

Molecular Networks

Structure

Dynamics

Applications



PIONEER®
A DUPONT COMPANY

Modeling in biology

- Mathematical biology
 - Biology: molecular not mathematical
 - Theoretical biology – any practical value?
- Engineering: artificial networks
 - Examples
 - Oscillations (Elowitz 2000)
 - Toggle switch (Gardner 2000)
 - Designed, implemented, characterized
- Discovery: models go mainstream
 - von Dassow 2000: robustness of fly development
 - Elowitz 2002: intrinsic noise of gene expression



Molecular networks?

- Generalized mass action model
 - Chemical kinetics legacy
 - Time evolution of molecule numbers
 - Instantaneous interactions between molecules
 - Rate of reaction depends on the molecule numbers
 - Encompass
 - Gene regulations
 - Metabolic pathway
 - Signal transduction cascades
- Other models of gene expression
 - Causal networks, Bayesian networks, Boolean networks
 - Neural networks



Outline

1. Structure of molecular networks
2. Dynamics of molecular networks
3. Applications



I. Structure

From artistic to textual to
hierarchical representations of
molecular interactions



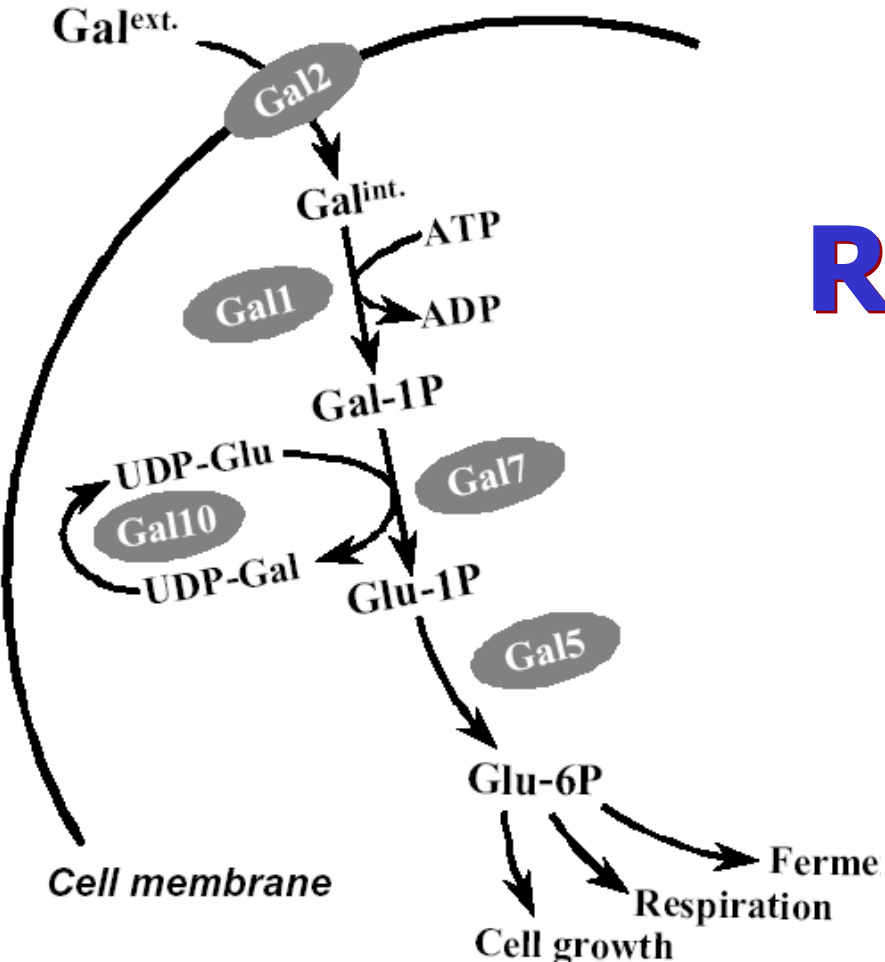
PIONEER®
A DUPONT COMPANY

Yeast Galactose Switch

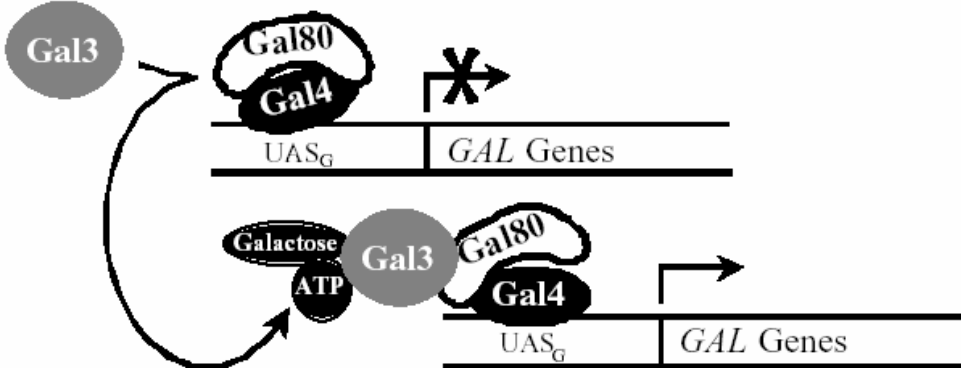
- Pathway ON if Galactose and no Glucose
- Four levels of integration
 - Environmental parameters: Glu & Gal
 - Signal transduction mechanism
 - Metabolic pathway
 - Protein-Protein interactions
 - Regulated gene expression
- Model system
 - Extensively described
 - Large number of mutants and observation



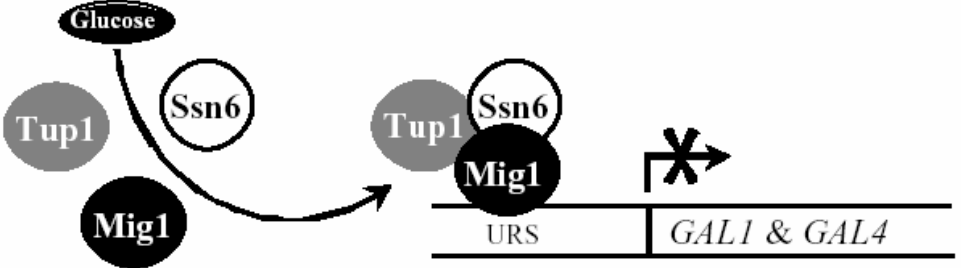
Traditional Representation



Galactose induction



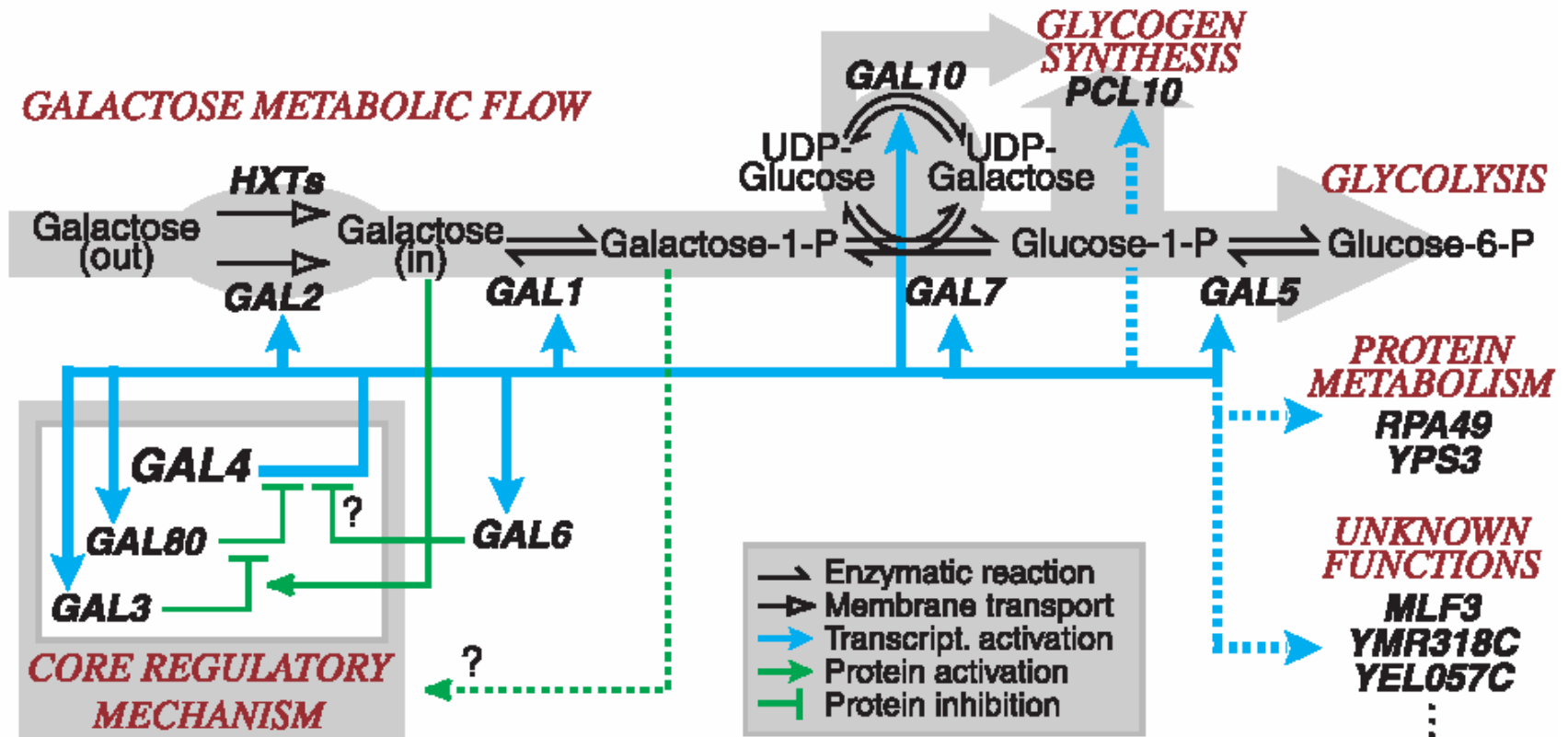
Glucose repression



Ostergaard,S (2000) Nat. Biotech.



