Systems Biology and Bioengineering

National Yang-Ming University, Taiwan

Lectures by Bernhard Palsson

University of California, San Diego
Dept. of Bioengineering

Goal

To give a brief (two-day) overview of the emerging field of systems biology and bioengineering

Schedule

- **Lecture #1** Bringing Genomes to Life:
  - The Use of *in silico* Models
- **Lecture #2** High-throughput Technology:
  - A (very) Brief Overview
- **Lecture #3** Reconstruction Methods:
  - Piecing together biochemical reaction networks
- **Lecture #4** Representing Networks Mathematically:
  - The Stoichiometric Matrix
- **Lecture #5** Extreme Pathways:
  - Basic Concepts
- **Lecture #6** Flux-balance Analysis
  - Basic Concepts
- **Lecture #7** Genome-scale Models:
  - Lessons Learned

Bringing Genomes to Life:
The Use of *in silico* Models

Bernhard Palsson

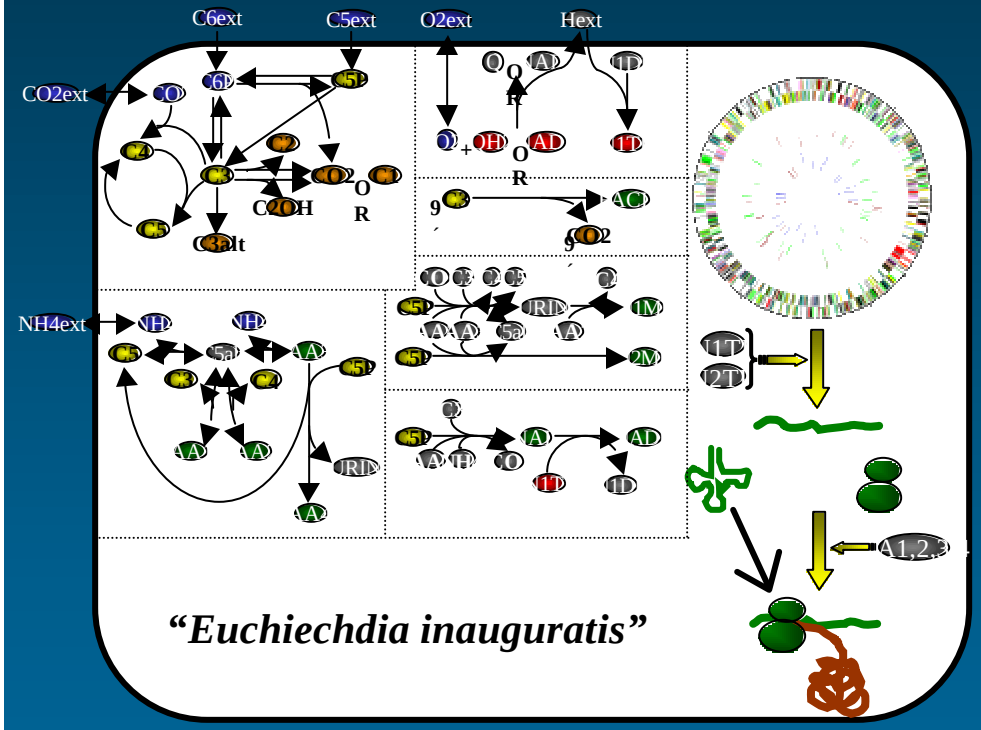
Lecture #1

September 15, 2003

University of California, San Diego
Department of Bioengineering

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

Outline



1) Introduction to Systems Biology

2) Constraint-based approach: *Analyzing complex biological systems*

3) Models driving experiments: *Genome-scale models as hypotheses*

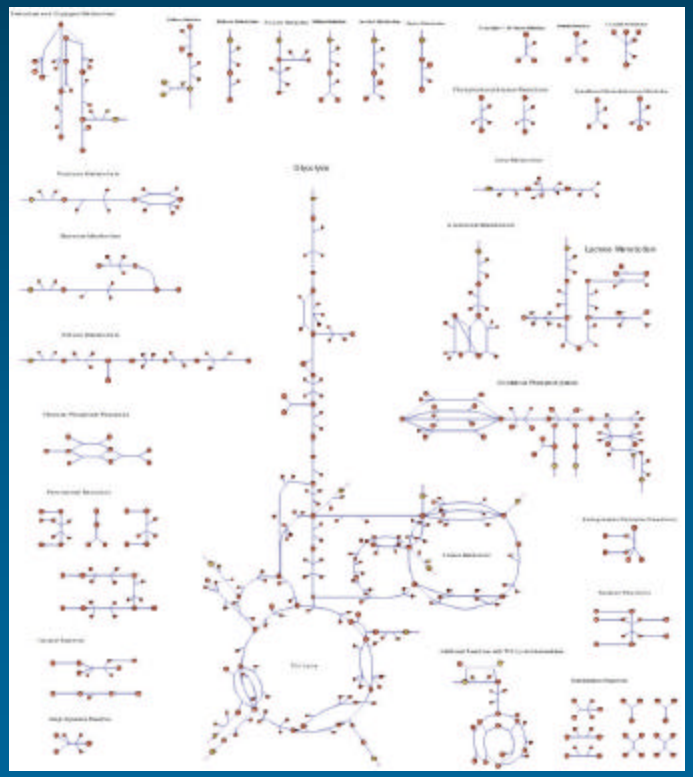
Introduction to Systems Biology

University of California, San Diego
Department of Bioengineering

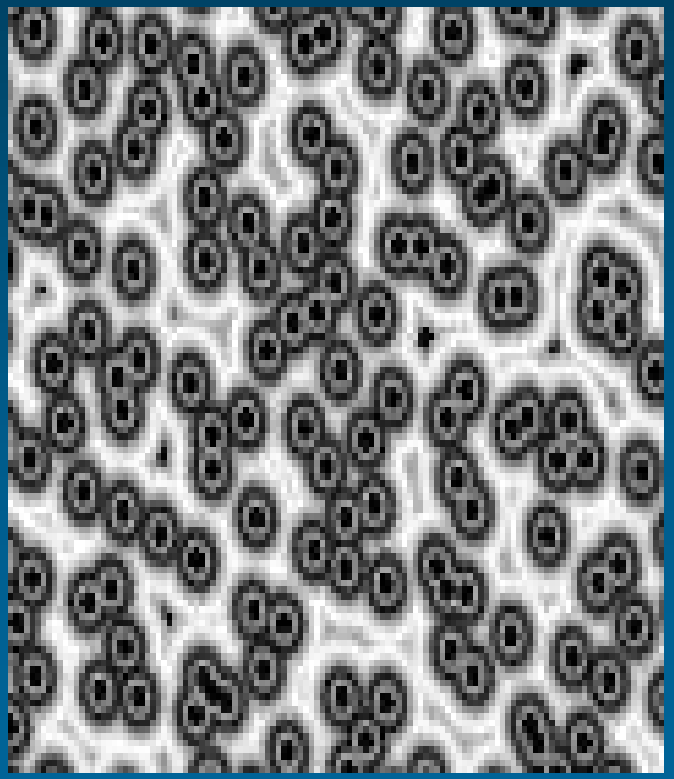
Systems Biology Research Group
http://systemsbiology.ucsd.edu

Biology is now asking:

If every molecule in a cell is replaced over time, is it still the same cell?



If every cell in an organism is replaced over time, is it still the same organism?

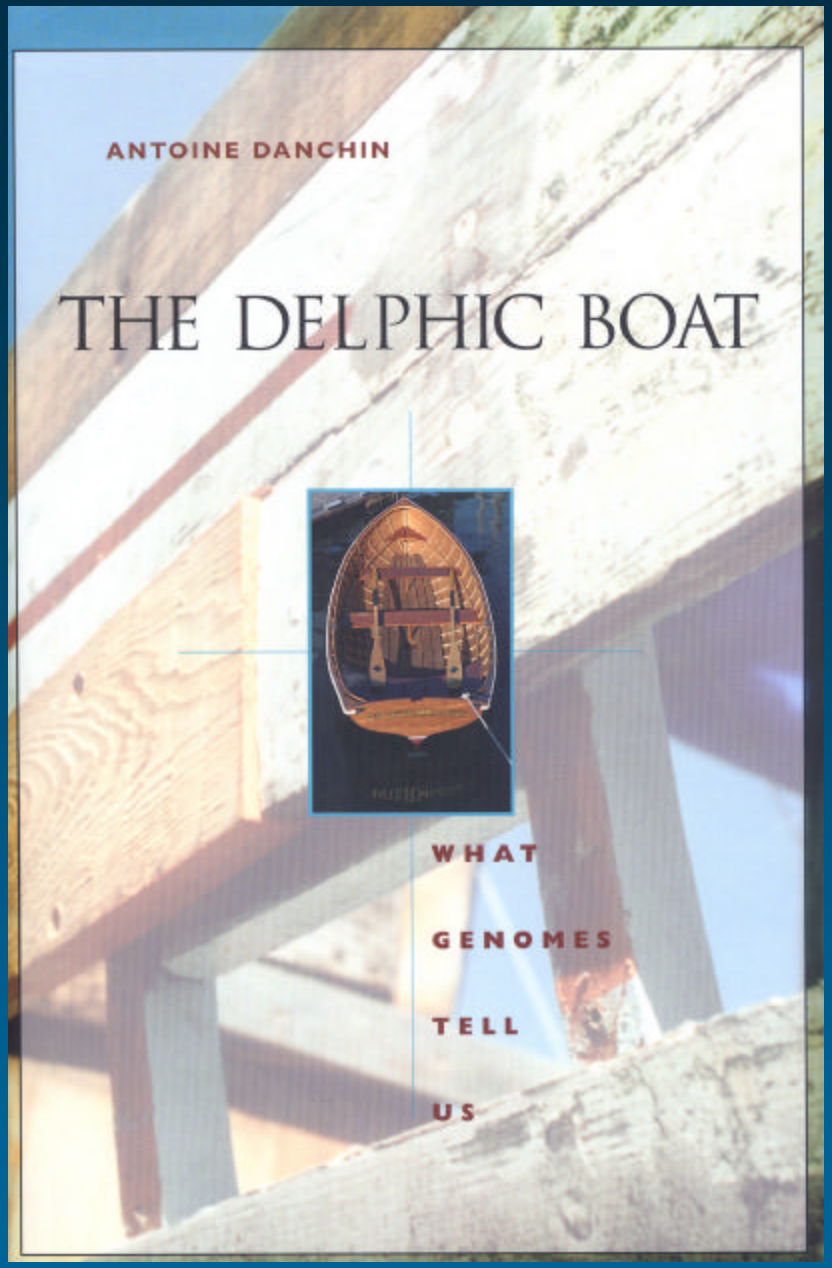


University of California, San Diego
Department of Bioengineering

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

The Oracle of Delphi asked:

If every plank in a boat is replaced over time, is it still the same boat?



