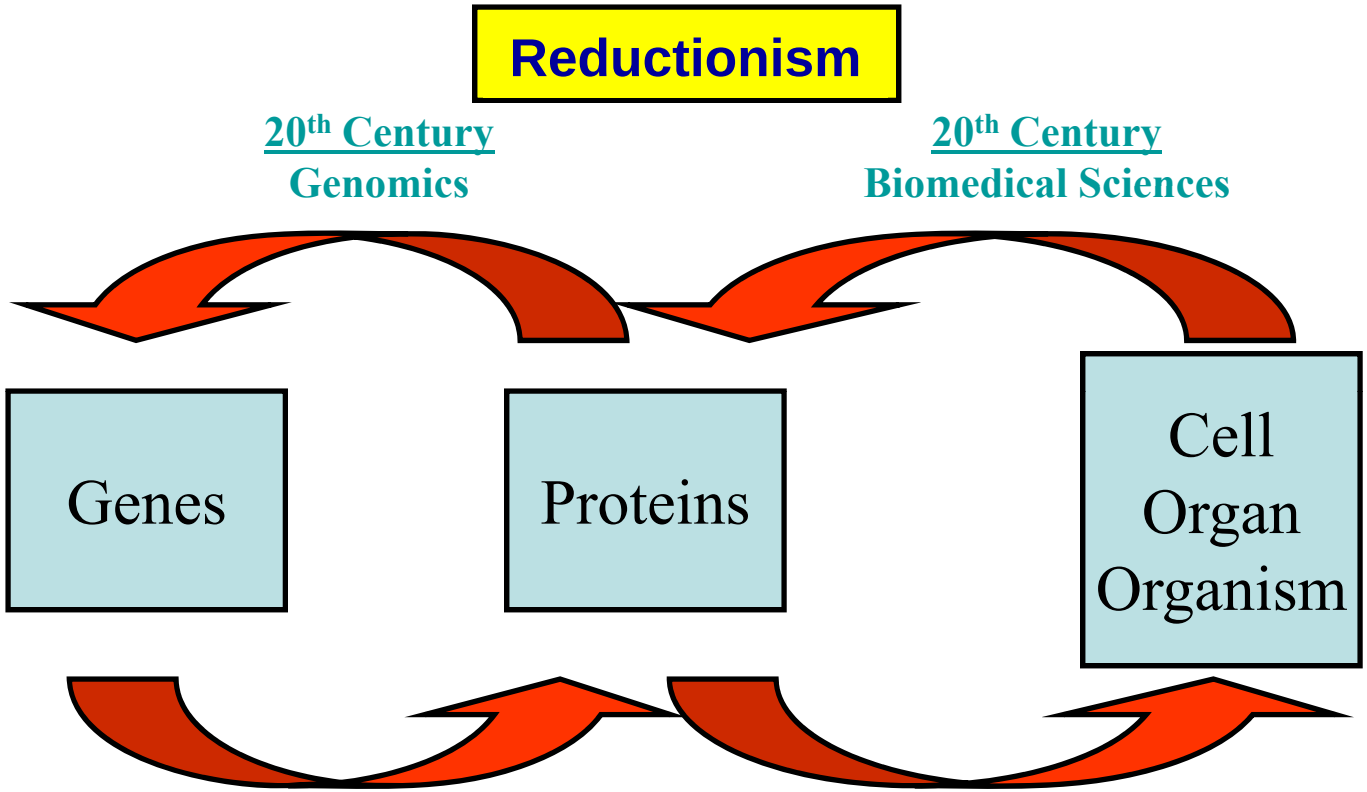
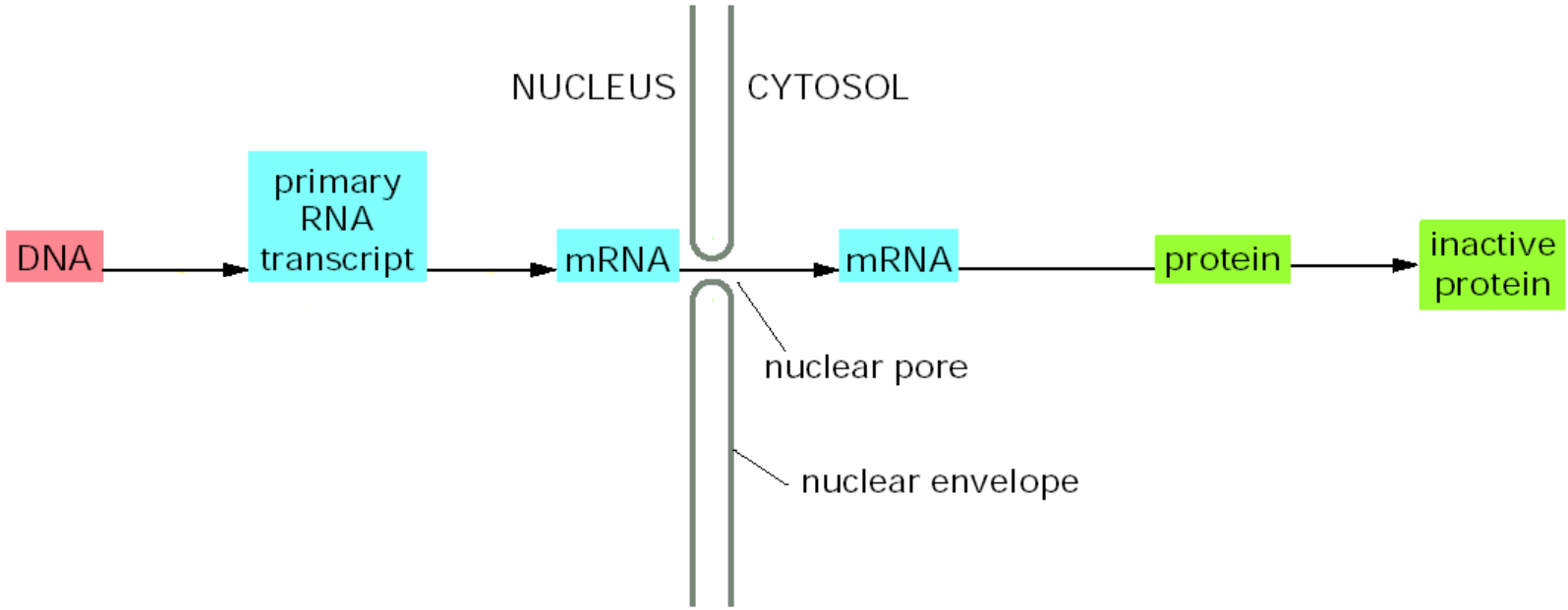
Overview – Systems Biology

Trends in Scientific Investigation



[adapted from (J. Weiss, 2003)]

Central Dogma and Levels of Control



[Alberts *et al.*, **Essential Cell Biology**, 1998]

SCIENCE JOURNAL
By SHARON BEGLEY

ADD TO PERSONALIZED HOME PAGE

FROM THE ARCHIVES: February 21, 2003

Biologists' New Approach: Do Not Shoot the Radio

How would a team of biologists fix a radio? First, they'd secure a large grant to purchase hundreds of identical working radios. After describing and classifying scores of components (metal squares, shiny circles with three legs, etc.), they'd shoot the radios with .22s.

Examining the corpses, the biologists would pick out those that no longer work. They'd find one radio in which a .22 knocked out a wire and triumphantly declare they had discovered the Key Component (KC) whose presence is required for normal operation.

But a rival lab would discover a radio in which the .22 left the Key Component intact but demolished a completely different Crucial Part (CP), silencing the radio. Moreover, the rivals would demonstrate that the KC isn't so "key" after all; radios can work fine without it.

Finally, a brilliant post-doc would discover a switch whose position determines whether KC or CP is required for normal operation. But the biologists still can't fix the blasted radios.

For those of you who haven't looked inside a radio lately, the Key Component is the wire connecting the external (FM) antenna to the innards of the radio, the Crucial Part is the internal (AM) antenna and the switch is the AM/FM switch.

Biologists can't repair radios because their part-by-part approach fails to describe the radio as a system -- what's connected to what and how one part affects another.

Biologists' affinity for the one-part-at-a-time approach, argues biologist Yuri Lazebnik of Cold Spring Harbor Laboratory on New York's Long Island, who dreamed up the radio analogy, is "a flaw of biological research today."

For that, thank the events of 50 years ago.

On Feb. 28, 1953, a Saturday, James Watson spent the morning at his Cambridge, England, lab piecing together cardboard representations of the "base pairs" in the DNA molecule. With that, he and Francis Crick realized that the master molecule of heredity is shaped like a spiral staircase, or double helix.

This discovery ushered in the era of the gene and gave birth to a new field: molecular biology. The study of living things became a science in which progress meant describing the smallest bits possible, usually one at a time -- one stretch of DNA, one RNA, one protein. The double helix, Harvard University naturalist E.O. Wilson once said, "injected into all of biology a new faith in reductionism" -- a "shoot the radio" approach

RECENT COLUMNS
March 7
▪ Global Warming May Chill the North
February 28
▪ Why DNA Isn't Destiny
February 14
▪ The Less Romantic Side of Mating
MORE

SEARCH PAST COLUMNS
Search for these words:
Display all columns

Can a biologist fix a radio?—Or, what I learned while studying apoptosis

As a freshly minted Assistant Professor, I feared that everything in my field would be discovered before I even had a chance to set up my laboratory. Indeed, the field of apoptosis, which I had recently joined, was developing at a mind-boggling speed. Components of the previously mysterious process were being discovered almost weekly, frequent scientific meetings had little overlap in their contents, and it seemed that every issue of *Cell*, *Nature*, or *Science* had to have at least one paper on apoptosis. My fear led me to seek advice from David Papermaster (currently at the University of Connecticut), who I knew to be a person with pronounced common sense and extensive experience. David listened to my outpouring of primal fear and explained why I should not worry.

David said that every field he witnessed during his decades in biological research developed quite similarly. At the first stage, a small number of scientists would somewhat leisurely discuss a problem that would appear esoteric to others, such as whether cell cycle is controlled by an oscillator or whether cells can commit suicide. At this stage the understanding of the problem increases slowly, and scientists are generally nice to each other, a few personal antipathies notwithstanding. Then, an unexpected observation, such as the discovery of cyclins or the finding that apoptosis failure can contribute to cancer, makes many realize that the previously mysterious process can be dissected with available tools and, importantly, that this effort may result in a miracle drug. At once, the field is converted into a Klondike gold rush with all the characteristic dynamics, mentality, and morals. A major driving force becomes the desire to find the nugget that will secure a place in textbooks, guarantee an unrelenting envy of peers, and, at last, solve all financial problems. The assumed proximity of this imaginary nugget easily attracts both financial and human resources, which results in a rapid expansion of the field. The understanding of the biological process increases accordingly and results in crystal clear models that often explain everything and point at targets for future miracle drugs. People at this stage are not necessarily nice, though, as anyone who has read about a gold rush can expect. This description fit the then current state of the apoptosis field rather well, which made me wonder why David was smiling so reassuringly. He took his time to explain.

At some point, David said, the field reaches a stage at which models, that seemed so complete, fall apart, predictions that were considered so obvious are found to be wrong, and attempts to develop wonder drugs largely fail. This stage is characterized by a sense of frustration at the complexity of the process, and by a sinking feeling that despite all that intense digging the promised cure-all may not materialize. In other words, the field hits the wall, even though the intensity of research remains unabated for a while, resulting in thousands of publications, many of which are contradictory or largely descriptive. The flood of publications is explained, in part, by the sheer amount of accumulated information (about 10,000 papers on apoptosis were published yearly over the last few years), which makes reviewers of the manuscripts as confused and overwhelmed as their authors. This stage can be summarized by the paradox that the more facts we learn the less we understand the process we study.

It becomes slowly apparent that even if the anticipated gold deposits exist, finding them is not guaranteed. At this stage, the Chinese saying that it is difficult to find a black cat in a dark room, especially if there is no cat, comes to mind too often. If you want to continue meaningful research at this time of widespread desperation, David said, learn how to make good tools and how to keep your mind clear under adverse circumstances. I am grateful to David for his advice, which gave me hope and, eventually, helped me to enjoy my research even after my field did reach the state he predicted.

At some point I began to realize that David's paradox has a meaning that is deeper than a survival advice. Indeed, it was puzzling to me why this paradox manifested itself not only in studies of fundamental processes, such as apoptosis or cell cycle, but even in studies of individual proteins. For example, the mystery of what the tumor suppressor p53 actually does seems only to deepen as the number of publications about this protein rises above 23,000.

The notion that your work will create more confusion is not particularly stimulating, which made me look for guidance again. Joe Gall at the Carnegie Institution, who started to publish before I was born, and is an author of an excellent series of essays on the history of biology (Gall, 1996), relieved my mental suffering by pointing out that a period of stagnation is eventually interrupted by a new development. As an example, he referred to the studies of cell death that took place in the nineteenth century (Gall, 1996, chapter 29), faded into oblivion, and reemerged a century later with about 60,000 studies on the subject published during a single decade. Even though a prospect of a possible surge in activity in my field was relieving, I started to wonder whether anything could be done to expedite this event, which brought me to think about the nature of David's paradox. The generality of the paradox suggested some common fundamental flaw of how biologists approach problems.

To understand what this flaw is, I decided to follow the advice of my high school mathematics teacher, who recommended testing an approach by applying it to a problem that has a known solution. To abstract from peculiarities of biological experimental systems, I looked for a problem that would involve a reasonably complex but well understood system. Eventually, I thought of the old broken transistor radio that my wife brought

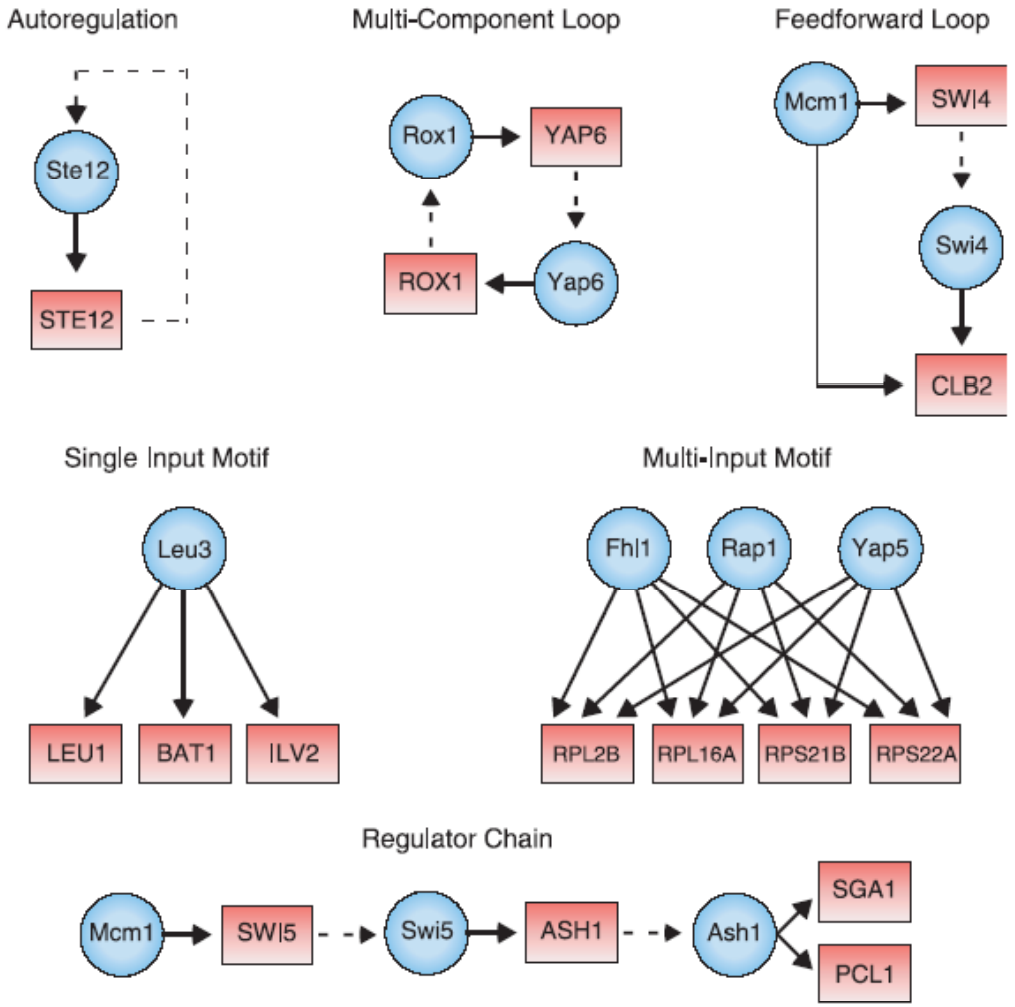
Figure 1. The radio that has been used in this study

“Engineering” of Biological Networks

[Alon, 2003]

