenase, Diacetyl reductase
b2137	1.1.1.5	putative oxidoreductase	
b2615	2.7.1.23	ORF	NAD+ kinase
b3718	3.1.1.31	putative isomerase	6-phosphogluconolactonase (Pgl)
b1511	2.7.1.47	putative sugar kinase (2nd module)	D-ribulokinase
b1524	3.5.1.2	putative glutaminase	glutaminase A,B

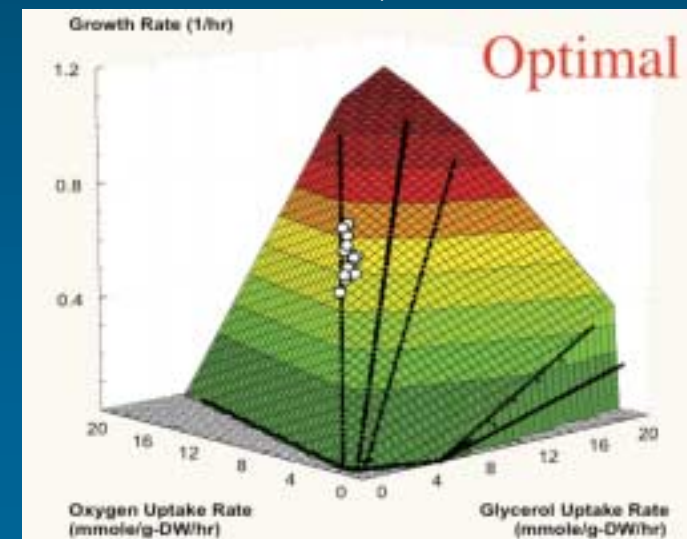
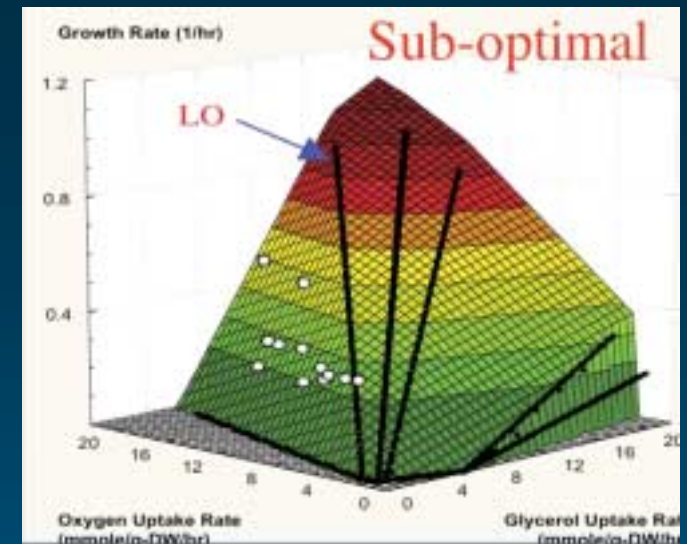
Enzymes acting on network gaps

Suggest alternate substrates for enzymes
Multiple hits for target enzymes

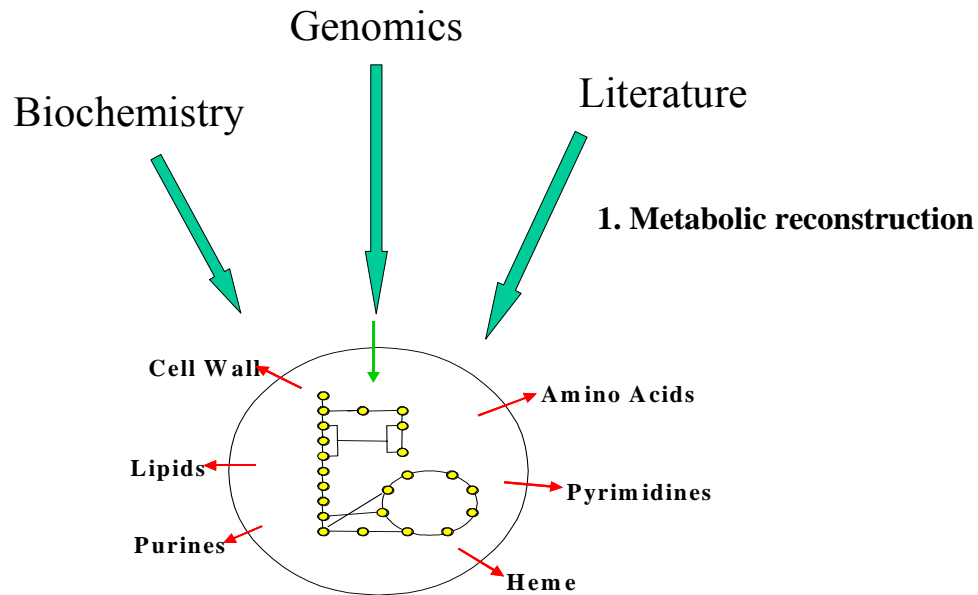
Enzymes in *E. coli* without locus assignments (EcoCyc)

Metabolic model makes growth predictions for knock-out strains (86%).
Regulated metabolic model increases accuracy of predictions (91.4%).

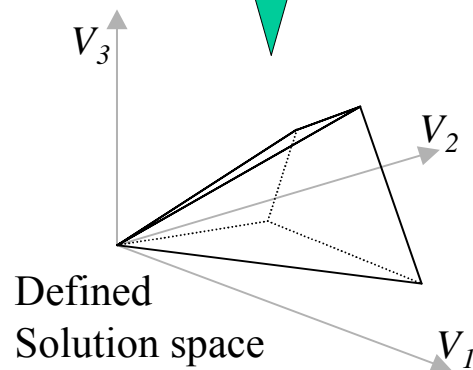
Example #5: Predicting complex biology; adaptive evolution and picking optimal growth states



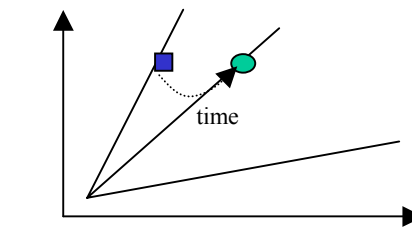
Using Adaptive Evolution



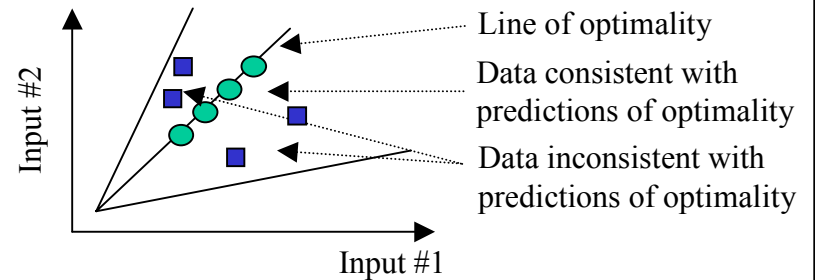
2. Constraints imposed on the metabolic network