University of California, San Diego
Department of Bioengineering

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

The answer basically is ‘yes’

Thus, the interconnections of biological components--the ‘blueprint,’ the ‘circuit diagrams’--of cells are taking center stage in biology:

and thus... we have the emergence of systems biology

Two key questions arise:

- What is the nature of the links between the components in a biological network?
  - Molecular biology: basic chemistry
  - Cell/tissue biology: “higher-order” chemistry
  - Structural/topological properties of networks
- What are the functional states and properties of biological network?
  - P/C nature of the intra-cellular environment
  - Spatio-temporal organization
  - Near crystalline state
  - Some Biological Network Properties:
    - Self-assembly and selection at all levels in biology
    - Definition of ‘self’ is fundamental in biology

Basic links in molecular biology

- The prototypical transformations are bi-linear:
  - X+Y <-> Z (covalent change)
  - X+Y <-> X:Y (association of molecules)
- Key properties
  - Stoichiometry fixed and constrained by elemental and charge balancing
  - Relative rates fixed by thermodynamics, but is condition dependent
  - Absolute rate highly variable and manipulable by cells
- Cells cannot just form new links at will: candidate links are constrained by the basic rules of chemistry

Absolute rates are key biological design variables

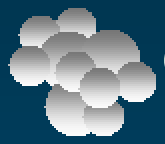
- Evolution -- by selection over time -- has discovered how “best” to perform a reaction
- The orientation of substrate molecules on a surface of an enzyme brings them into the right spatial relationship to increase (and fix) the probability of a reaction occurring
- “Similar” reactions” and evolution (zero to something)

From The Touchstone of Life,
W.R. Lowenstein,
Oxford Universiy Press, New York, 1999.

Links to networks

- Reactions link together to form a network
- The basic structural features are bi-linear and they are a ‘tangle of cycles’ representing key chemical properties
- As network size grows, the number or possible functional states grows faster than the number of components
- Multi-functionality of networks leads to multiple possible states (‘behaviors’ or ‘phenotypes’)
- Multiple possible phenotypes call for the selection of ‘states’ (‘optimal’) based on network history and survival
- Built-in mechanisms (‘regulation’) are needed to choose (‘express’) the selected states

Components View



Component

Function

S + E <–> X <–>E + P

Time-dependent Concentration

Compute Flux for Function

Systems View

Needed Homeostasis

Reaction Network

Steady State Flux Map

Calculate k

Calculate C

Kinetics determine C

C useful for systems dynamics

Some key features of biological networks

- They evolve--i.e., they are time-variant
  - Key difference from P/C sciences
  - Principally through kinetics and changing available/active links
  -
- They have a ‘sense of purpose’ (objective) which fundamentally is ‘survival’

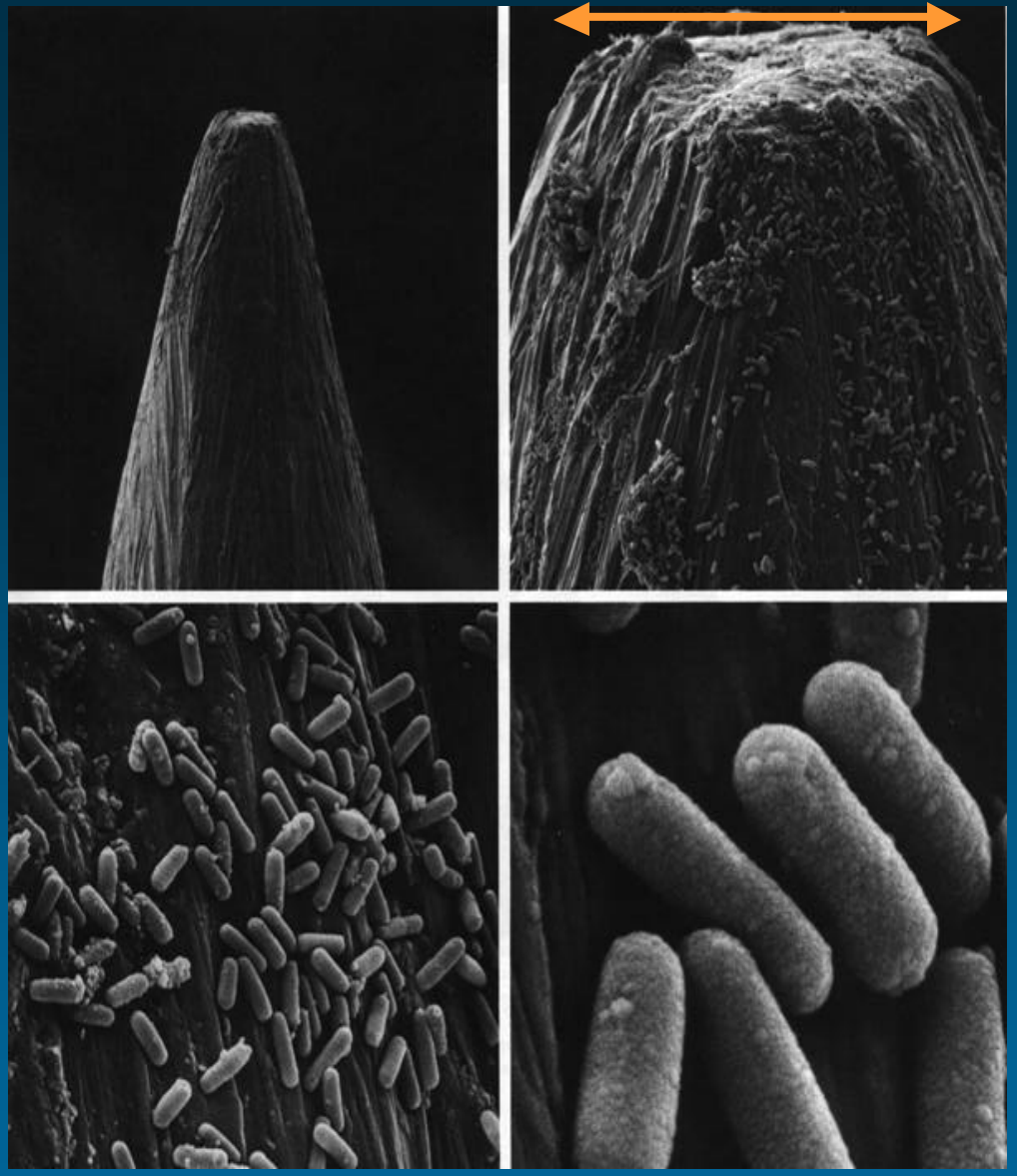
The constraint-based approach (theory?) to analysis of complex biological systems

Are boundary conditions more important than the model equations?

University of California, San Diego
Department of Bioengineering

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

E. coli on a pin



Alberts et al, 1st edition

University of California, San Diego
Department of Bioengineering

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

Modern Modeling Approaches

- Need to integrate diverse data types (genomic, trancriptomic, proteomic, metabolomic, phenomic,...)
- Must be easily scalable to cell or genome-scale
- Account for inherent biological uncertainty

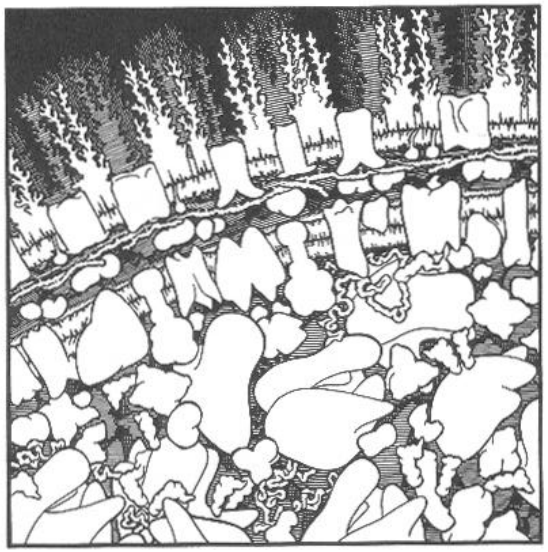


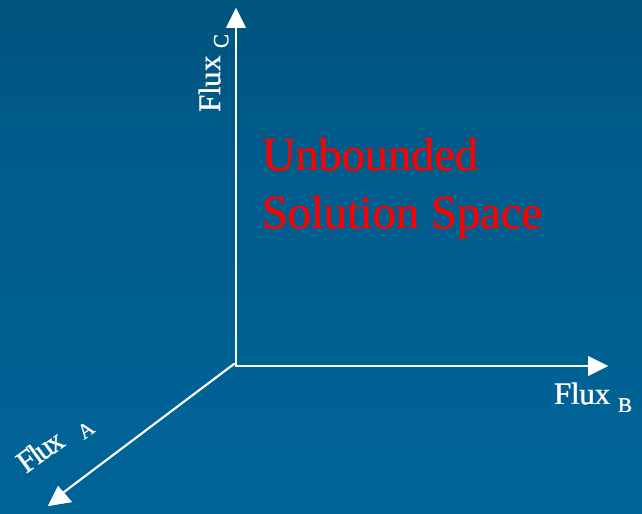
Figure 4.3 Cell Wall

The *Escherichia coli* cell wall, seen in cross section, is composed of several concentric layers. The water outside the cell is at top, in black, and the densely packed cytoplasm is at bottom. The layers of the cell wall from top to bottom (outside to inside) are: the outer membrane, with its gluey polysaccharides extending outward and porin pores spanning the bilayer; the thin layer of crosslinked peptidoglycan strands, connected to the outer membrane by small lipoproteins; the periplasmic space, containing a few small proteins; and the complex inner membrane, studded with many different proteins. (1,000,000 ×)

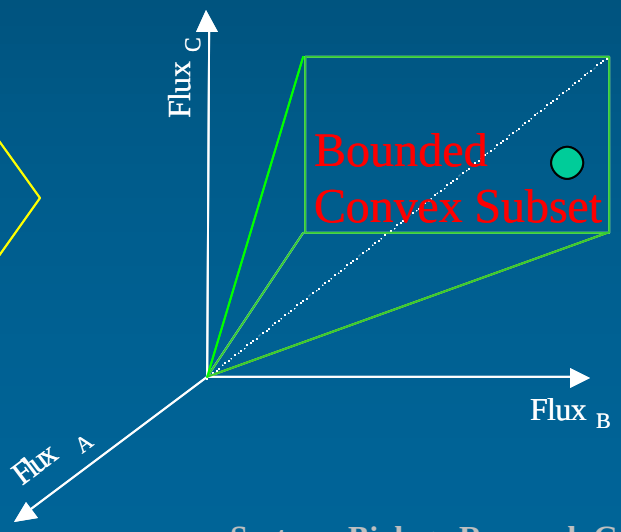
Constraints-based Analysis

How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?

