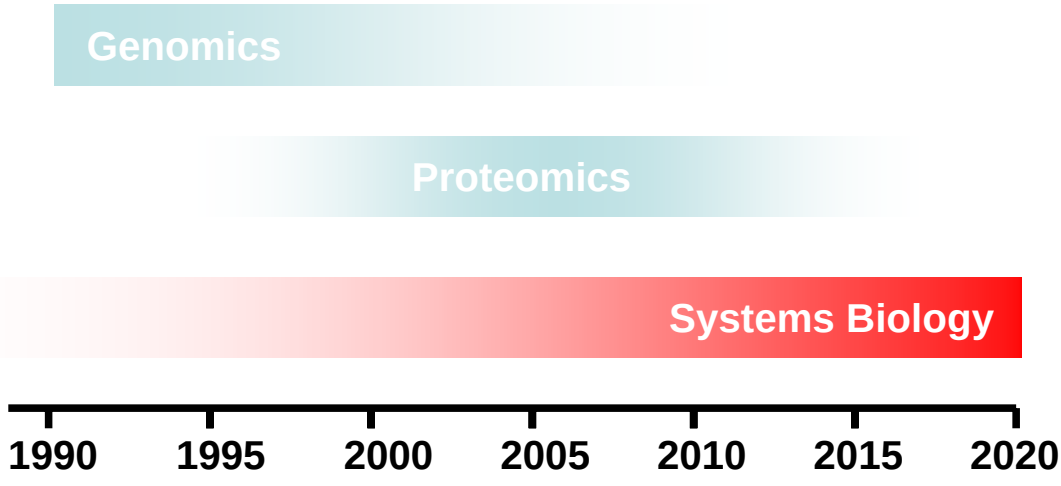
Systems Biology

Micro 343
David Wishart Rm. Ath 3-41
david.wishart@ualberta.ca

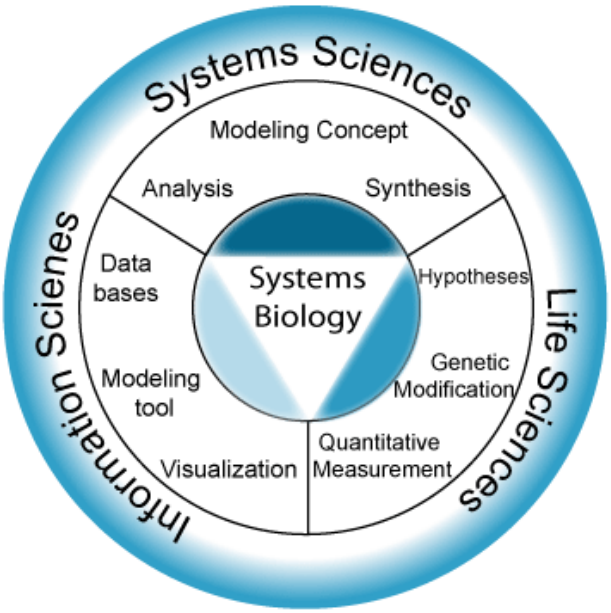
Genomics, Proteomics & Systems Biology

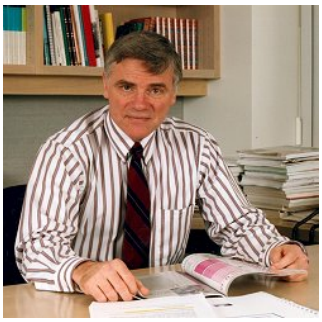


What is Systems Biology?

- **Systems Biology** - ***The study of the mechanisms underlying complex biological processes as integrated systems of many interacting components. Systems biology involves (1) collection of large sets of experimental data (2) proposal of mathematical models that might account for at least some significant aspects of this data set, (3) accurate computer solution of the mathematical equations to obtain numerical predictions, and (4) assessment of the quality of the model by comparing numerical simulations with the experimental data.***
- ***First described in 1999 by Leroy Hood***

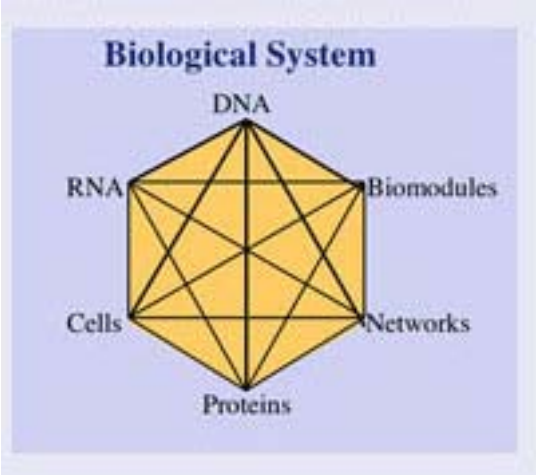
System Biology

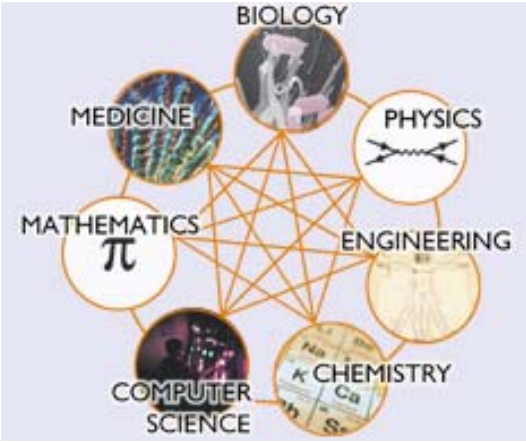




Lee Hood – director of the Institute for System Biology

Systems Biology





Institute for System Biology

http://www.systemsbiology.org/

What's it good for?

- Basic Science/"Understanding Life"
- Predicting Phenotype from Genotype
- Understanding/Predicting Metabolism
- Understanding Cellular Networks
- Understanding Cell-Cell Communication
- Understanding Pathogenicity/Toxicity
- "Raising the Bar" for Biologists

Making Biology a Predictive Science

Are We Ready?

- 100's of completed genomes
- 1000's of known reactions
- 10,000's of known 3D structures
- 100,000's of protein-ligand interactions
- 1,000,000's of known proteins & enzymes
- Decades of biological/chemical know-how
- Computational & Mathematical resources

The Push to Systems Biology

The Technologies of Systems Biology

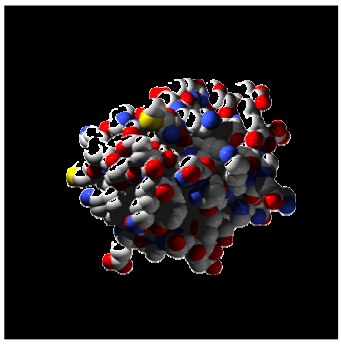
- Genomics (HT-DNA sequencing)
- Mutation detection (SNP methods)
- Transcriptomics (Gene/Transcript measurement, SAGE, gene chips, microarrays)
- Proteomics (MS, 2D-PAGE, protein chips, Yeast-2-hybrid, X-ray, NMR)
- Metabolomics (NMR, X-ray, capillary electrophoresis)

Going From Technology to Systems Biology

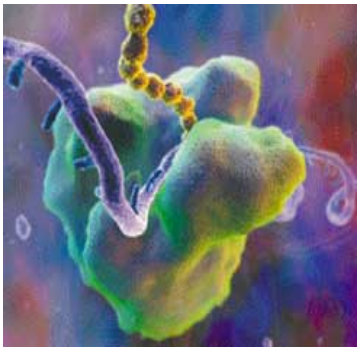
- Genomics → Genometrics
- Proteomics → Proteometrics
- Metabolomics → Metabometrics
- Phenomics → Phenometrics
- Bioinformatics → Biosimulation
- ***Quantify, quantify, quantify***

How to Do it?

Three Types of Simulation



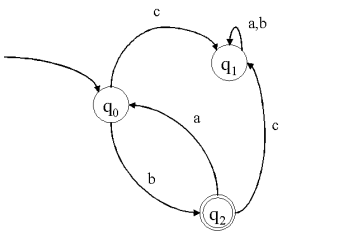
Atomic Scale
0.1 - 1.0 nm
Coordinate data
Dynamic data
0.1 - 10 ns
Molecular dynamics



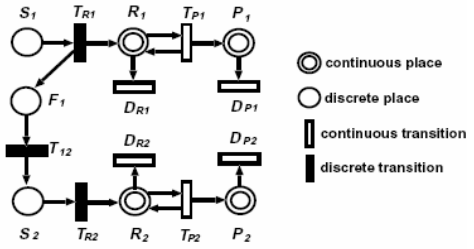
Meso Scale
1.0 - 10 nm
Interaction data
Kon, Koff, Kd
10 ns - 10 ms
Mesodynamics

Continuum Model
10 - 100 nm
Concentrations
Diffusion rates
10 ms - 1000 s
Fluid dynamics

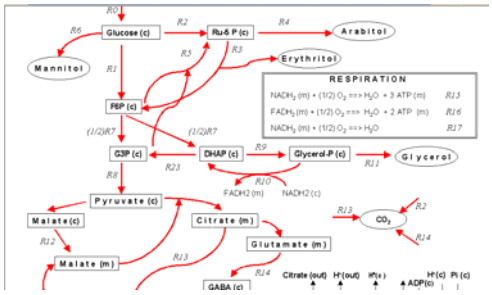
How To Do it? (Computationally)