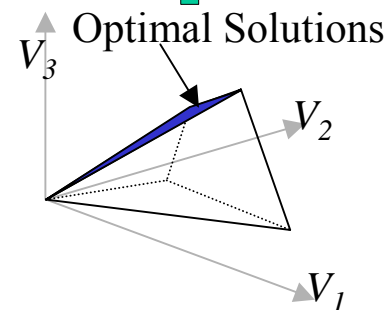
3. Linear optimization to Find 'best solution



5. Whole Genome Directed Evolution

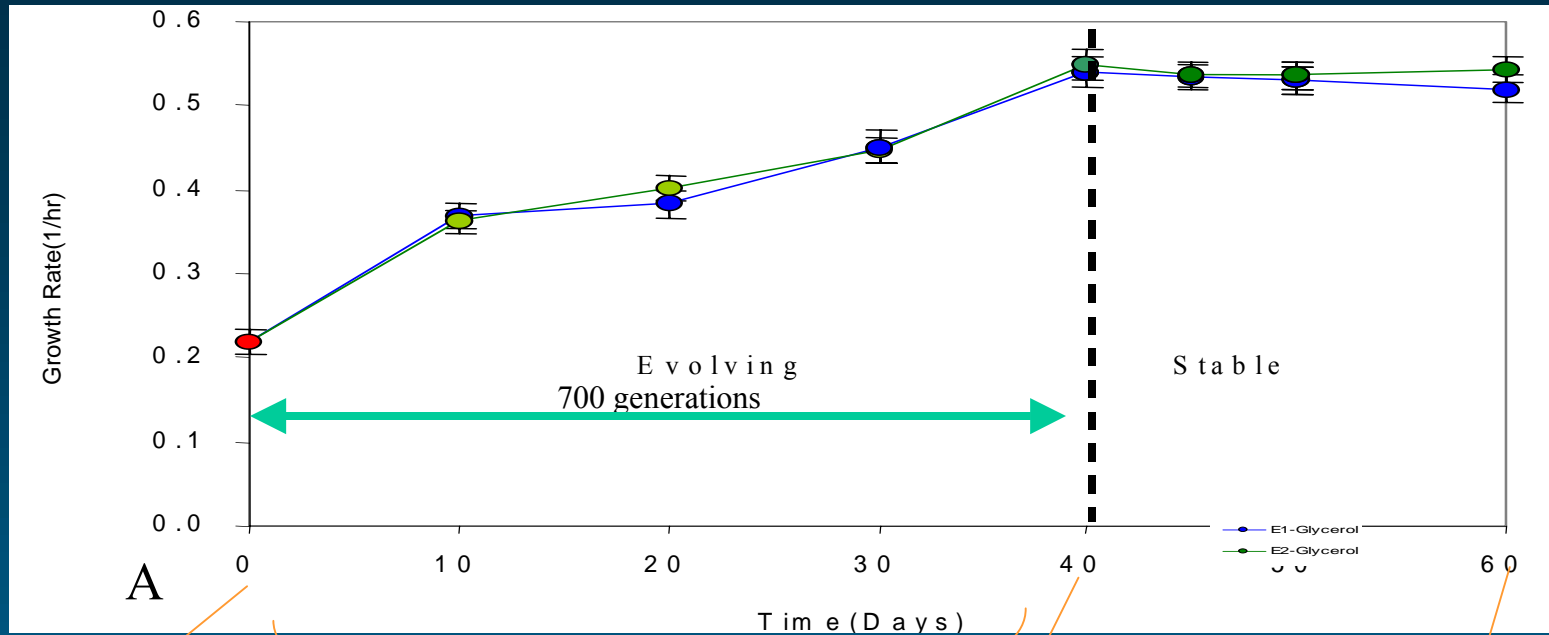


4. Phenotype phase plane analysis: Displays all optimal solutions simultaneously



Evolving Growth Rate on Glycerol

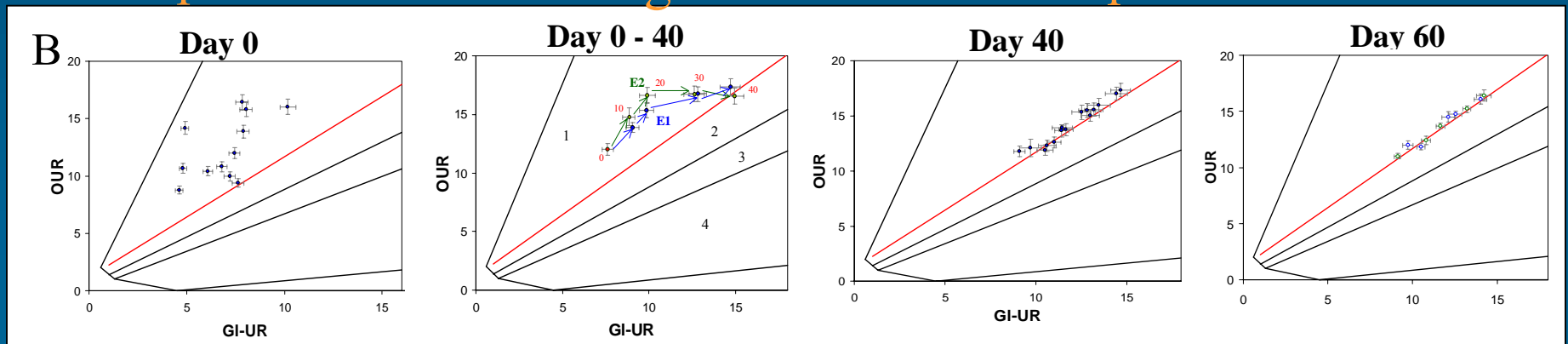
Growth rate



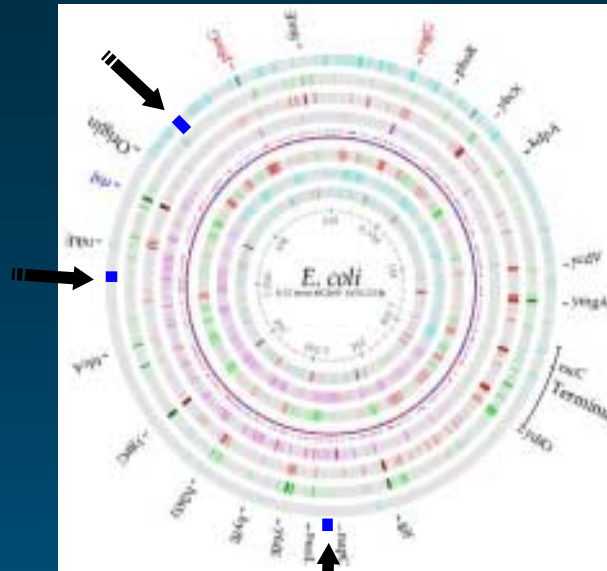
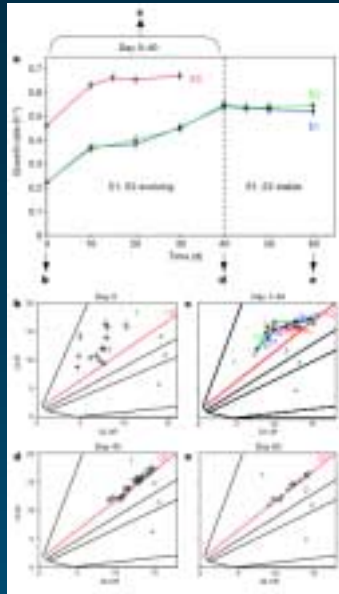
pre

during

post



Genome location of glycerol metabolic and regulatory genes selected for re-sequencing