Cybernetic Representation



Ideker, T (2001) Science



Stoichiometry

- Chemical equations

- Suitable for limited number of interactions
- Convenient for verification
- Difficult to get a global perspective

- Software

- Gepasi,
- WinScamp, Jarnac

```
E01: GAL-4 --> GAL-4 + Gal80p Kf=2
E02: GAL-4 --> GAL-4 + Gal3p Kf=5
E03: GAL-4-80-3 --> GAL-4-80-3 + Gal80p Kf=5
E04: GAL-4-80-3 --> GAL-4-80-3 + Gal3p Kf=7
E05: gal4g --> gal4g + Gal4p Kf=9
E06: GAL-4-80-3 --> GAL-4-80-3 + Gal2 Kf=1
E07: GAL-4-80-3 --> GAL-4-80-3 + Gal1 Kf=1
E08: GAL-4-80-3 --> GAL-4-80-3 + Gal7 Kf=1
E09: GAL-4-80-3 --> GAL-4-80-3 + Gal10 Kf=1
E10: GAL-4-80-3 --> GAL-4-80-3 + Gal5 Kf=1
E11: GAL-4 --> Gal5 + GAL-4 Kf=10
E12: GAL-4 --> GAL-4 + Gal10 Kf=1
E13: GAL-4 --> GAL-4 + Gal7 Kf=1
E14: GAL-4 --> GAL-4 + Gal1 Kf=1
E15: GAL-4 --> GAL-4 + Gal2 Kf=1
ED1: Gal2 --> EnzCost Kf=698
ED2: Gal1 --> EnzCost Kf=47
ED3: Gal7 --> EnzCost Kf=967
ED4: Gal10 --> EnzCost Kf=1056
ED5: Gal5 --> EnzCost Kf=167
F1*: GalExt --> Gal Kf=1
F2*: GluExt --> Glu Kf=10
F3*: Glu --> Flux Kf=1000
F4*: Glu-6P --> Glu Kf=1000
P1: GalExt + Gal2 <-> ES1 Kf=8539 Kr=2182
P2: ES1 --> Gal + Gal2 Kf=632
P3: Gal + Gal1 <-> ES2 Kf=3193 Kr=7358
P4: ES2 --> Gal1-1P + Gal1 Kf=133
...
```



Stoichiometry

- Chemical equations

- Suitable for limited number of interactions
- Convenient for verification
- Difficult to get a global perspective

- Software

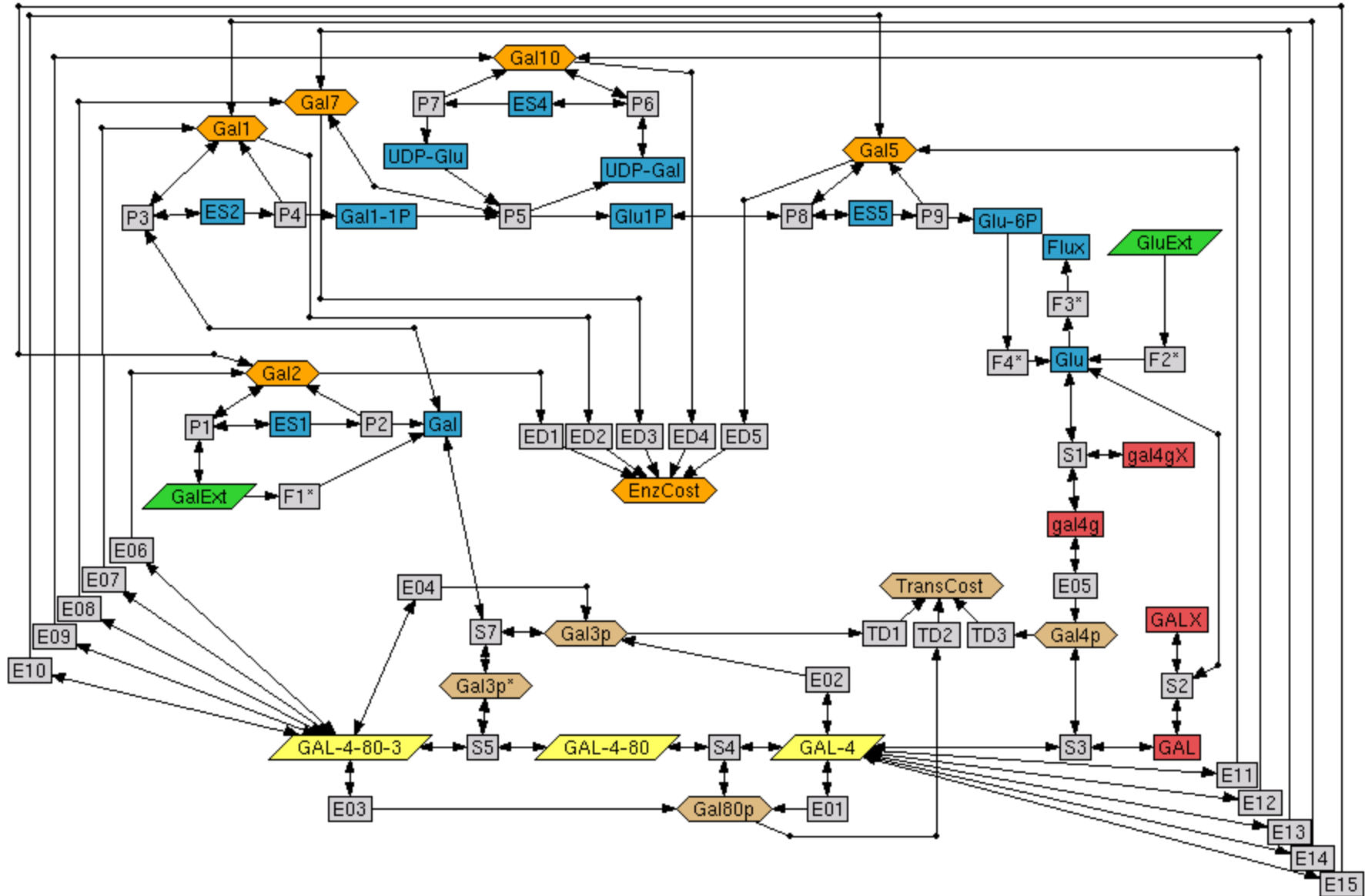
- Gepasi
- WinScamp

Intractable!

```
E01: GAL-4 --> GAL-4 + Gal80p Kf=2
E02: GAL-4 --> GAL-4 + Gal3p Kf=5
E03: GAL-4-80-3 --> GAL-4-80-3 + Gal80p Kf=5
E04: GAL-4-80-3 --> GAL-4-80-3 + Gal3p Kf=7
E05: gal4g --> gal4g + Gal4p Kf=9
E06: GAL-4-80-3 --> GAL-4-80-3 + Gal2 Kf=1
E07: GAL-4-80-3 --> GAL-4-80-3 + Gal1 Kf=1
E08: GAL-4-80-3 --> GAL-4-80-3 + Gal7 Kf=1
E09: GAL-4-80-3 --> GAL-4-80-3 + Gal10 Kf=1
E10: GAL-4-80-3 --> GAL-4-80-3 + Gal5 Kf=1
E11: GAL-4 --> Gal5 + GAL-4 Kf=10
E12: GAL-4 --> GAL-4 + Gal10 Kf=1
E13: GAL-4 --> GAL-4 + Gal7 Kf=1
E14: GAL-4 --> GAL-4 + Gal1 Kf=1
E15: GAL-4 --> GAL-4 + Gal2 Kf=1
ED1: Gal2 --> EnzCost Kf=698
ED2: Gal1 --> EnzCost Kf=47
ED3: Gal7 --> EnzCost Kf=967
ED4: Gal10 --> EnzCost Kf=1056
ED5: Gal5 --> EnzCost Kf=167
t: GalExt --> Gal Kf=1
... --> Glu Kf=10
... --> Flux Kf=1000
... --> Glu Kf=1000
GalExt + Gal2 <-> ES1 Kf=8539 Kr=2182
ES1 --> Gal + Gal2 Kf=632
P3: Gal + Gal1 <-> ES2 Kf=3193 Kr=7358
P4: ES2 --> Gal1-1P + Gal1 Kf=133
...
```



Flat diagram



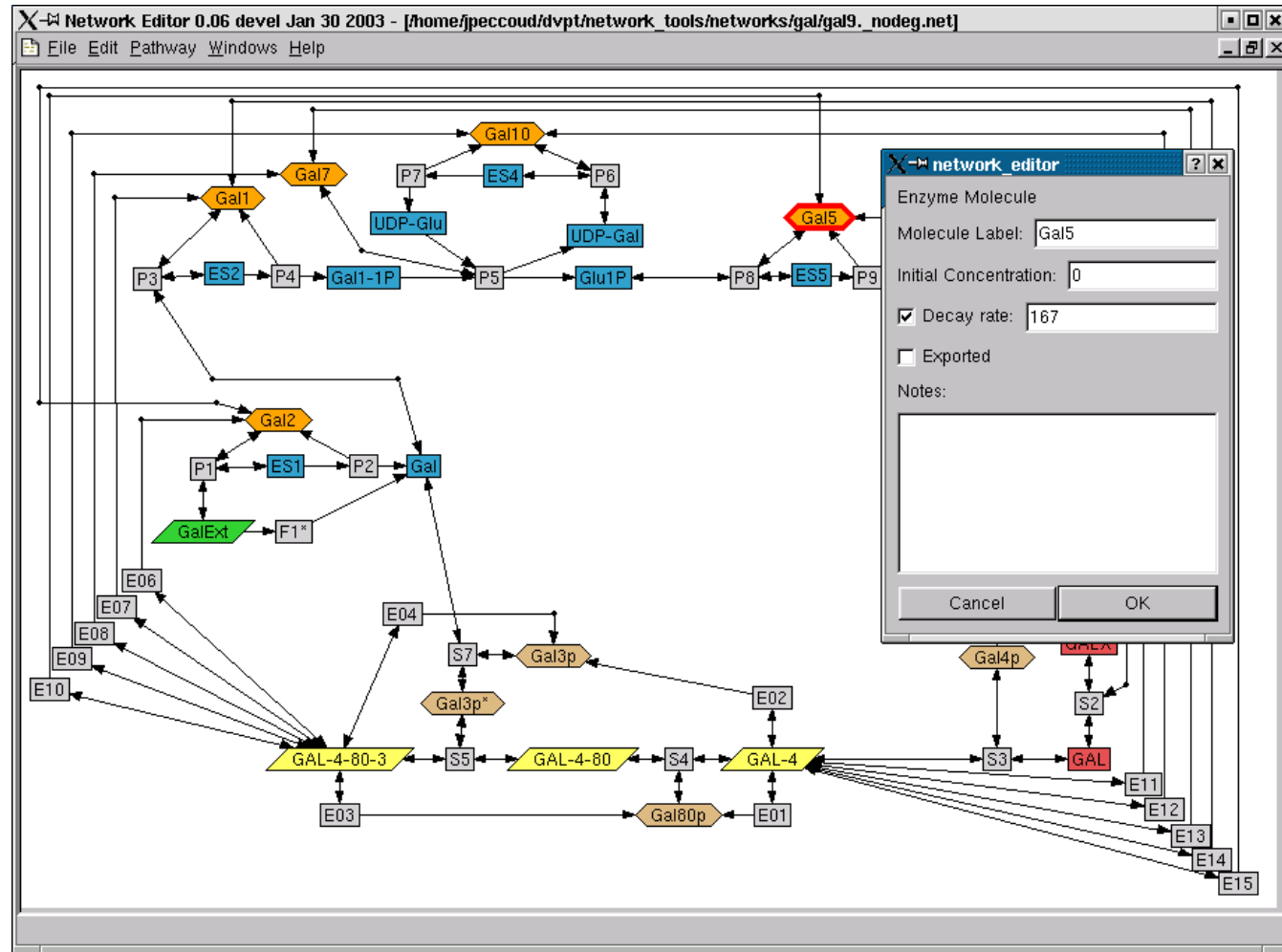
Flat diagram

- Editing
 - Biologists are used to diagrams
 - Flexibility: variables and interactions
- Visualization
 - Global perspective on the network
 - Color coding of various components
- Limitations
 - Size of the monitor
 - Connections between remote objects
 - The logic of the network gets lost
 - Need to simplify the graphical representation