–Sherlock Holmes, A Study in Scarlet



Constraints
(i) Stoichiometric
(ii) Thermodynamic
(iii) Capacity

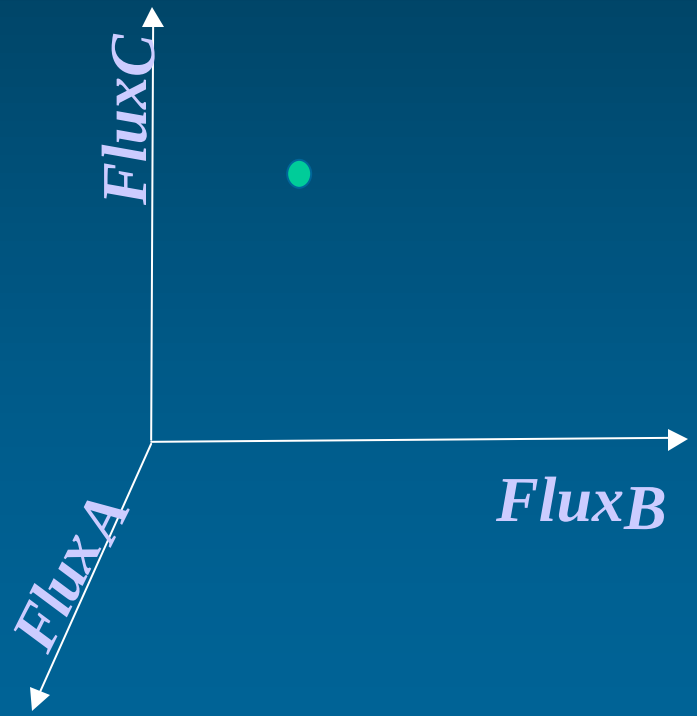


University of California, San Diego
Department of Bioengineering

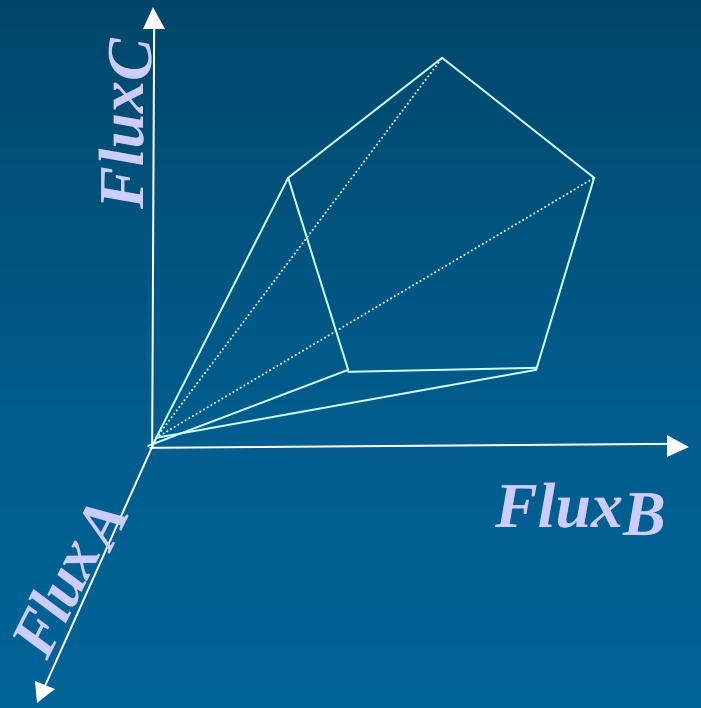
Systems Biology Research Group
http://systemsbiology.ucsd.edu

Theory-based vs. Constraint-based models: Single points vs. Solution spaces

- Complete knowledge
- Solution a single point



- Incomplete knowledge
- Solution space

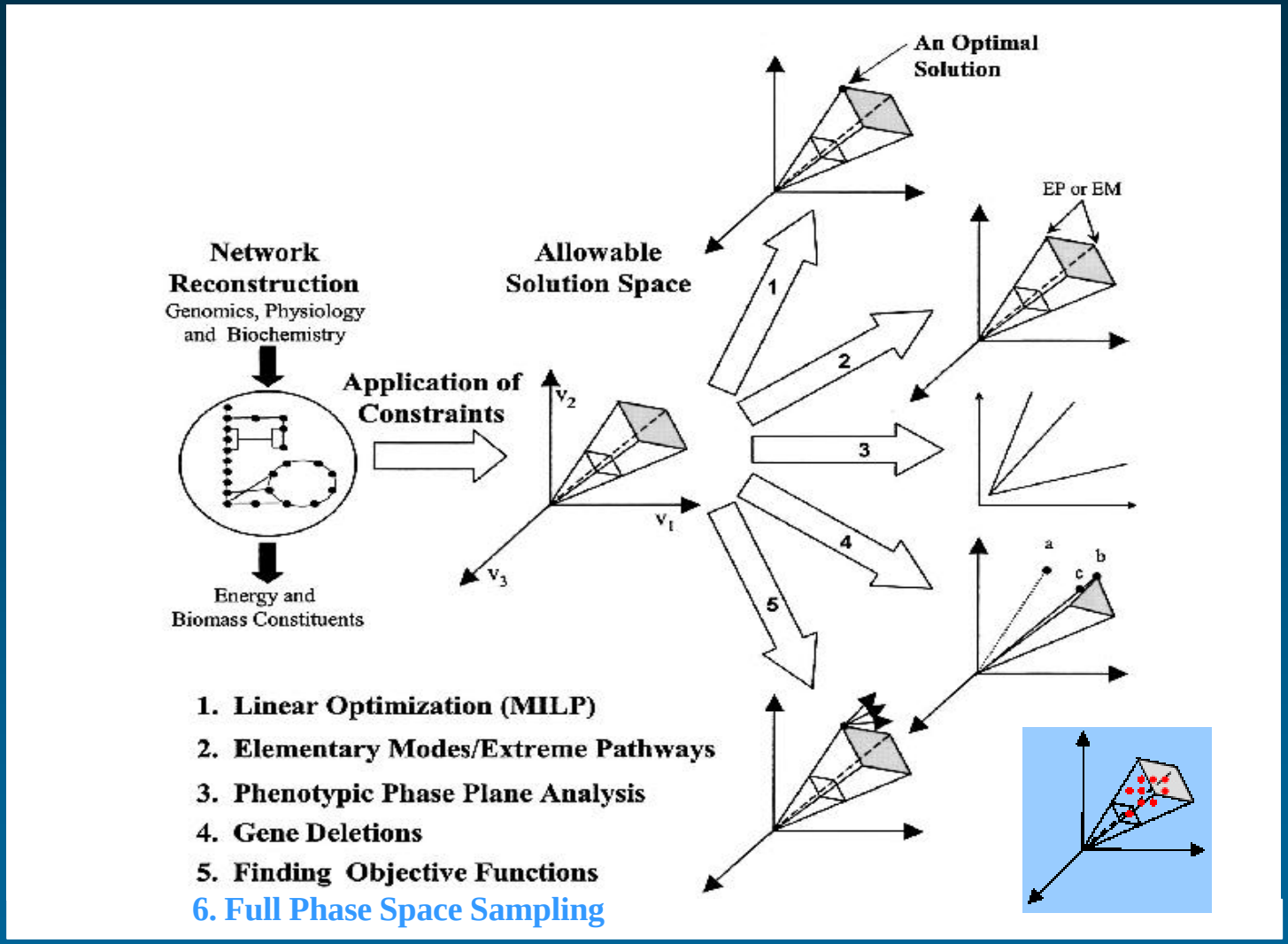


University of California, San Diego
Department of Bioengineering

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

Constraint-Based Modeling Methods

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



Current Constraint-Based Models

- *Haemophilus influenzae* 362/488/343* JBC 1999
- *Escherichia coli* 695/720/436, PNAS 2000
- *Helicobacter pylori* 291/388/339, J. Bact 2002
- *Saccharomyces cereviciae* 957/1294/801. Gen. Res 2003
- Other metabolic networks have been reconstructed for:
  - *Bacillus subtilis*,
  - *Streptococcus pneumoniae*,
  - *Pseudomonas aeruginosa*,
  - *Geobacter sulfurreducens*,
  - *Mycobacterium tuberculosis*,
  - *Anabaena*,
  - *Plasmodium falciparum.*

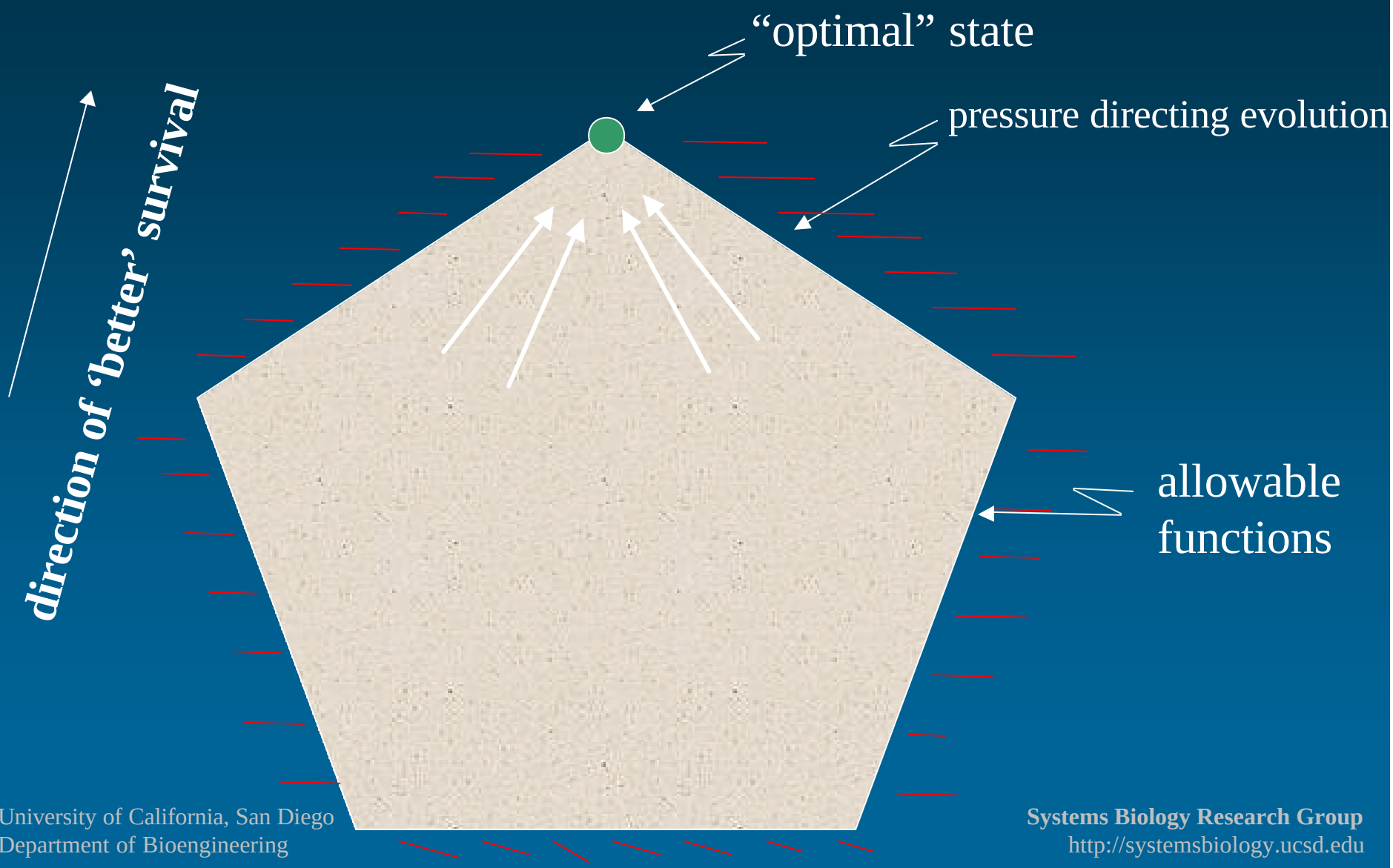
*(genes/reaction/metabolites)