Fructose 1,6 bisphosphate (FBP) binding site on *glpK*

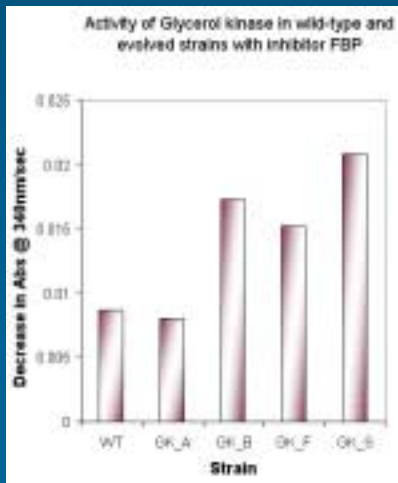
FBP binding loop (230-236) : IGGKGGTR

Mutation in FBP binding site

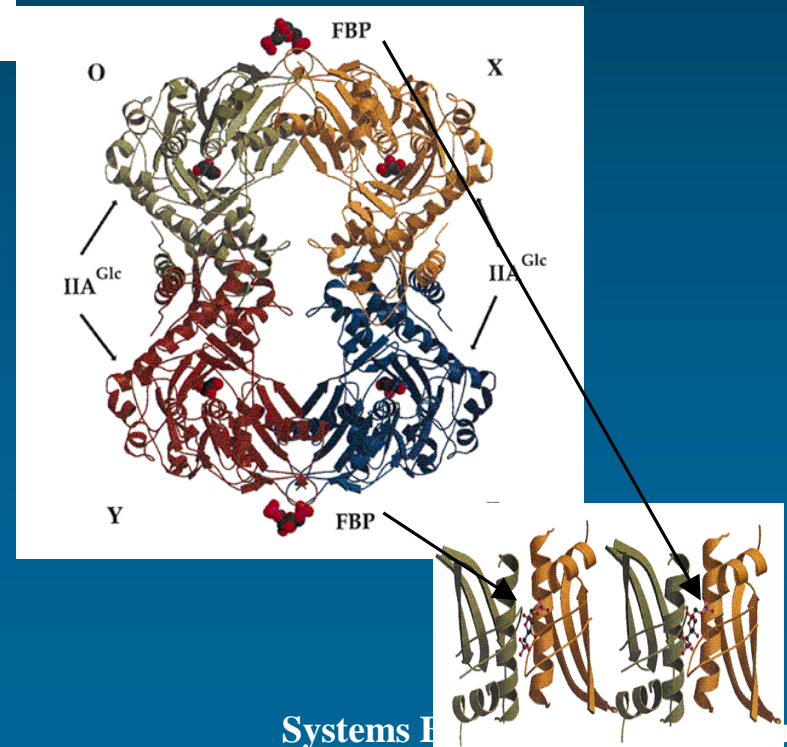
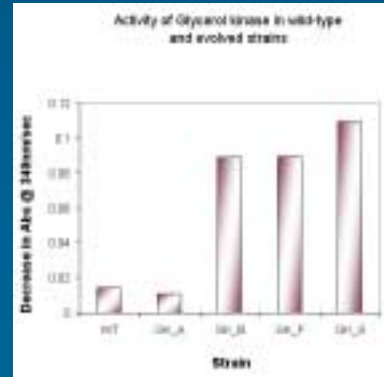
(*ggc* → *gac*)

		<i>ggc</i>	
wt	SEVYGTNIC	GKGGTRIPIS	
		<i>gac</i>	
mut	SEVYGTNID	GKGGTRIPIS	
AA #	-----		
		230	236

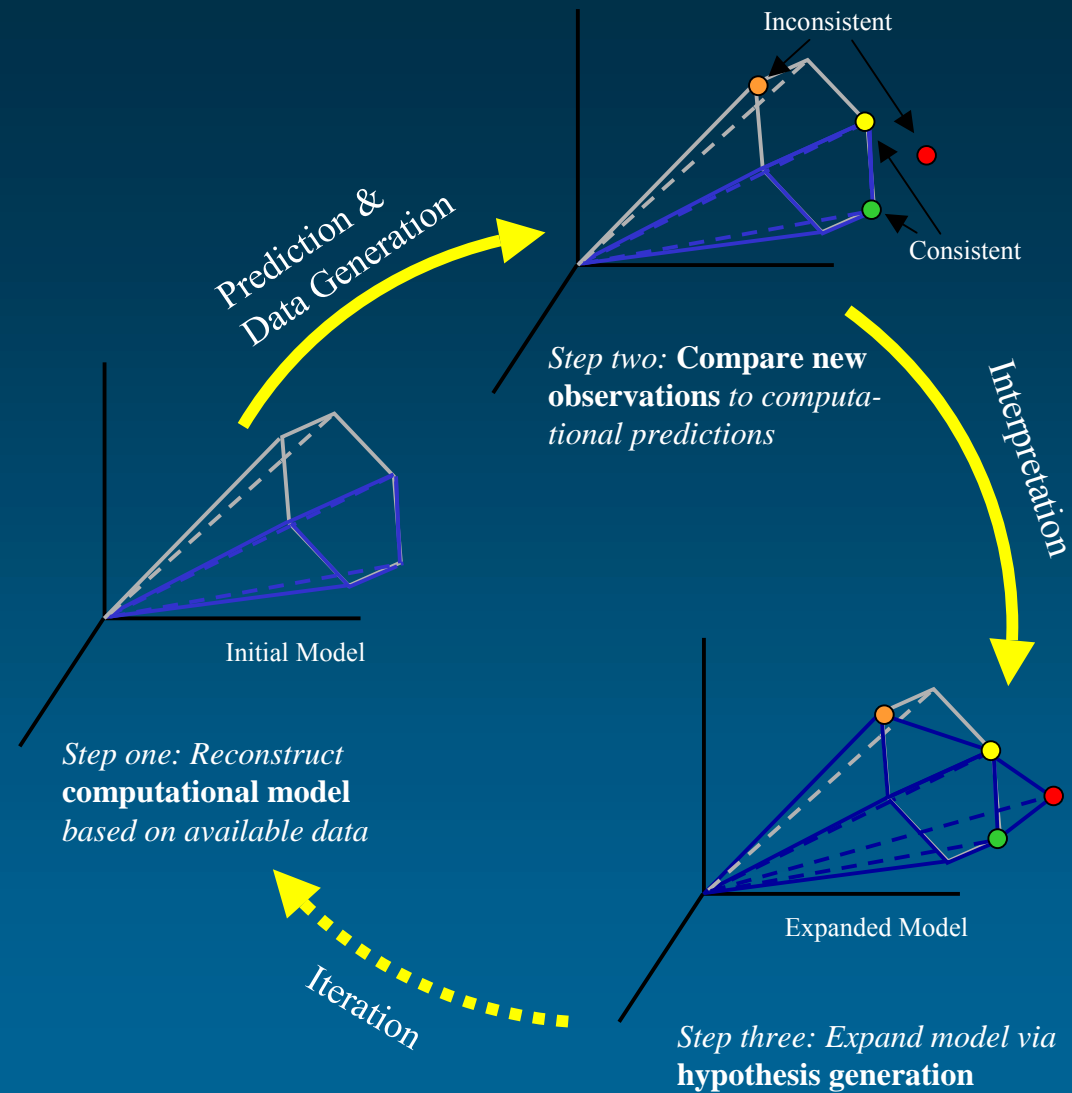
2-3 fold decrease in inhibition by 2mM FBP on activity of *glpK* with mutation