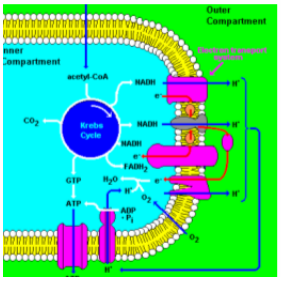
Pi Calculus



Petri Nets



Flux Balance Analysis



$$\frac{dx_1}{dt} = k_{11}x_1 + k_{21}x_2 + k_{31}x_3 + \ldots$$

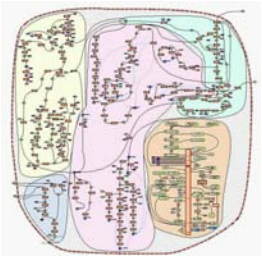
$$\frac{dx_2}{dt} = k_{12}x_1 + k_{22}x_2 + k_{32}x_3 + \ldots$$

$$\frac{dx_3}{dt} = k_{13}x_1 + k_{23}x_2 + k_{33}x_3 + \ldots$$

$$\frac{dx_4}{dt} = k_{13}x_1 + k_{24}x_2 + k_{34}x_3 + \ldots$$

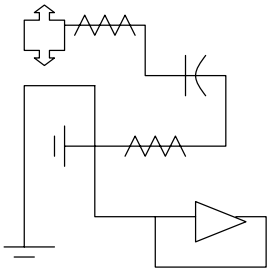
Differential Eqs

How To Do it? (Computationally)

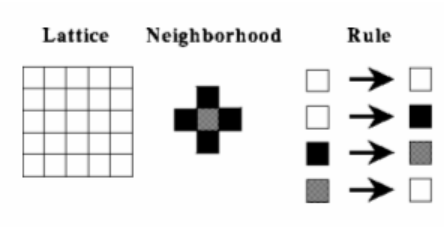


Boolean Networks

Reservoir Analysis



Electrical Circuit Model

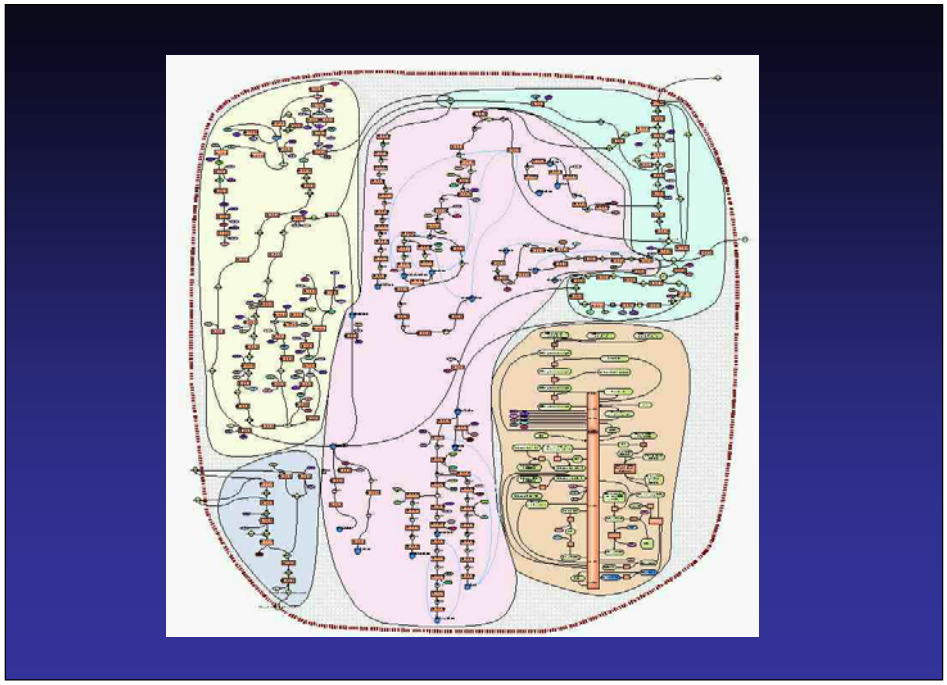


Cellular Automata

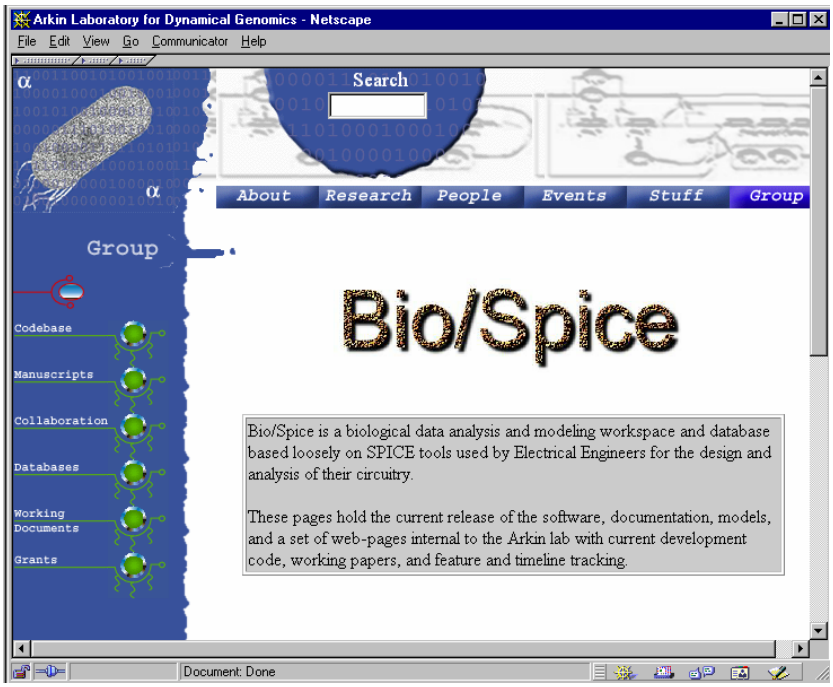
Who's Doing It?

- E-cell Project (Keio University, Japan)
- BioSpice Project (Arkin, Berkeley)
- Metabolic Engineering Working Group (Palsson & Church, UCSD, Harvard)
- Silicon Cell Project (Netherlands)
- Virtual Cell Project (UConn)
- Gene Network Sciences Inc. (Cornell)
- Project CyberCell (Edmonton/Calgary)

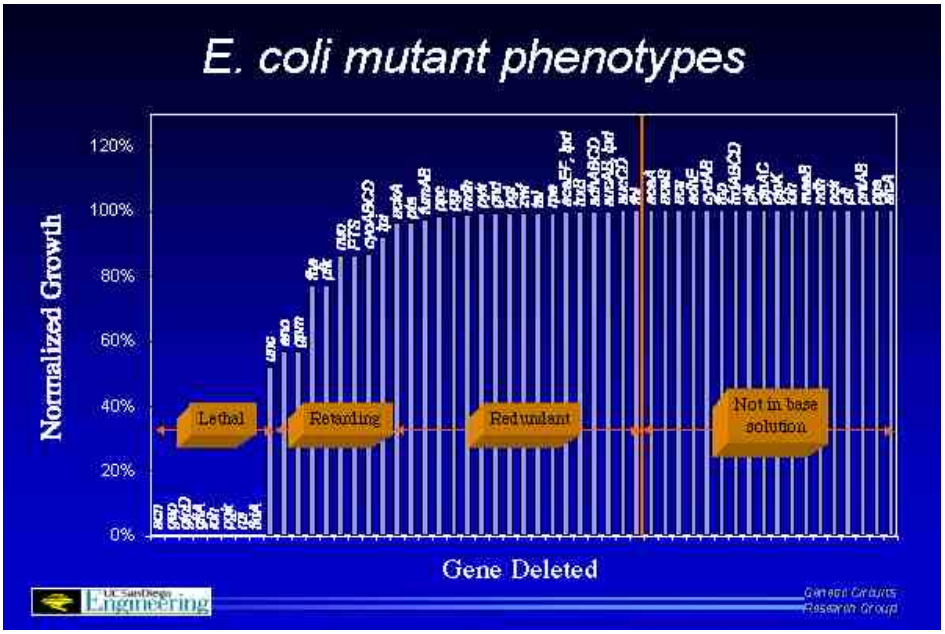
www.e-cell.org



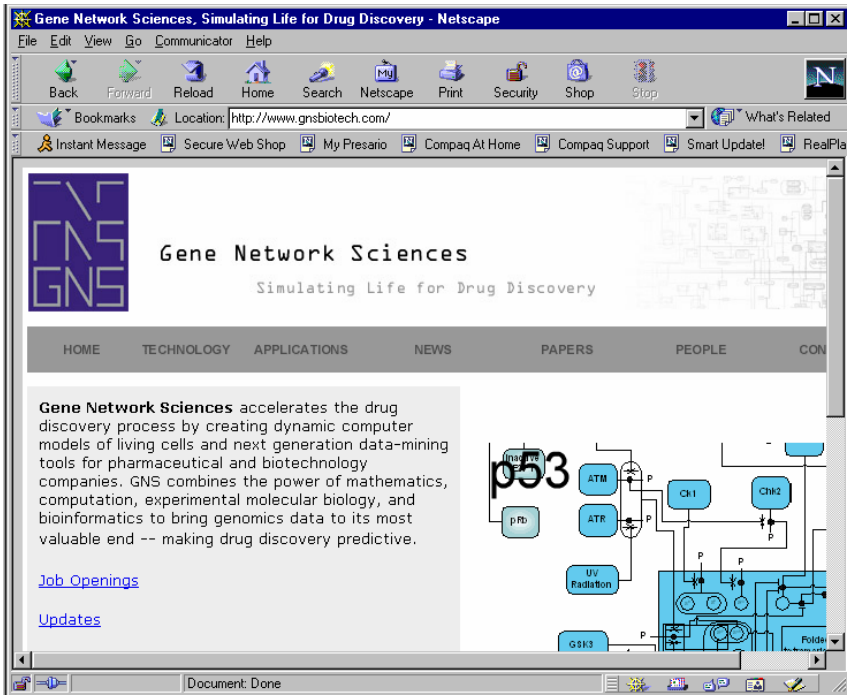
Adam Arkin



B. Palsson-Genetic Circuits

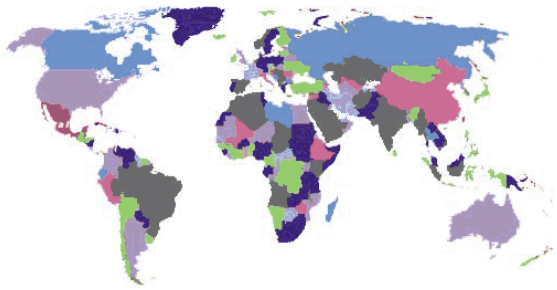


Gene Network Sciences



Nationalism in Simulation

- **Petri Nets – Germany, Japan**
- **Flux-Balance Analysis – USA**
- **Pi Calculus – France**
- **ODE's and PDE's – Japan, UK**
- **Agent-Based methods (CA) - Canada**