- Modularity
  - (in network) set of nodes that have strong interactions and a common function
  - has defined input nodes and output nodes that control the interactions with the rest of the network
  - has internal nodes that do not significantly interact with nodes outside the module
- Robustness to component tolerances
- Recurring circuit elements

Motifs in Biological Regulation - Yeast



[Lee et al., 2002]

Motifs in Biological Regulation – E. coli

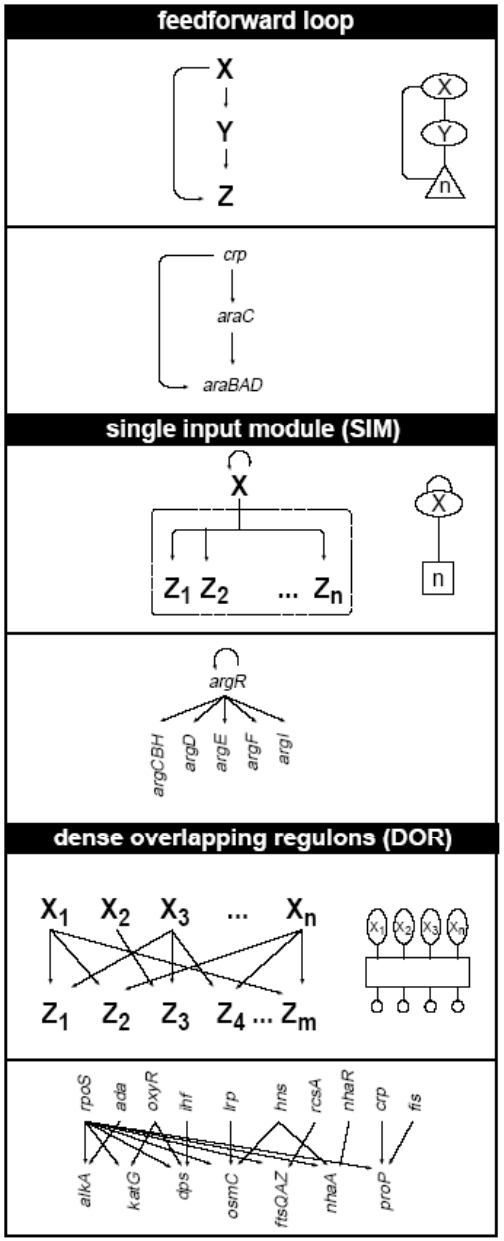


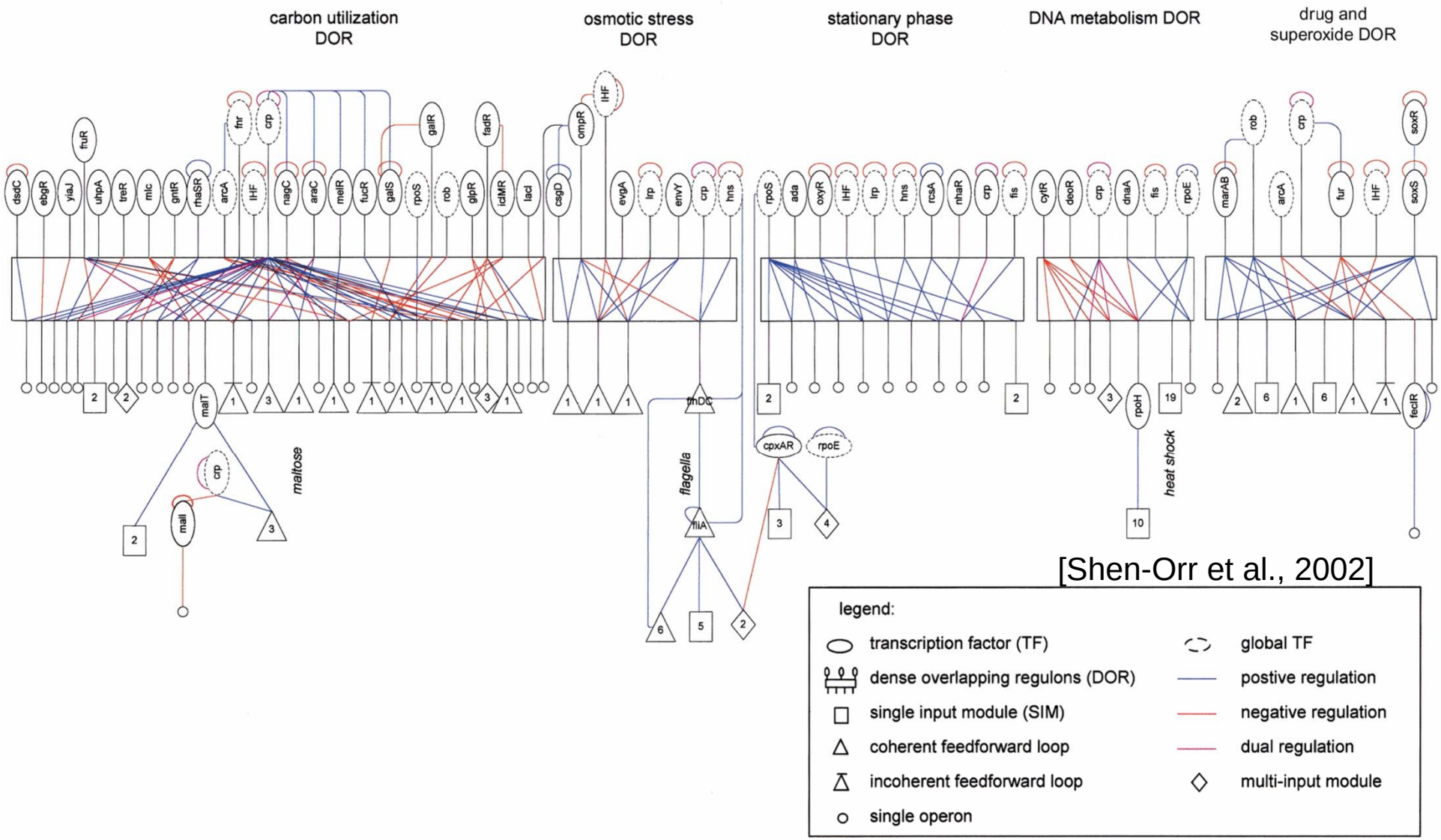
Table 1 • Statistics of occurrence of various structures in the real and randomized networks

Structure	Appearances in real network	Appearances in randomized network (mean ± s.d.)	*P* value
Coherent feedforward loop	34	4.4 ± 3	$P < 0.001$
Incoherent feedforward loop	6	2.5 ± 2	$P \sim 0.03$
Operons controlled by SIM (>13 operons)	68	28 ± 7	$P < 0.01$
Pairs of operons regulated by same two transcription factors	203	57 ± 14	$P < 0.001$
Nodes that participate in cycles*	0	0.18 ± 0.6	$P \sim 0.8$

*Cycles include all loops greater than size 1 (autoregulation). *P* value for cycles is the probability of networks with no loops.

[Shen-Orr et al., 2002]

E. Coli Transcriptional Network

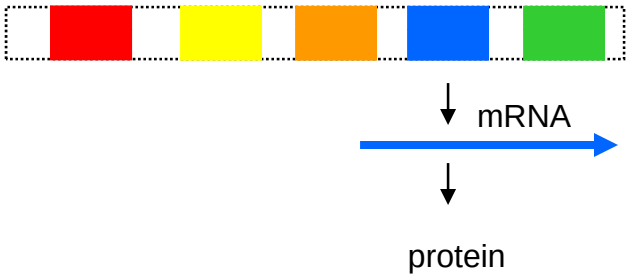


Gene Regulation

Hierarchy of Biological Regulation

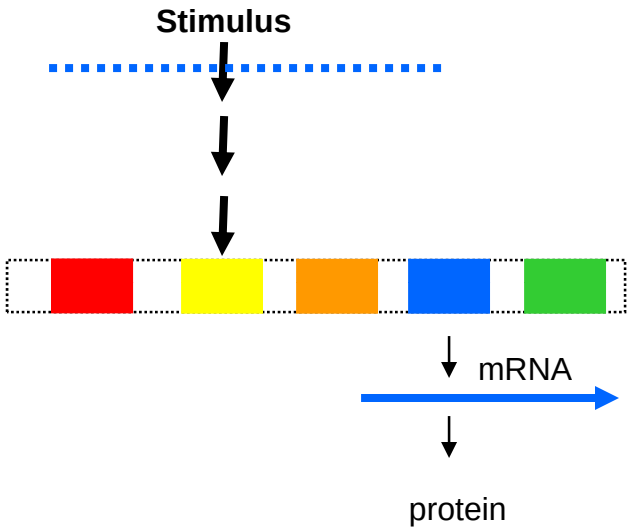
[Savageau, *Chaos*, 2001; Alberts et al., **Mol. Biol. Of the Cell** 4th ed., etc.]

- Transcriptional Units



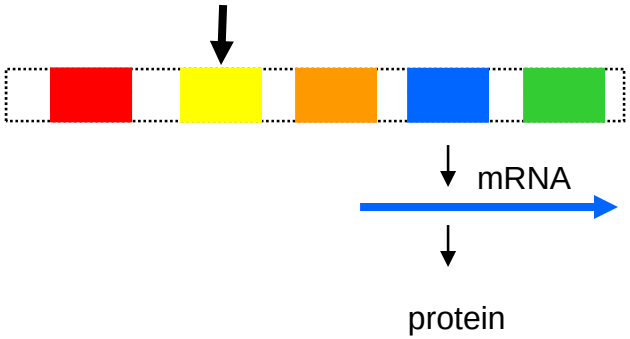
Hierarchy of Biological Regulation

- Transcriptional Units
- Input Signal



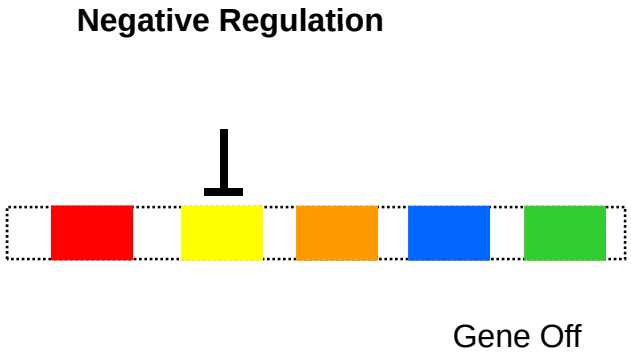
Hierarchy of Biological Regulation

- Transcriptional Units
- Input Signal
- Mode



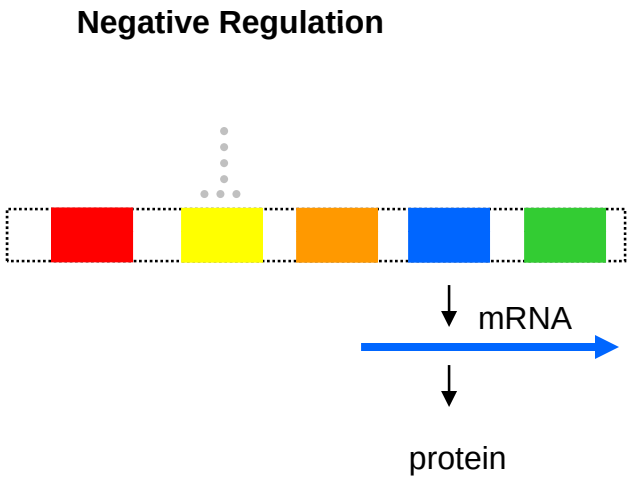
Hierarchy of Biological Regulation

- Transcriptional Units
- Input Signal
- Mode



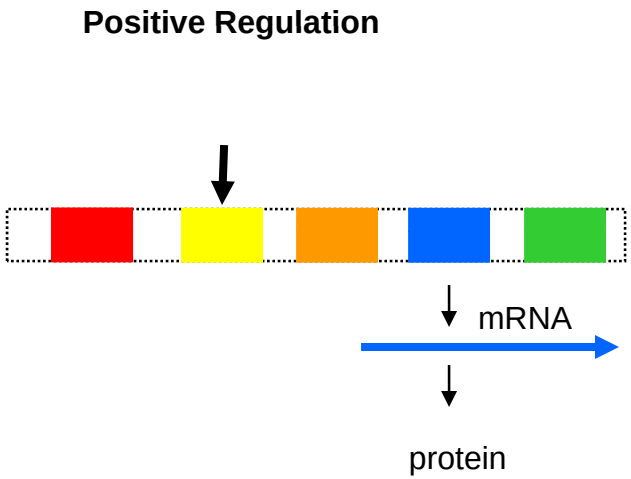
Hierarchy of Biological Regulation

- Transcriptional Units
- Input Signal
- Mode



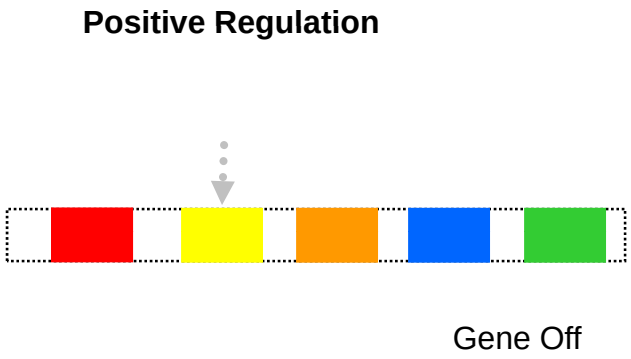
Hierarchy of Biological Regulation

- Transcriptional Units
- Input Signal
- Mode



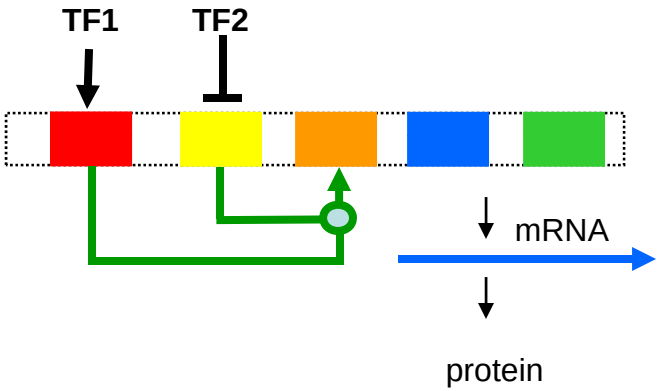
Hierarchy of Biological Regulation

- Transcriptional Units
- Input Signal
- Mode



Hierarchy of Biological Regulation

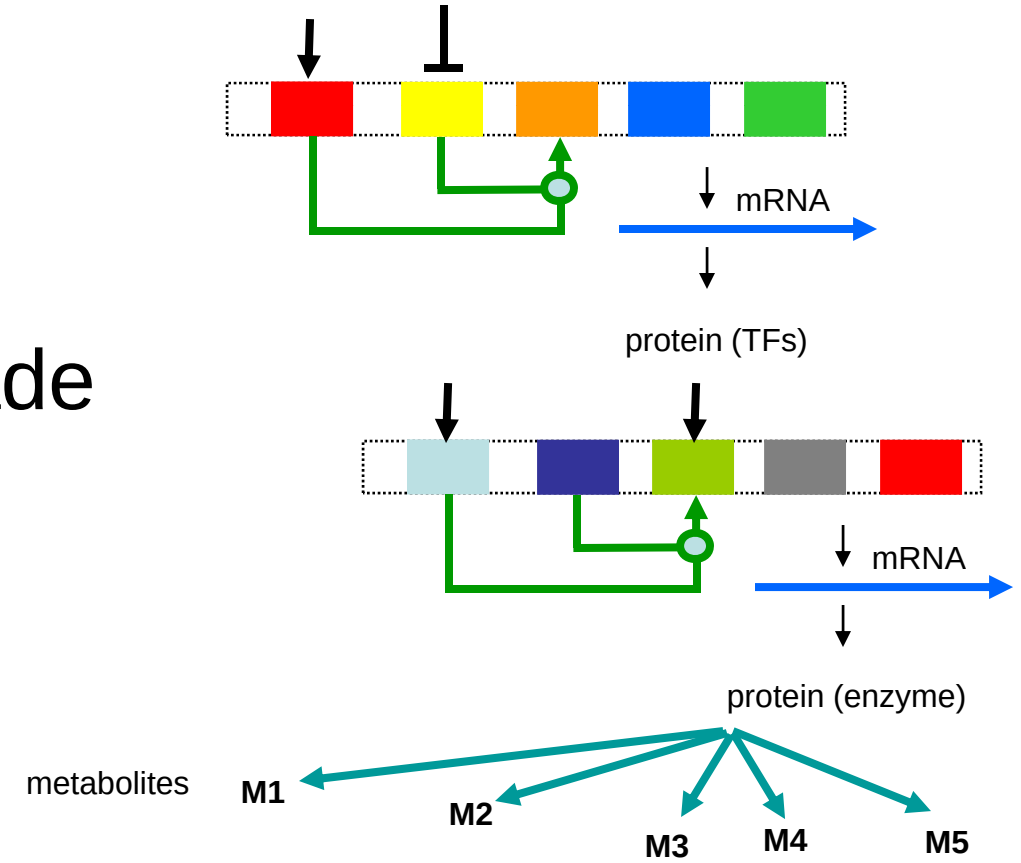
- Transcriptional Units
- Input Signal
- Mode
- Logical Unit