~ 10 fold increase in activity of *glpK* with mutation (G231D: GLY → ASP)

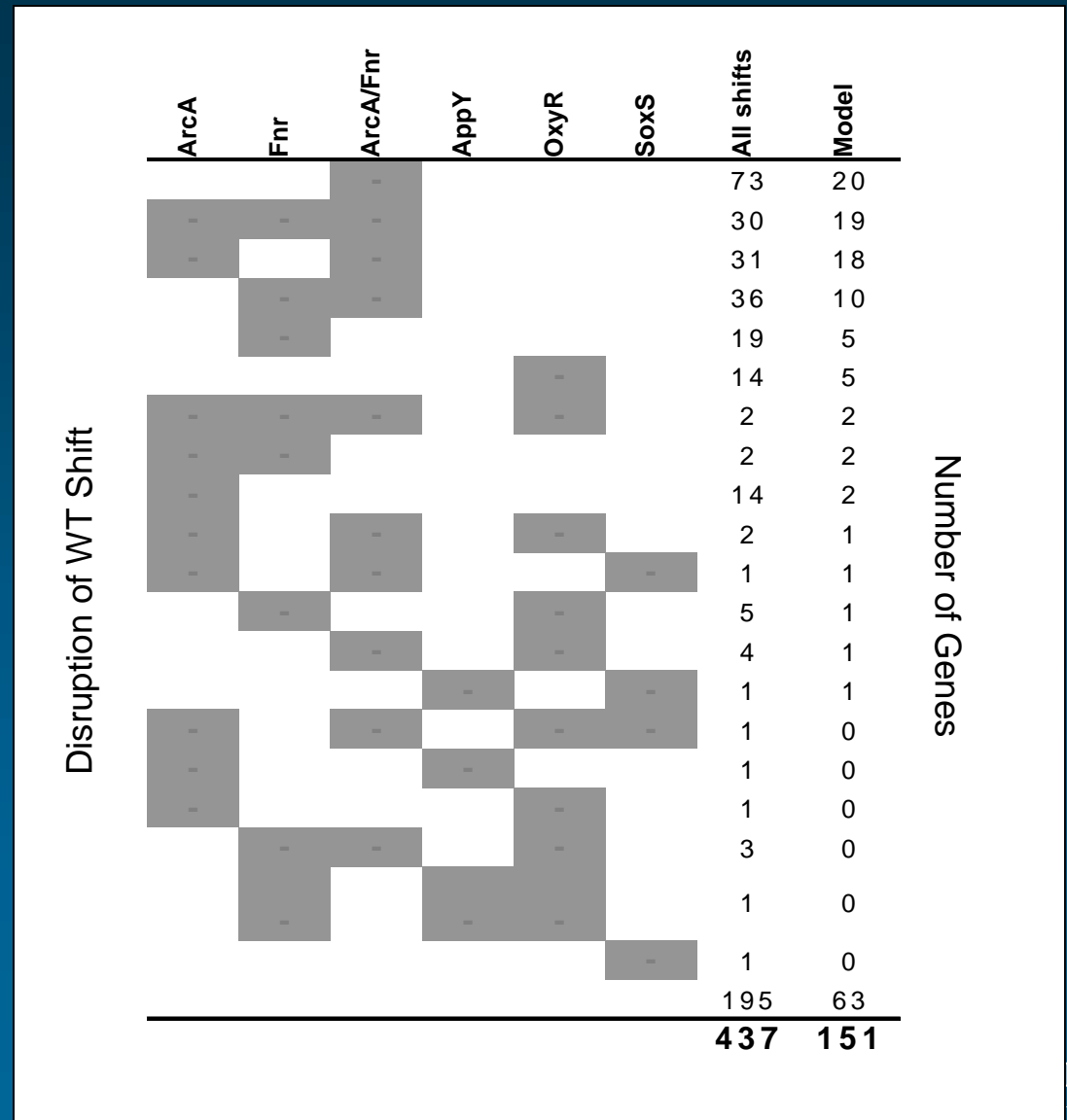


Example #6: Hypothesis generation: transcriptional regulation in *E. coli*



Model-Centric Hypothesis Generation

- **Genome-scale regulatory/metabolic model of *E. coli***
 - 1,008 genes
- **Systematic network perturbation analysis**
 - ArcA, Fnr, ArcA/Fnr, AppY, OxyR, SoxS
- **Generate new rules for model**
- **Hypotheses generation**



Model-driven hypothesis generation

Gene Expression Study

- Added new rules for 78 genes
- Removed old rules for 27 genes
- Changed old rules for 10 genes
- Total of 115 changes in regulatory rules

Prediction & Data Generation

Step two: Compare new observations to computational predictions

Interpretation

iMC1010^{v1}

- Phenotypic Predictions
 - 79% (10828/13750) accuracy
- Expression Predictions
 - 49% (23/47) accuracy
 - 15% (23/151) coverage

Step one: Reconstruct computational model based on available data

iMC1010^{v2}

- Phenotypic Predictions
 - 79% (10833/13750) accuracy
- Expression Predictions
 - 98% (100/102) accuracy
 - 66% (100/151) coverage