Some Problems…

- **Almost all simulation systems are ultimately based on solving either ordinary differential equations (ODEs), partial differential equations (PDEs) or stochastic differential equations (SDEs)**
- **Differential equations are "hard" to work with when simulating spatial phenomena, when dealing with discrete events (binding, switching), non continuous variables (low copy number) or when key parameters are unknown or unknowable**

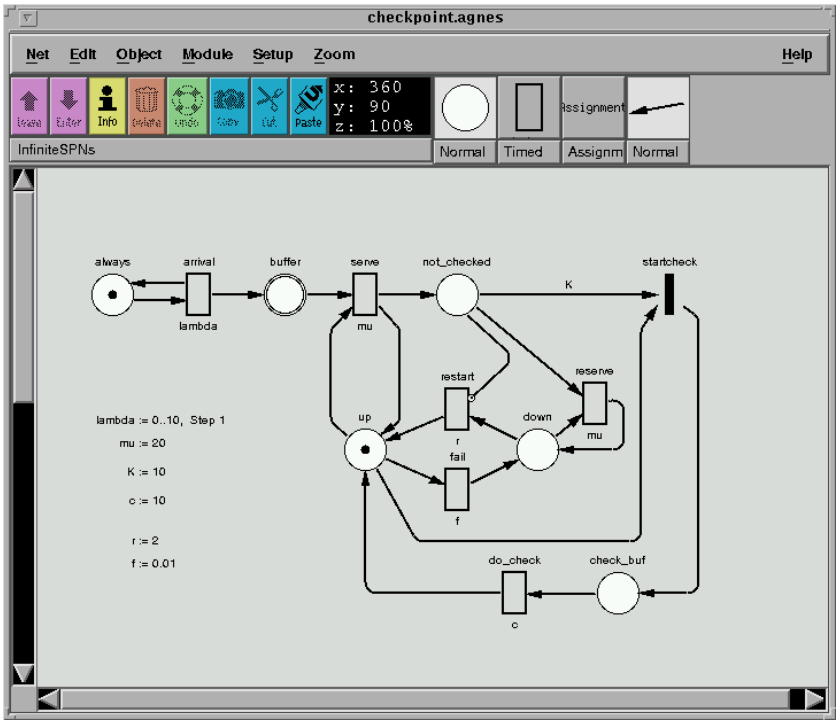
Some Problems…

- DEs are notorious for instabilities or situations where small rounding errors lead to singularities or chaotic behavior
- DE methods are not conducive to visualization or interactive “movies”
- DE methods require considerable mathematical skill and understanding (not common among biologists)
- DE methods don’t easily capture stochasticity or noise (common in biology)
- *Issue of realism – cells don’t do calculus*

Is There a Better Way?

- Sidney Brenner calls it “biological arithmetic – not calculus”
- Needs to accommodate the discrete (binding, signaling) and continuous (substrate concentration) nature of many cellular phenomena
- Two new approaches which avoid DEs
  - Petri Nets (stochastic and hybrid)
  - Cellular automata or agent based methods

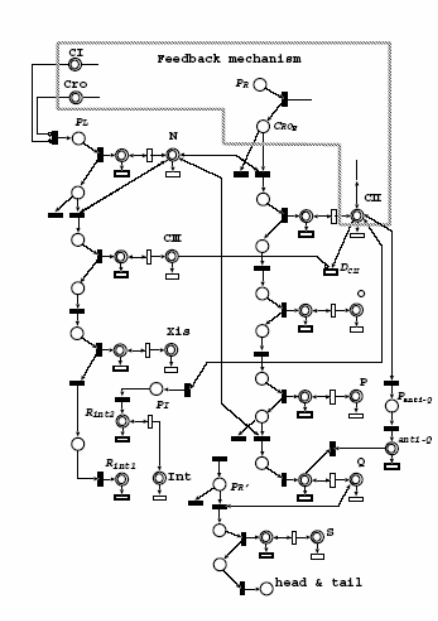
Petri Nets



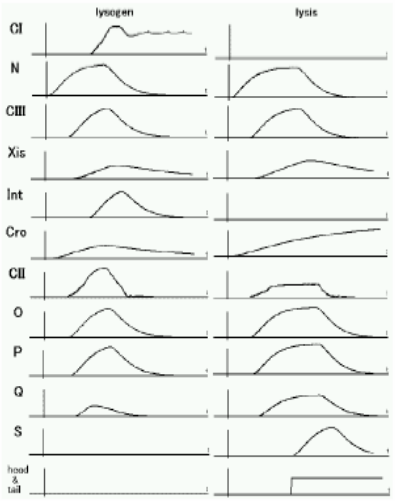
Petri Nets

- A directed, bipartite graph in which nodes are either "places" (circles) or "transitions" (rectangles)
- A Petri net is marked by placing "tokens" on linked or connected places
- When all the places have a token, the transition "fires", removing a token from each input place and adding a token to each place pointed to by the transition (its output places)
- Petri nets are used to model concurrent systems, particularly network protocols w/o differential eqs.
- Hybrid petri nets allow modelling of continuous and discrete phenomena

Hybrid Petri Nets

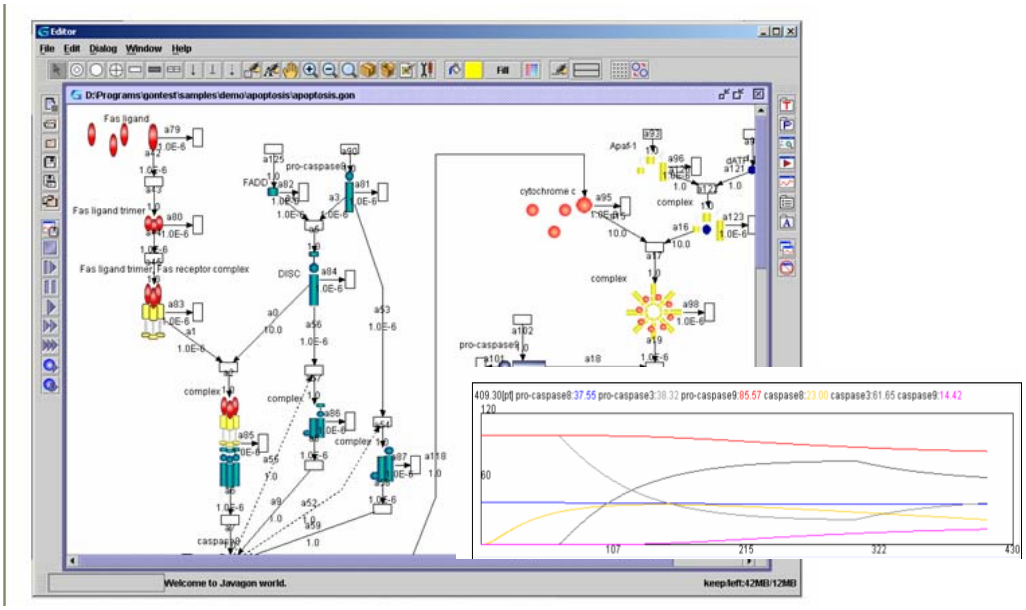


λ phage control circuit



Predicted protein expression

Cell Illustrator – An HPN with a GUI *www.genomeobject.net*



Now sold as a product by Gene Networks International - http://www.gene-networks.com

Petri Nets - Limitations

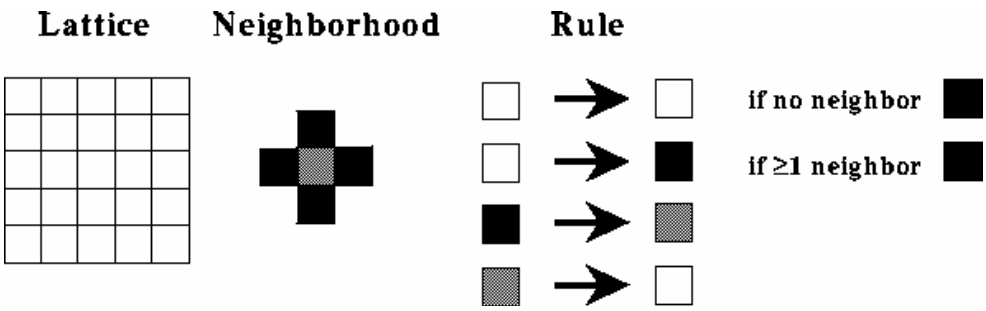
- **Not designed to handle spatial events or spatial processes easily**
- **Stochasticity is "imposed", it does not arise from underlying rules or interactions**
- **Does not reproduce physical events (brownian motion, collisions, transport, binding, etc.) that might be seen in a cell – Petri Nets are more like a plumbing and valving control system**

What about Cellular Automata?