TF1(+)	TF2(-)	EXP
ON	ON	OFF
ON	OFF	ON
OFF	ON	OFF
OFF	OFF	OFF

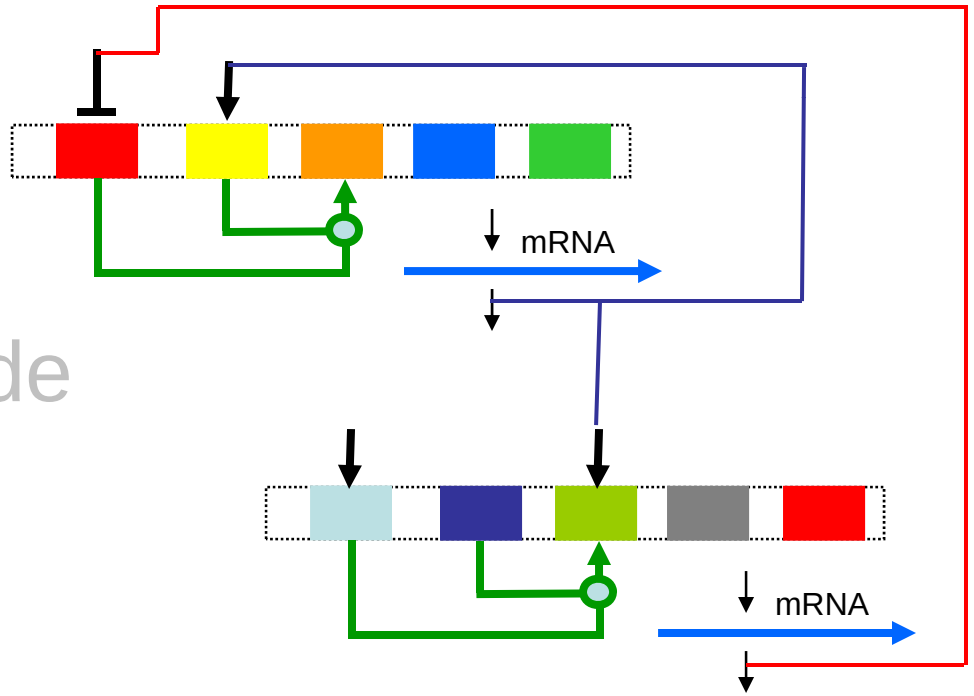
Hierarchy of Biological Regulation

- Transcriptional Units
- Input Signal
- Mode
- Logical Unit
- Expression Cascade



Hierarchy of Biological Regulation

- Transcriptional Units
- Input Signal
- Mode
- Logical Unit
- Expression Cascade
- Connectivity



Gene Regulatory Networks

- Whole genome sequencing → (all) potential macromolecular players
- High throughput methods are at relatively mature state
- GRN models → DNA-specific predictions (which can be validated)
- Transcription and translation are slow (cf. protein-protein and enzymatic rxns). May suggest switch-like behavior (Boolean) for latter.

Eukaryotes – More Complex Story

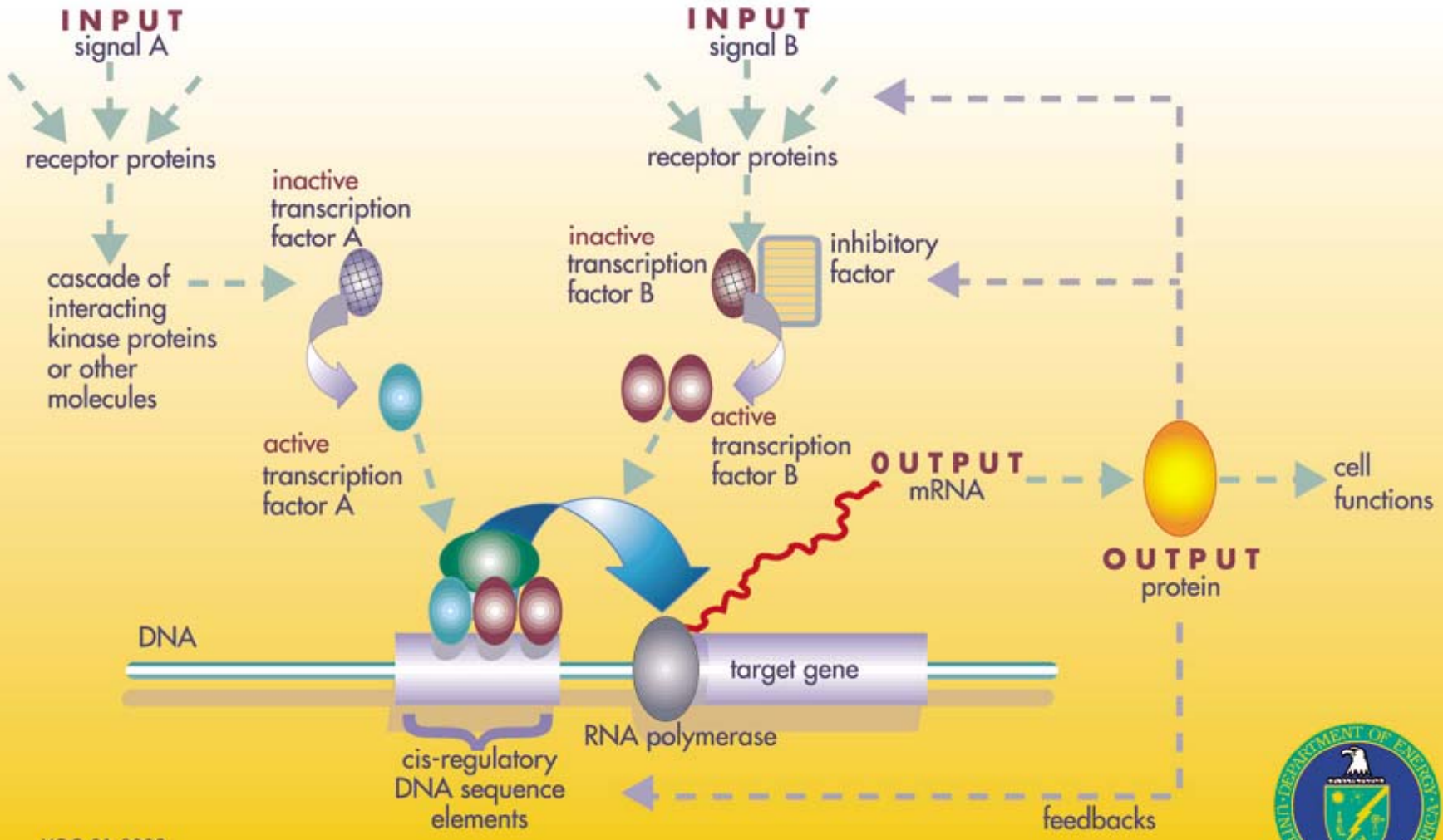
- Gene regulatory proteins can influence from a long distance (thousands of bp away from promoter) → single promoter influenced by virtually unlimited number of regulatory sequences scattered along DNA

- RNA polymerase II (transcribes all protein-coding genes) cannot initiate transcription alone. Requires “general” transcription factors to be assembled.

- Packing of eukaryote DNA into chromatin provides additional layers of regulation (not available to bacteria)

GENOMES to LIFE

A GENE REGULATORY NETWORK



Integrated Circuits

- Multi-gene interactions
- Already studied: genes are regulated
- Will study: proteome is dynamic – changing w/ environment
  - But promoters don't change…
  - How do cells turn "on" and "off"?…
  - [and not in a wildly fluctuating manner…]
- One interesting motif: toggle switch (turn genes on and off)
  - Ramifications for development, cell cycle, cancer, etc.
  - Interesting attribute: robust probabilistic switching under uncertainty

Noise & Uncertainty

- Cell division/mitosis & cytokenesis → how to divide TFs
  - "randomness" → captured w/ binomial frequency function
  - e.g., 50 TF, 2 daughters, → p=0.50
  - 50/50 split → 0.112
  - 25+/-5 → 0.880, etc.

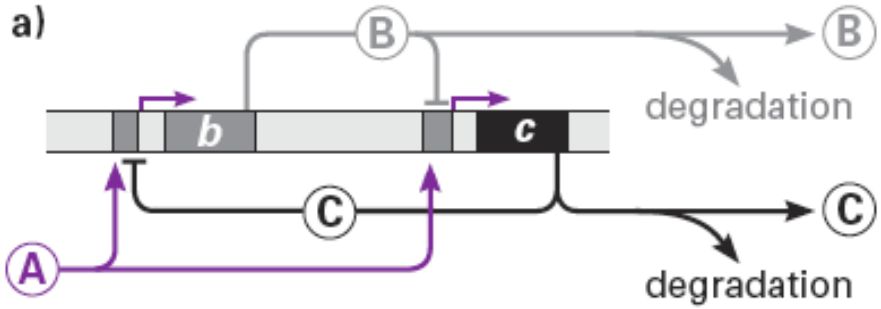
$$\binom{n}{k} p^k (1-p)^{n-k}$$

- Few binding sites for each protein, slow rates of binding
  - "randomness" → right protein/time/place
  - also, once bound, variable delay for activation of transcription
  - protein produced from one gene obeys normal distribution

  (i.e., random variable that is composite of many small random events)

- Environmental uncertainty
  - development, morphology, concentrations, etc.

Toggle Switch Circuit

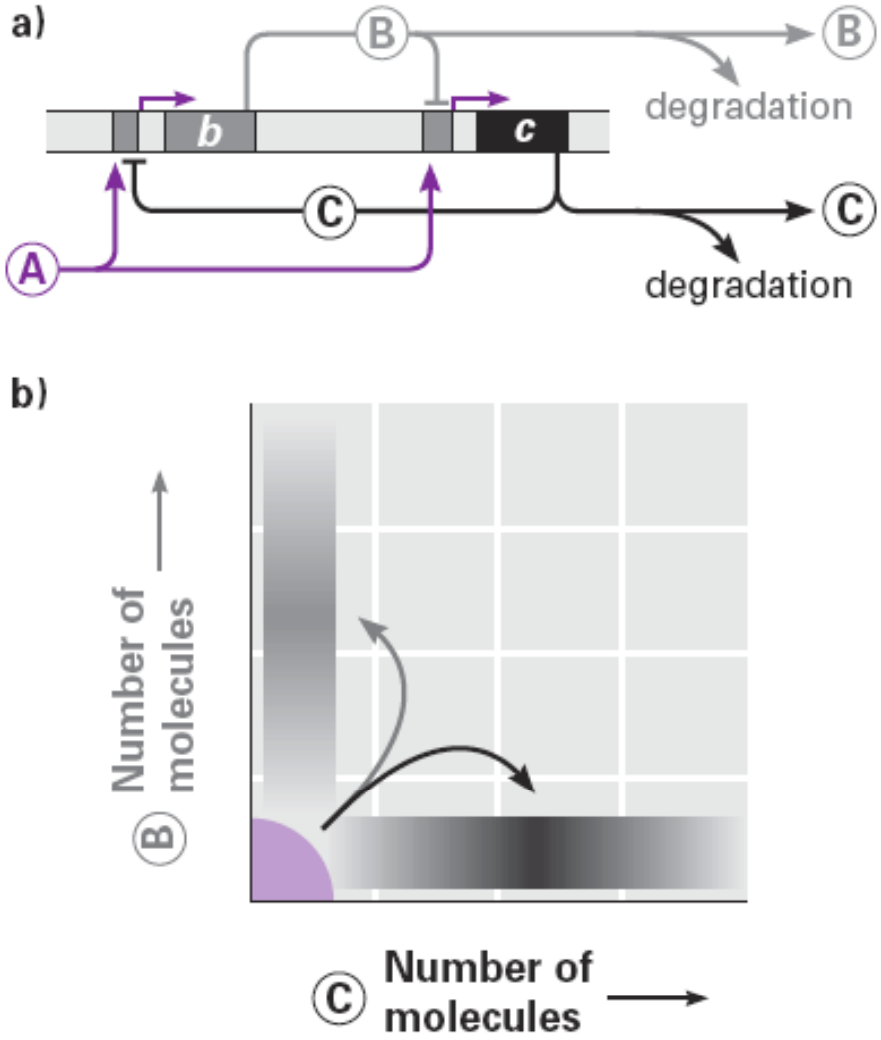


Protein A: TF for genes b & c

Protein B: degraded by cell, performs other functions, represses expression of c

Protein C: degraded by cell, perform other functions, represses expression of b

Toggle Switch Circuit



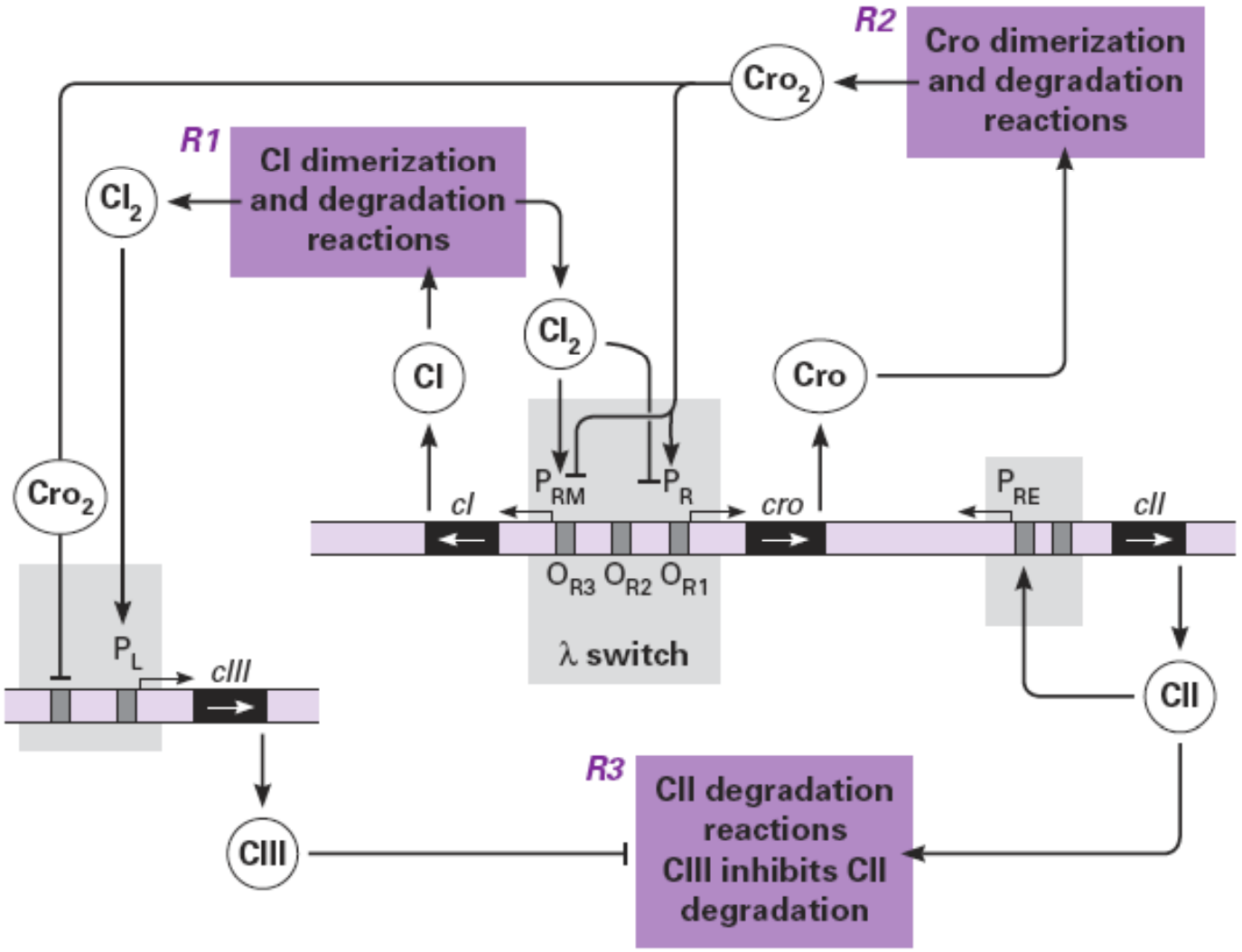
[Campbell & Heyer, 2006]

λ Phage Switch – Introduction

[Arkin *et al*., Genetics, 1998]

- Pathogen (virus) that can switch exterior to fool immune system
- Two life domains
  - Live quietly in E. coli (lysogenic)
  - Replicate quickly, kill host, launch progeny (lytic)
- Choice determined by single protein: CII
- Very similar to simple circuit just considered