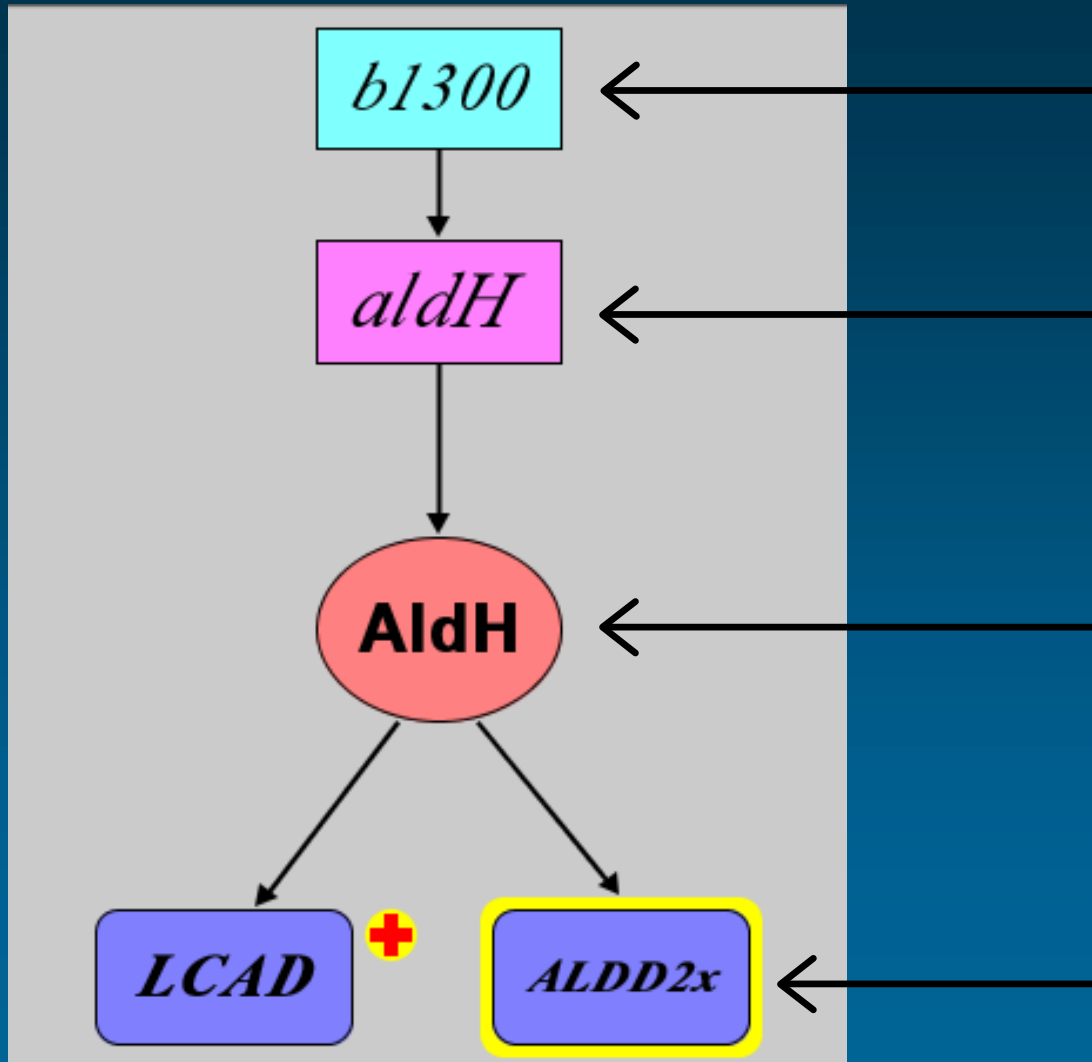
Step three: Expand model via hypothesis generation

110 new regulatory hypotheses overall

Iteration

Example #7: Integration of multiple data sets: periodicity in gene expression

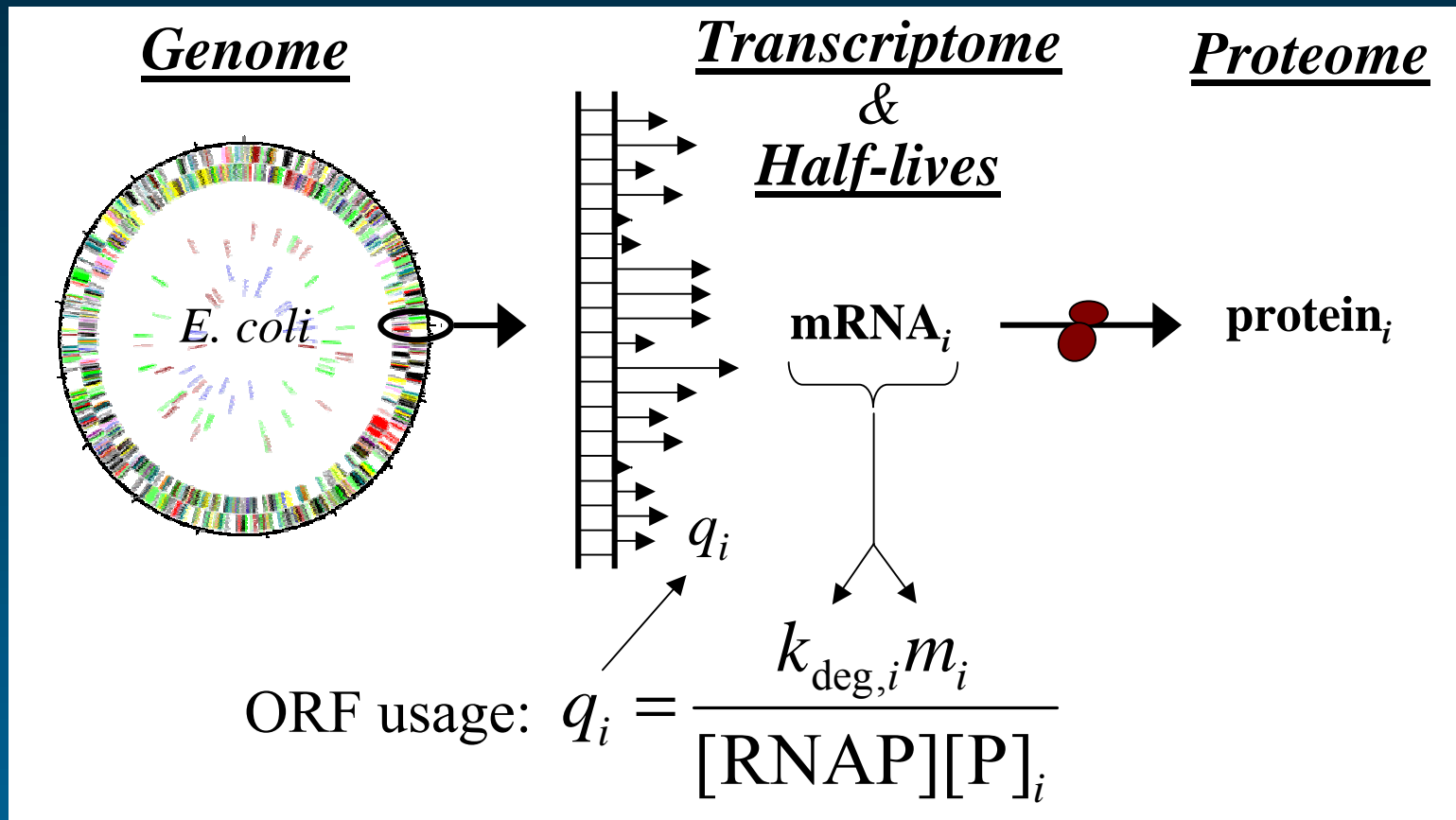
Integrating Heterogeneous Datasets



LEVELS	GPR ASSOCIATIONS
Gene	Succinate Dehydrogenase b0724 b0723 b0722 b0721
mRNA	sdhB sdhA sdhD sdhC
Protein	(b0721 and b0722 and b0723 and b0724) Sdh (Sdh) (Sdh)
Reaction	sdhA sdhD
Gene	D-Xylose ABC Transporter b3568 b3567 b3566
mRNA	xylH xylG xylF
Protein	(XylF and XylG and XylH) XylH XylG XylF
Reaction	XylH
Gene	Glyceraldehyde 3-Phosphate Dehydrogenase b1779 b1417 b1416
mRNA	gapA gapC gapC2
Protein	(b1779) (b1416 and b1417) GapA GapC
Reaction	(GapA) or (GapC) GAPD