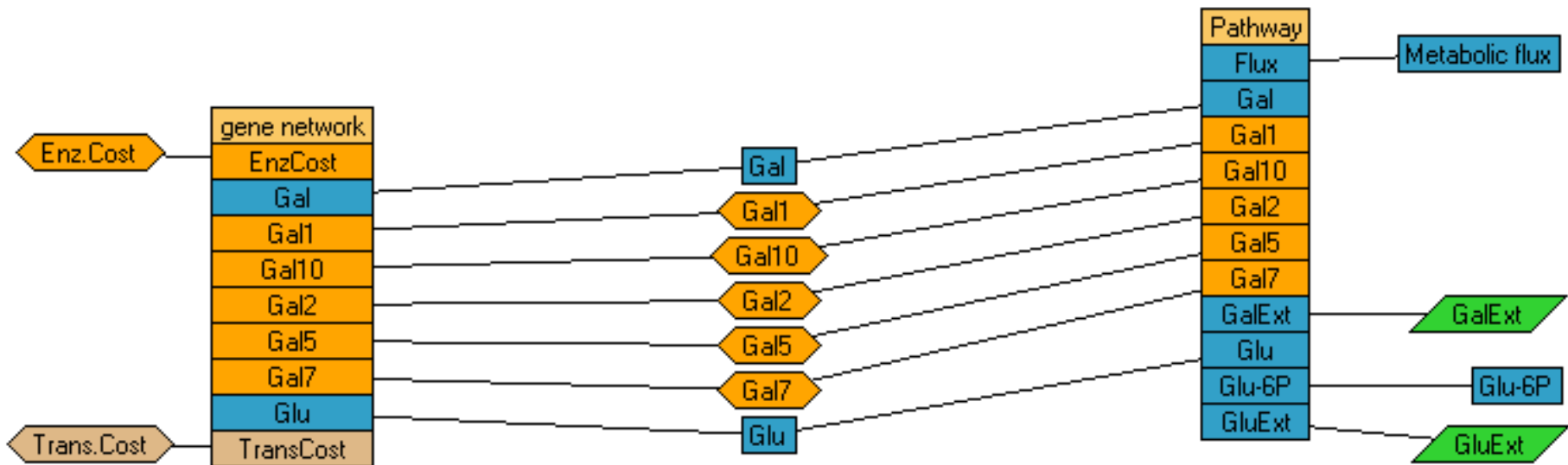
Protein degradation

Degradation reactions are specified in the molecule dialog box



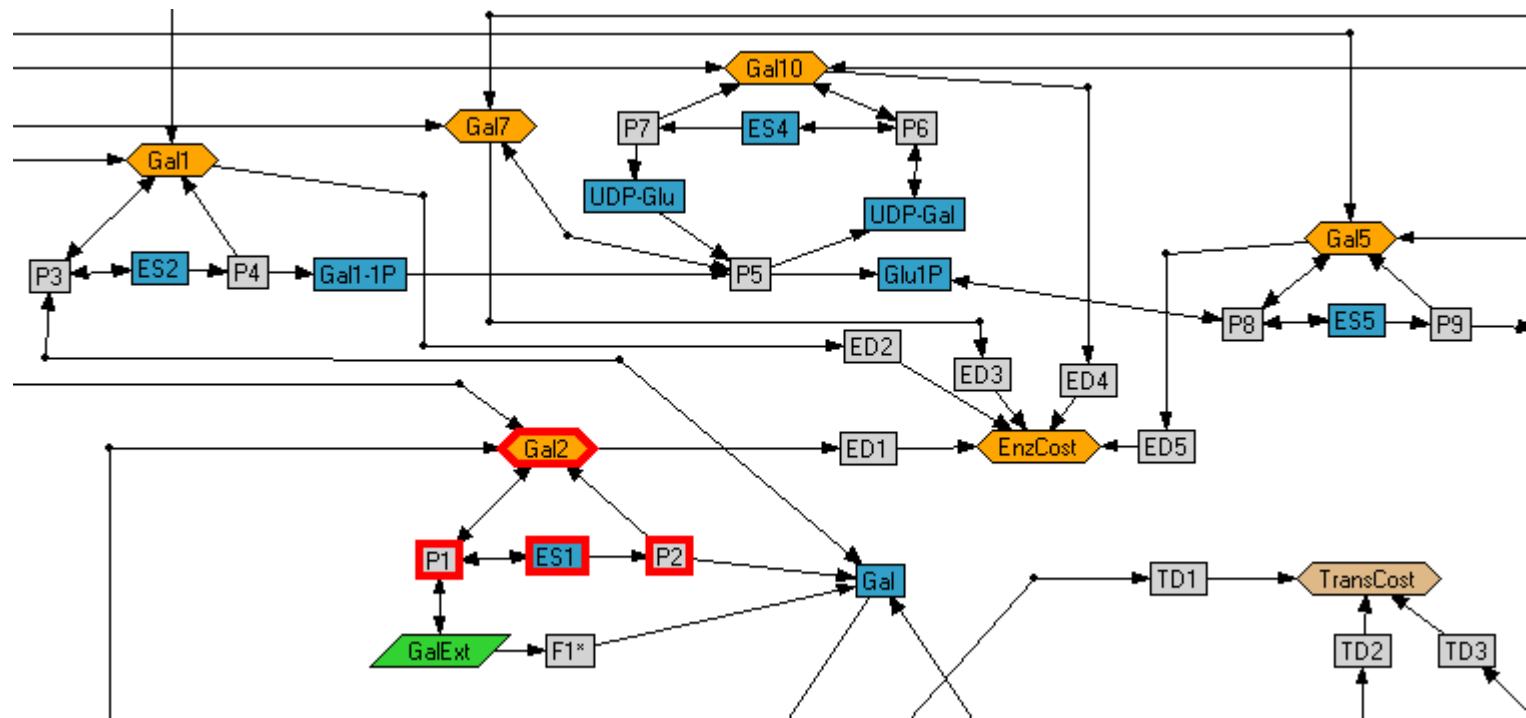
Sub-networks

- Identify I/O (exported variables)
- Save network in file
- Use sub-network to build composed model



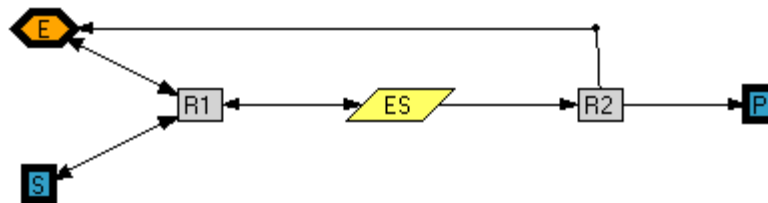
Canonic Mechanisms

- Enzymatic reactions
 - Mechanism: $E + S \leftrightarrow ES \rightarrow E + P$
 - Save mechanism in one file
 - Use it to build pathways
- Gene inductions, repression, etc...



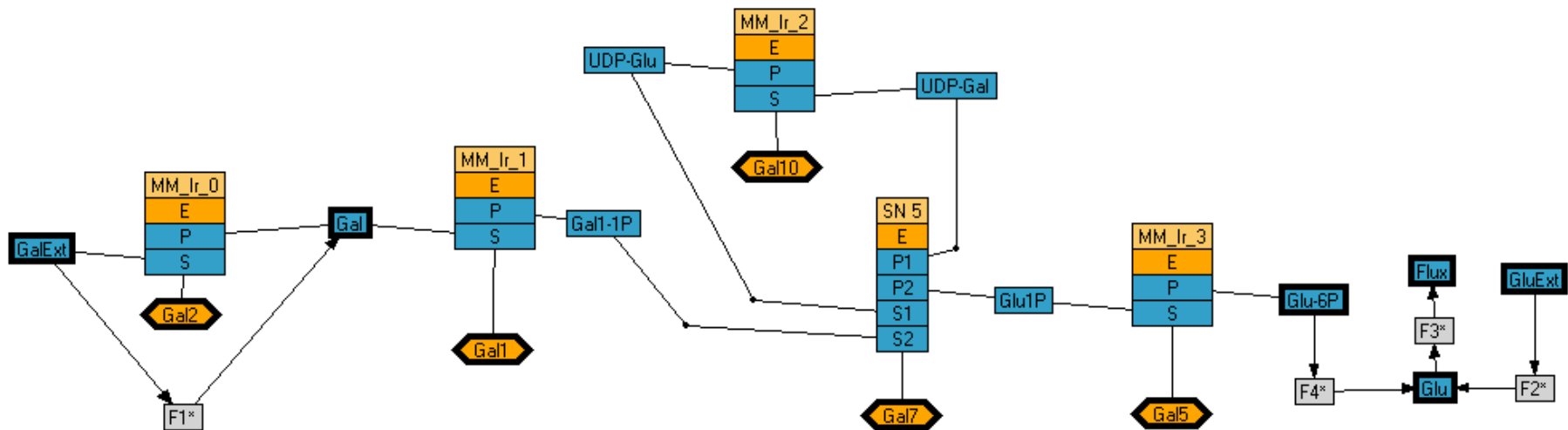
Canonical Mechanisms

- Enzymatic reactions
 - Mechanism: $E + S \rightleftharpoons ES \rightarrow E + P$
 - Save mechanism in one file
 - Use it to build pathways
- Gene inductions, repression, etc...



Canonical Mechanisms

- Enzymatic reactions
 - Mechanism: $E + S \rightleftharpoons ES \rightarrow E + P$
 - Save mechanism in one file
 - Use it to build pathways
- Gene inductions, repression, etc...