λ Phage Switch Circuit



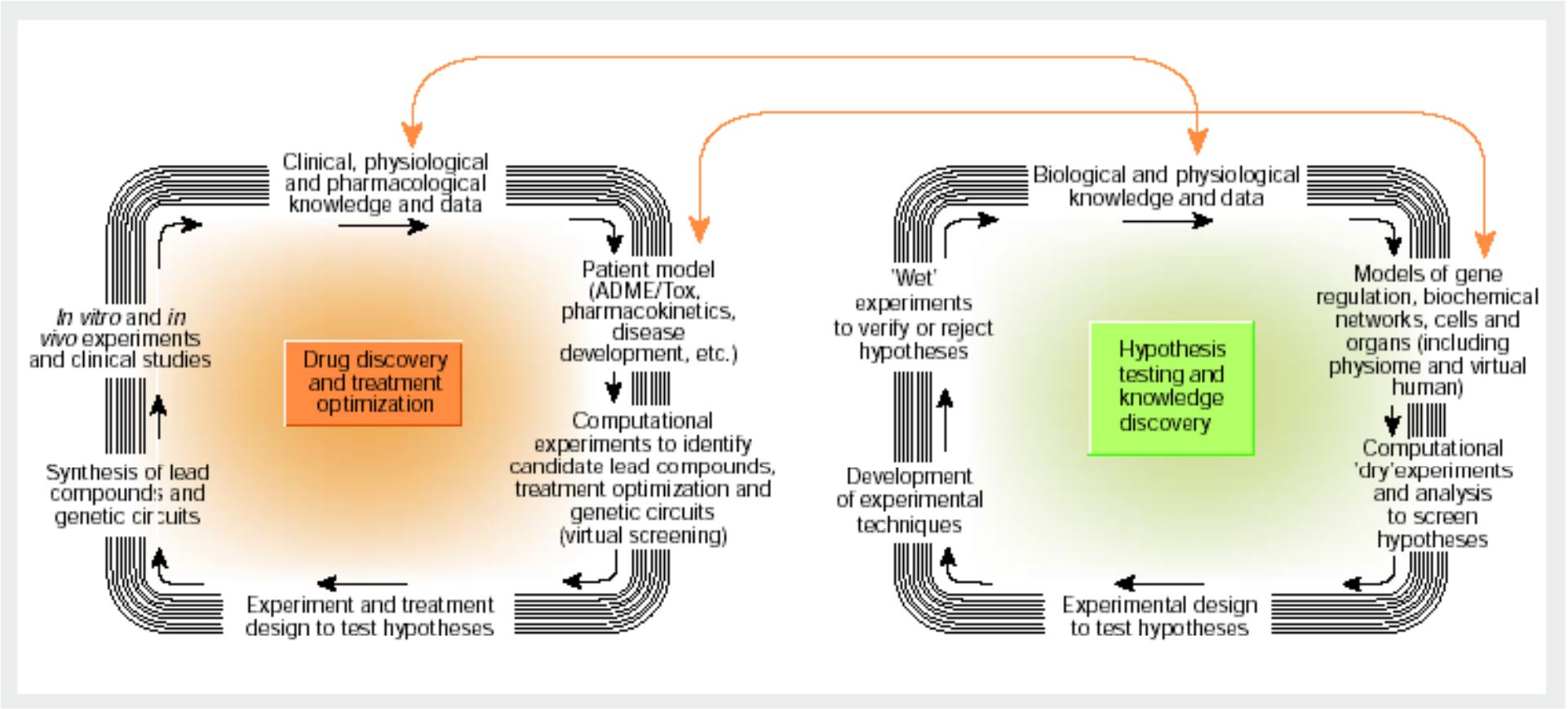
[Campbell & Heyer, 2006]

How do Cells Cope?

- Usual control tricks
  - Cascades and relays (low pass filters)
  - Negative feedback
  - Integral feedback
  - Redundancy

Modeling Issues

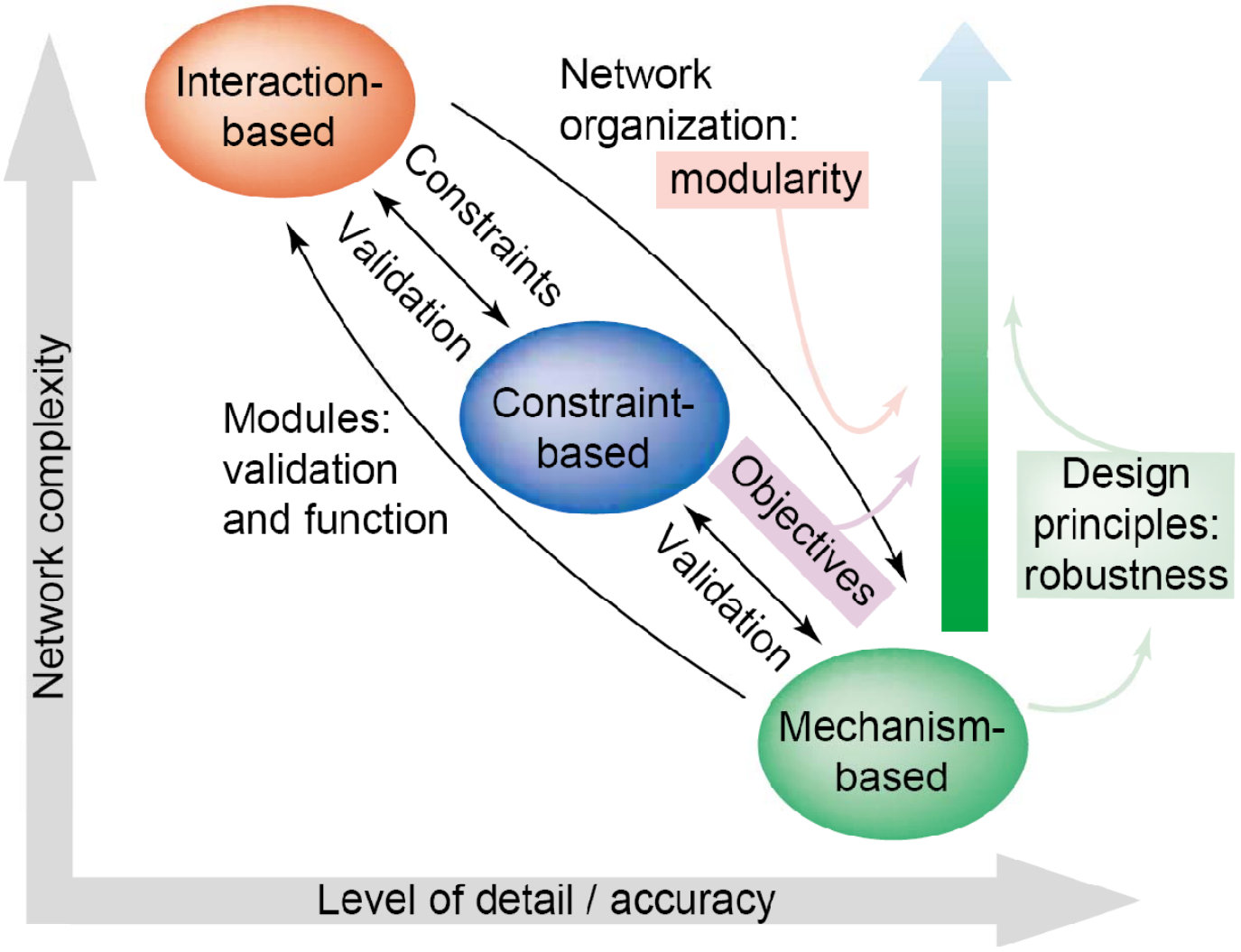
Iterations for Model and Hypothesis



[Kitano, Nature, 2002]

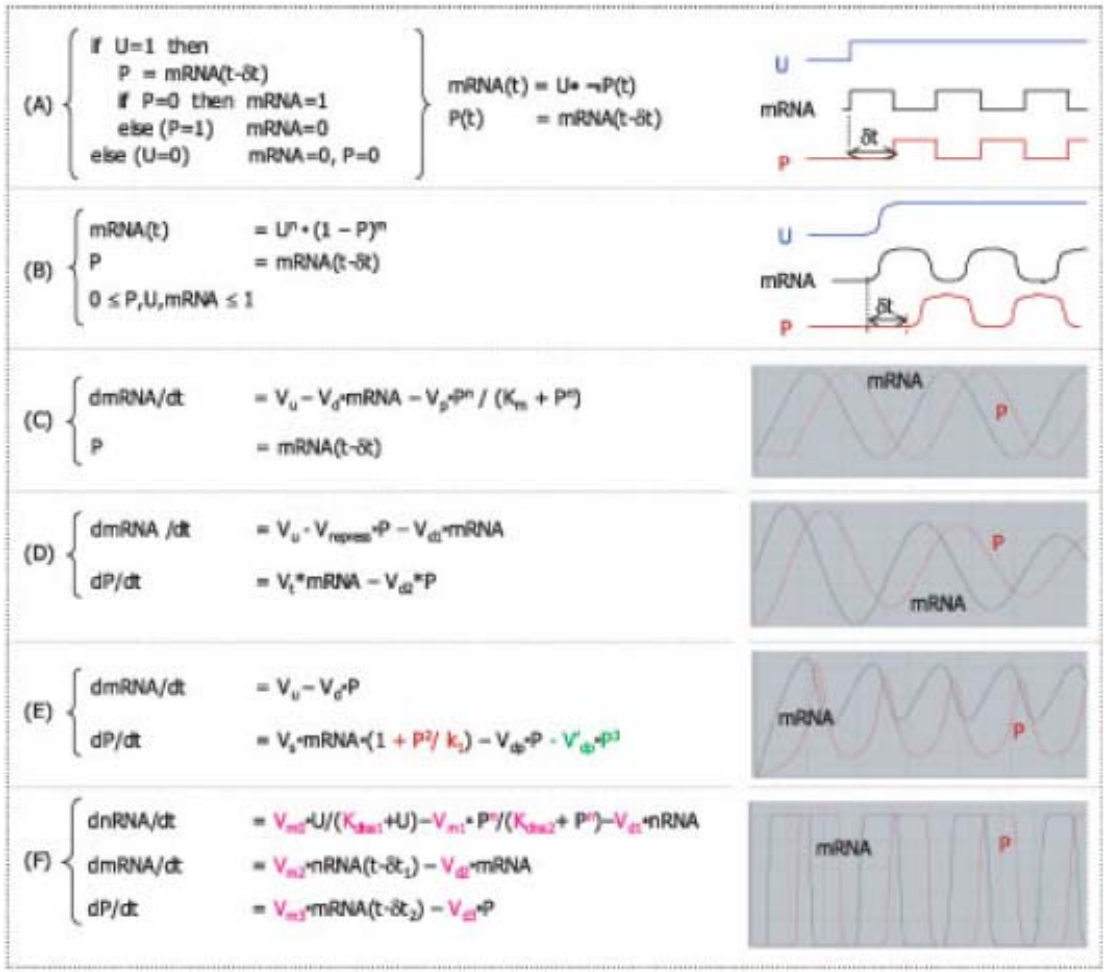
The Spectrum of Descriptions

[Stelling, 2004]



Current Opinion in Microbiology

Nice Example of Modeling Scope



[Bolouri & Davidson, 2002]

Nonlinear ODE Representations

- Gene regulation captured by rate equations: $\frac{dx_i}{dt} = f_i(x) \qquad 1 \le i \le n$

Enzymatic Kinetics

Michaelis-Menten Kinetics

$$E + S \underset{k_{-1}}{\overset{k_1}{\Leftrightarrow}} ES \xrightarrow{k_2} E + P$$

Assumptions:

1) [S] >> [E]
2) Steady-state ([ES] ~ constant)
3) Total enzyme concentration is constant

A2 + algebra:

$$\frac{[E][S]}{K_M} = [ES]$$

$$K_{M=} \frac{k_{-1} + k_2}{k_1}$$

A3 + algebra:

$$[E_0] = [E] + [ES]$$

$$v = \frac{k_2[E_0][S]}{K_M + [S]}$$

$$v = \frac{v_{\max}[S]}{K_M + [S]}$$

Enzymatic Kinetics

Hill Kinetics

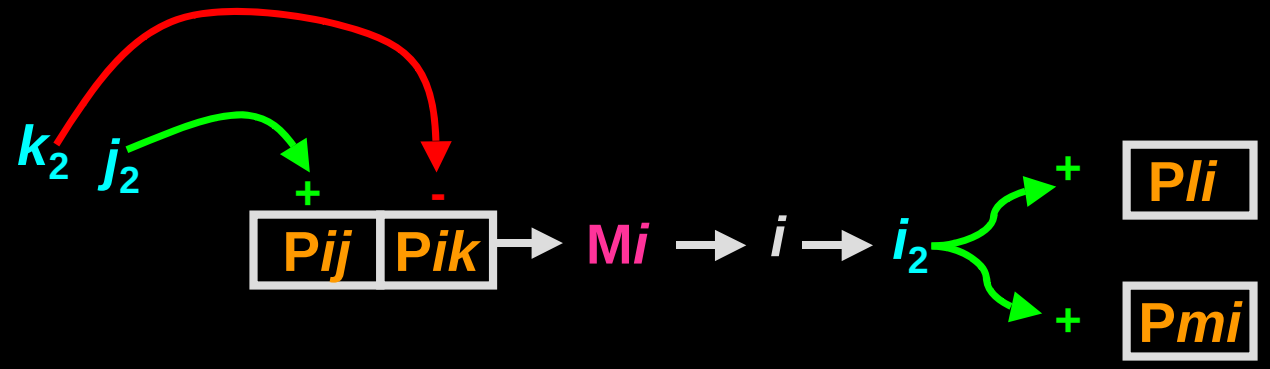
(n substrate molecules bind to enzyme)

$$E + S \Leftrightarrow C_1 \rightarrow P$$
$$S + C_{1`} \Leftrightarrow C_2 \rightarrow C_1 + P$$
$$S + C_{2`} \Leftrightarrow C_3 \rightarrow C_2 + P$$
$$\vdots$$
$$S + C_{n-1`} \Leftrightarrow C_n \rightarrow C_{n-1} + P$$

$$v = \frac{v_{\max}[S]^n}{K_M^n + [S]^n}$$

Transcriptional Regulation Module

(Adapted from [Barkai and Leibler, 2000])



$$\frac{d[Mi]}{dt} = k_{RPij}[Pij] + k_{Rj2Pij}[j_2Pij] + k_{RPik}[Pik] + k_{Rk2Pik}[k_2Pik] - k_{dMj}[Mi]$$

$$\frac{d[Pij]}{dt} = -k_{pij2}[Pij][j_2] + k_{upij2}[j_2Pij]$$

$$\frac{d[j_2Pij]}{dt} = k_{pij2}[Pij][j_2] - k_{upij2}[j_2Pij]$$

$$\frac{d[i]}{dt} = k_{Ti}[Mi] - k_{di}[i] - 2k_{i2}[i]^2 + 2k_{ui2}[i_2]$$

$$\frac{d[i_2]}{dt} = k_{i2}[i]^2 - k_{ui2}[i_2] - k_{di2}[i_2] - \{promoter\ binding\} + \{promoter\ unbinding\}$$

Comments

- Overall transcription = linear sum of bound and unbound promoter sites

Pros/Cons of Nonlinear ODE Models

- Not analytically solvable
- Host of powerful integration engines
- Time delay adds challenge (especially variable time delay)
- Often get lots of parameters

- Challenge: combinatorial explosion in parameters
- Opportunity: identify sensitivity

Computational Models of Chemical Reacting Systems

- **Continuous and deterministic** – (*rate equations*) Described by ordinary differential equations (ODE). Huge numbers of molecules.

- **Continuous and stochastic** – (*Langevin regime*) Valid under certain conditions. Described by Stochastic Differential Equations (SDE). Large numbers of molecules.

- **Discrete and stochastic** – Finest scale of representation for well stirred molecules. Exact description via Stochastic Simulation Algorithm (SSA) [Gillespie, 1977]. The *only* algorithm for small numbers of molecules.