ics

Integrating “Omics” Data

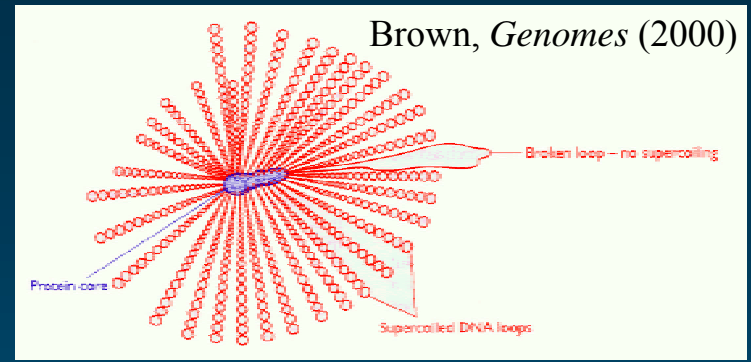


$\mathbf{q} = (q_1 \dots q_{4290})$ constitutes the “transcription state” of the genome

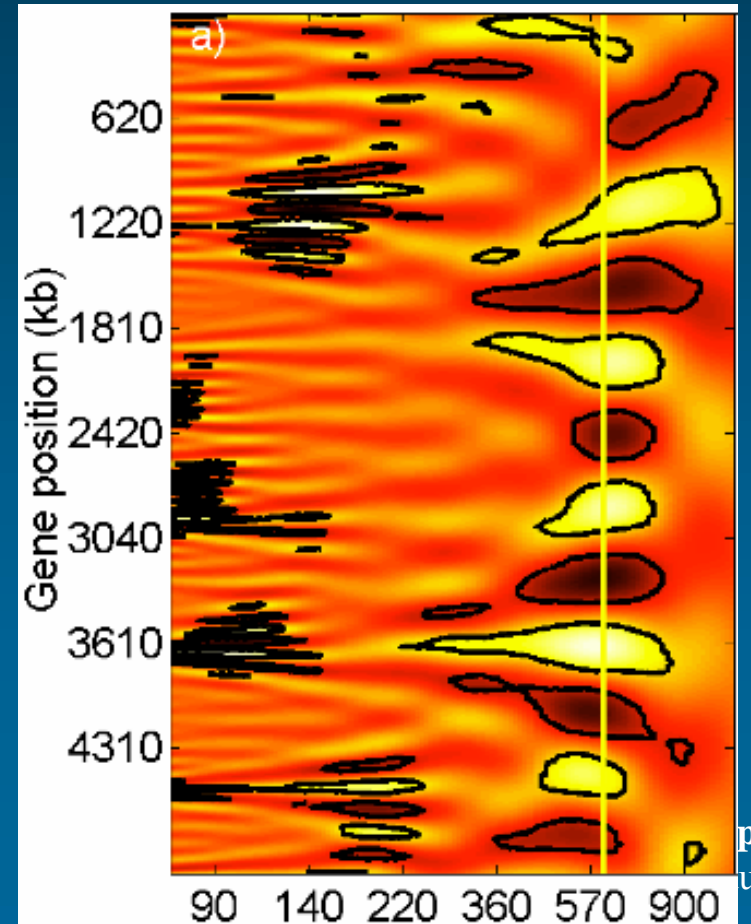
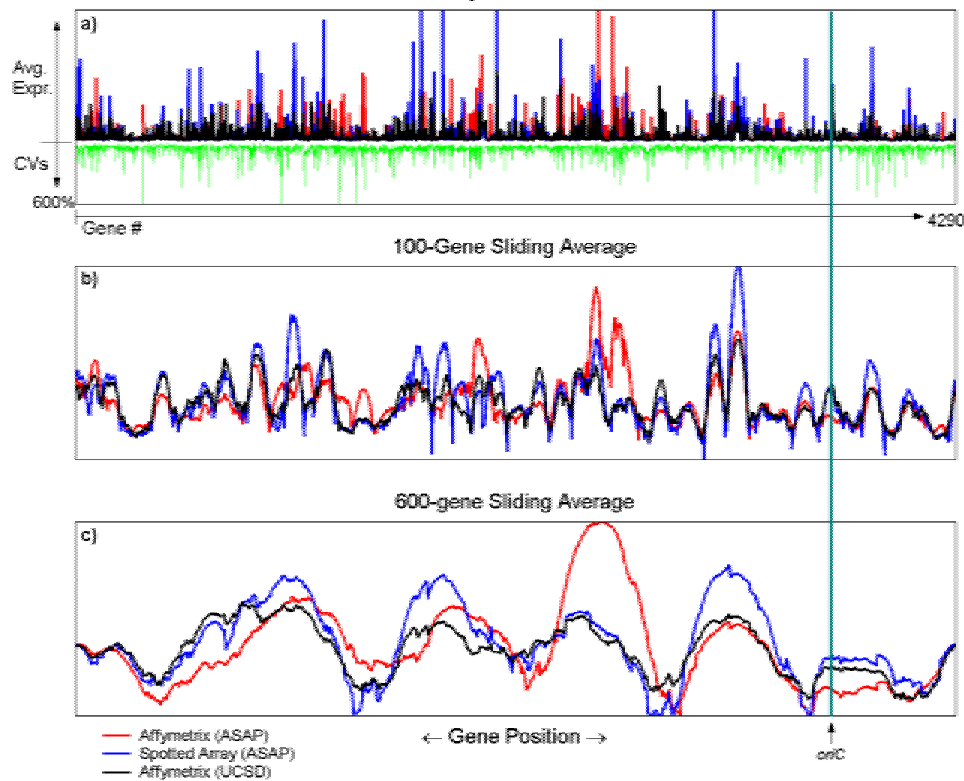
$\mathbf{t} = (t_1 \dots t_{4290})$ can be calculated on a per codon basis and account for relative tRNA abundance to give the state of the proteome