Cellular Automata

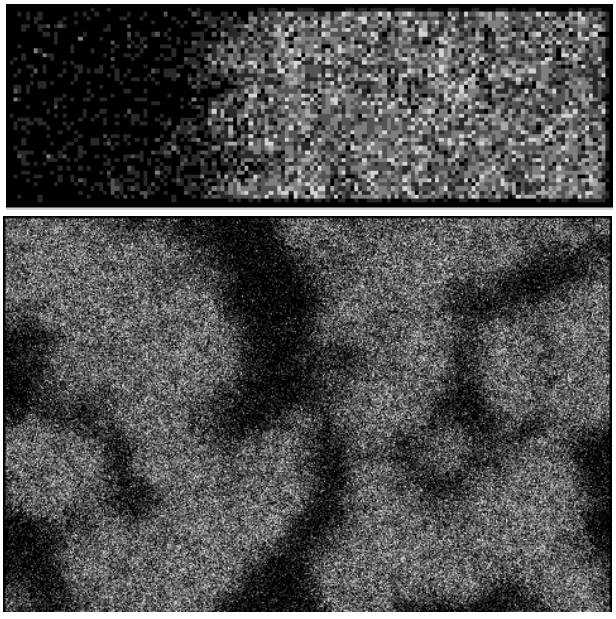
- **Computer modelling method that uses lattices and discrete state "rules" to model time dependent processes – a way to animate things**
- **No differential equations to solve, easy to calculate, more phenomenological**
- **Simple unit behavior -> complex group behavior**
- **Used to model fluid flow, percolation, reaction + diffusion, traffic flow, pheromone tracking, predator-prey models, ecology, social nets**
- ***Scales from 10^{-12} to 10^{+12}***

Cellular Automata



Can be extended to 3D lattice

Reaction/Diffusion with Cellular Automata



CA Methods in Games

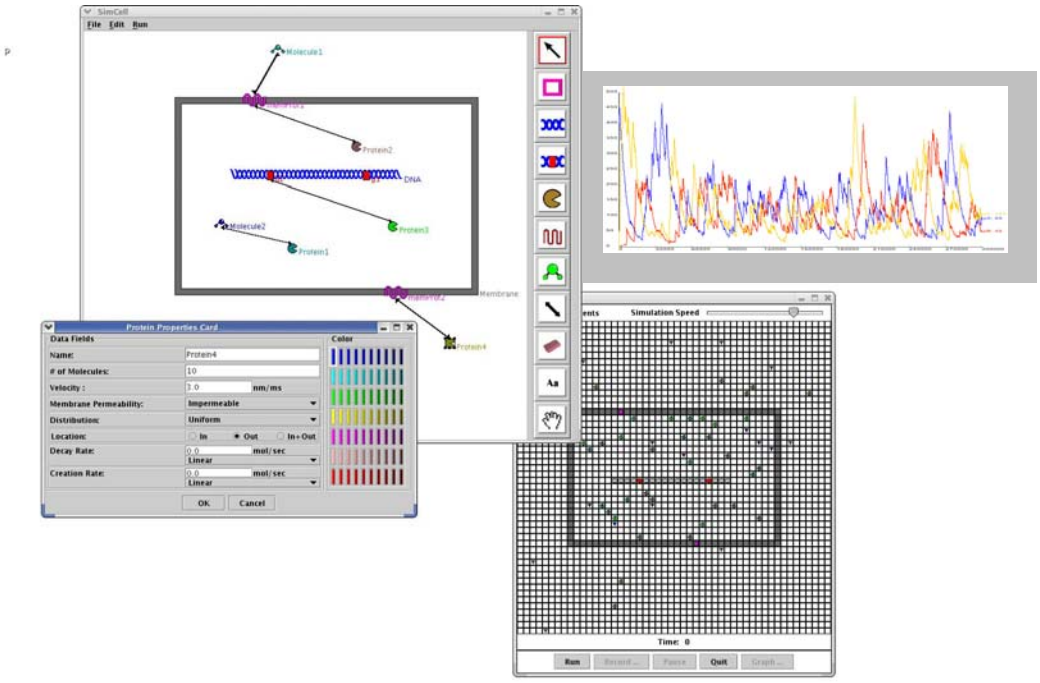
SimCity 2000

The SIMS

Dynamic Cellular Automata

- A novel method to apply Brownian motion to objects in the Cellular Automata lattice (mimics collisions)
- Takes advantage of the scale-free nature of Brownian motion and the scale-free nature of heterogeneous mixtures to allow simulations to span many orders of time (nanosec to hours) and space (nanometers to meters)

SimCell



SimCell

- **Java application that uses Dynamic Cellular Automata (DCA) to model motions, interactions, transport and transformations at the meso-scale (10^{-8} to 10^{-6} m)**
- **Uses a square, 2D lattice to model processes, lattice squares are equivalent to 3x3 nm regions**
- **Molecular objects are moved randomly and interactions determined according to a set of interaction rules that are only applied when objects are in contact (collision detection)**

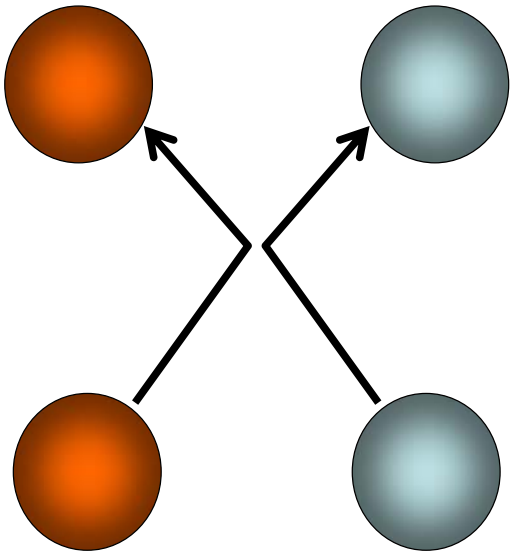
SimCell Interactions

- User defines interaction rules between molecular objects using a simple GUI according to biological observations and measurements
- Interaction rules framed internally as logical boolean operations ('if-then-else' and 'do while') that respect boundaries and cellular barriers

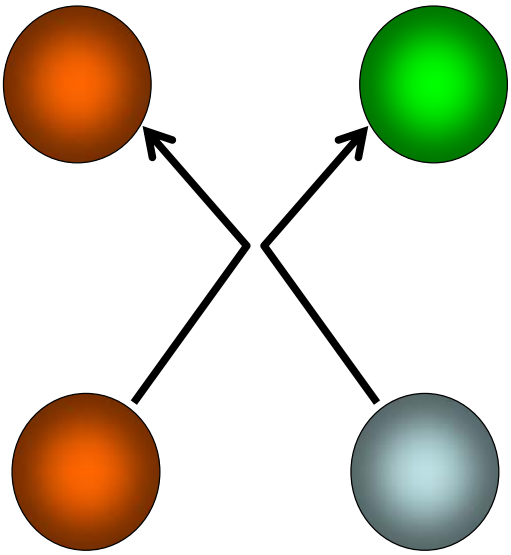
SimCell Interactions

- Five different types of molecules or objects allowed in SimCell: 1) small molecules, 2) soluble proteins, 3) membrane proteins, 4) DNA molecules, and 5) membranes
- Protein-ligand interactions reduced to relatively small number of possibilities
  - Touch and Go (T&G)
  - Bind and Stick (B&S)
  - Transport (TRA)