Continuum Versus Stochastic Simulation

Continuum

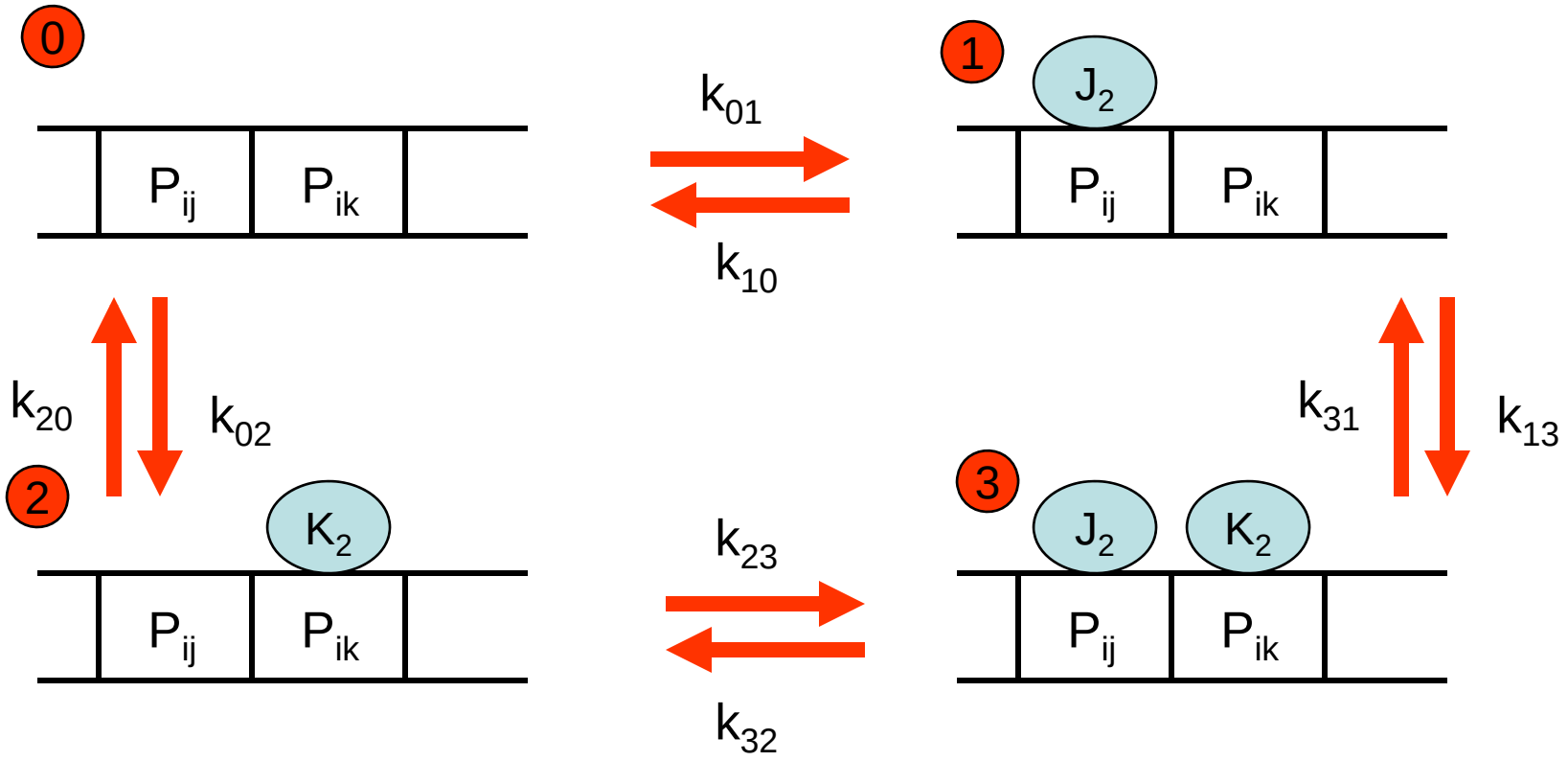
- Molar (mole/L)
- Reaction rates

- ODEs describing the changes in states (reactants, products)
- Initial value problem solver

Stochastic

- # of molecules
- Propensity functions (probability of rxns)
- ODEs describing the changes in probabilities of states (Master equation)
- Stochastic simulation algorithm (SSA)

4 Possible States



“Conservation” Equations

$$1J_2 + 1\left(P_{ij}\right)\left(P_{ik}\right) \underset{k_{10}}{\overset{k_{01}}{\rightleftarrows}} 1\left(J_2 \bullet P_{ij}\right)\left(P_{ik}\right)$$

$$1K_2 + 1\left(P_{ij}\right)\left(P_{ik}\right) \underset{k_{20}}{\overset{k_{02}}{\rightleftarrows}} 1\left(K_2 \bullet P_{ik}\right)\left(P_{ij}\right)$$

$$1J_2 + 1\left(K_2 \bullet P_{ik}\right)\left(P_{ij}\right) \underset{k_{32}}{\overset{k_{23}}{\rightleftarrows}} 1\left(J_2 \bullet P_{ij}\right)\left(K_2 \bullet P_{ik}\right)$$

$$1K_2 + 1\left(J_2 \bullet P_{ik}\right)\left(P_{ij}\right) \underset{k_{31}}{\overset{k_{13}}{\rightleftarrows}} 1\left(J_2 \bullet P_{ij}\right)\left(K_2 \bullet P_{ik}\right)$$

How to Simulate

- Monte Carlo methods to simulate molecule state changes

- Alternative: capture probabilities directly
  - Molecules represent configuration
  - State is probability distribution over all configurations
  - # probabilities scales with # configurations
  - Obeys differential equation

Simple Example

$$\overline{P}(t+\Delta t) = \begin{bmatrix} P(0,t+\Delta t) \\ P(1,t+\Delta t) \\ P(2,t+\Delta t) \\ P(3,t+\Delta t) \end{bmatrix} = \overline{\overline{A}}\ \overline{P}(t)$$

$$= \begin{bmatrix} 1-k_{01}[J_2]\Delta t - k_{02}[K_2]\Delta t & k_{10}\Delta t & k_{20}\Delta t & 0 \\ k_{01}[J_2]\Delta t & 1-k_{10}\Delta t - k_{13}[K_2]\Delta t & 0 & k_{31}\Delta t \\ k_{02}[K_2]\Delta t & 0 & 1-k_{20}\Delta t - k_{23}[J_2]\Delta t & k_{32}\Delta t \\ 0 & k_{13}[K_2]\Delta t & k_{23}[J_2]\Delta t & 1-k_{31}\Delta t - k_{32}\Delta t \end{bmatrix} \overline{P}(t)$$

ODE Version

- Take limit of small time step $$\frac{d\overline{P}}{dt} = \overline{\overline{B}}\,\overline{P}(t)$$
  - No such thing as equilibrium configuration
  - There is an equilibrium in probability distribution
  - Multiple algorithms available to solve

Chemical Master Equation (CME)

- Key assumption: well-stirred system

- Reactions are discrete random events with probability given by the propensity function

  $a_j(\boldsymbol{x})dt$: the probability, given $\boldsymbol{X}(t) = \boldsymbol{x}$, that one R_j reaction happen insides Ω in between the time interval $[t,t+dt)$

- No exact prediction of states, but can track the probability

  $P(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)$: the probability of $\boldsymbol{X}(t) = \boldsymbol{x}$, given the initial condition $\boldsymbol{X}(t_0) = \boldsymbol{x}_0$ $(t>t0)$

Chemical Master Equation Derivation

$$P\left(\boldsymbol{x},t+dt \mid \boldsymbol{x}_0,t_0\right) = P\left(\boldsymbol{x},t \mid \boldsymbol{x}_0,t_0\right)\left[1-\sum_{j=1}^{M} a_j\left(\boldsymbol{x}\right)dt\right] + \sum_{j=1}^{M} P\left(\boldsymbol{x}-\nu_j,t \mid \boldsymbol{x}_0,t_0\right) a_j\left(\boldsymbol{x}-\nu_j\right)dt$$

the system is already in state $\boldsymbol{x}$ and no reaction occurs

the system is in state $\boldsymbol{x}-\nu_j$ and reaction R_j occurs

Chemical Master Equation

$$\frac{\partial P\left(\boldsymbol{x},t \mid \boldsymbol{x}_0,t_0\right)}{\partial t} = \sum_{j=1}^{M}\left[a_j\left(\boldsymbol{x}-\nu_j\right)P\left(\boldsymbol{x}-\nu_j,t \mid \boldsymbol{x}_0,t_0\right) - a_j\left(\boldsymbol{x}\right)P\left(\boldsymbol{x},t \mid \boldsymbol{x}_0,t_0\right)\right]$$

… in general, there exists no analytical solution

Stochastic Simulation Algorithm (SSA)

[Gillespie, 1976]

- Main idea:

  $p(\tau,j|\boldsymbol{x},t)d\tau$: the probability, given $\boldsymbol{X}(t) = \boldsymbol{x}$, that the next reaction will occur in within $[t+\tau, t+\tau+d\tau)$, and will be an R_j reaction

- Joint probability function of "time to next reaction" (τ) and "index of next reaction" (j)

  $p(\tau,j|\boldsymbol{x},t)d\tau$ is a function of the propensities

Stochastic Simulation Algorithm

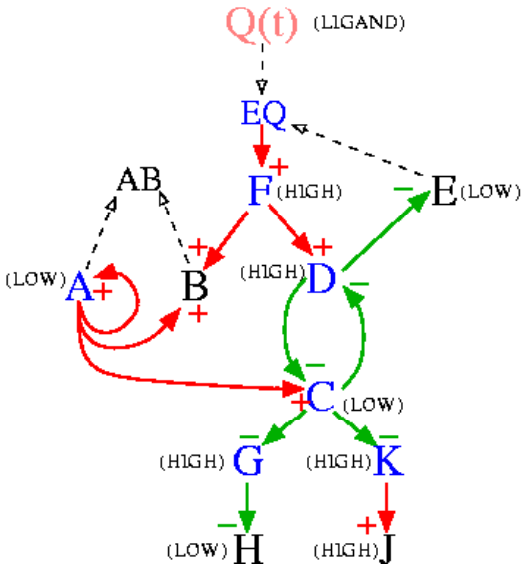
1. Initialize the time $t = t_0$ and the state $\boldsymbol{x} = \boldsymbol{x}_0$
2. Evaluate the propensities $a_j(\boldsymbol{x})$
3. Pick two random numbers from uniform distribution
4. Compute τ and j

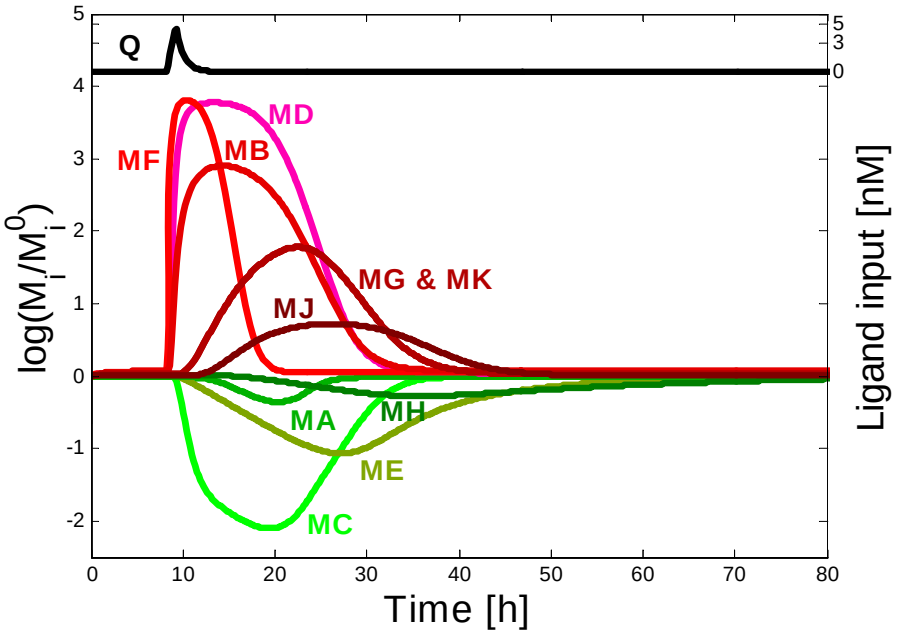
$$\tau = \frac{1}{\sum a_i(\boldsymbol{x})} \ln\left(\frac{1}{r_1}\right)$$

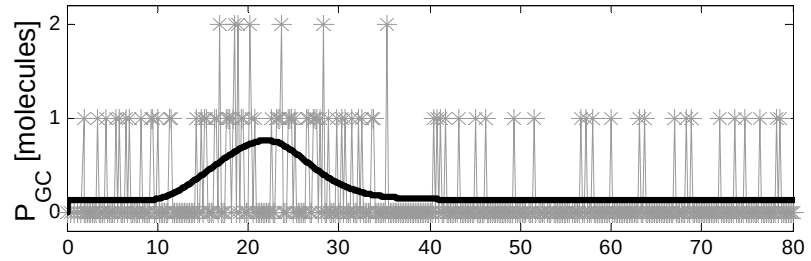
$$j = \text{ smallest integer : } \sum_{i=1}^{j} a_i(\boldsymbol{x}) \geq r_2 \sum a_i(\boldsymbol{x})$$

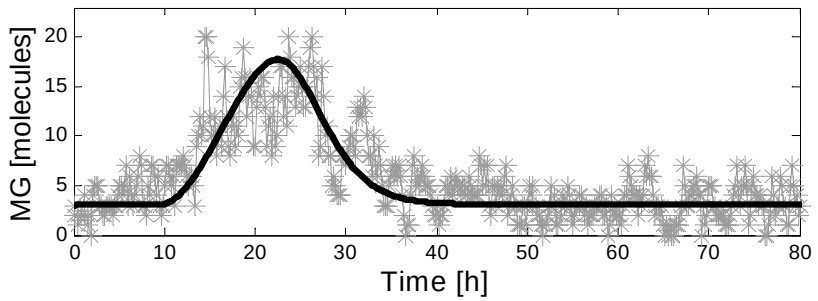
5. Step forward in time by τ and update the states
   $x(t+\tau) = x+\nu_j$
6. Repeat from beginning