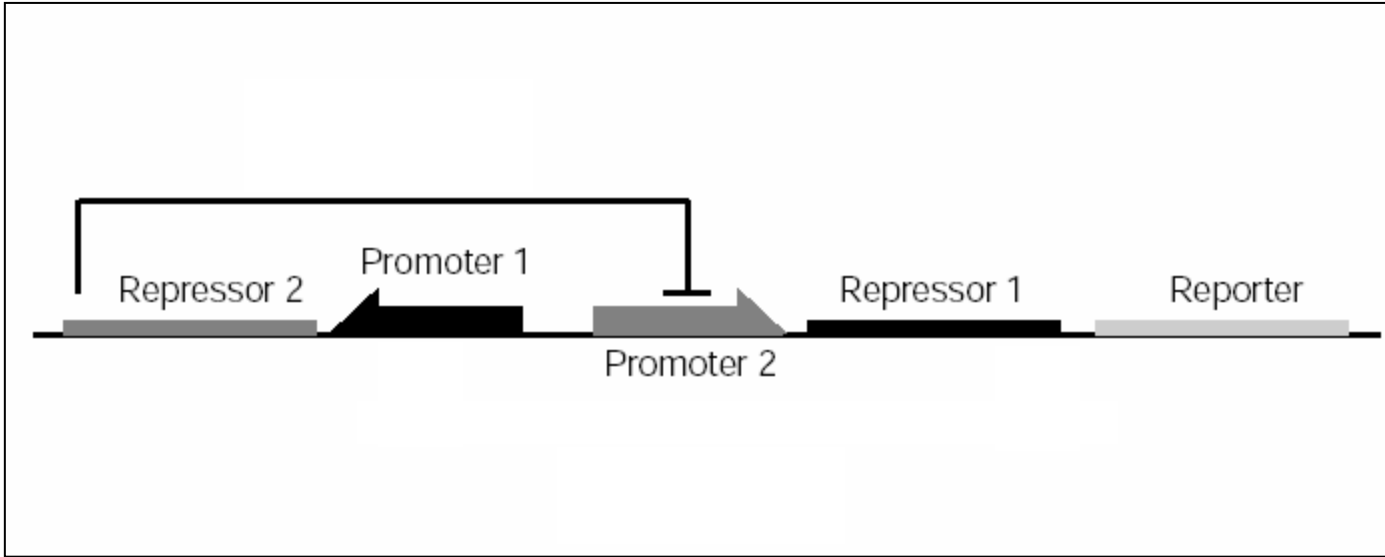
II. Dynamics

Analyzing the time-evolution of
molecular networks



PIONEER®
A DUPONT COMPANY

A Bi-stable Switch

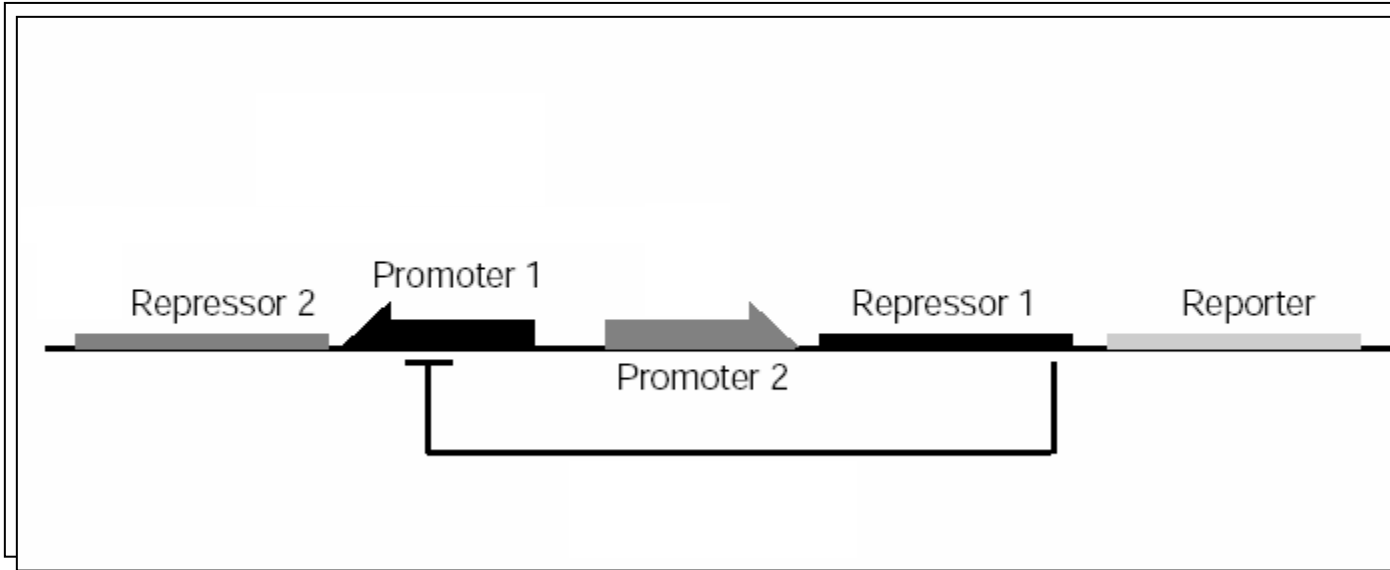


- Two repressors repressing each other leads to bistability
- E. Coli
 - LacI + TetR
 - LacI + cI



(Gardner, 2000)

A Bi-stable Switch

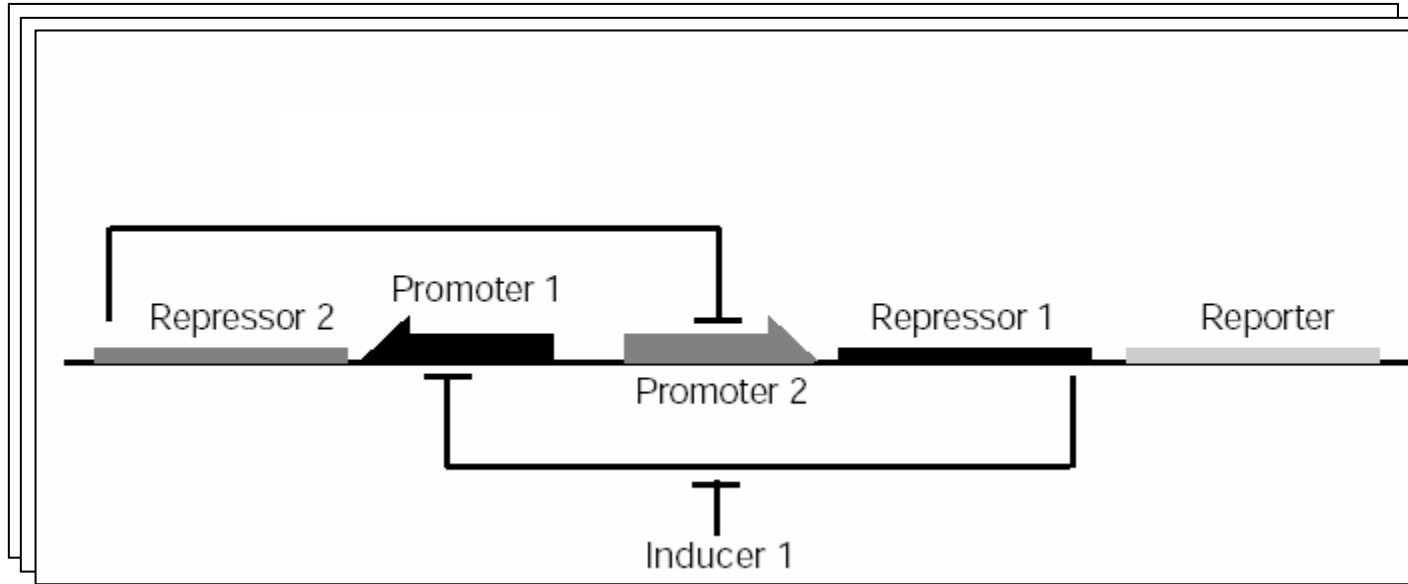


- Two repressors repressing each other leads to bistability
- E. Coli
 - LacI + TetR
 - LacI + cI



(Gardner, 2000)

A Bi-stable Switch

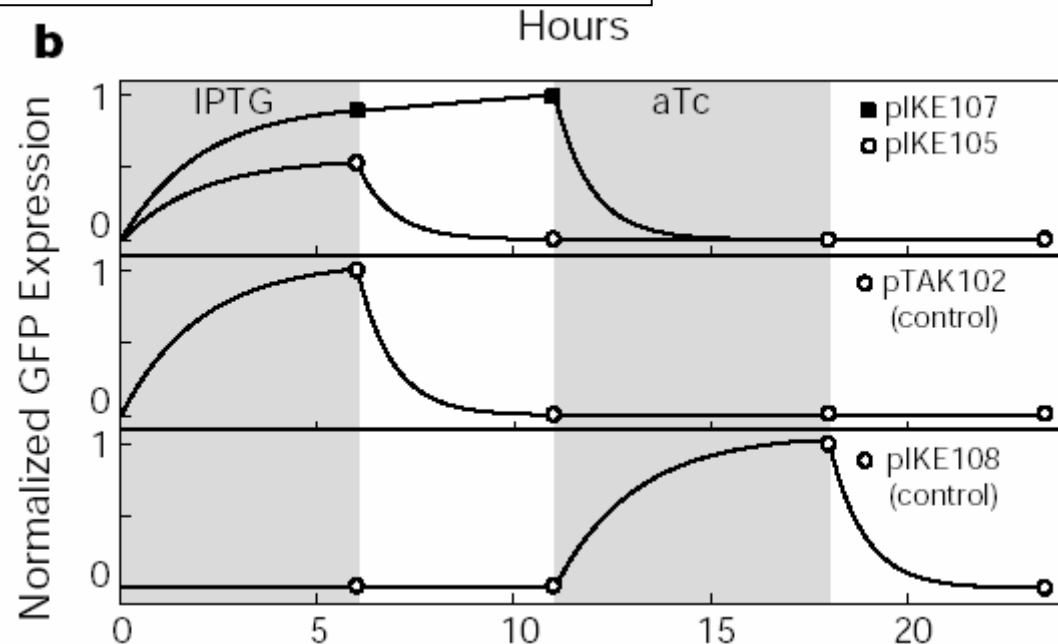


- Two repressors repressing each other leads to bistability

- E. Coli

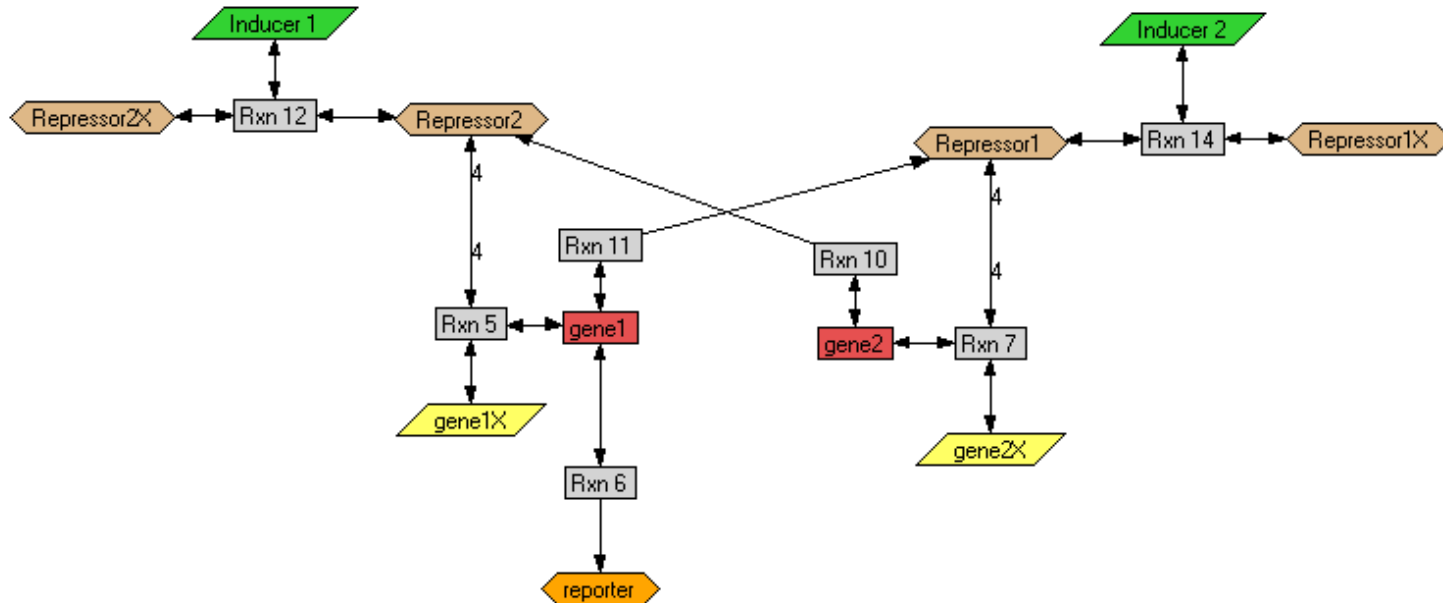
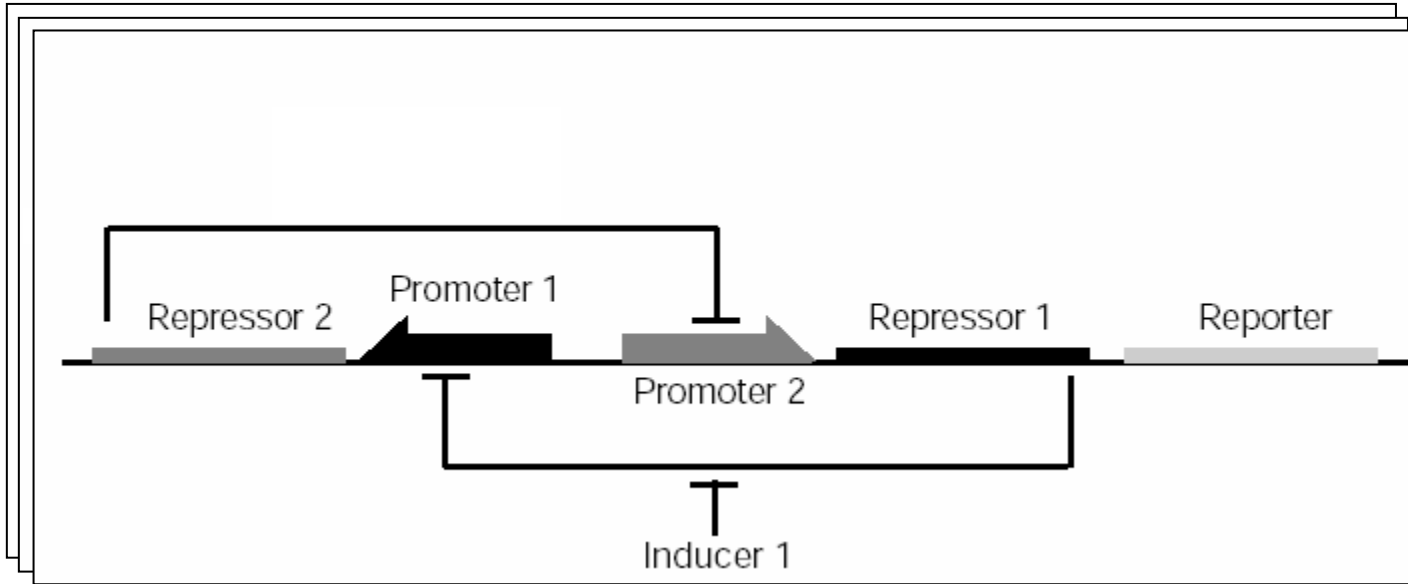
- LacI + TetR