Touch & Go

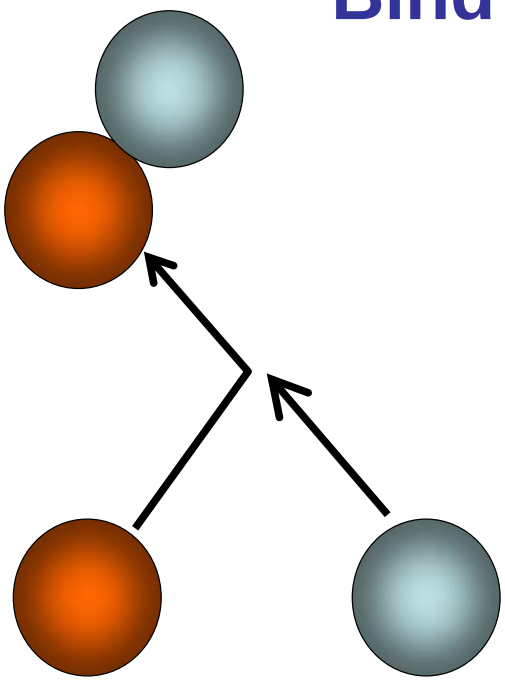


No interaction

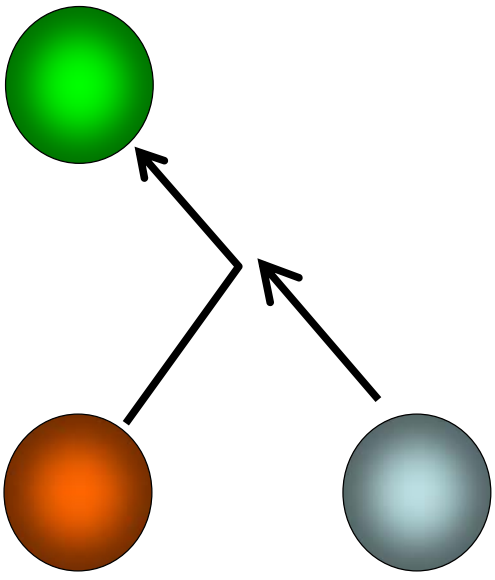


Interaction/catalysis

Bind & Stick

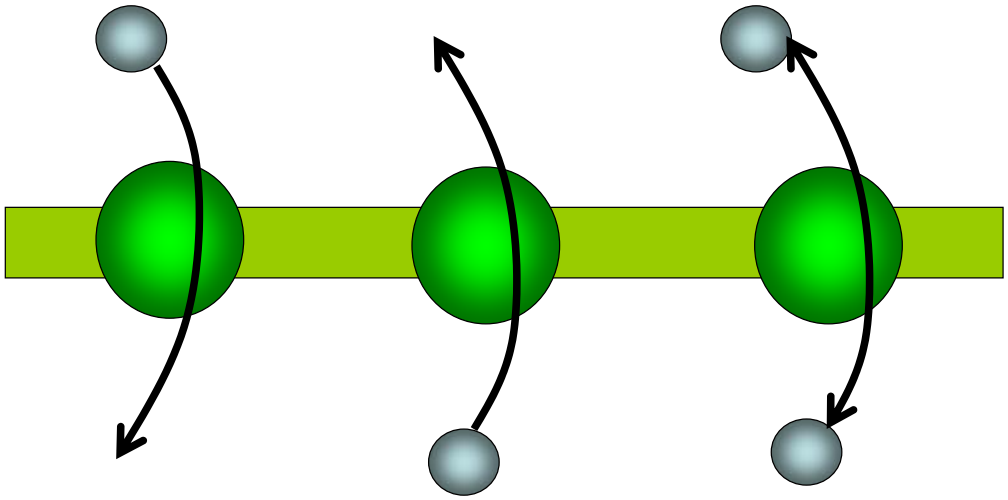


Preserve ID



Interaction/catalysis

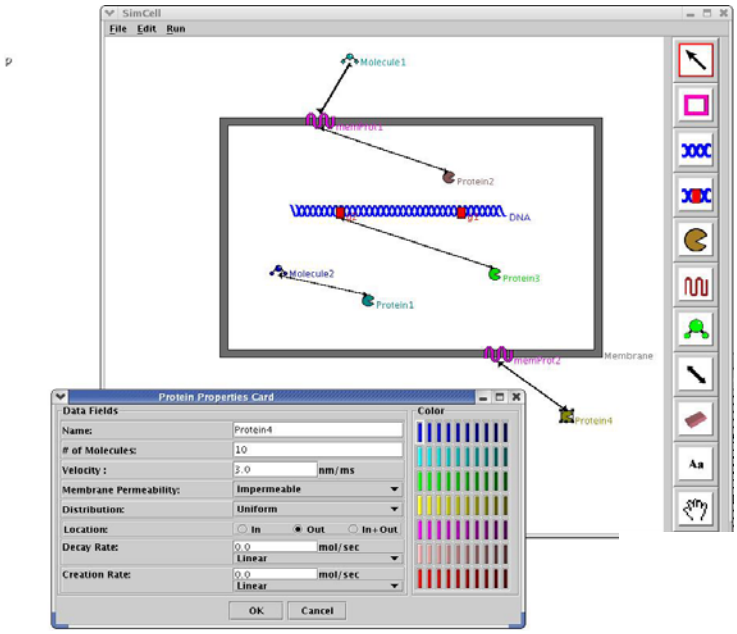
Transport



Interaction Rules

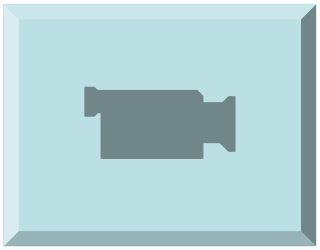
	Small Molecule	Membrane Protein	Soluble Protein	DNA Molecule	Membrane
Small Molecule	T&G |0|-|-|	T&G |0|1|2| B&S |0|-|-| TRA |0|1|2|	T&G |0|1|2| B&S |0|-|-|	T&G |0|-|-|	T&G |0|1|-| TRA |0|1|2|
Membrane Protein	T&G |0|1|2| B&S |0|-|-| TRA |0|1|2|	T&G |0|-|2| B&S |0|-|-|	T&G |0|1|2| B&S |0|-|-| TRA |0|1|2|	N/A	N/A
Soluble Protein	T&G |0|-|2| B&S |0|-|-|	T&G |0|1|2| B&S |0|-|-| TRA |0|1|2|	T&G |0|1|2| B&S |0|-|-|	T&G |0|-|-| B&S |0|1|2|	T&G |0|1|2| B&S |0|-|-| TRA |0|1|2|
DNA Molecule	T&G |0|-|-|	N/A	T&G |0|-|-| B&S |0|1|2|	N/A	N/A
Membrane	T&G |0|1|-| TRA |0|1|2|	N/A	T&G |0|1|2| B&S |0|-|-| TRA |0|1|2|	N/A	N/A

Drawing Interaction Rules with SimCell



Some Examples of SimCell in Action

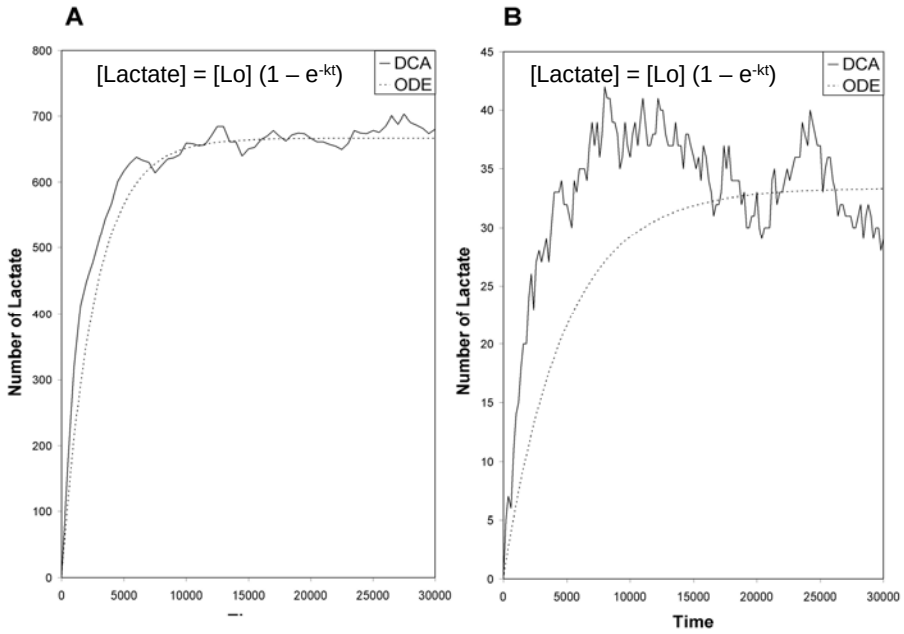
Diffusion in Cytoplasm



Explaining Protein Diffusion in Cells via SimCell

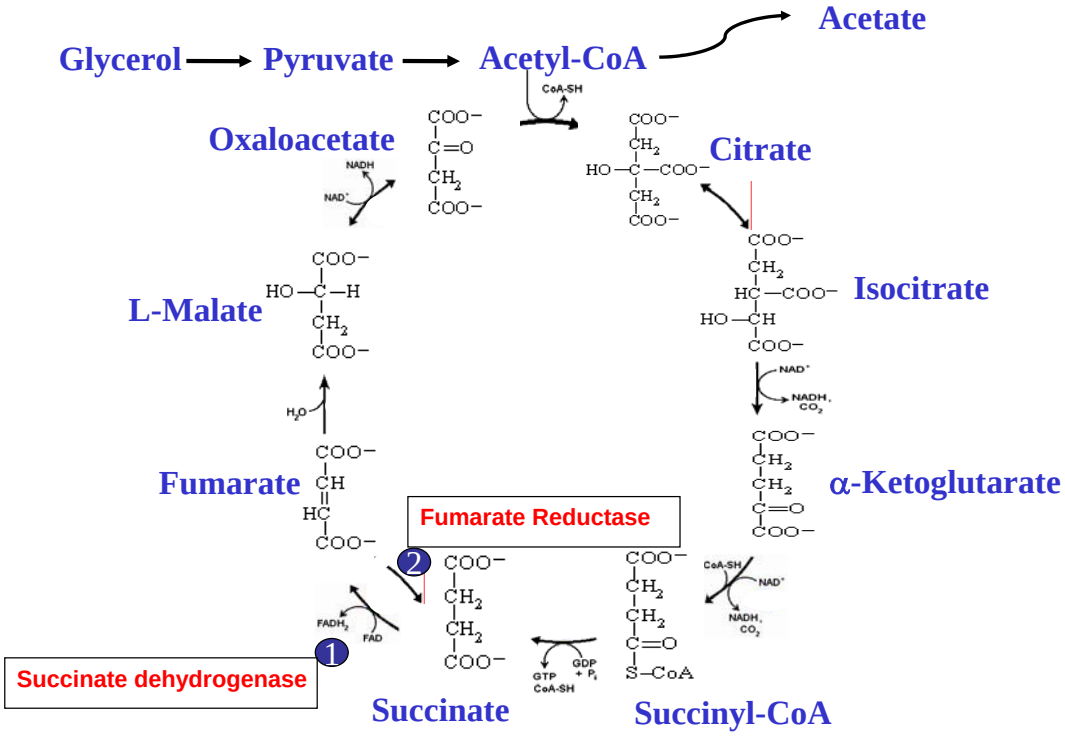
	# Moves	Collision	Sticky	Viscosity	Average Diffusion Rate (nm/ms)		
					0%	25%	50%
A	8	No	No	No	30.4	30.4	30.2
B	9	No	No	No	26.9	26.9	26.7
C	9	No	No	Yes	18.8	18.8	18.6
D	9	Yes	No	No	26.7	20.1	13.4
E	9	Yes	Yes	No	26.7	6.4	1.6
F	9	Yes	Yes	Yes	18.7	4.5	1.1
Observed					27.0	3.0	1.0