Stochastic Gene Network Response

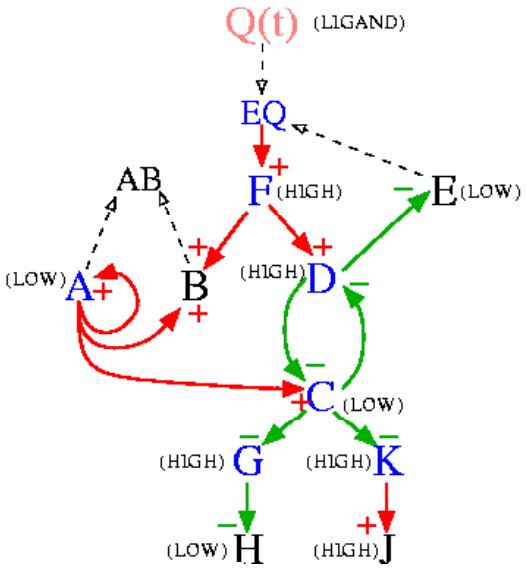


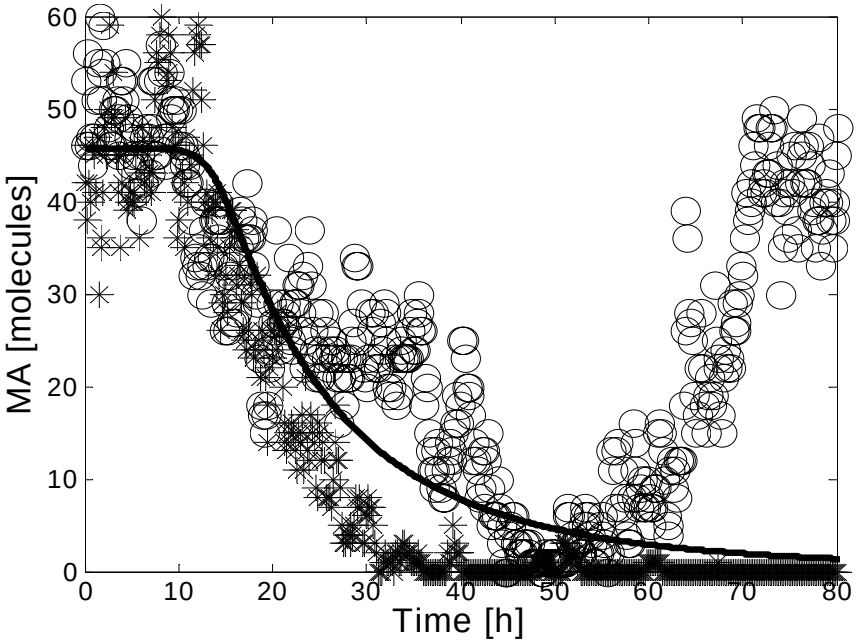


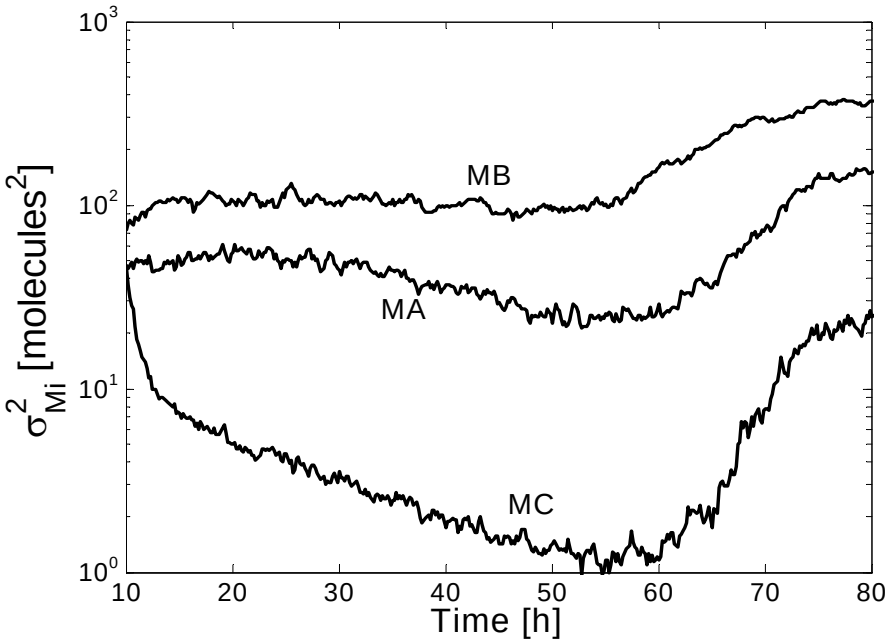




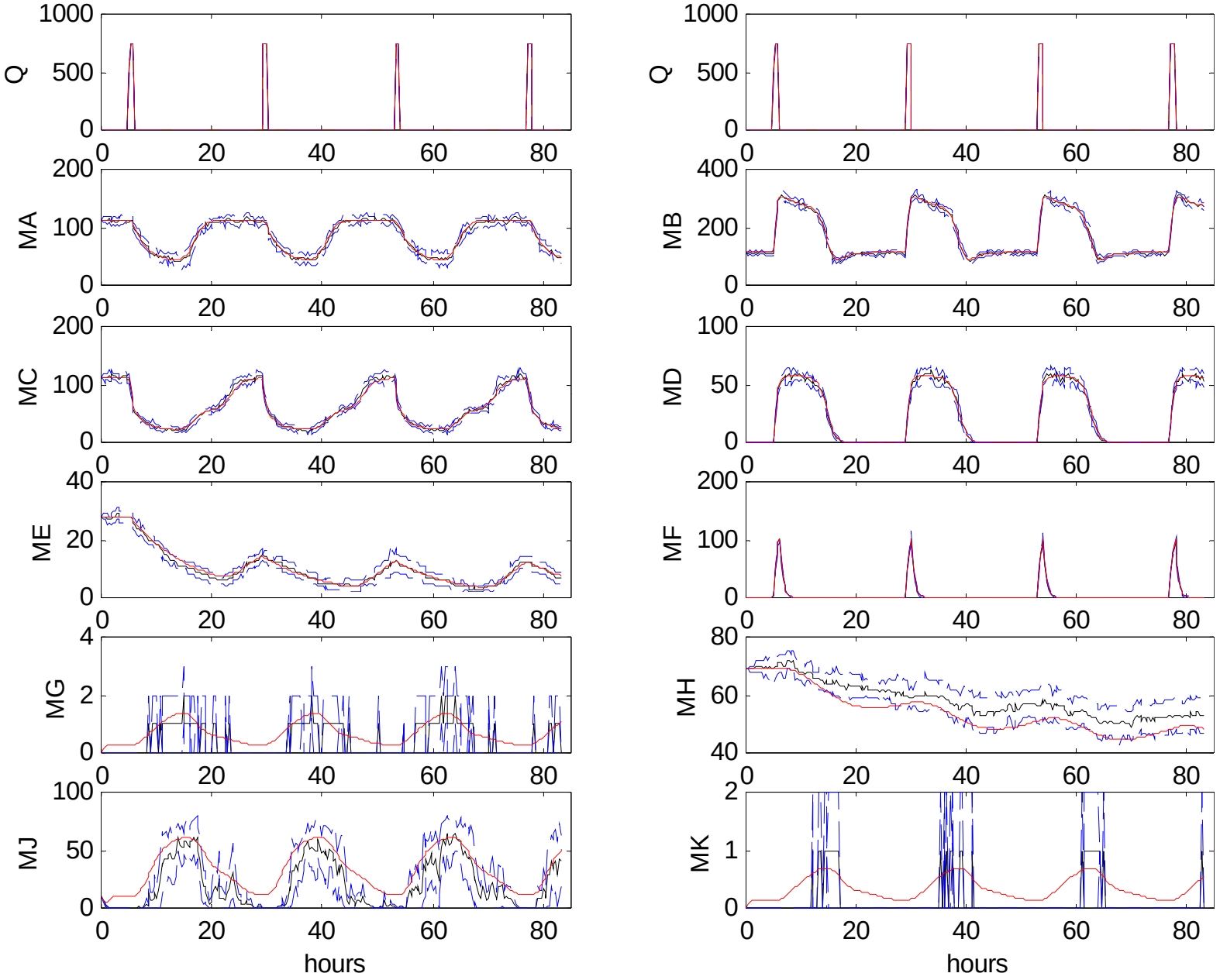
Stochastic Gene Network Response



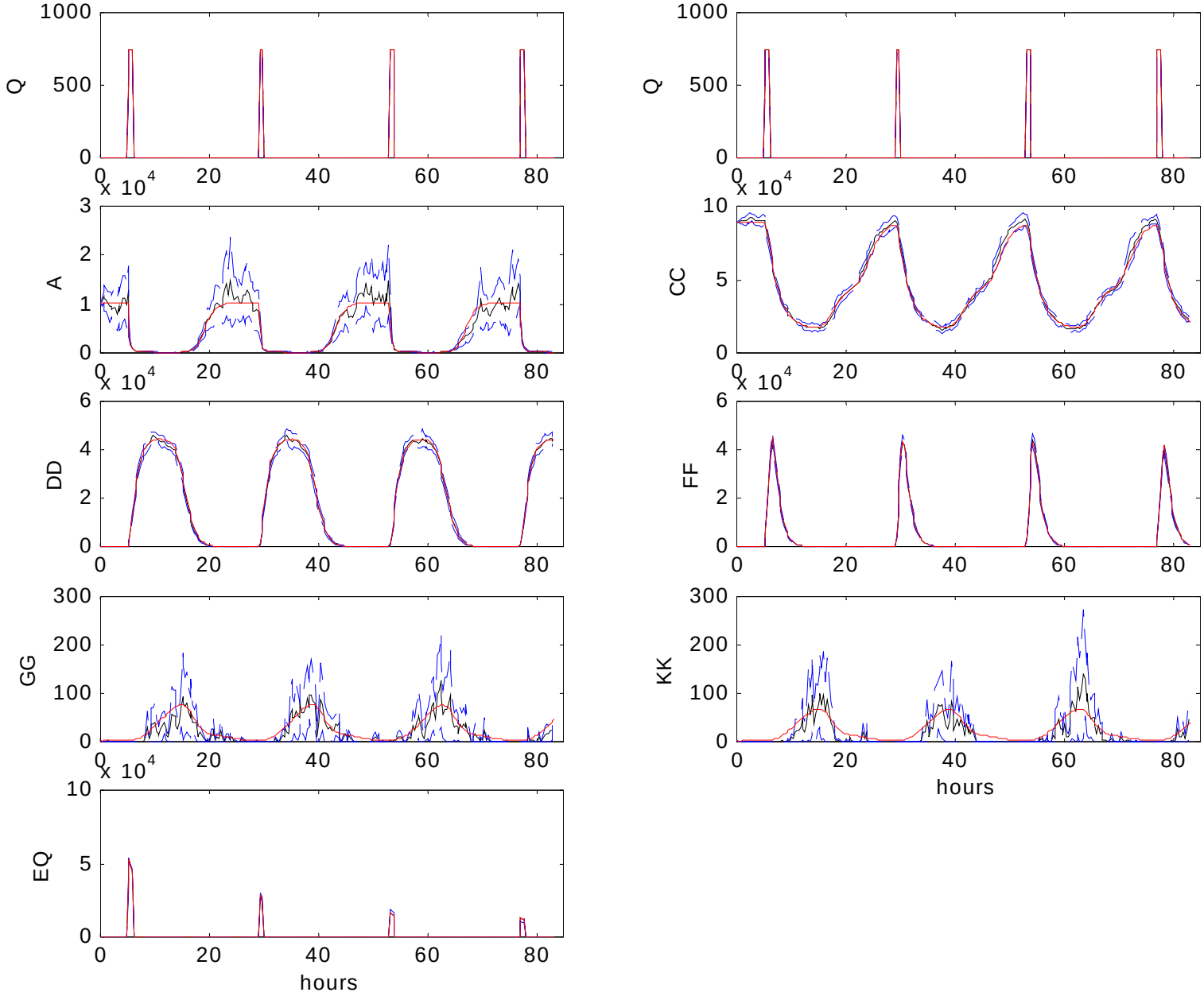




mRNAs (molecules) vs. time (hours)



Transcription factors (molecules) vs. time (hours)



Observations on Stochastic Simulation

- One can go even deeper
  - Molecular dynamics
  - Quantum mechanical description
- But not tractable for gene regulation (nor insightful)
- Guidelines:

1. # molecules sufficiently high that single molecule changes can be approximated by change in continuous concentration → ODE
2. Fluctuations about mean << mean → ODE
3. Otherwise, and if solution well mixed locally → stochastic

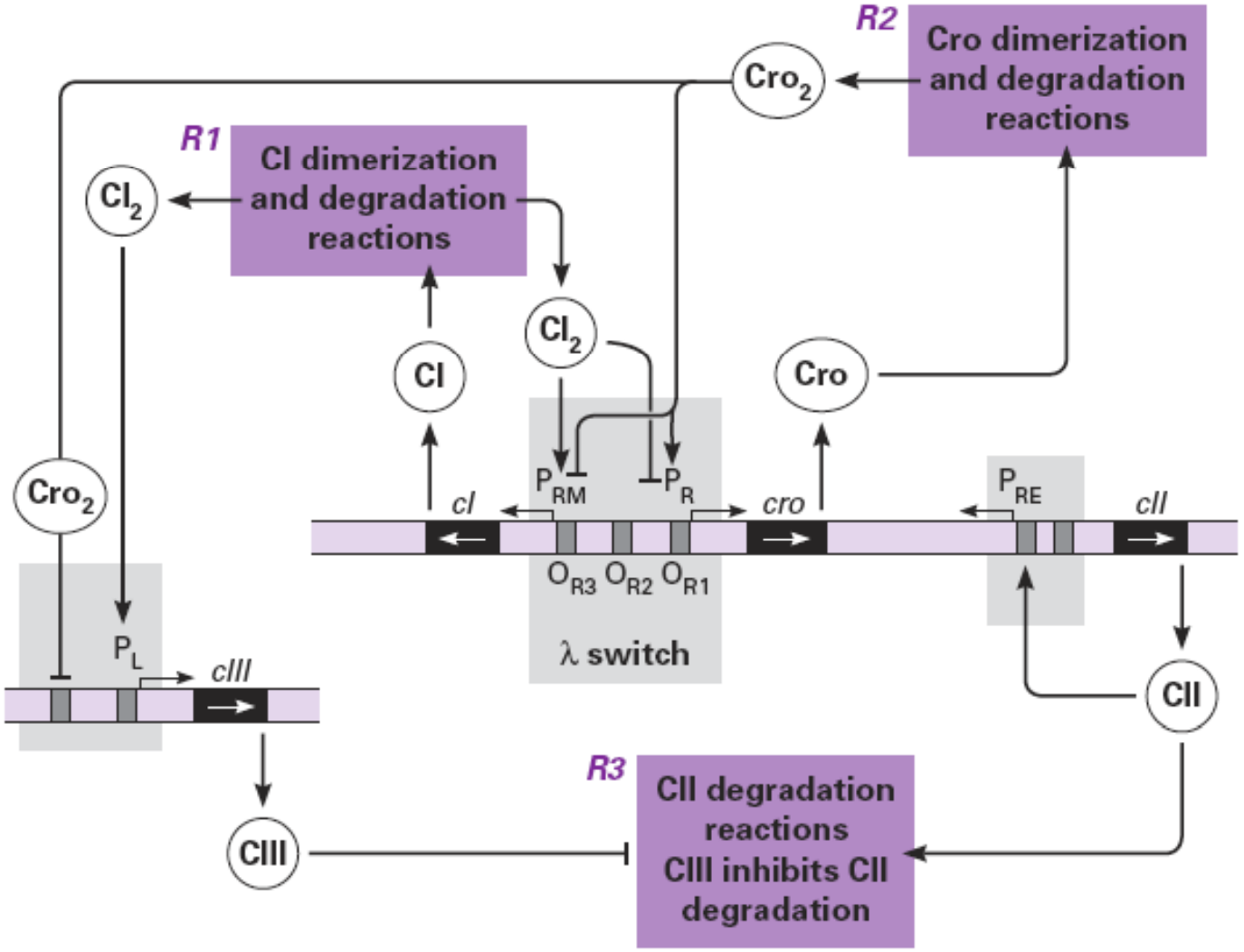
Lambda Phage

Lambda Phage Revisited

- Virus that infects E. Coli
- 2 developmental pathways
  - Replicate/lyse
  - Lysogeny
- Simple developmental switch
- 50,000 bp genome sequenced early on, lots of regulatory knowledge
- Architecture suggests two approaches to analysis:
  - Boolean circuits/switches
  - Stochastic distribution – bistable switch

- References:
  - [Bower & Bolouri, Ch. 2]
  - [McAdams & Shapiro, 1995]
  - [Arkin et al., 1998]

Lambda Phage “Circuit”



[Campbell & Heyer, 2006]

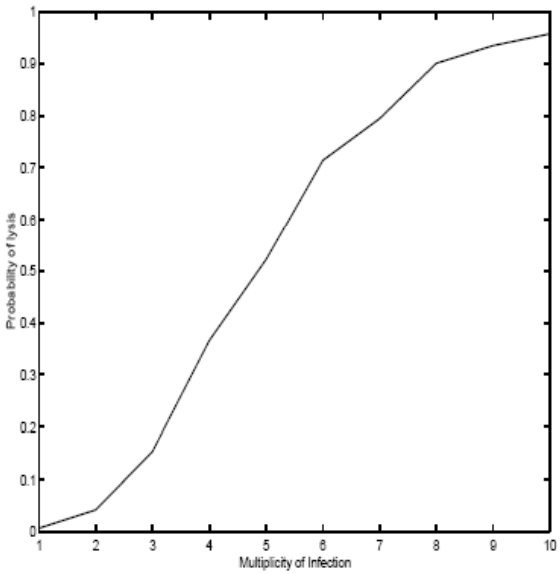
Discrete Stochastic Model

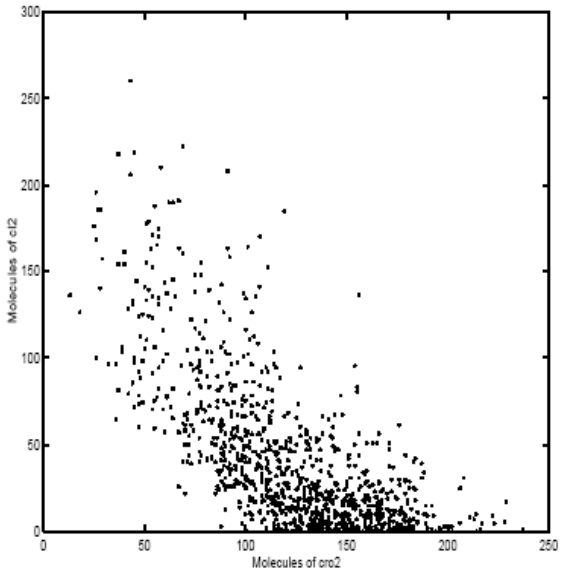
[Gibson/Bruck]

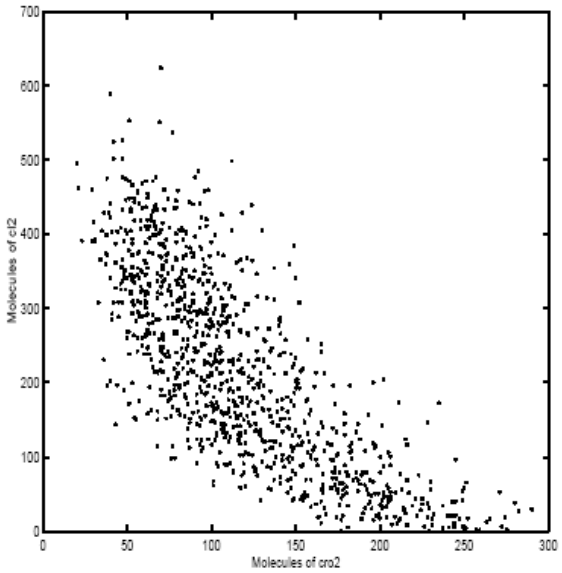
- Focus on N, Cro_2, and CI_2
  - CI_2 is present at high levels in lysogens and represses expression of all other genes
  - Cro is key in lytic pathway: inhibits production of CI_2 and controls production of key proteins (cell lysis, replication of λ DNA)
  - CI_2 and Cro are mutually inhibitory
  - N produced early in life cycle, production halted after fate choice

- Model: Regulation of N as function (Cro and CI_2)