- LacI + cI



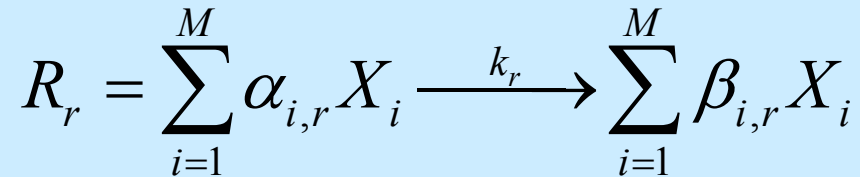
(Gardner, 2000)

A Bi-stable Switch



Chemical kinetics

- Mass-action rate law
 - Rate depends on the reactant concentrations
 - Stoichiometric matrices
- Control variables
 - Environment
 - Operator control
- Numerical solutions
 - Speed/Stability
 - CVODE



$$v_r = k_r \prod_{i=1}^M [X_i]^{\alpha_{i,r}}$$

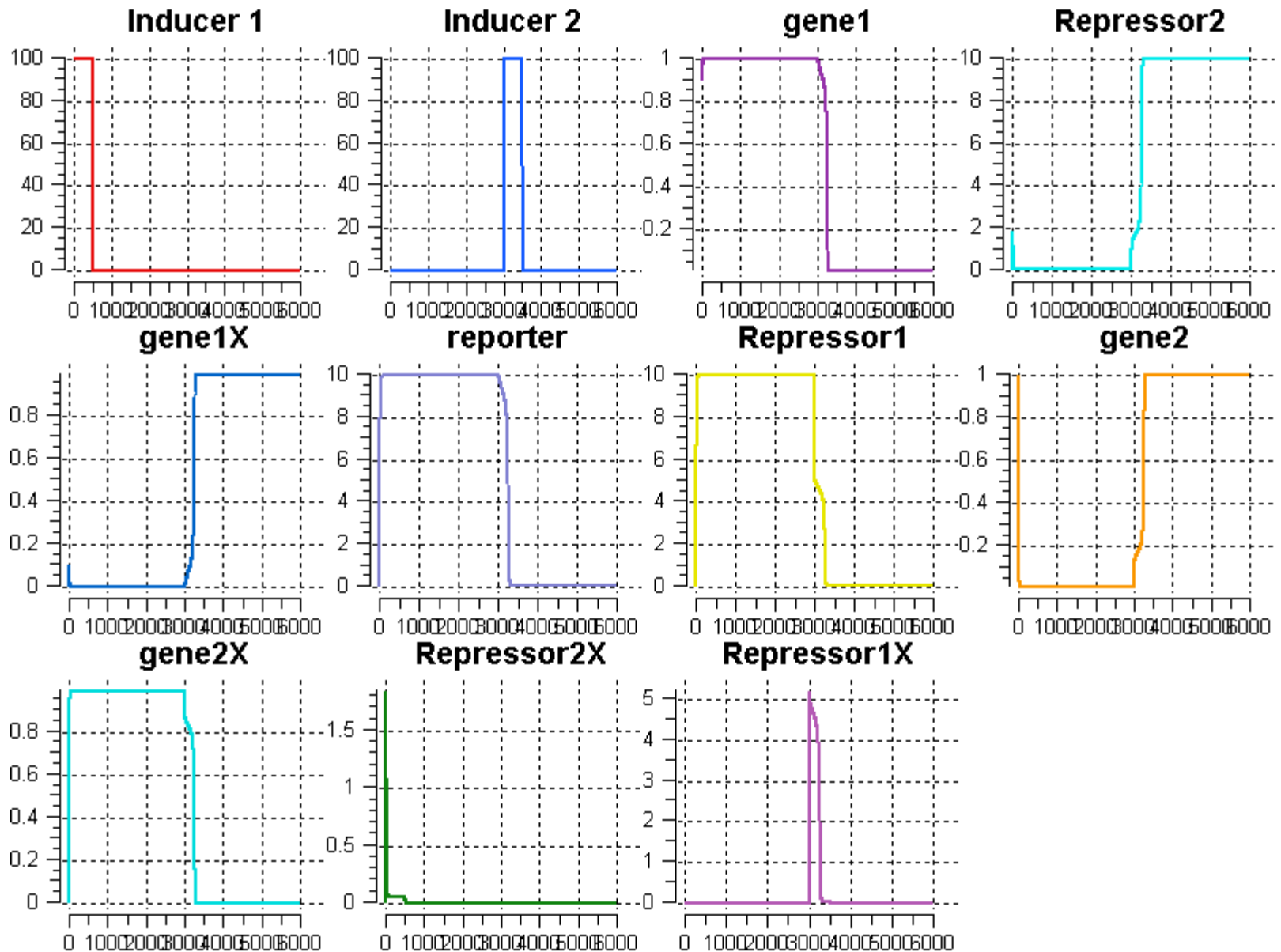
$$\frac{dX_i}{dt} = \sum_r \beta_{i,r} v_r - \sum_r \alpha_{i,r} v_r$$

$$= \sum_r \gamma_{i,r} v_r$$

(Erdi & Toth, 1989)



Simulation results



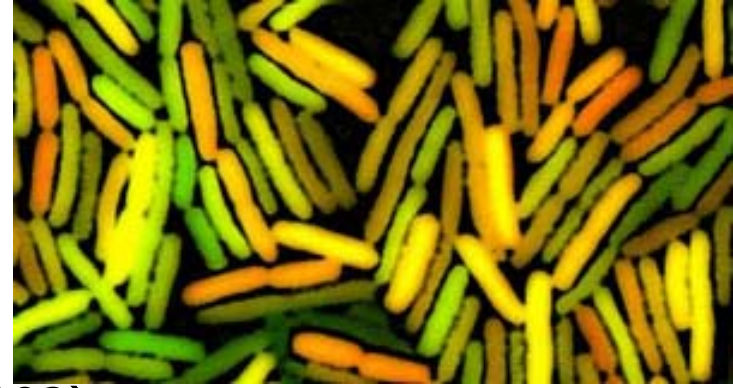
Molecular noise

$$\frac{X}{V} \xrightarrow{X \rightarrow \infty} [X]$$

- Thermodynamic limit
 - Concentrations are limits
 - Fluctuations is expected for small copy numbers
- Molecules in small copy numbers
 - Genes: 1, 2 or more copies / cell (plasmids...)
 - RNAs: TIGR Arabidopsis microarrays
 - 360,000 mRNA copies / cell
 - Rare transcripts: 5-10 copies/cell
 - Moderate: 100-500 copies/cell
 - Abundant: 10,000 copies/cell
 - Proteins
 - 80% of *E. Coli* genes express fewer than 100 copies/cell (Guptasarma, 1995)



Phenotypic noise



- Gene expression
 - Random phenotype (Elowitz, 2002)
 - Fluctuations of protein expression controlled by rates of transcription and translation (Ozbudak, 2002)
- Cellular Differentiation:
 - Lambda lysis/lysogeny decision
 - Lateral specification:
 - Hair cell differentiation in plants (Schiefelbein, 2003)
 - Notch/Delta systems (Artavanis-Tsakonas, 1999)
- Organism level
 - Haploinsufficiency: NF1(Kemkemer, 2002)



Stochastic kinetics

- Stochastic reaction rates

Kinetic constant: volume
when order > 1

Combinatorial term

$$R_r = \sum_{i=1}^M \alpha_{i,r} X_i \xrightarrow{k_r} \sum_{i=1}^M \beta_{i,r} X_i$$
$$\lambda_r(X) = \frac{k_r}{V \sum_{i=1}^M \alpha_{i,r}^{-1}} \prod_{i=1}^M \frac{X_i!}{(X_i - \alpha_{i,r})!}$$
$$\lambda(X) = \sum_r \lambda_r(X)$$

$$0 \leq s < \tau : X_{t=s} = X_{t=0}$$

$$P_X(\tau > s) = e^{-\lambda(X)s}$$

$$P(X_{t=\tau} = X_{t=0} + \gamma_r) = \frac{\lambda_r(X_{t=0})}{\lambda(X_{t=0})}$$