Generations of constraints-based models: use of ‘omics’ data

- 1st generation
  - Hard constraints
  - Determine capabilities (what)
- 2nd generation
  - Regulation of expression
  - Determine choices (why)
- 3rd generation
  - Regulation of activity
  - Determine trajectories (how)

- Genomics
  - annotated sequence
  - legacy data
- Expression profiling
  - transcriptomics
  - proteomic
- Concentration data
  - metabolomics
  - proteomic

Evolution and governing constraints



University of California, San Diego
Department of Bioengineering

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

Engineering Design

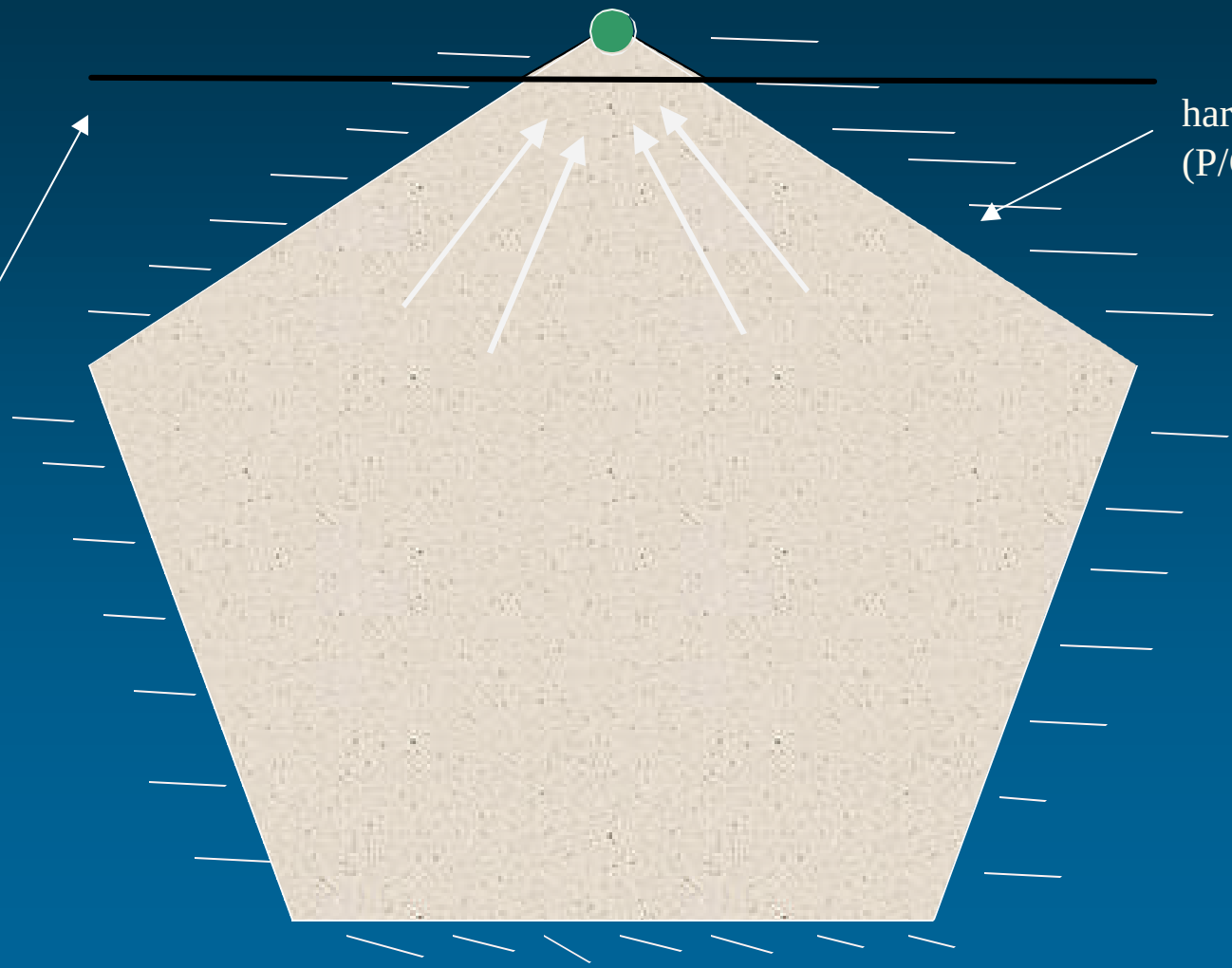
- Objective
  - separation of protein, building a bridge, designing a car, etc
- Constraints:
  - geometry, materials, diffusion constants, cost, time
- Design envelope
- Optimize design using free design variables
  - optimal engineering designs do evolve
  - see Detroit's industrial history museum

Engineering vs. Biological Design

- Objective
  - separation of protein
- Constraints:
  - Geometry
  - Materials
  - Diffusion constants
- Design envelope
- Optimize design using free design variables

• Objective
– Survival, growth, pH
• Constraints:
– Max fluxes
– Connectivity
– P/C factors
• Solution space
• Optimize design using kinetic and regulatory variables

Imposition of combined constraints to produce the “optimal” phenotype



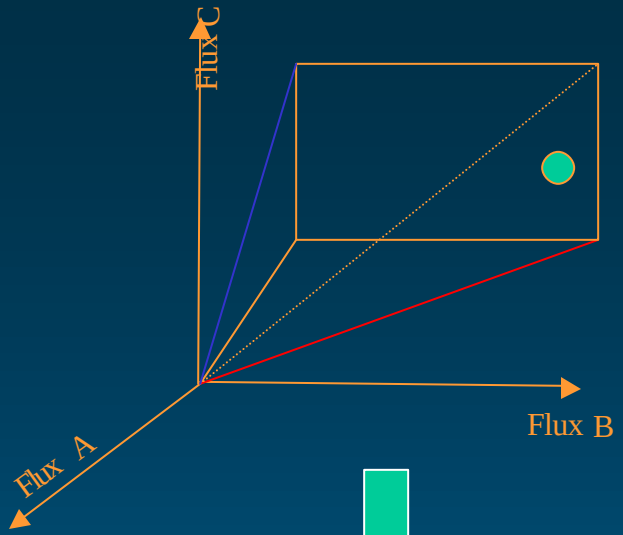
hard constraints
(P/C environment, etc.)

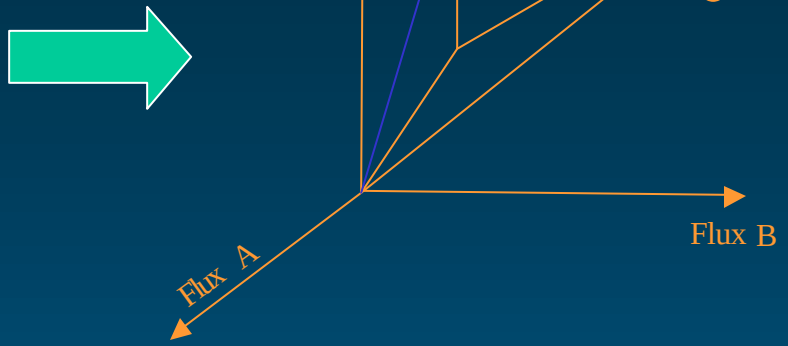
regulatory
constraints;
self-imposed
restraints

Biological Design

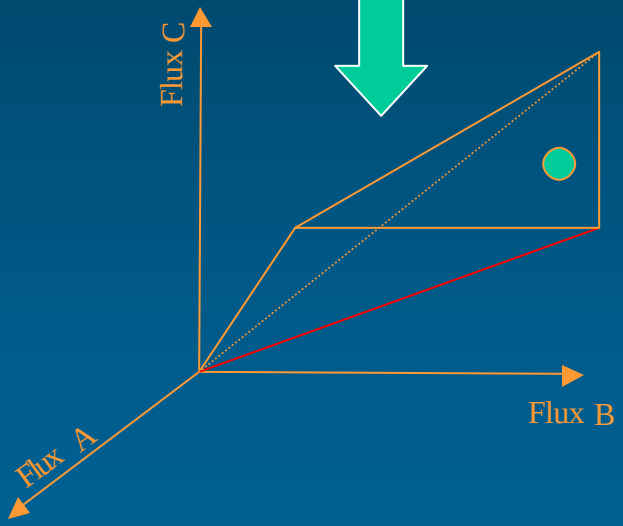
Regulation of expression:
shrinking solution space