Enzyme-Substrate Progress Curves

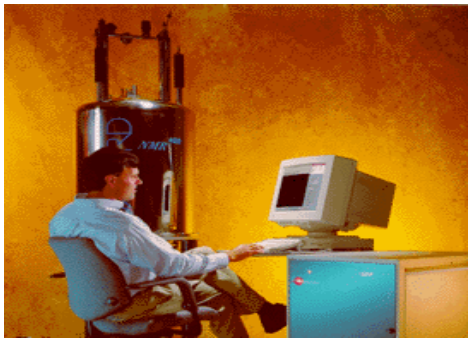


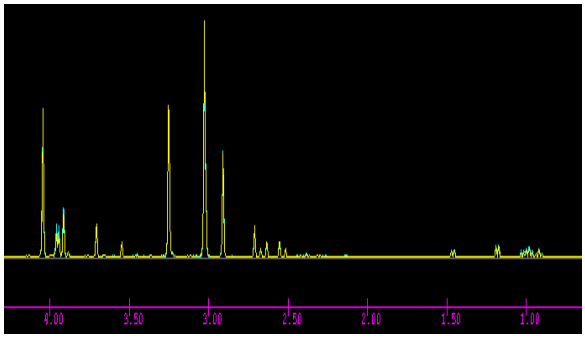
pyruvate + NADH → lactate + NAD+

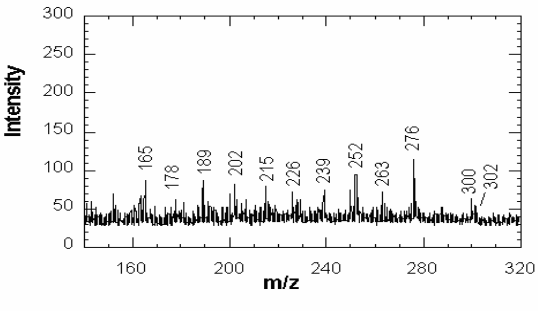
The TCA Cycle & SimCell



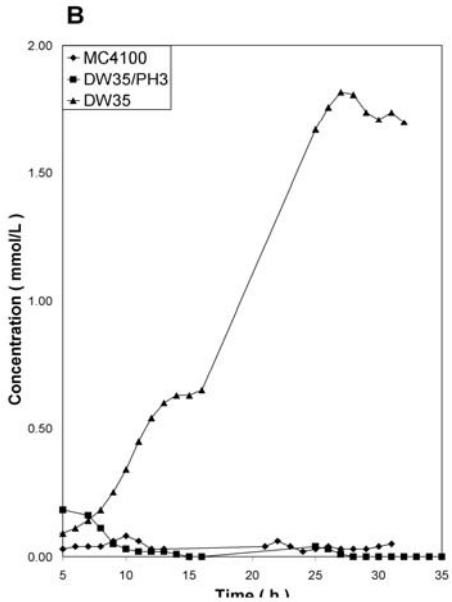
Metabolic Profiling



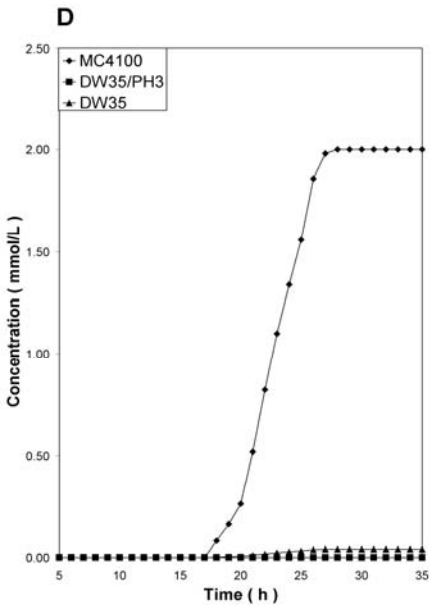




Succinate Production

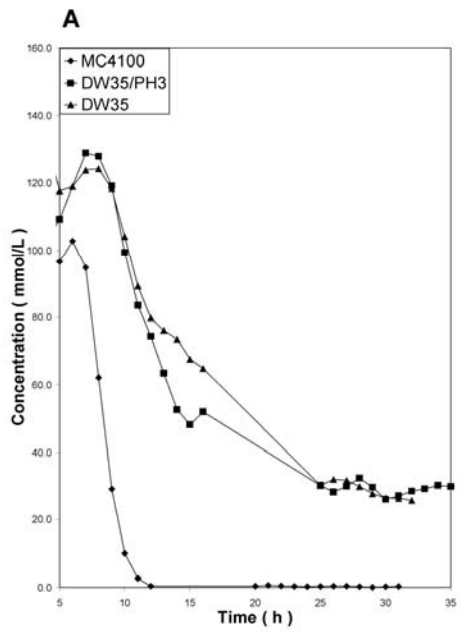


Observed

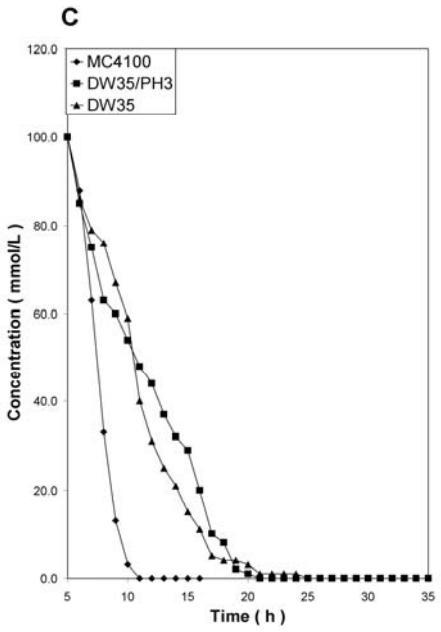


Predicted (SimCell)

Glycerol Consumption

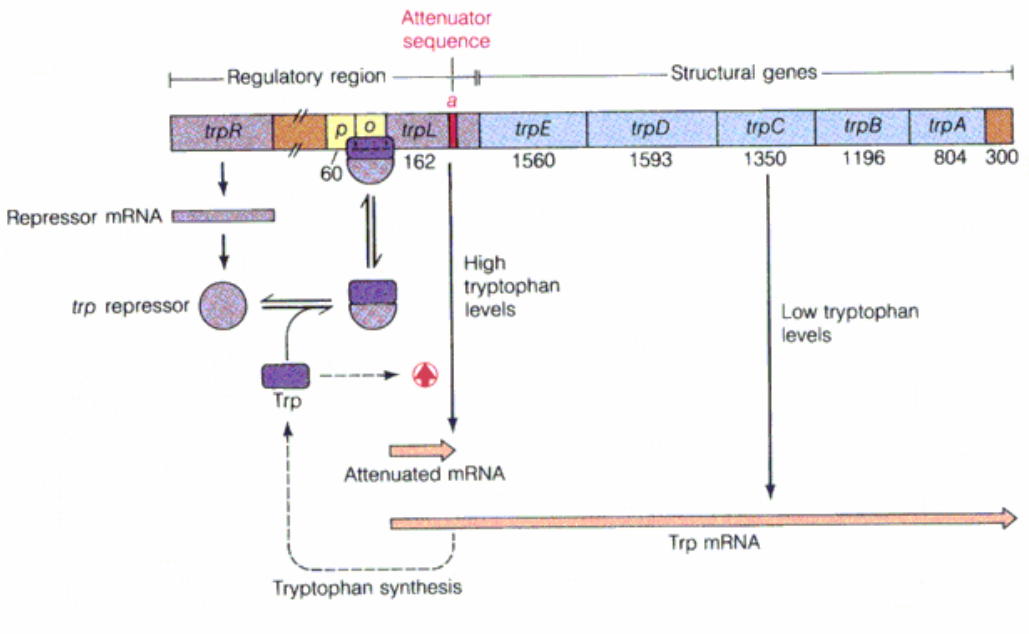


Observed

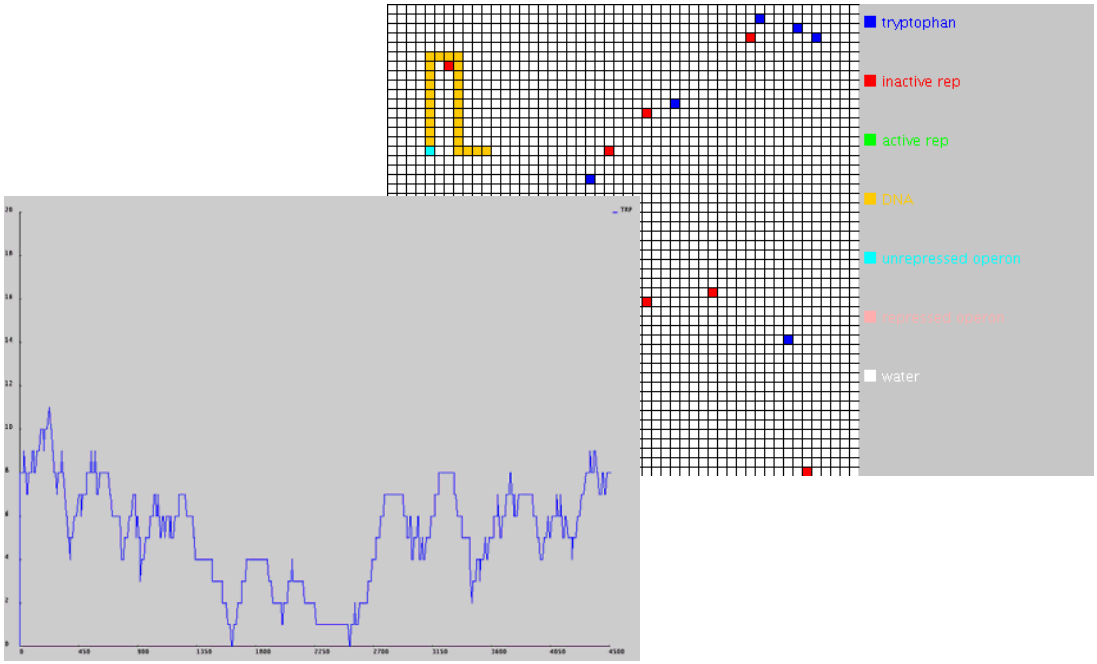


Predicted (SimCell)