- Jump process dynamics

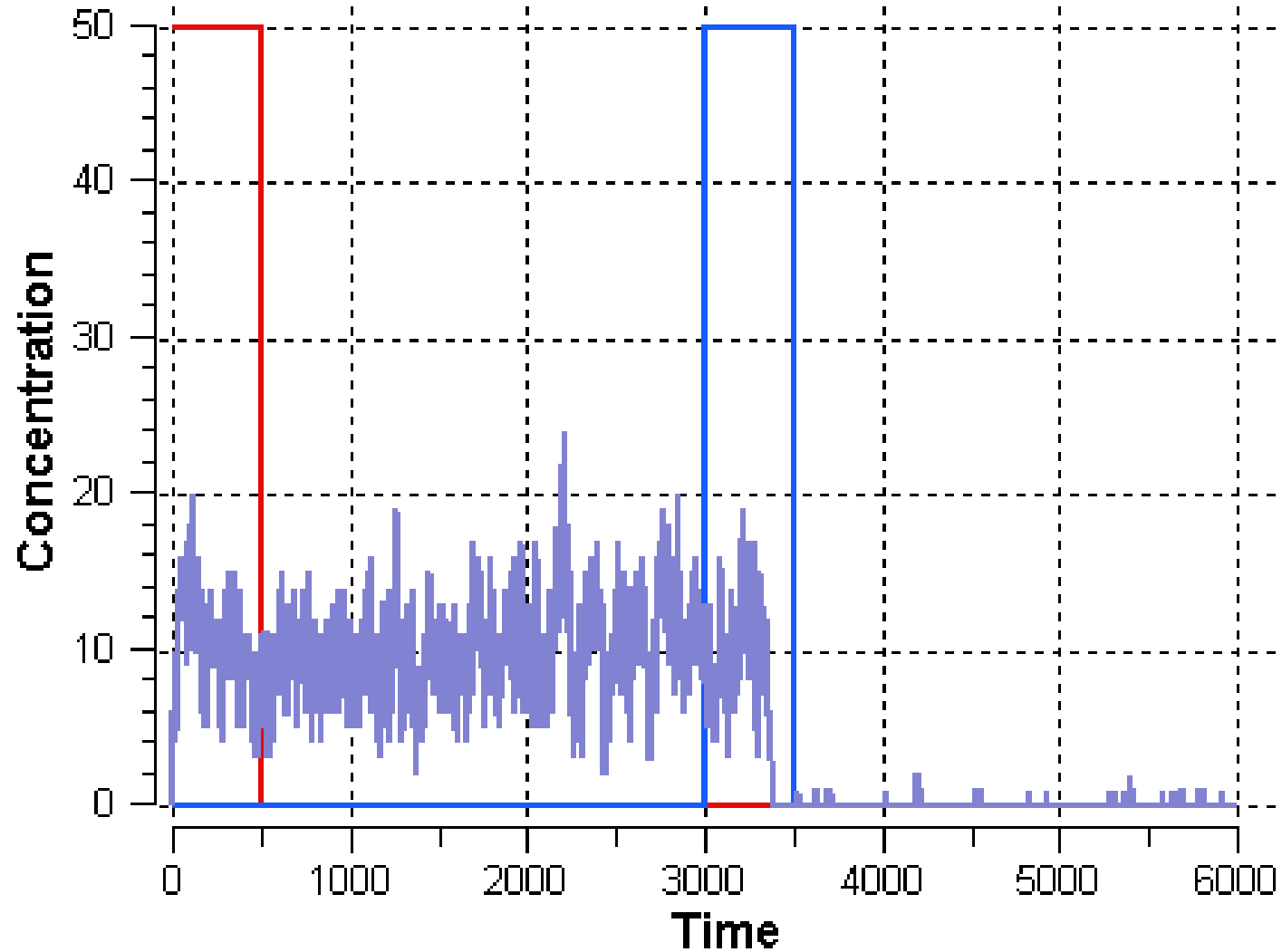
Instant reactions

Next instant of
reaction

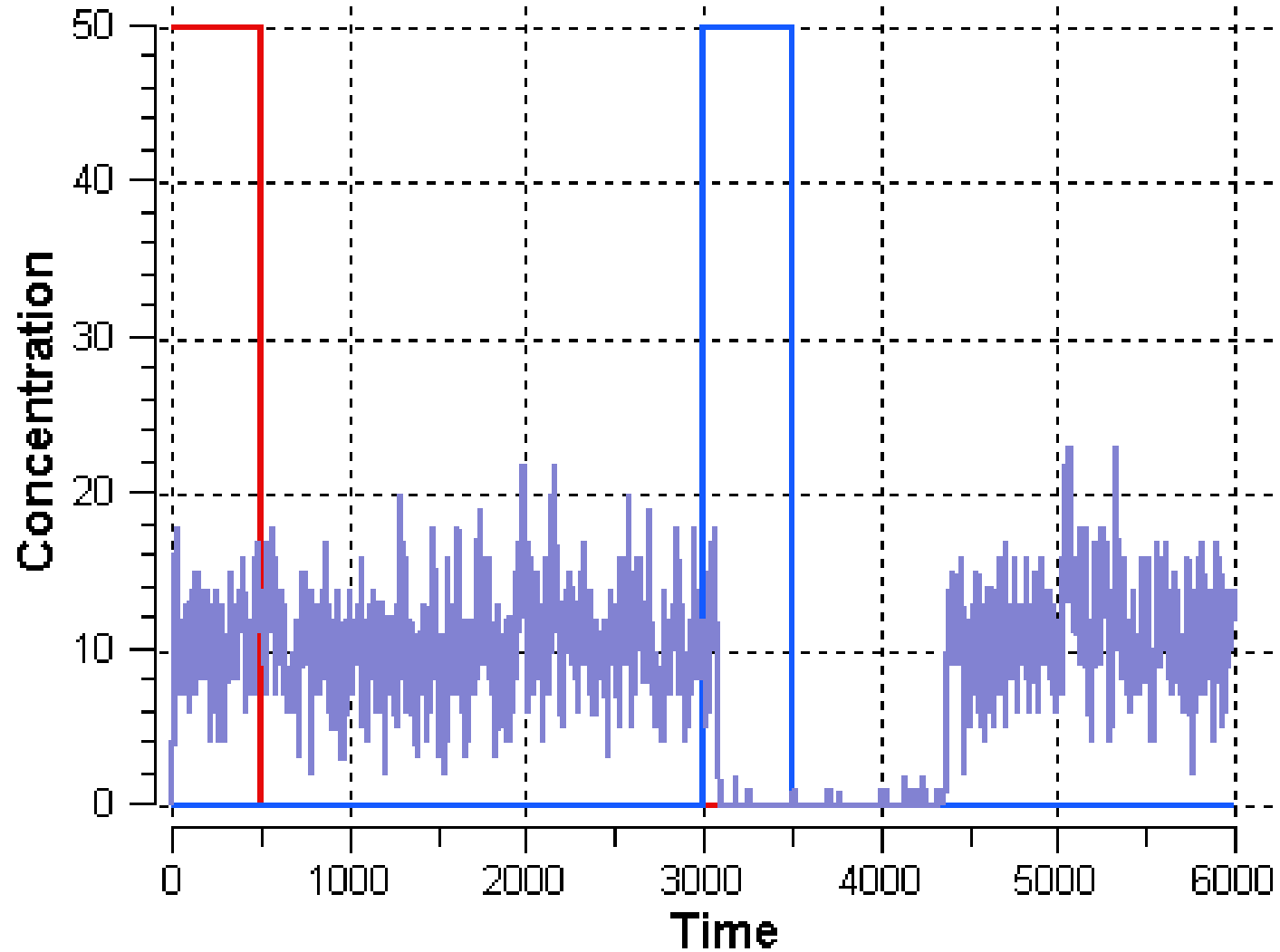
Next reaction



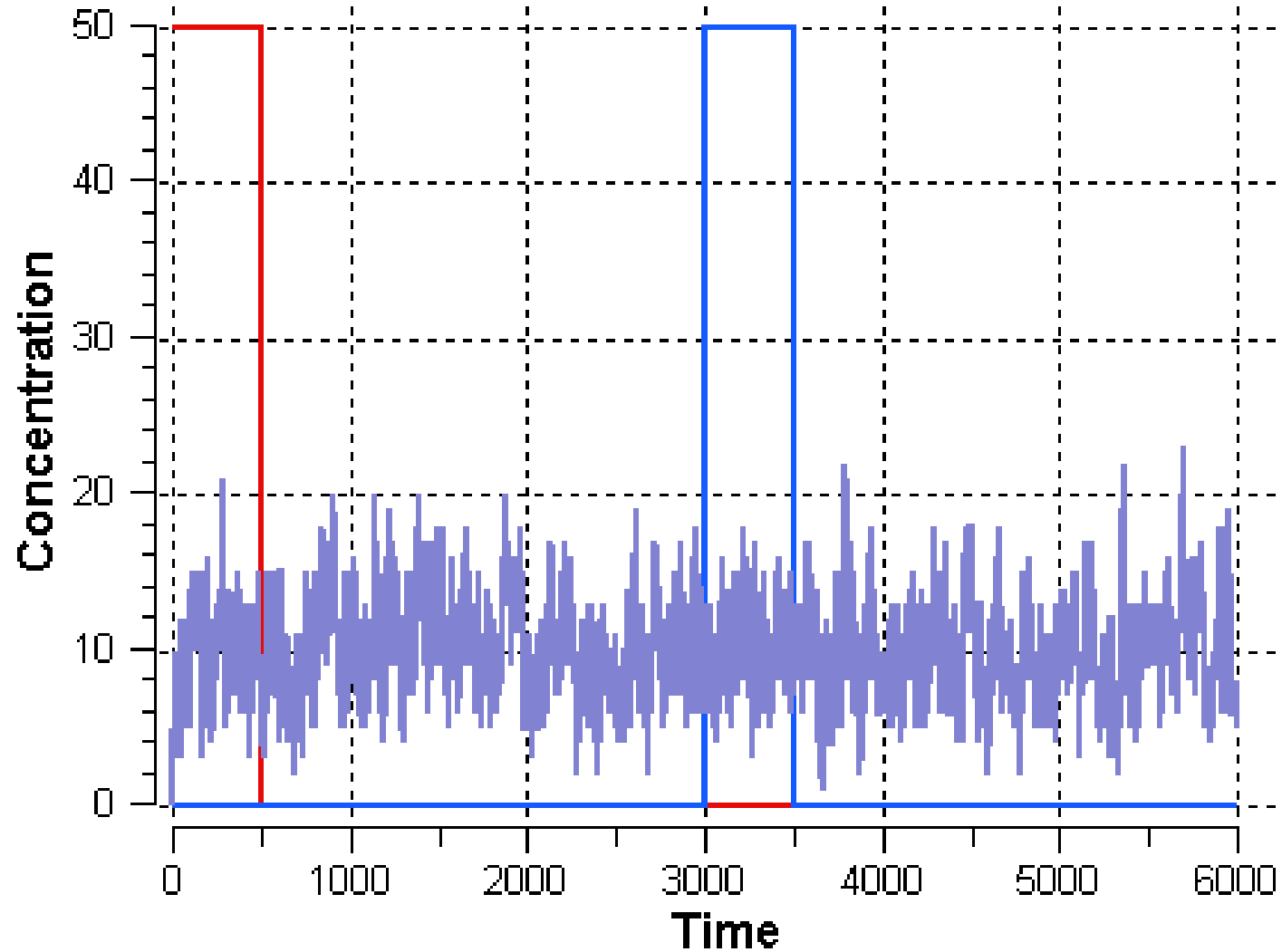
Stochastic trajectories



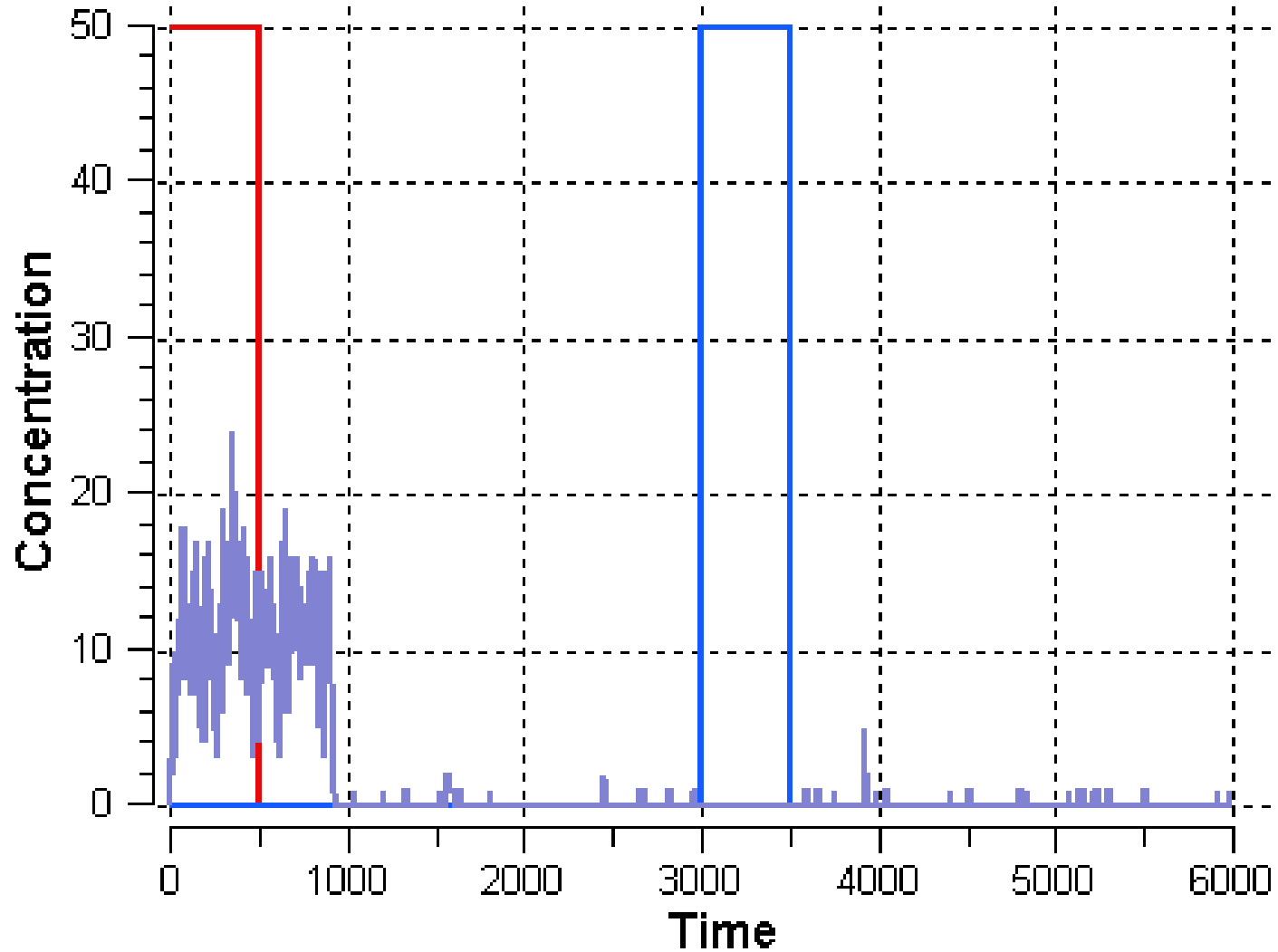
Stochastic trajectories



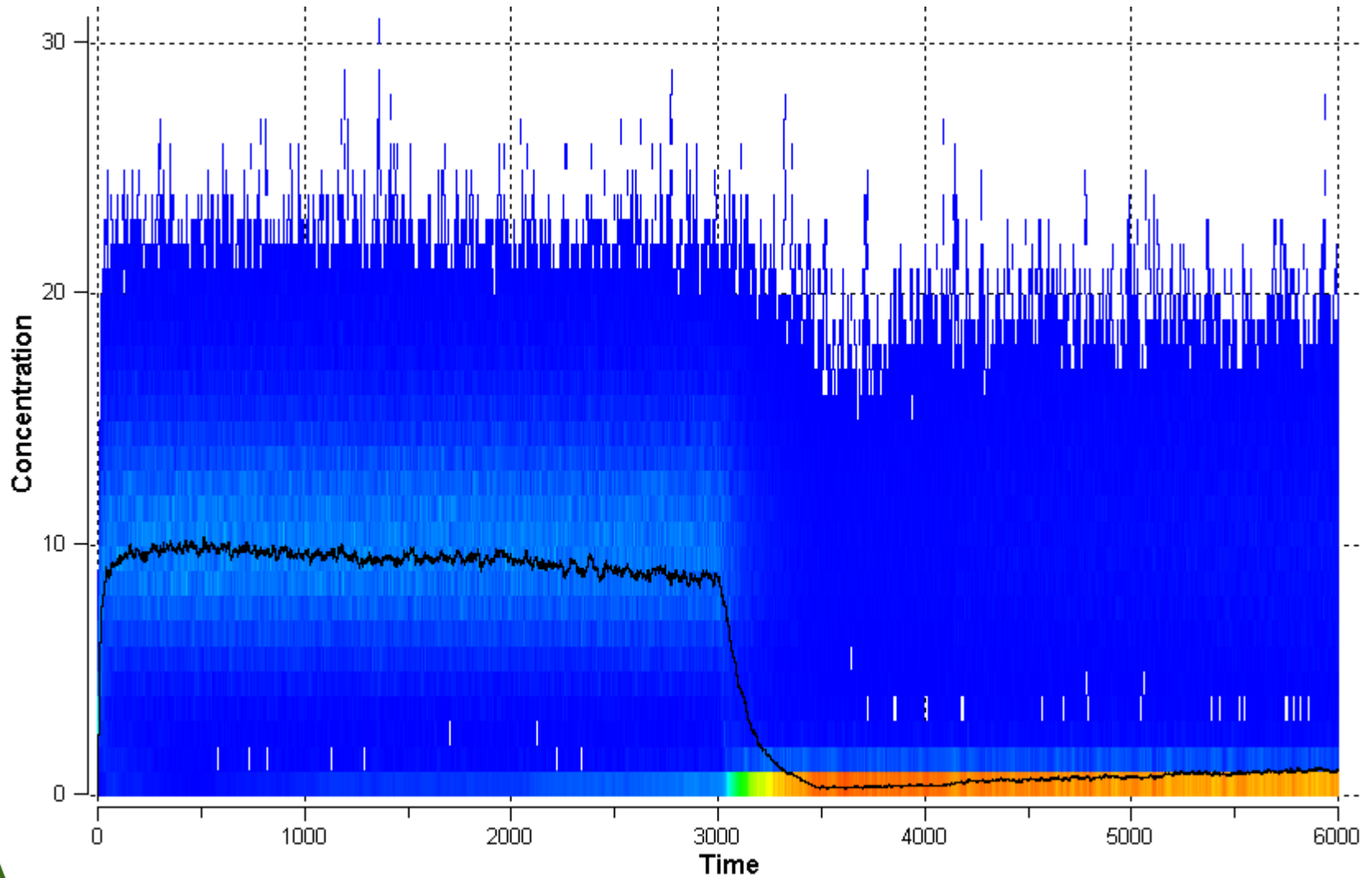
Stochastic trajectories



Stochastic trajectories



Reporter distribution



Solving stochastic kinetics

- Exact solutions

- $P(X_t)$ solution of the Chapman-Kolmogorov equation
- Tractable when the state space is finite and small (10^7 states)
- Cost is a function of the state space size

- Monte-Carlo simulations

- Not exact – only estimates
- Cost is a function (a^2) of accuracy (a) and intensity
- Rare events can be missed

- Simulation environment

- Stochastic activity networks: Mobius
- Goss & Peccoud (1998) *Proc. Nat. Acad. Sci.*
- Limitation: biological systems are open and stiff



Overcoming stiffness

- Collaboration with Bill Sanders group
- Approximation techniques
 - Approximate stochastic process with a simpler one
 - “Skip over” several event occurrences using a simplified process
 - E.g. Poisson approximation (Solari 2001, 2002)
- Hybrid simulation
 - Partition state-space into discrete and continuous parts based on rates, population size, application
 - Represent continuous by a Langevin process
 - Simulate discrete part



Poisson Approximation

- Naïve approximation

- Approximate reaction process by independent Poisson processes
- State derived from event process
- Zero-order approximation: rate constant at each time-step
- Converges to exact process when

- Enhanced approximation

$$\Delta t \rightarrow 0$$

- Poisson processes are non-stationary with varying rates
- Rates modeled by a coupled system of differential equations
- Boundary conditions / Error analysis



Hybrid Simulation

- Key ideas

- Partition reactions based on rate
- Fast rates: differential equation, Langevin process
- Slow rates: non-stationary Poisson process

- Partitioning

- How? Population size? Reaction rates?
- When? Static partition or dynamic partition

- Questions

- Should the continuous partition have a stochastic component?
- How to determine the partitioning threshold?
- Can the threshold be selected to bound the error to within a specified value?