Regulation of activity: location
within a shrunk solution space





No ‘red’: solution not achievable



No ‘blue’: solution achievable but solution space smaller

Shaping the Solution Space

Covert, M.W., Schilling, C.H., Palsson, B.O.
JTB Nov 2001

University of California, San Diego
Department of Bioengineering

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

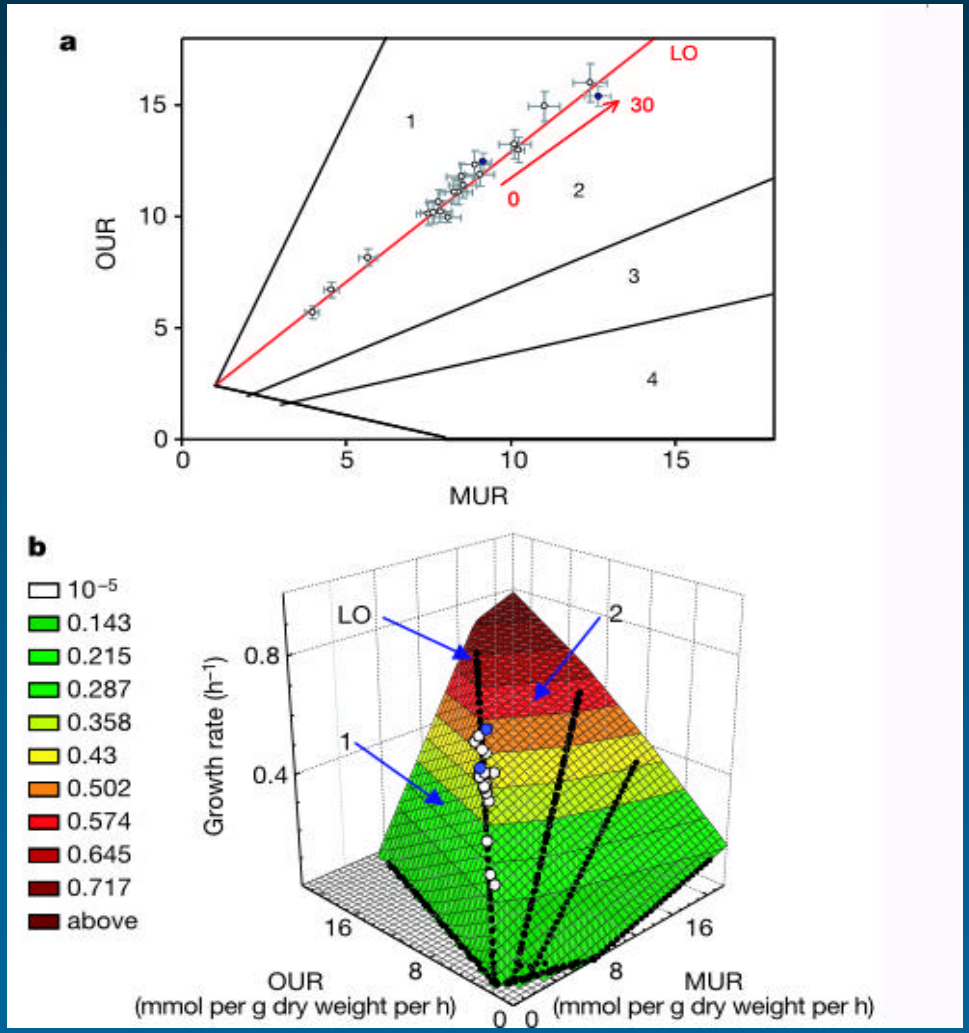
Models driving experiments

University of California, San Diego
Department of Bioengineering

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

Complex biological processes: growth and adaptive evolution

The optimal growth hypothesis



Nature, Nov 2002

University of California, San Diego
Department of Bioengineering

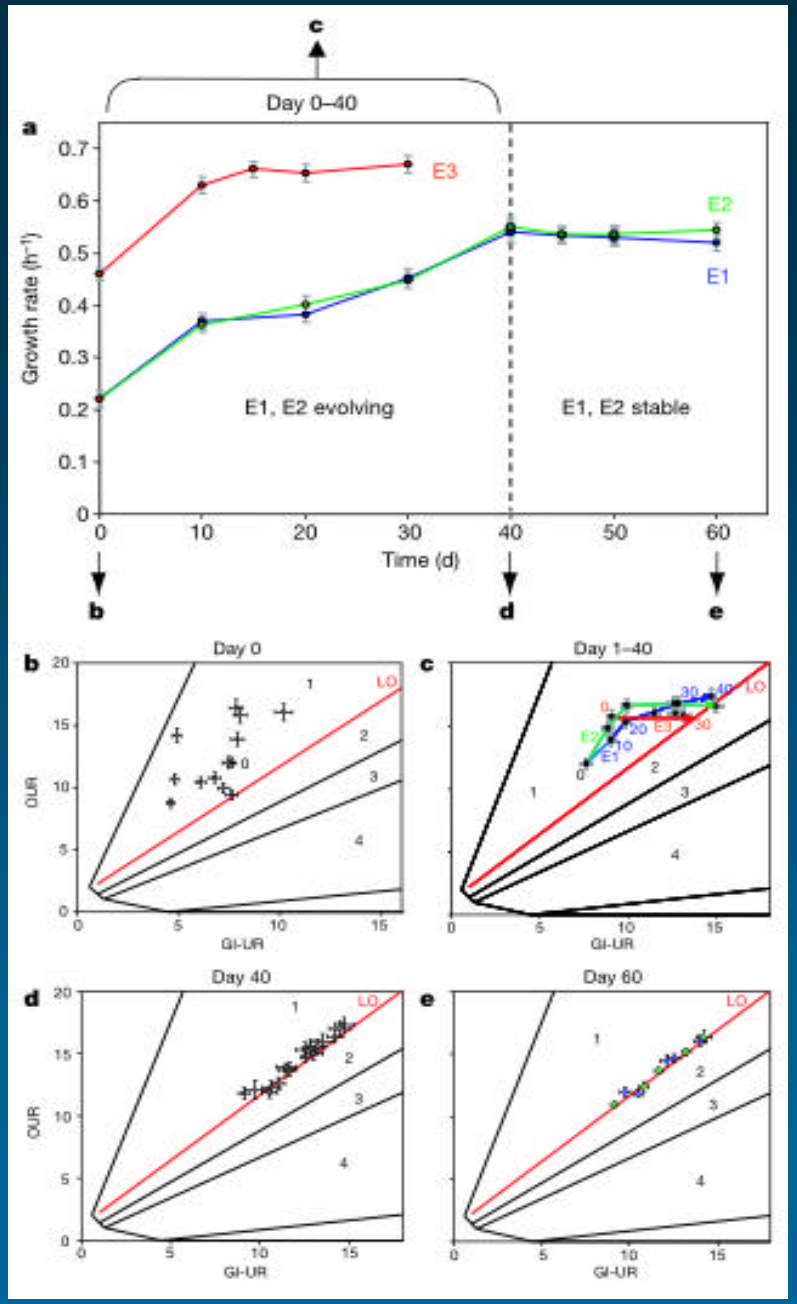
Systems Biology Research Group
http://systemsbiology.ucsd.edu

Why non-optimal growth?

Adaptive evolution

- *E. coli* K-12 MG1655 exhibits sub-optimal growth on glycerol
- Adaptive evolution over 700 generations more than doubles growth rate
- Adaptation is reproducible
- Expression profiling shows that a few dozen genes have >3x change in expression

Nature, Nov 2002



University of California, San Diego
Department of Bioengineering

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

Other lessons

BASIC SCIENCE

- *E. coli* can ‘learn’ new optimal behaviors in 500-1000 generations
- *E. coli* has silent phenotypes, or alternative optimal solutions
- *E. coli* ‘forgets’
- *E. coli* has coupled objectives

ENGINEERING DESIGN

- *E. coli* knock-outs follow the same patterns
- *E. coli* phenotypes can be designed in a computer a priori and the designs ‘implemented’ through adaptive evolution

Mechanisms of (adaptive) Evolution:

multi-time scale process--some features

- Change in expression -- Instantaneous
  - look up entry in 'look-up-table'
- If 'looked up' solution is not the best one (Nature 420:186, 2002)
  - then mutations & selection ensues to modify network properties
  - 500 -1000 generations or about 40 days
- Long-term: Pathway genesis & evolution (TIBS 28:336 2003)
  - Enzyme recruitment for similar chemical functions
  - Centered around highly-connected metabolites
- Longer-term: modification of gene content (GR 13:1589 2003)
  - Gene deletion (GD), horizontal gene transfer (HGT), de novo gene genesis(GG)
  - 3*GD = HGT + 2* GG

Evolving knock-out strains

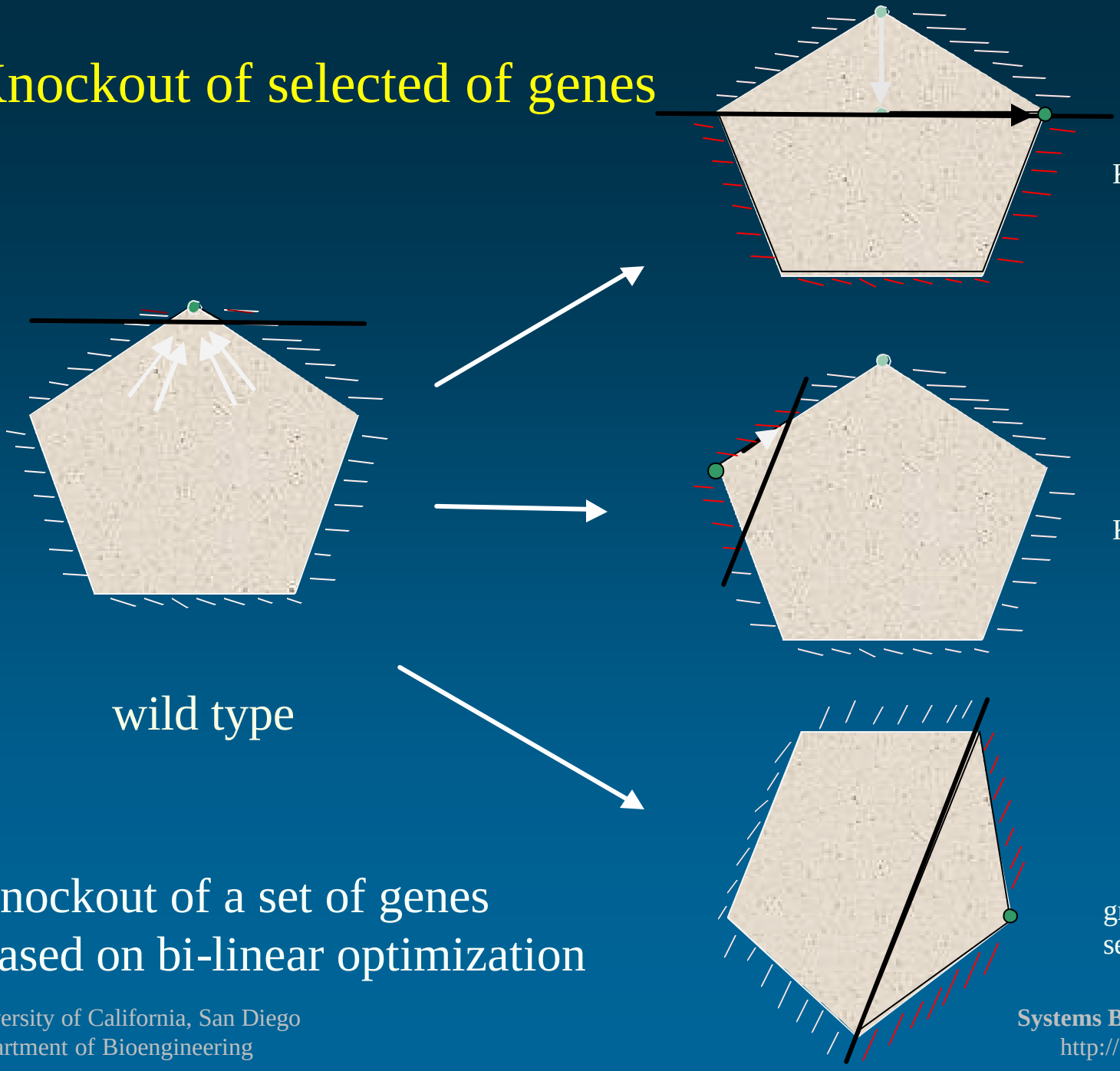
Probe gene usage in metabolic network

Full growth recovery should be achieved after deletion of genes with a reduced cost of zero