Periodicity in genome usage

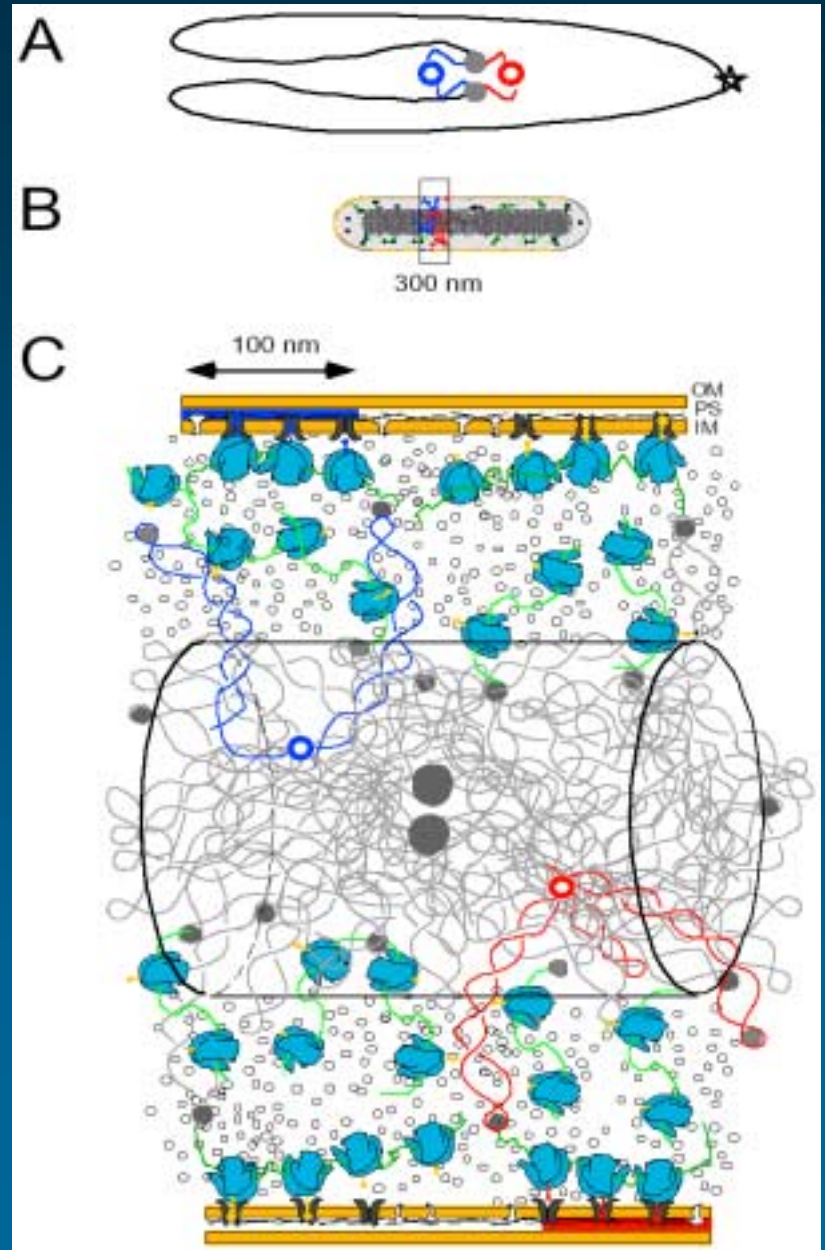
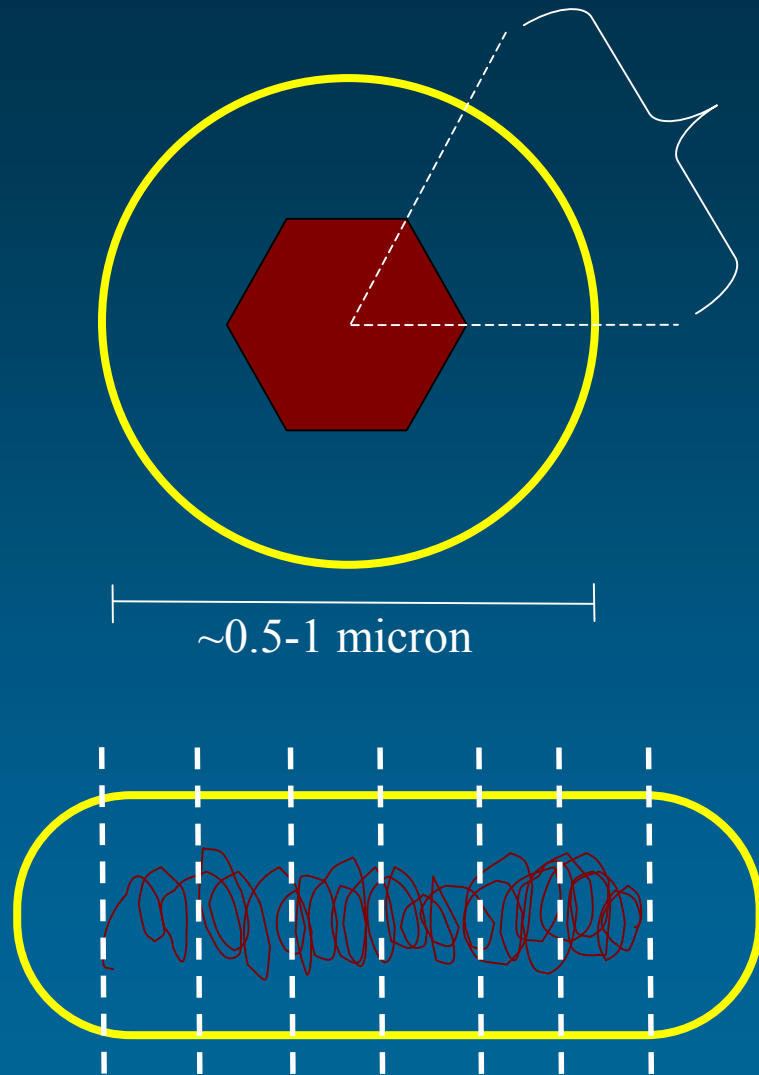
- Periodicity in *E. coli* expression of ~100 and ~600 genes
- Appear to be distinct 6 regions of genome usage