III. Application: Genetic Analysis

Genotype & Environment

Phenotype, trait, and fitness functions

Response to selection



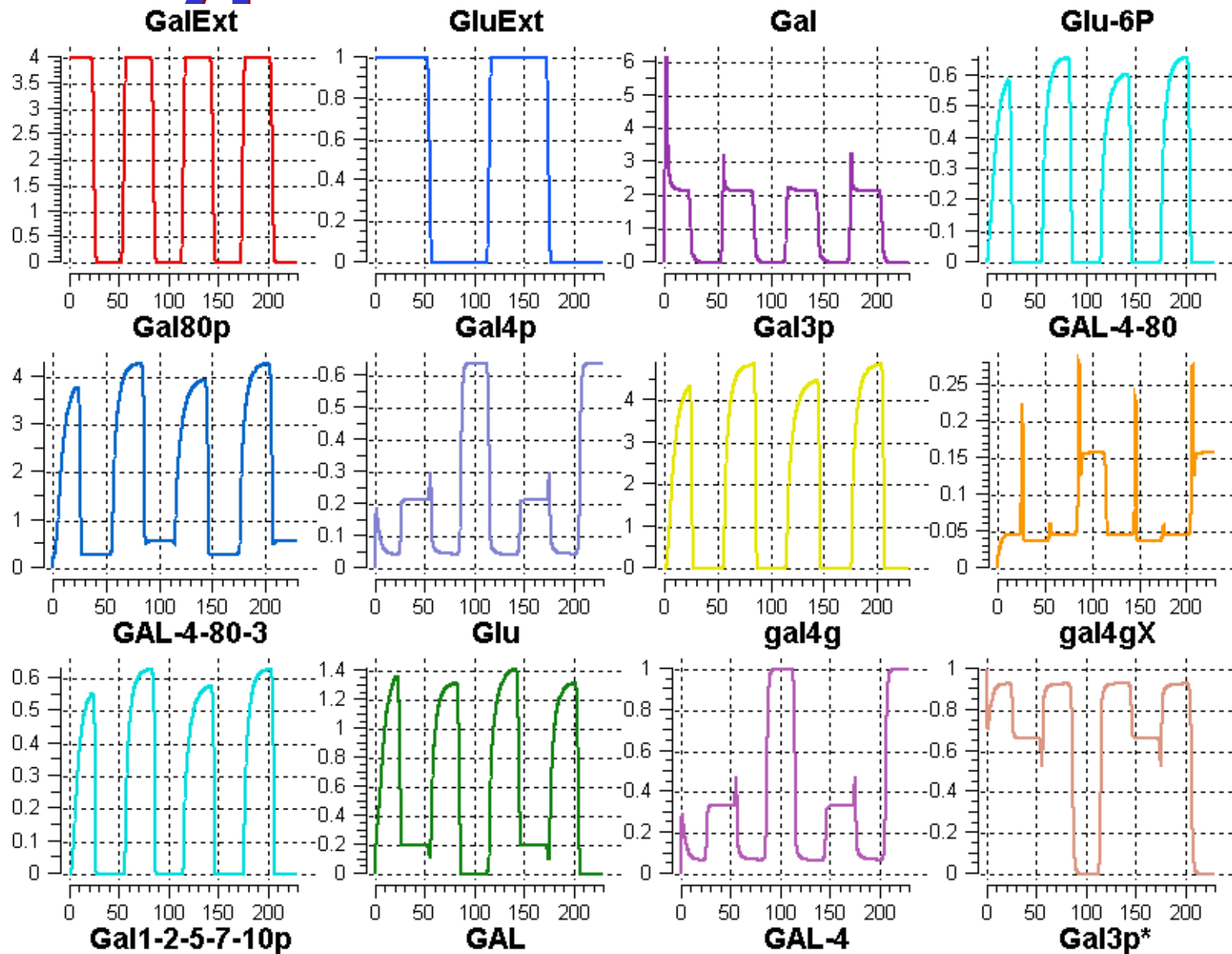
PIONEER®
A DUPONT COMPANY

Genotypes & Environment

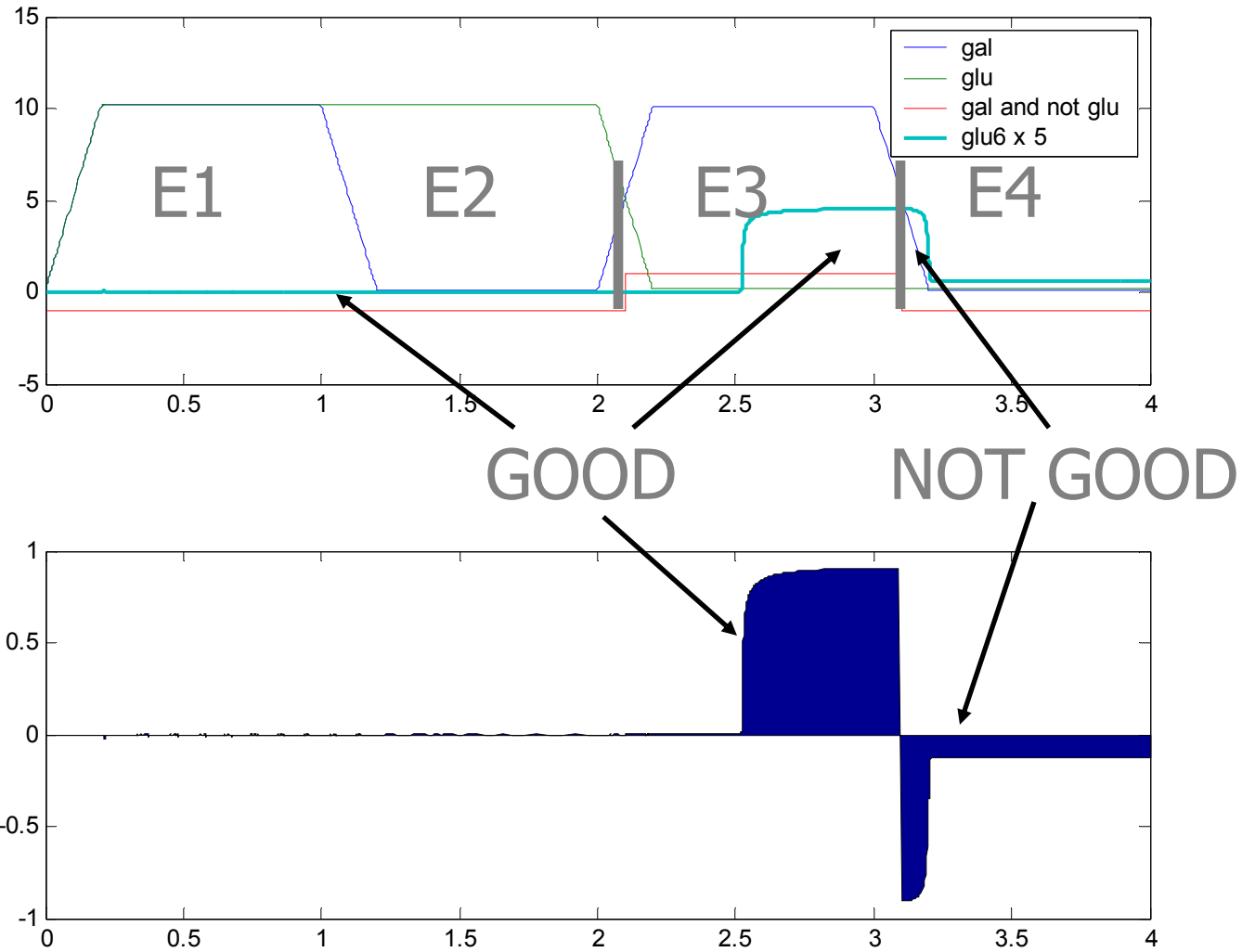
- Allele = parameter value
- Locus = parameter = reaction
- Co-dominance
 - $k_{Aa} = (k_{AA} + k_{aa})/2$
- Environment:
 - State variable
 - Impulse function
- Several loci associated to a gene
 - Promoter
 - Functional domain
- Independence
 - Protein stability and protein function may not be independent



Phenotype



Phenotype



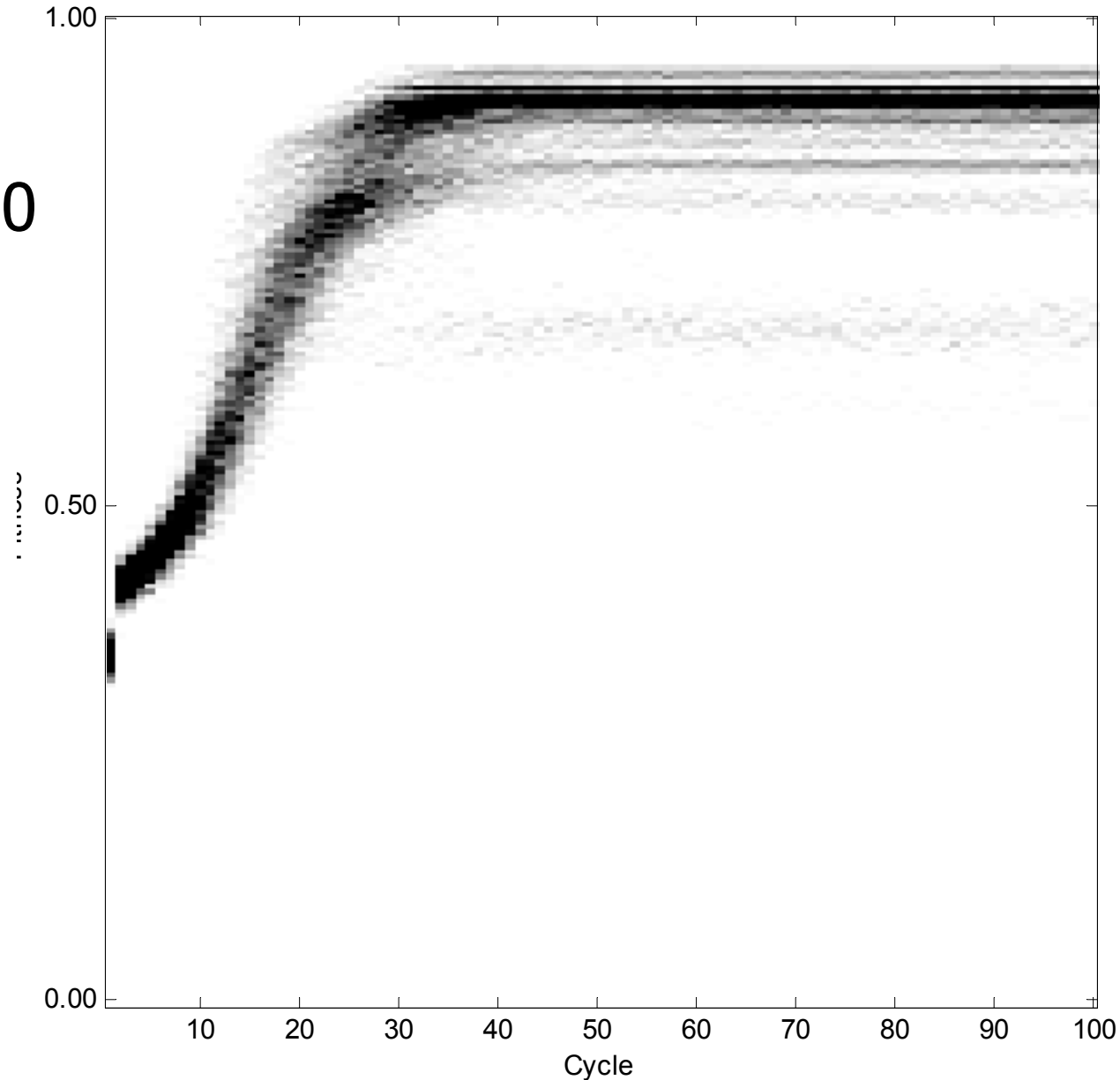
Trait and Fitness Functions

- Trait functions = performance variables
 - Trait value can depend on
 - The environment
 - The initial condition
 - The model parameters
 - Subjective and difficult to express
- Fitness function
 - Selection applies on fitness
 - Combines several trait values into a single score
 - Equivalent when a single trait is considered

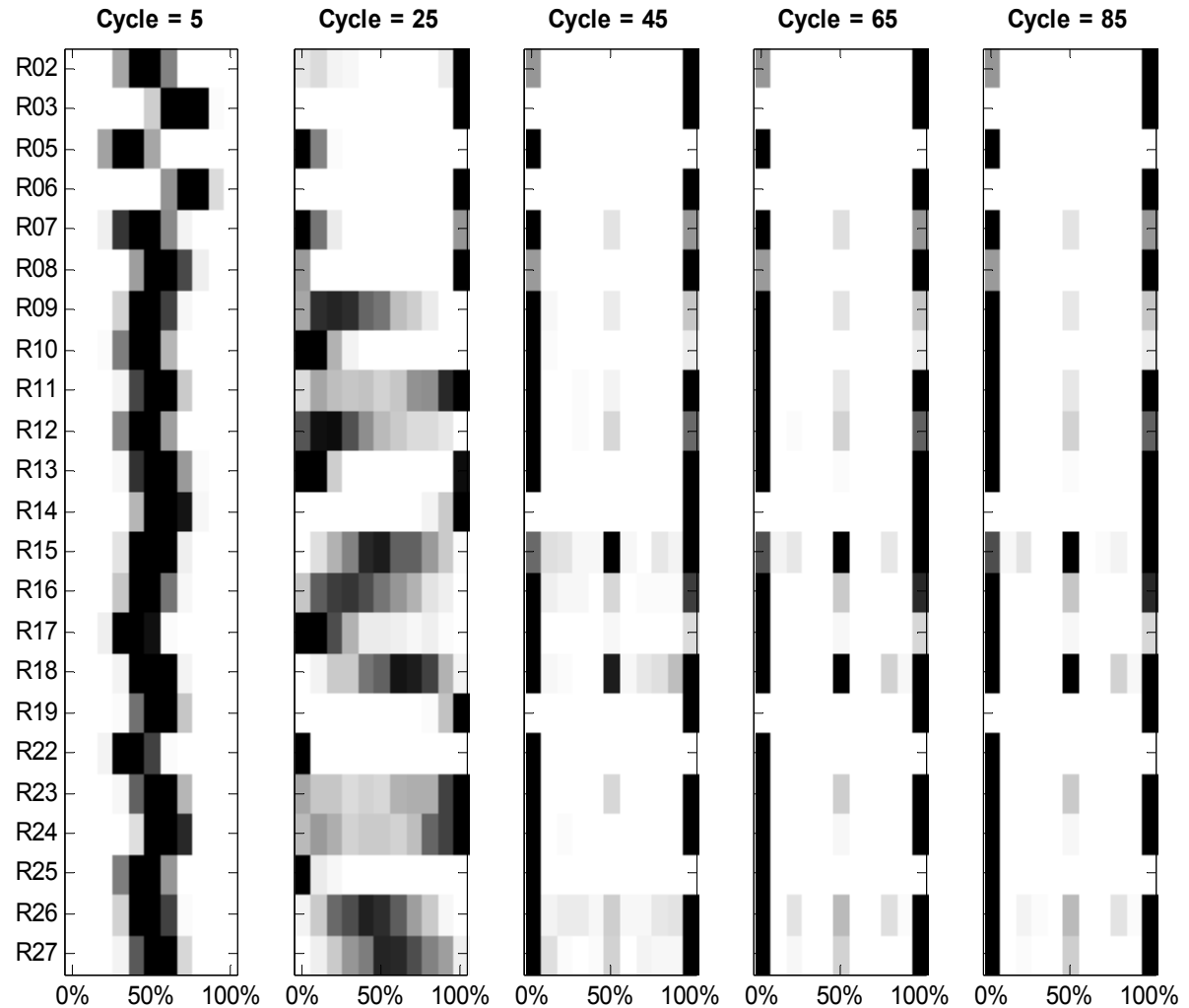


Fitness Response to Selection

- Several peaks
- $P(\text{Trapped}) \approx 0.30$
- Jump backward
- Early additivity



Allele Frequency Response



Summary of Results

- Correspondence between molecular networks and quantitative genetics models
- Can simulate the response of molecular networks to selection
 - Finite genetic space
 - Continuous genetic space using mutation operators
- Fitness landscape
 - Population level: multiple local maxima of fitness
 - Network level: plateaus - robustness
- Many loci do not have any favorable allele
- Despite interaction at the molecular level, the V_a is the largest component of V_g



Sharing our vision

- Models as knowledge representation
 - Capture the biologist expertise
 - Integrate the expertise of several biologists
 - Knowledge extraction from biological databases
- Discovery
 - Interpretation of experimental data
 - Checking the consistency of models, hypotheses with experimental data
 - Automatic exploration of data interpretations
- Engineering: CAD for genetic engineering
 - Construct design
 - Construct verification
 - Construct optimization
 - Construct implementation



Credits

- RIM-Bioinformatics

- Jean Peccoud
- Kent Vander Velden
- Chris Winkler
- Maury Leysens
- Lane Arthur

- Associative Genetics

- Dean Podlich
- Mark Cooper



Thank you!



PIONEER®
A DUPONT COMPANY