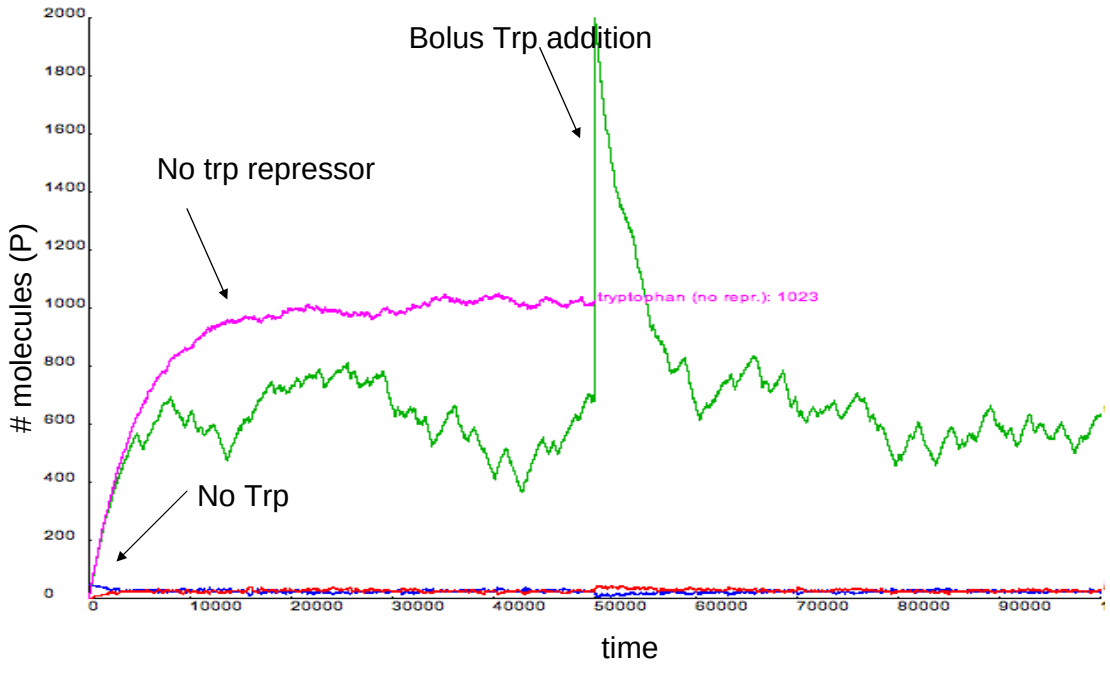
Trp Repressor



CA for Trp Repressor



More Trp Repressor



Repressilator

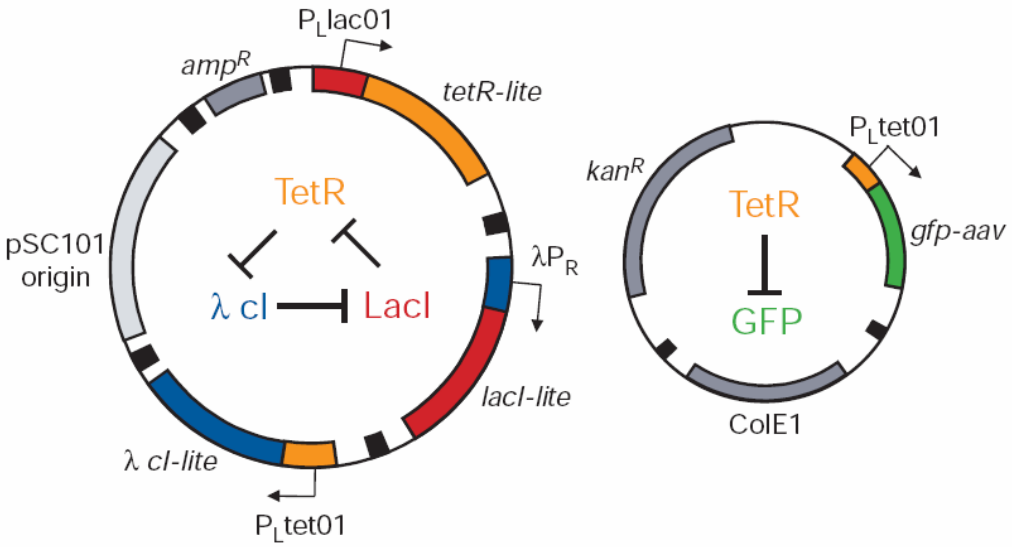
A synthetic oscillatory network of transcriptional regulators

Michael B. Elowitz & Stanislas Leibler

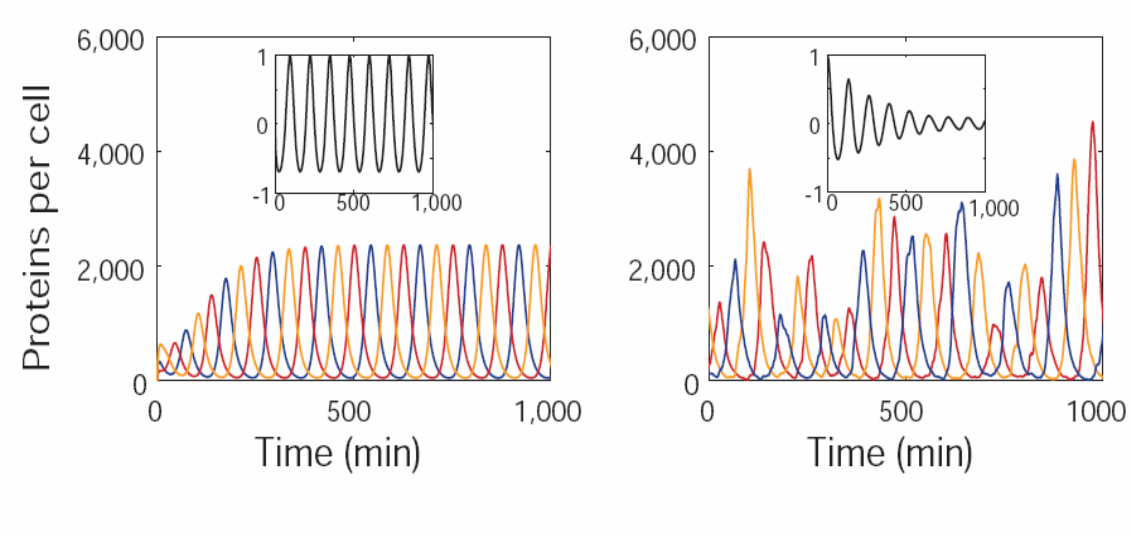
Departments of Molecular Biology and Physics, Princeton University, Princeton, New Jersey 08544, USA

Nature, **403**: 335-338 (2000)