Design of strains

Metabolic phenotypes can be prospectively determined *in silico*

Knockout of selected of genes



KNO of structural gene

KNO of regulatory gene

wild type

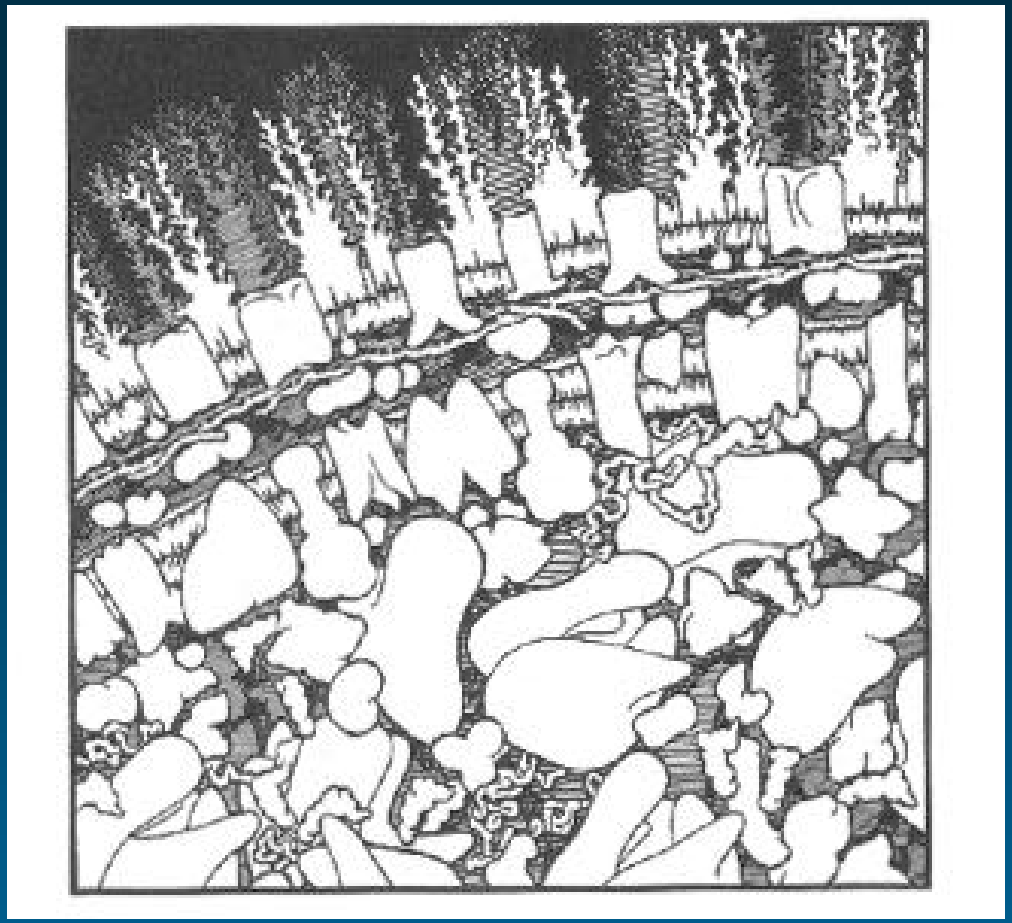
growth and by-product secretion

Knockout of a set of genes
Based on bi-linear optimization

University of California, San Diego
Department of Bioengineering

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

The intracellular environment is very complex and counter-intuitive



Viscosity ~ 1000 x H_2O

Pressure (osmotic) < 150 atm

Electrical gradient ~ 300,000 V/cm

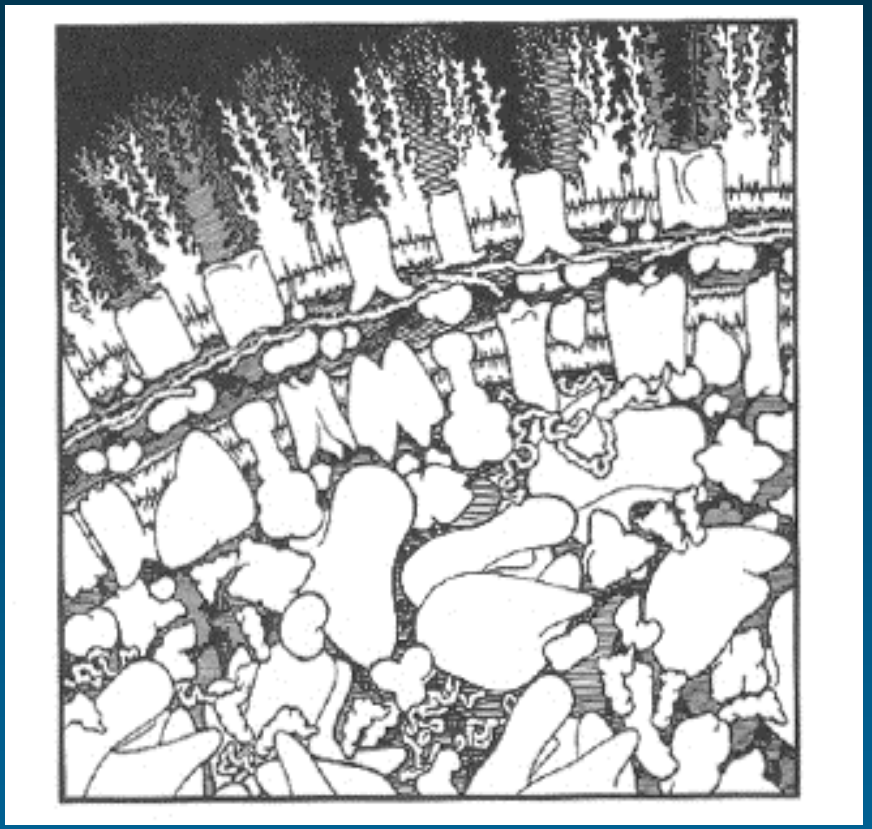
From The Machinery of Life,
David S. Goodsell,
Springer-Verlag, New York, 1993.

University of California, San Diego
Department of Bioengineering

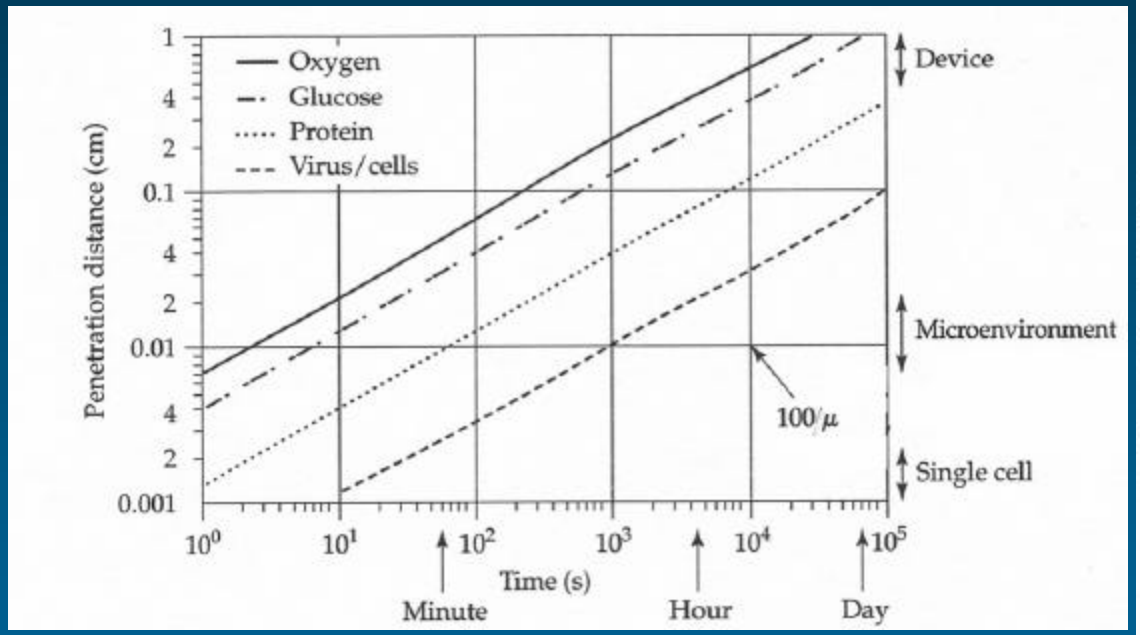
Systems Biology Research Group
http://systemsbiology.ucsd.edu

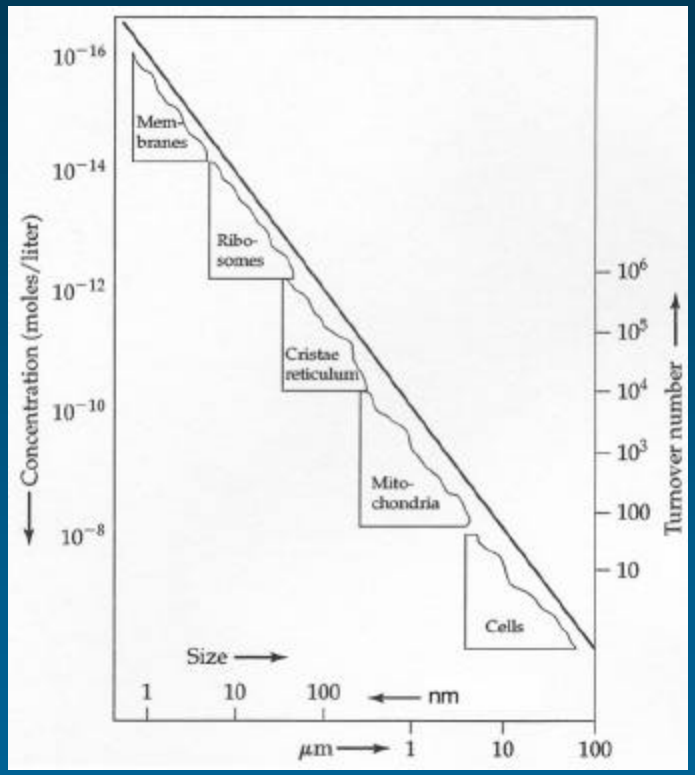
P/C Considerations

- Thermal forces
  - The special role of diffusion
  - Molecular noise
- Spatial hindrances
  - Structure of the cell and the genome
  - Compartmentalization and volume regulation
- Electro-chemical
  - Seems like a somewhat unexplored territory