Final time distribution of cro2 and repressor







[Gibson & Bruck, Caltech Tech Report 26]

Basic Model for N Production

$$\mathrm{RNAP{\bullet}DNA_{closed}} \xrightarrow{k_1} \mathrm{RNAP{\bullet}DNA_{open,\ 0}}$$

$$\mathrm{RNAP{\bullet}DNA_{open,\ n}} \xrightarrow{k_2} \mathrm{RNAP{\bullet}DNA_{open,\ n+1}}$$

$$\mathrm{RNAP{\bullet}DNA_{open,\ MAX}} \xrightarrow{k_3} \mathrm{RNAP_{free} + DNA_{free} + mRNA_{free}}$$

$$\mathrm{RNase + mRNA_{free}} \xrightarrow{k_4} \mathrm{RNase}$$

$$\mathrm{Ribosome + mRNA_{free}} \xrightarrow{k_5} \mathrm{Ribosome{\bullet}mRNA_0}$$

$$\mathrm{Ribosome{\bullet}mRNA_n} \xrightarrow{k_6} \mathrm{Ribosome{\bullet}mRNA_{n+1}}$$

$$\mathrm{Ribosome{\bullet}mRNA_{MAX}} \xrightarrow{k_7} \mathrm{Ribosome_{free} + mRNA_{free} + protein}$$

$$\mathrm{protein} \xrightarrow{k_8} \mathrm{no\ protein}$$

$$\mathrm{protein + protein} \xrightarrow{k_9} \mathrm{protein{\bullet}protein}$$

$$\mathrm{protein{\bullet}protein} \xrightarrow{k_{10}} \mathrm{protein + protein}$$

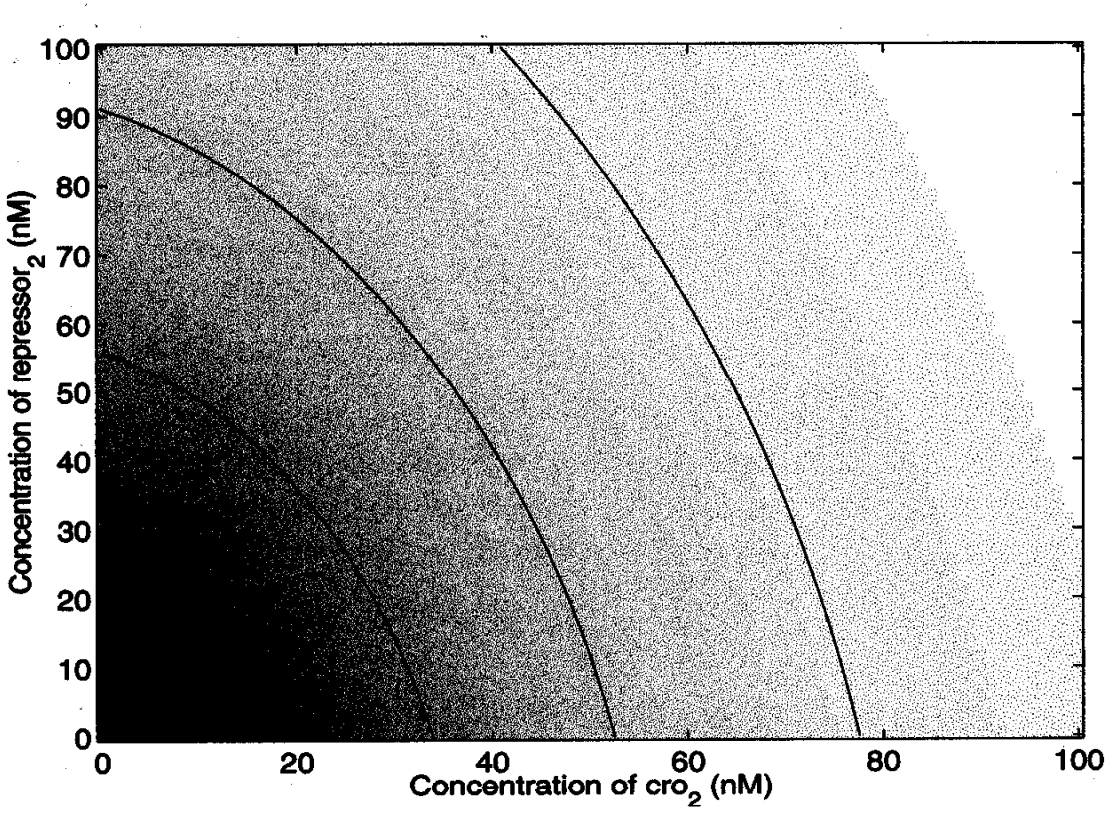
Basic Model – TF-DNA Binding

TABLE 1

Parameters for promoters P_{RE} and P_L

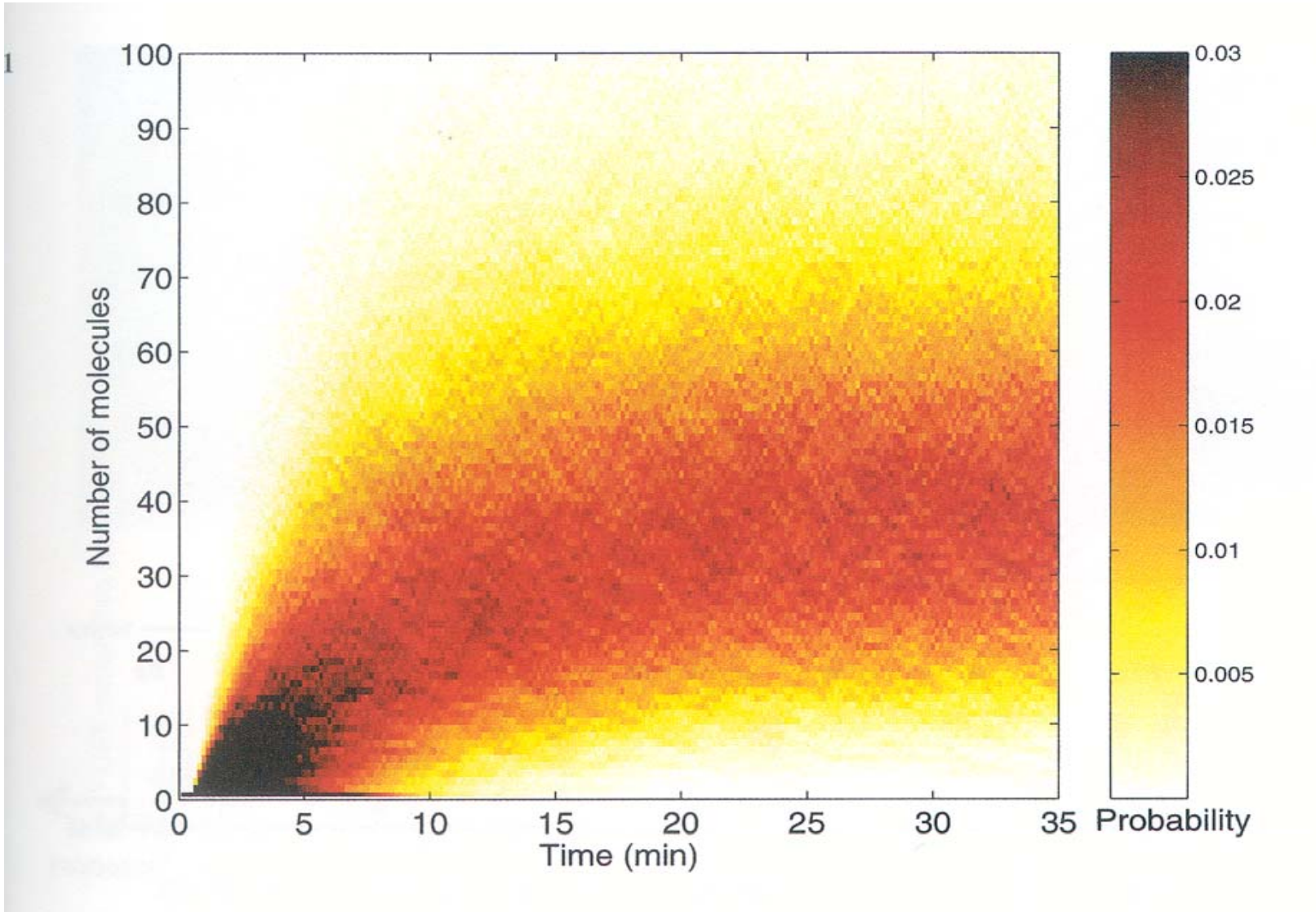
No.	State O_1	State O_2	ΔG (kcal mol^{-1})	k_{oc} (sec^{-1})
		Promoter P_{RE}		
1	—	—	0.0	0.0
2	—	RNAP	−9.9	0.00004
3	CII	—	−9.7	0.0
4	CII	RNAP	−21.5	0.015
		Promoter P_L		
1	—	—	0.0	0.0
2	Cro_2	—	−10.9	0.0
3	—	Cro_2	−12.1	0.0
4	CI_2	—	−11.7	0.0
5	—	CI_2	−10.1	0.0
6	—	RNAP	−12.5	0.011
7	Cro_2	Cro_2	−22.9	0.0
8	Cro_2	CI_2	−20.9	0.0
9	CI_2	Cro_2	−22.8	0.0
10	CI_2	CI_2	−23.7	0.0

Dependence of Transcription Initiation



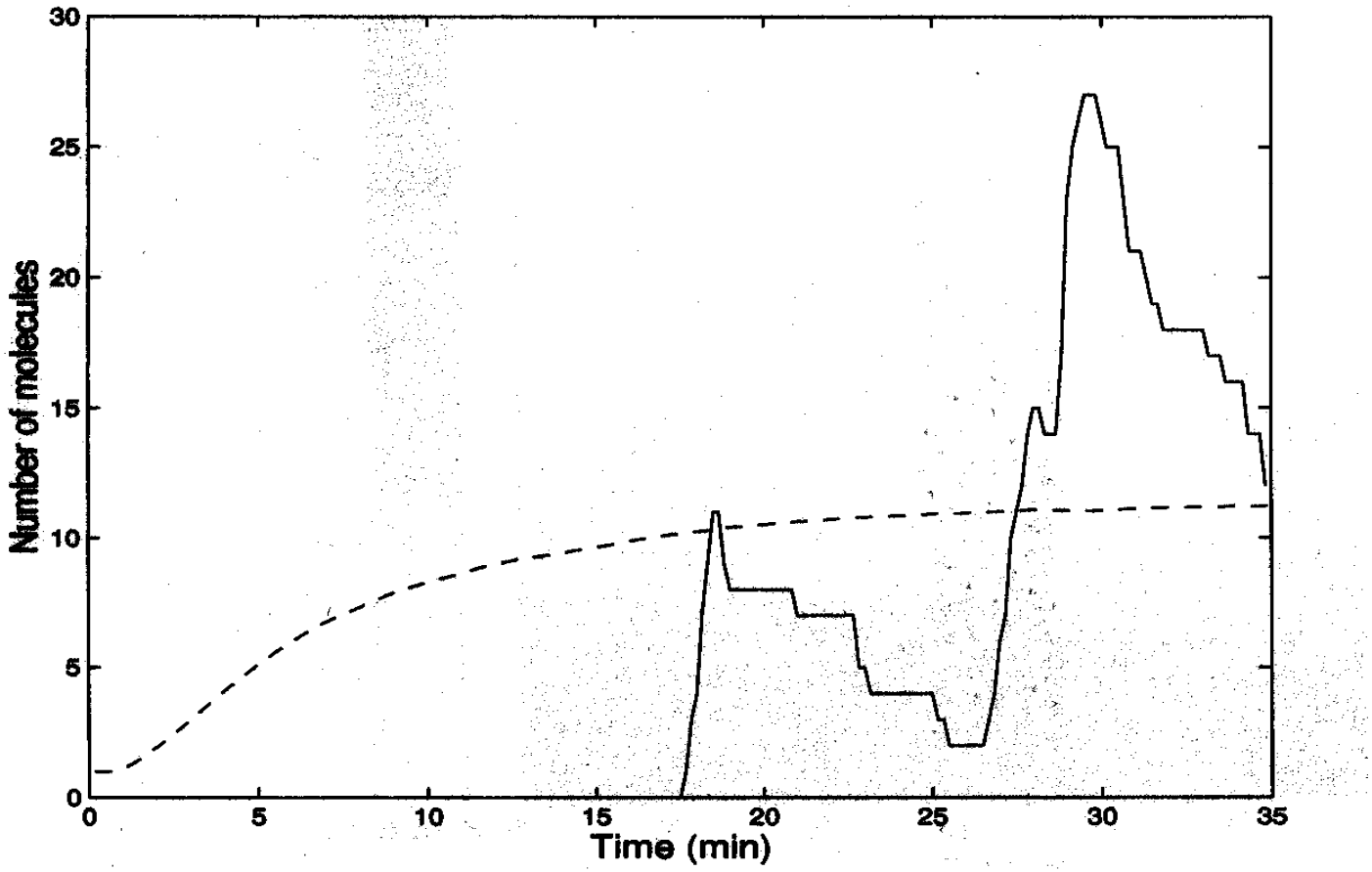
[Bower & Bolouri, 2002]

Probability of N at time t



[Bower & Bolouri, 2002]

Impact of Repressor



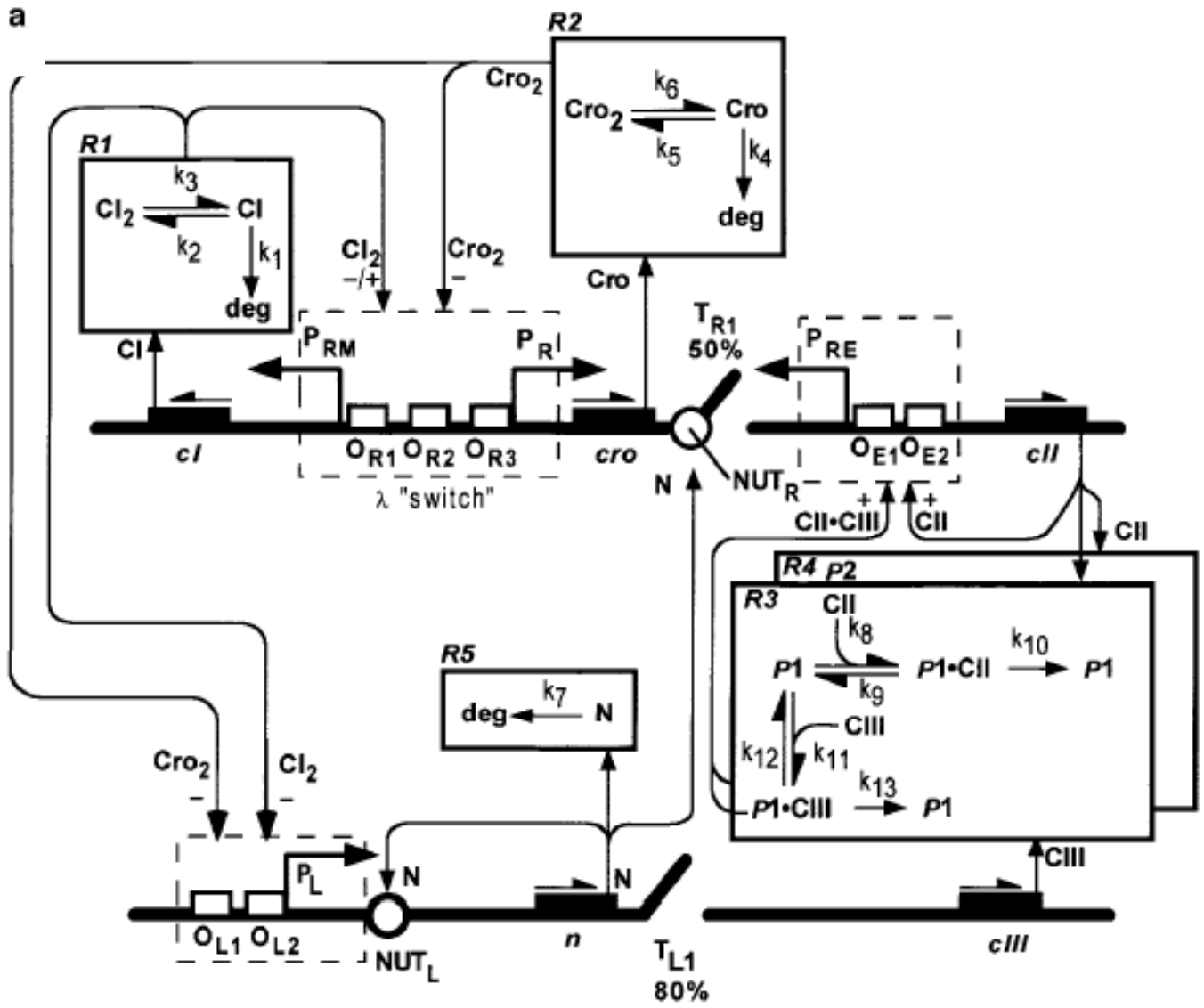
[Bower & Bolouri, 2002]

Questions to Address

- What is probability that the given mRNA will start translation rather than be degraded & vice versa?
- What is the probability that n proteins will be produced from one mRNA molecule before it degrades?
- What is the average number of proteins produced/mRNA transcript?