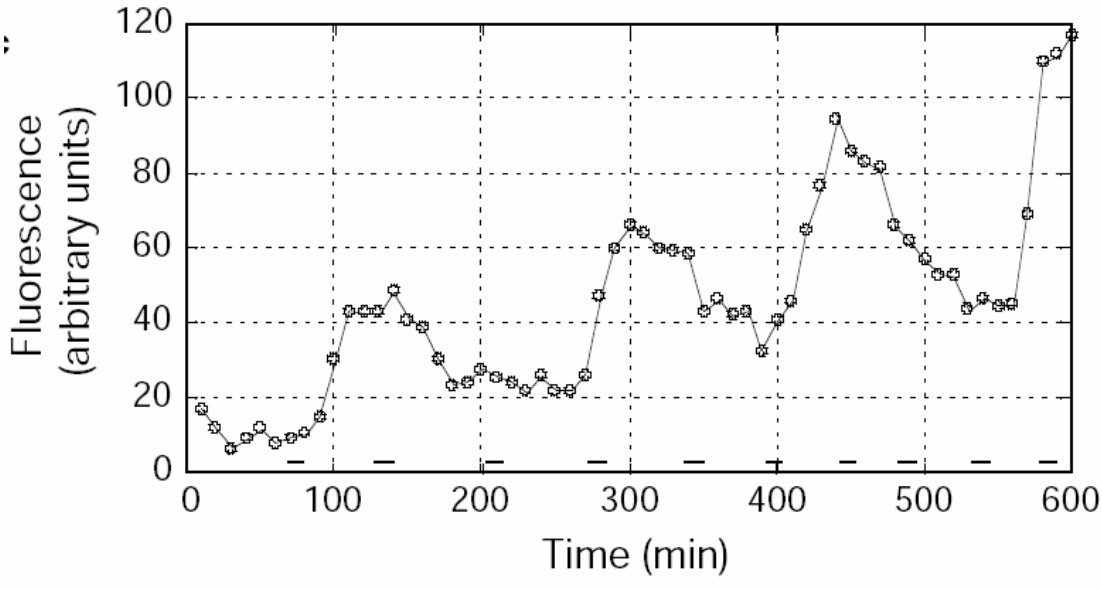
Repressilator



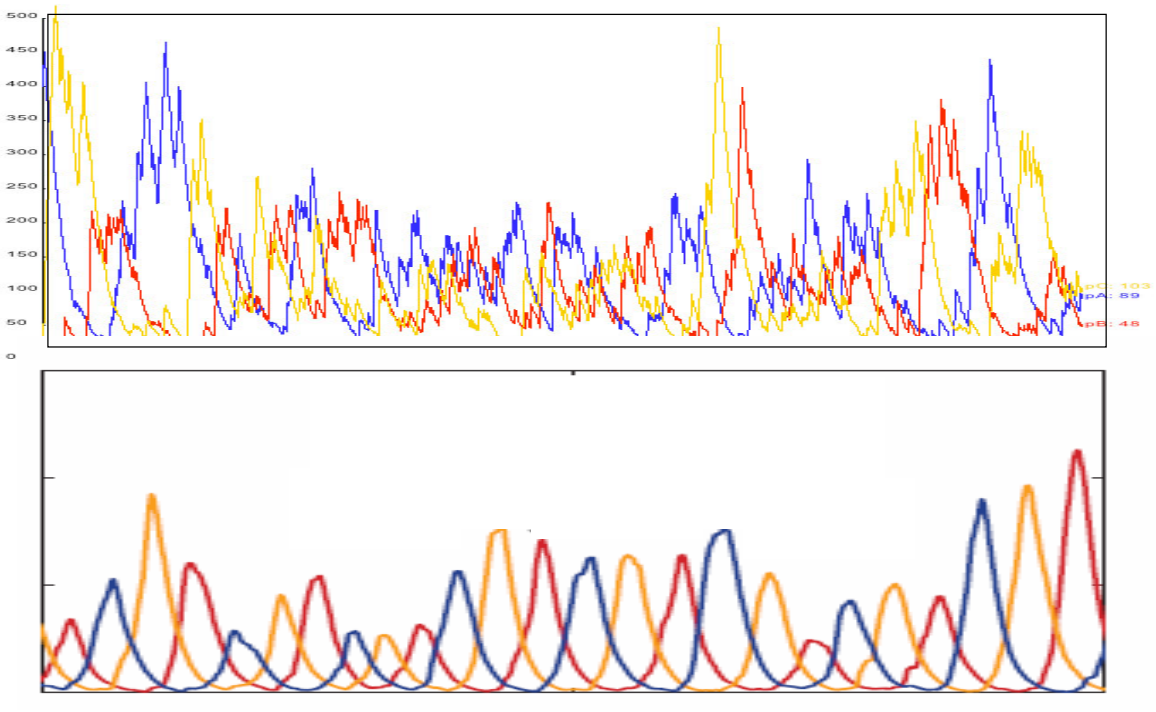
Repressilator



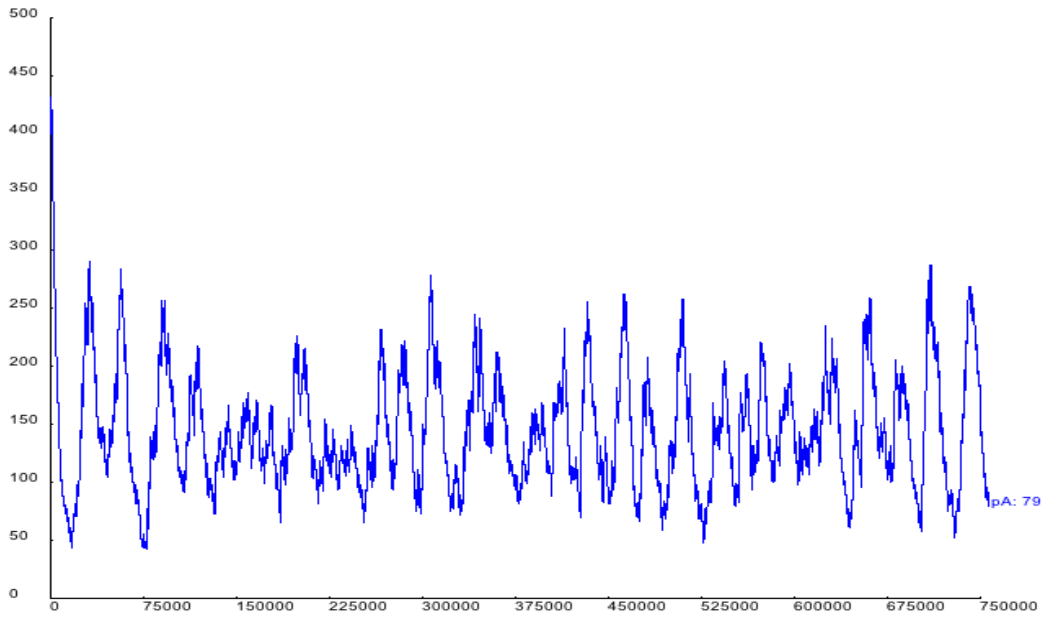
Repressilator



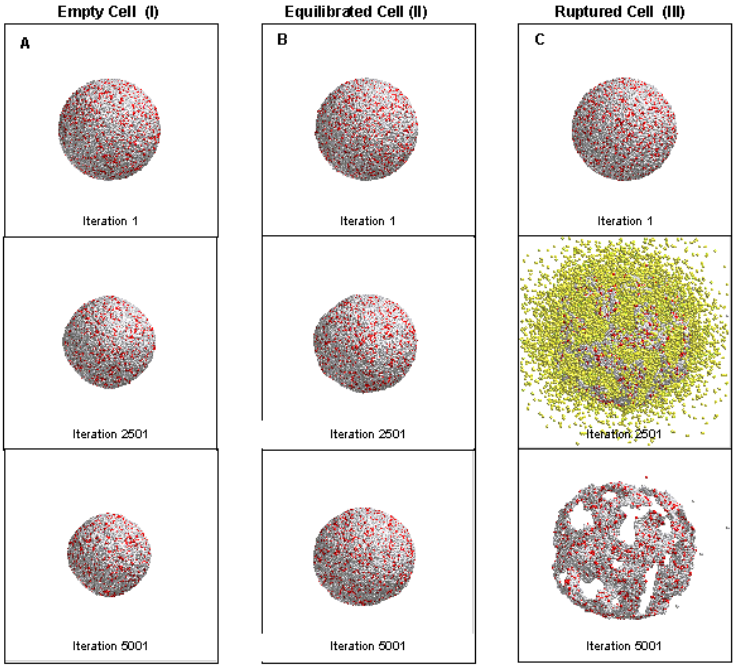
SimCell vs. DE



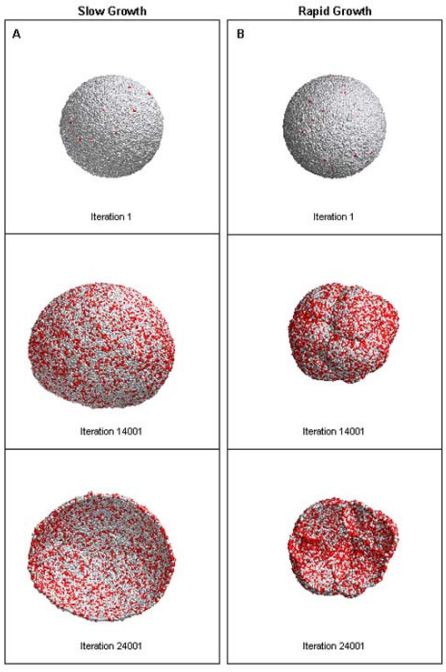
Repressilator Oscillations



Simulating Membranes & Osmotic Shock



Simulating Membrane Growth



SimCell and Cell Simulation

- **Ideal for model checking and validation**
- **Conceptually equivalent to spatially dependent stochastic Petri nets**
- **Universally applicable: Enzyme kinetics, diffusion, excluded volume, binding, vesicle fusion, osmotic lysis, osmotic pressure, genetic circuits, metabolism, transport, repression, signalling, cell division, embryo gene expression… *All from one tool!***

Summary

- **Systems Biology requires the integration of data archiving, experimentation and novel computational approaches**
- **There is a clear need for bioinformatics to step up from the static "stamp collection" phase to thinking about systems in dynamic/interactive/integrated terms**
- **New tools are needed to make this possible – consider DCA & Petri Nets**

Final Exam

- Short answer to long answer format
- Bring calculators
- Typically one question from each of the lectures in the last ½ of the course
- Some questions/answers will involve recall
- Most questions require analysis or some thinking or explaining

Typical Questions

- What is the correlation between protein expression and transcript expression? Provide three reasons to explain the difference
- Describe the algorithm or diagram a flow chart for XXXXX
- Explain the differences and similarities between systems biology and computational biology

Typical Questions

- **Here is some YYYY data from some XXXX experiment – interpret it and explain what it means**
- **Explain the difference between the XXX algorithm and the YYY algorithm. Give some examples or provide an illustration**