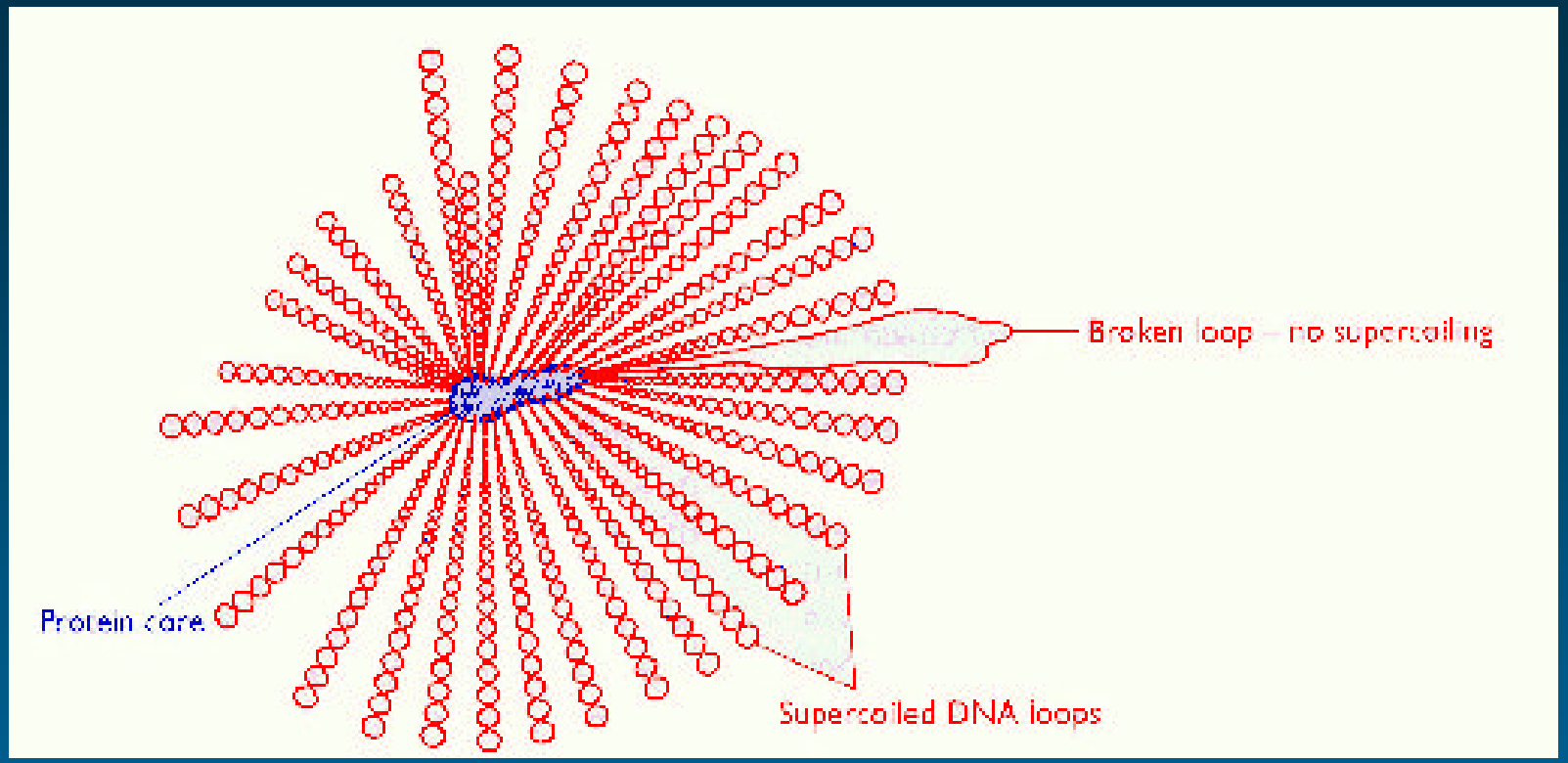
Diffusion is a dominant rate constraint





$$\tau = l^2 / D$$

Ultrastructure of the *E. coli* genome

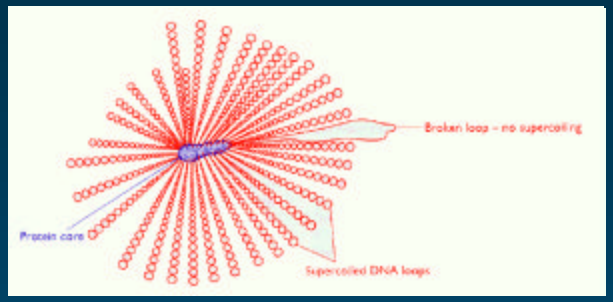


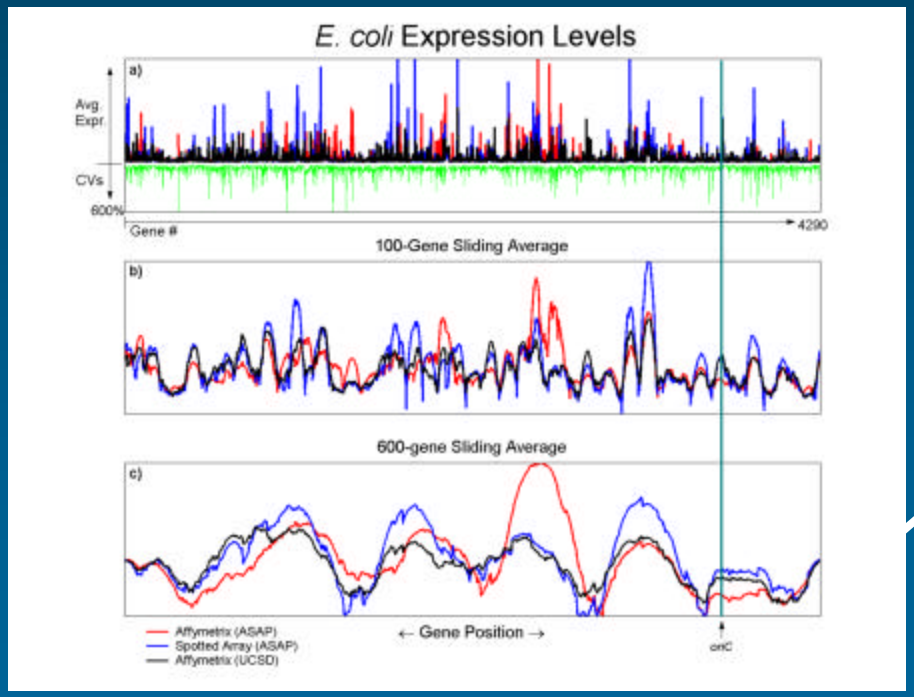
Brown, *Genomes* (1999)

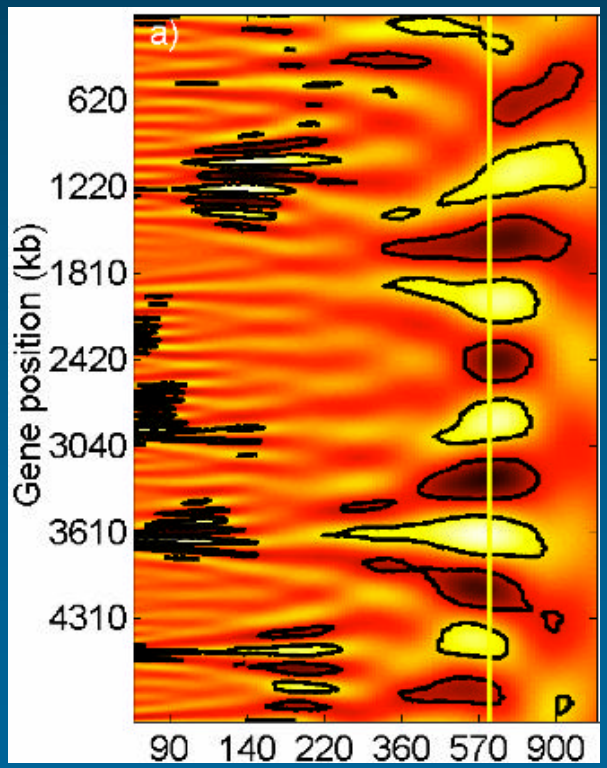
- Linear length of DNA = 1 mm = 1000 times cell size
- Protein core
- ~40-45 supercoiled loops of DNA radiate from central core

Periodicity in genome usage

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







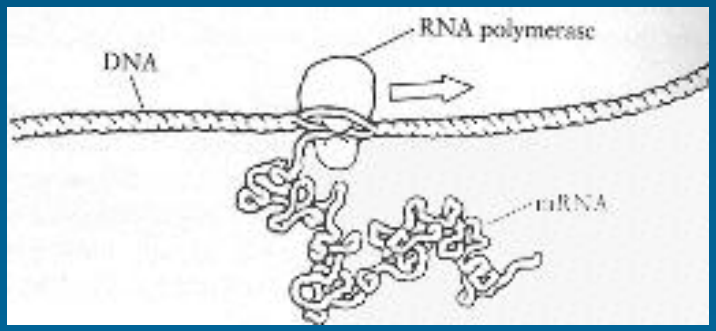
University of California, San Diego
Department of Bioengineering