E. coli Expression Levels



Topobiology of *E. coli* Genome



Integration of “Omics” Data

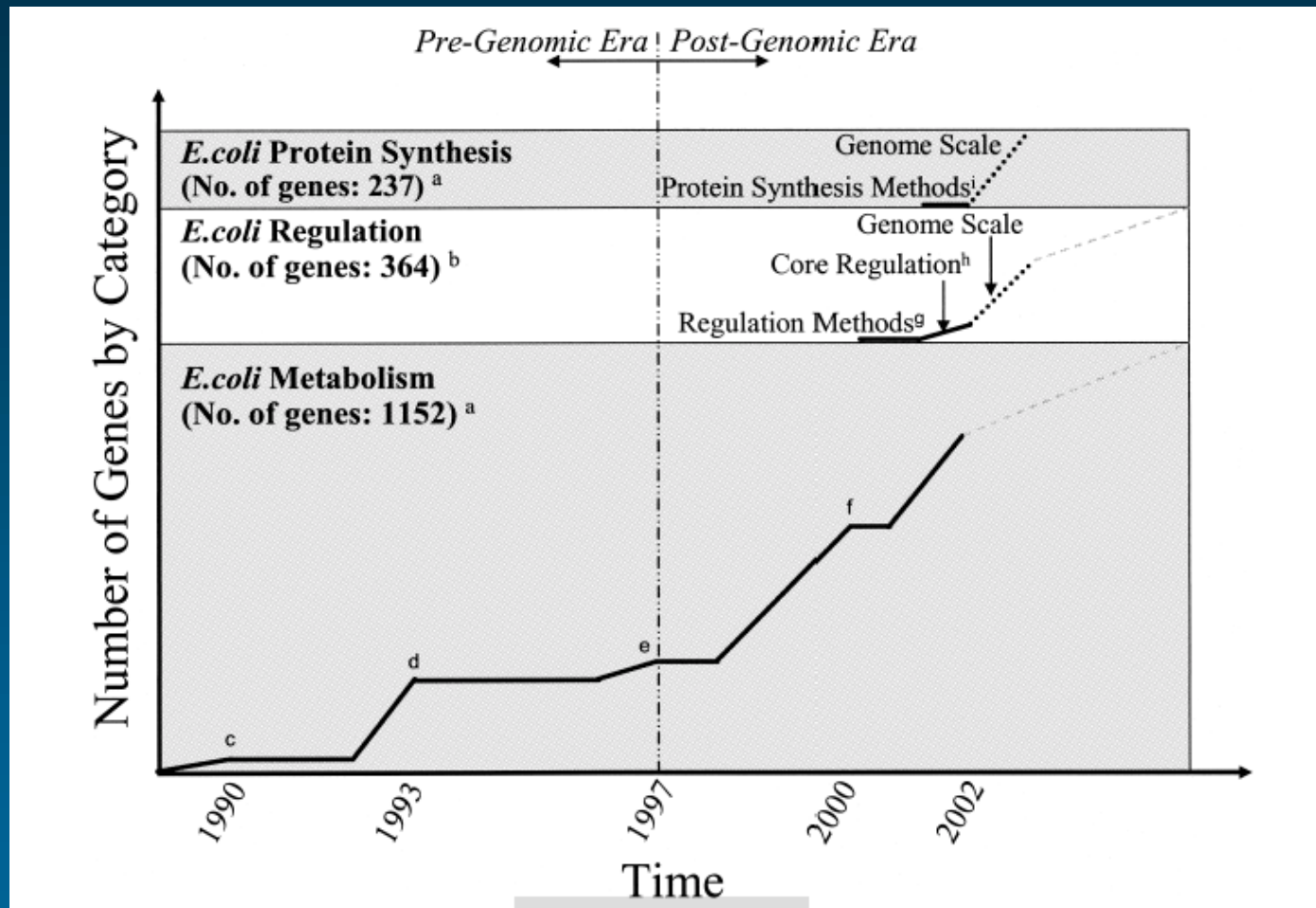
- Simultaneous analysis of multiple “-omics” data sets leads to new insights
- Topobiology at the ~ 200 nm scale seems to be important
- Means of accounting for 3D structural constraints is needed in whole-cell reconstructions going forward

Constraint-based Modeling: An Expanding Field

University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
<http://systemsbiology.ucsd.edu>

Development of the *E. coli* Model

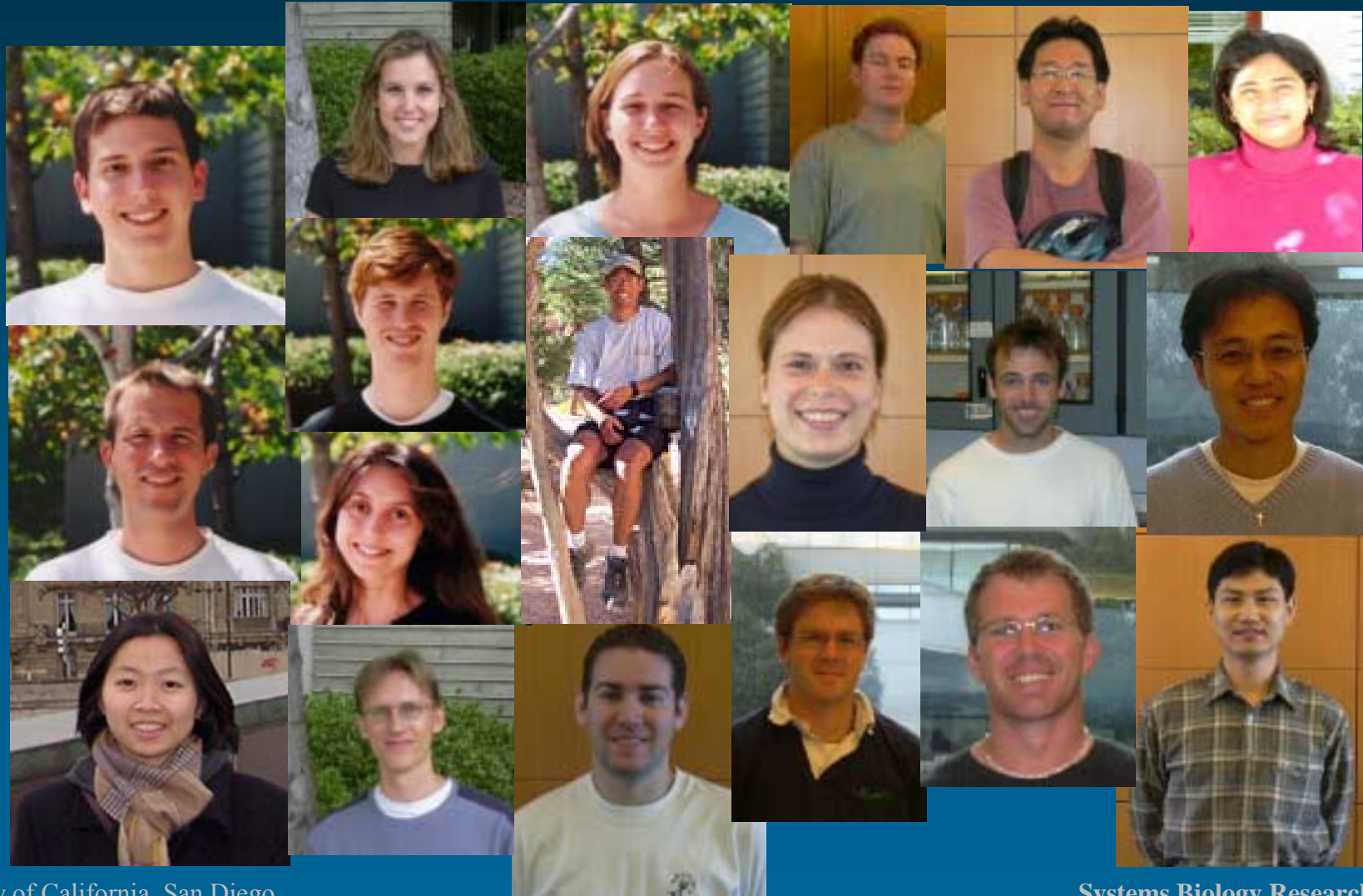


"Thirteen years of constraint-based model building of *E. coli*" *J Bacti*, May 2003

*(Several slides deleted due to
copyright issues...)*

Systems Biology Research Group

Principal Investigator: Bernhard Ø. Palsson, Ph.D.



University of California, San Diego
Department of Bioengineering

Systems Biology Research Group
<http://systemsbiology.ucsd.edu>