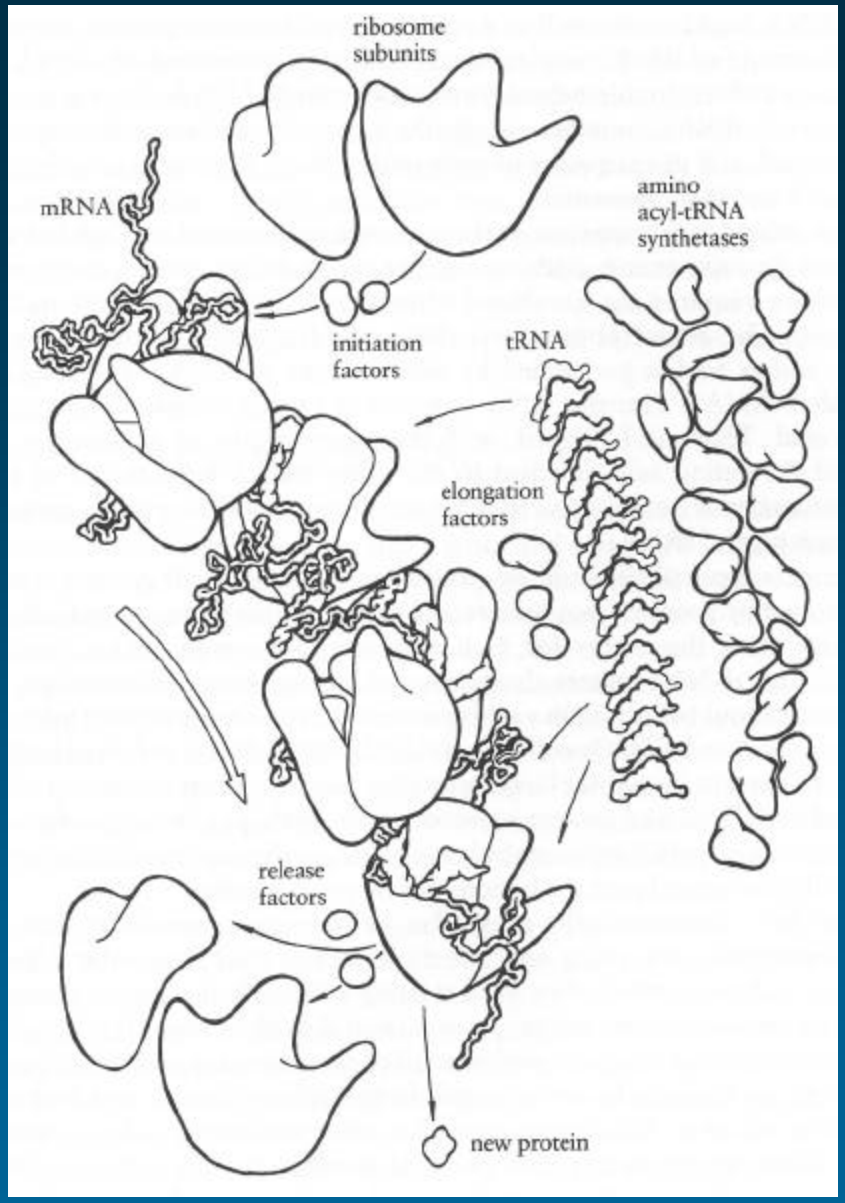
Allen et al, J. Bact 2003

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

A look inside *E. coli*

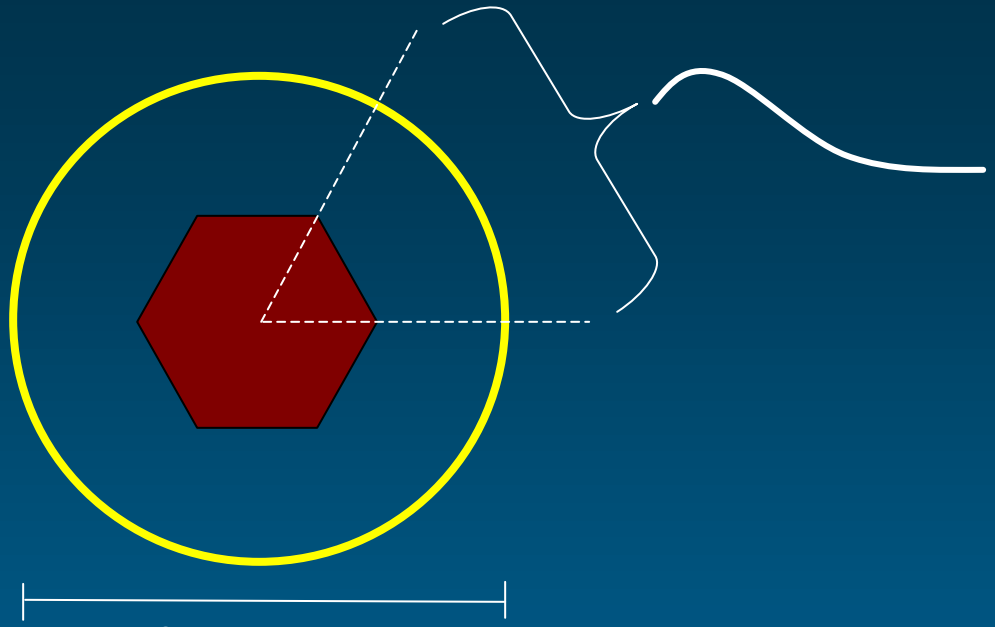
- DNA
  - 1mm total (~200nm/ORF)
- Ribosomes
  - 15-20 nm
  - ~5 ribosomes / ORF
- tRNA
  - 5 nm
  - 200,000 / cell (10 tRNA/ribosome)

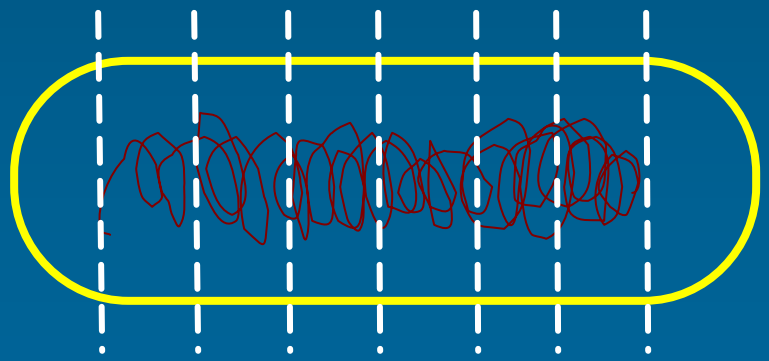




Goodsell, *The Machinery of Life* (1998)

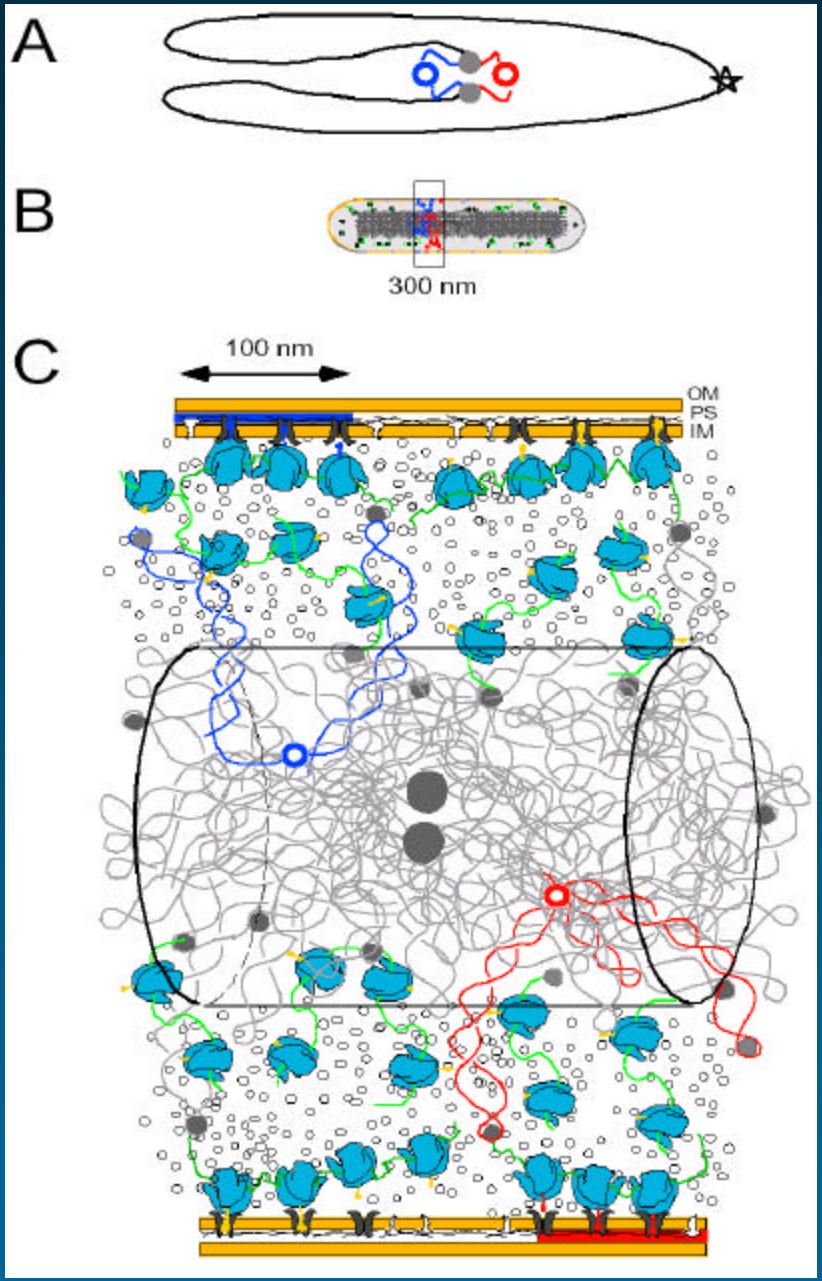
Topobiology of *E. coli* Genome

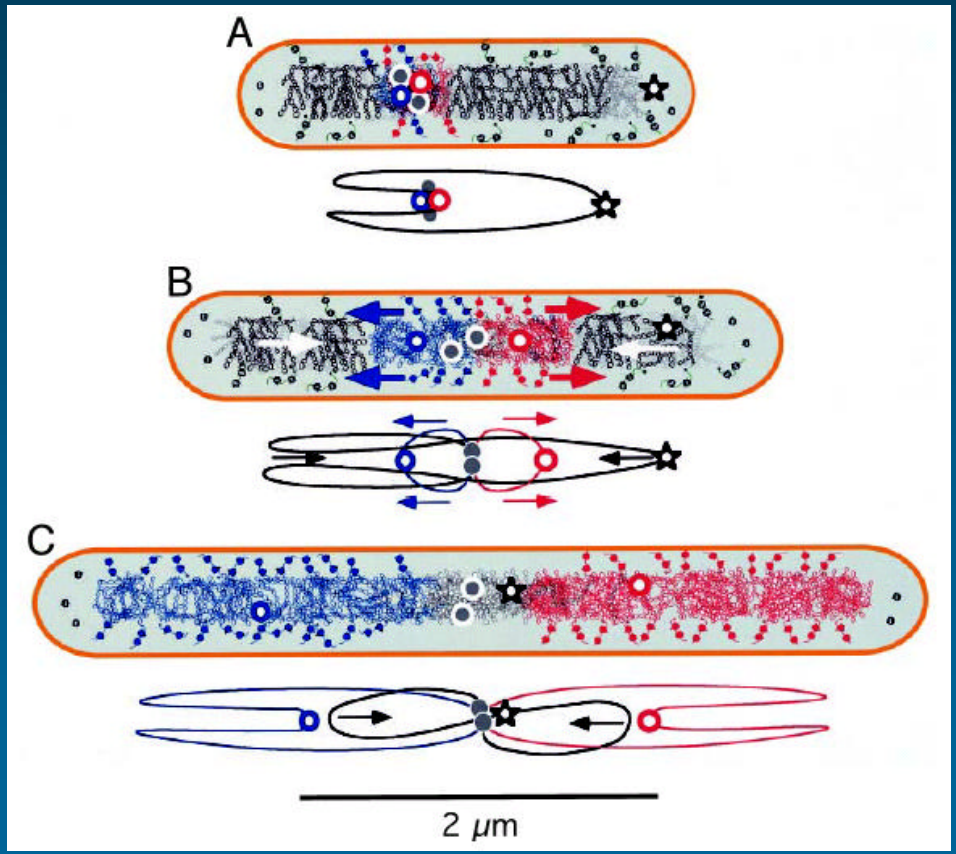




- Nucleoid can be divided into ~36-42 “coil-somes”
  - 25 folds/coil
  - ~ $(200\ nm)^3$
- Approximate composition per “coil-some”
  - 500 ribosomes
  - 5000 tRNA
  - 10 kbp DNA (100 ORFs)
  - 70 mRNA
  - 60 RNAP

Chromosome Segregation





University of California, San Diego
Department of Bioengineering

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

Genome-scale models as hypotheses in systems biology

- Hypothesis driven vs. discovery driven research
- Hypothesis represents ‘best guesses’ based on your current knowledge
- *In silico* models are the most compact representation of complex data sets
- Thus they represent highly structured hypotheses
- Models, like hypotheses, are to be dis-proven
- Never dis-proven in their entirety, but in parts
- Models give emergent properties and thus represent hypotheses about biological (systems) functions

Summary

- Systems biology and bioengineering is an emerging science
- Deals with the interaction of multiple components
- Many links form a network
- Moderately complex networks have many possible functional states
- Regulation selects “appropriate” states
- Genome-scale models have been constructed to describe whole cell functions
- *In silico* models will serve as hypothesis in systems biology