Detailed Circuit Schematic

[Arkin et al., *Genetics*, 1998]



Key Assumptions

1. Cell generation time is deterministic
2. Linear growth in volume
3. Housekeeping molecules constitutively expressed
4. .
5. .
6. .
7. Gene expression is stochastic
8. .
9. .
10. Target cells are infected simultaneously
11. Well mixed (cell)

Detailed Model

Transcription/Translation

Parameters for transcription and translation reactions

Reaction/event	Parameter	References and comments
Transcription reactions		
$RNAP{\cdot}DNA_n \xrightarrow{k_{22}} RNAP{\cdot}DNA_{n+1}$	$k_{22} = 30$ nt sec^{-1}	Selected as an average rate. Measured elongation rates vary widely, depending on DNA template and cell state (GOTTA *et al.* 1991; KENNELL and RIEZMAN 1977; KORNBERG and BAKER 1992; VOGEL and JENSEN 1994)
$RNAP{\cdot}DNA_{Nut(L,R)} \xrightarrow{k_{23}} RNAP{\cdot}DNA_{Nut(L,R)+1}$	$k_{23} = 5$ nt sec^{-1}	
$RNAP{\cdot}DNA_{Nut(L,R)} + N \underset{k_{25}}{\overset{k_{24}}{\rightleftharpoons}} RNAP{\cdot}N{\cdot}DNA_{Nut(L,R)+1}$	$k_{24} = 0.145$ (M sec)$^{-1}$ $k_{25} = 0.1$ sec^{-1}	Selected to produce termination and antitermination consistent with LI *et al.* (1992) and WHALEN *et al.* (1988)
$RNAP{\cdot}N{\cdot}DNA_{Nut(L,R)} \xrightarrow{k_{26}} RNAP{\cdot}N{\cdot}DNA_{Nut(L,R)+1}$	$k_{26} = 30$ nt sec^{-1}	
$RNAP{\cdot}DNA_{T_{R1}} \xrightarrow{k_{27}} RNAP{\cdot}DNA_{T_{R1}+1}$	$k_{27} = 15$ nt sec^{-1}	Selected to yield 50% termination at $N = 0$ nM (DAMBLY-CHAUDIERE *et al.* 1983; FRIEDMAN and GOTTESMAN 1983)
$RNAP{\cdot}DNA_{T_{R1}} \xrightarrow{k_{28}} RNAP + DNA_{T_{R1}}$	$k_{28} = 15$ sec^{-1}	
$RNAP{\cdot}N{\cdot}DNA_{T_{R1}} \xrightarrow{k_{29}} RNAP{\cdot}N{\cdot}DNA_{T_{R1}+1}$	$k_{29} = 30$ nt sec^{-1}	Assumption that antiterminated RNAP passes terminator freely
$RNAP{\cdot}DNA_{T_{L1}} \xrightarrow{k_{31}} RNAP{\cdot}DNA_{T_{L1}+1}$	$k_{31} = 5$ nt sec^{-1}	Selected to yield 80% termination at $N = 0$ nM
$RNAP{\cdot}DNA_{T_{L1}} \xrightarrow{k_{32}} RNAP + DNA_{T_{L1}}$	$k_{32} = 25$ sec^{-1}	Selected to yield 80% termination at $N = 0$ nM
$RNAP{\cdot}N{\cdot}DNA_{T_{L1}} \xrightarrow{k_{33}} RNAP{\cdot}N{\cdot}DNA_{T_{L1}+1}$	$k_{33} = 30$ nt sec^{-1}	Assumption: antiterminated RNAP passes terminator freely
Translation reactions		
$Ribosome + RNA_{RBS} \xrightarrow{k_{34}} Ribosome{\cdot}RNA_{RBS}$	$k_{34} = 0.002$ (M sec)$^{-1}$	(KENNELL and RIEZMAN 1977; SORENSEN and PEDERSEN 1991)
$Ribosome + RNA_n \xrightarrow{k_{35}} Ribosome{\cdot}RNA_{n+1}$	$k_{35} = 100$ nt sec^{-1}	(ADHYA and GOTTESMAN 1982; KENNELL and RIEZMAN 1977; SORENSEN and PEDERSEN 1991)
$RNase + RNA_{RBS} \xrightarrow{k_{36}} RNase$	$k_{36}{\cdot}RNase = 0.2$ sec^{-1}	Adjusted to get an average of 10 proteins per transcript
Average number of proteins per transcript (all transcripts)	10	(KEPES 1963; YARCHUK *et al.* 1992)

Detailed Model

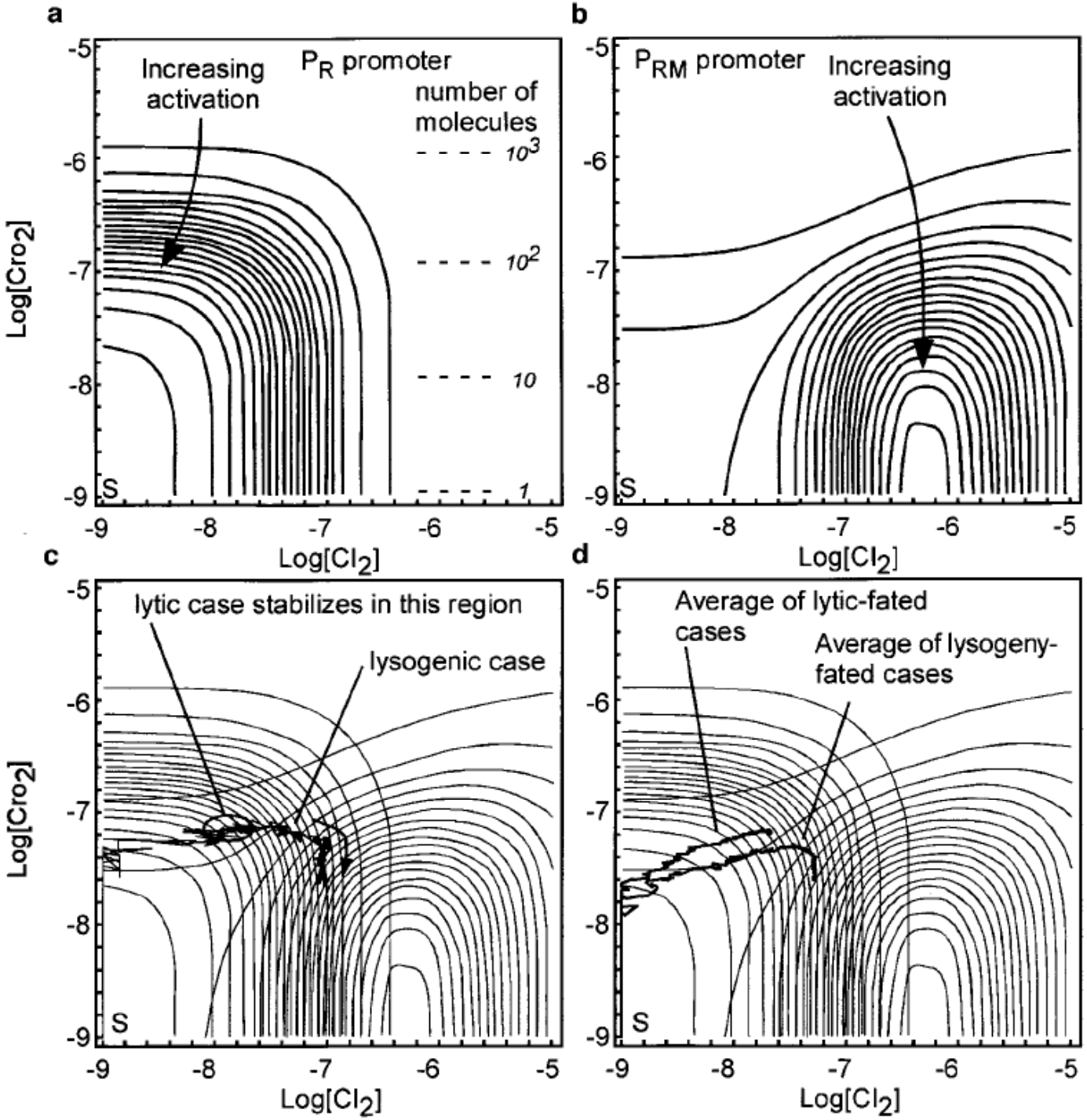
Housekeeping/Nongenetic Elements

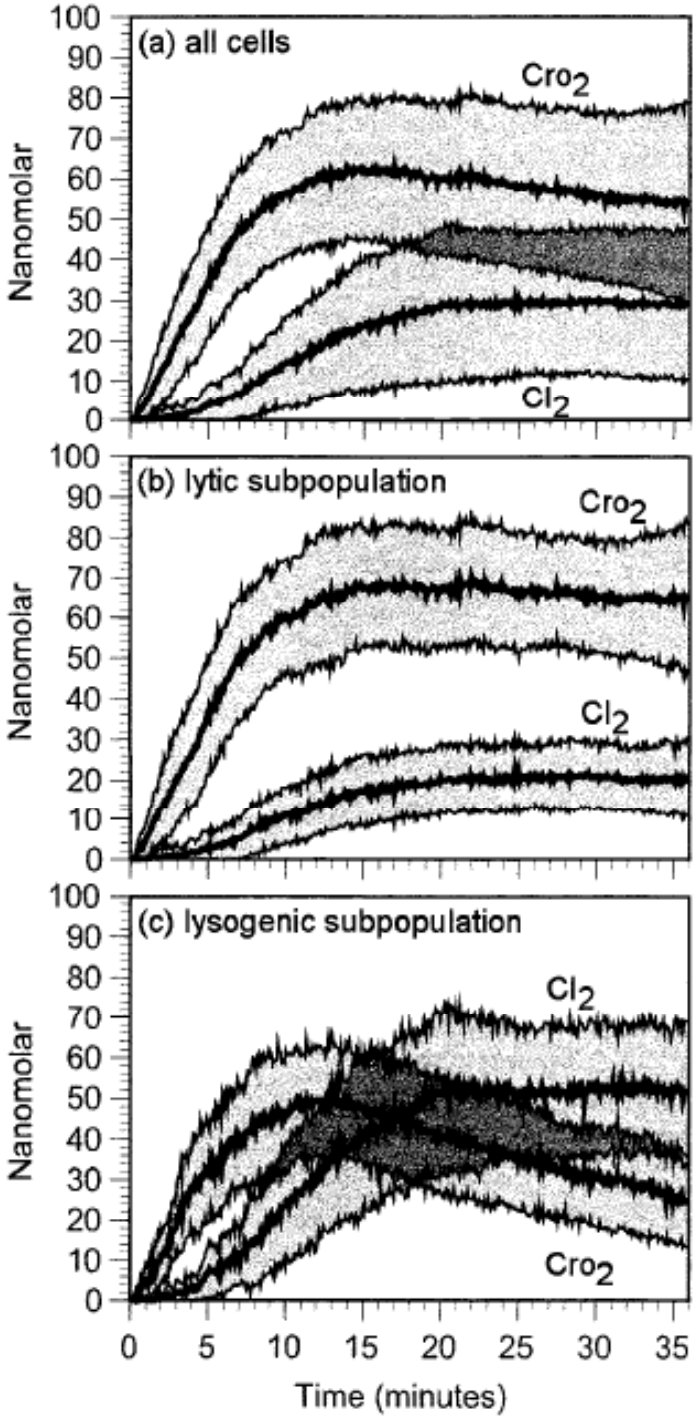
Parameters for housekeeping and nongenetic reactions

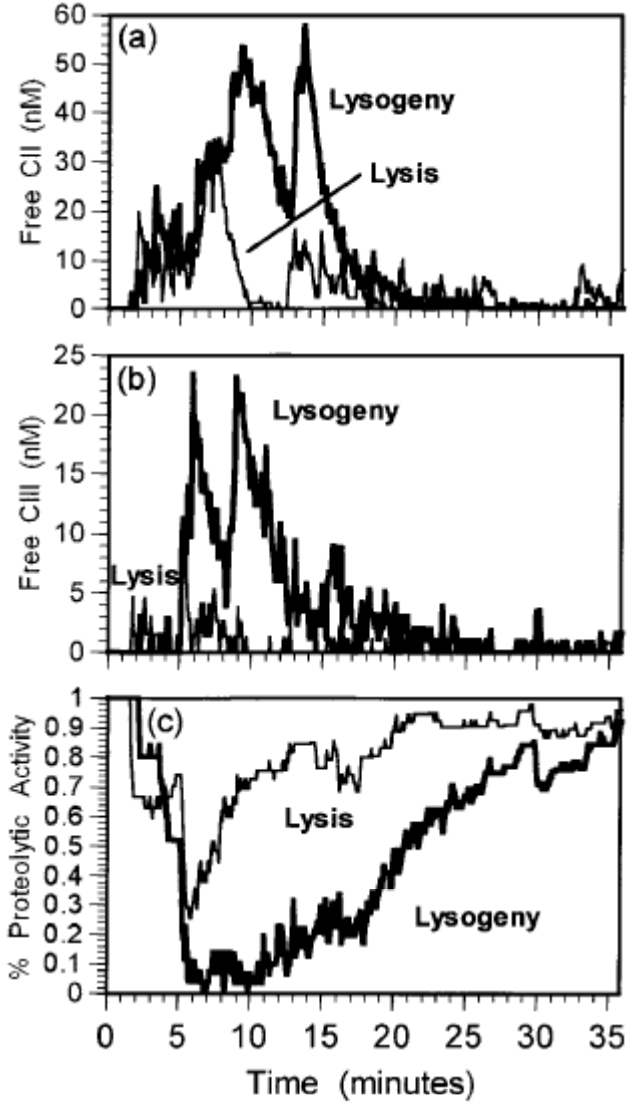
Reaction/event	Parameter	References and comments
Housekeeping reactions		
Available RNAP	RNAP = 30 nM	McClure (1980, 1983)
Available ribosomes	Ribosomes = 500 nM	
Cell volume (t) = $(1 + k_0 * t) \times 10^{-15}$ liters	$k_0 = 4.76 \times 10^{-18}$ liters sec^{-1}	To double initial cell volume of 10^{-15} liters in 35 min
Nongenetic reactions[a]		
CI $\xrightarrow{k_1}$ ( )	$k_1 = 0.0007\ sec^{-1}$	Selected to yield a CI/Cl_2 life time of approximately 40 min (Reinitz and Vaisnys 1990) in the concentration range between 20 and 100 nM
2·CI $\underset{k_3}{\overset{k_2}{\rightleftarrows}}$ CI_2	$k_2 = 0.05\ M^{-1}\ sec^{-1}$	Burz *et al.* (1994); Shea and Ackers (1985)
	$k_3 = 0.5\ sec^{-1}$	
Cro $\xrightarrow{k_4}$ ( )	$k_4 = 0.0025\ sec^{-1}$	Selected to match Cro/Cro_2 lifetime of approximately 30 min (Reinitz and Vaisnys 1990) in the concentration range between 20 and 100 nM
2·Cro $\underset{k_6}{\overset{k_5}{\rightleftarrows}}$ Cro_2	$k_5 = 0.05\ M^{-1}\ sec^{-1}$	Reinitz and Vaisnys (1990); Sauer (1979)
	$k_6 = 0.5\ sec^{-1}$	
N $\xrightarrow{k_7}$ ( )	$k_7 = 0.00231\ sec^{-1}$	Gottesman and Gottesman (1981)
P1 concentration[b]	*P1* = 35 nM	Adjusted to match the % lysogeny *vs.* API data (Kourilsky 1973)
CII + *P1* $\underset{k_9}{\overset{k_8}{\rightleftarrows}}$ *P1*·CII	$k_8 = 0.01\ M^{-1}\ sec^{-1}$	Selected to match CII half-life in Gottesman and Gottesman (1981)
P1·CII $\xrightarrow{k_{10}}$ *P1*	$k_9 = 0.01\ sec^{-1}$	
	$k_{10} = 0.002\ sec^{-1}$	
CIII + *P1* $\underset{k_{12}}{\overset{k_{11}}{\rightleftarrows}}$ *P1*·CIII	$k_{11} = 0.01\ M^{-1}\ sec^{-1}$	Selected to match CIII protection of CII degradation (Hoyt *et al.* 1982; Rattray *et al.* 1984) and CIII half-life Kornitzer *et al.* (1991a,b)
P1·CIII $\xrightarrow{k_{13}}$ *P1*	$k_{12} = 0.001\ sec^{-1}$	
	$k_{13} = 0.0001\ sec^{-1}$	
P2 concentration	*P2* = 140 nM	
CII + *P2* $\underset{k_{15}}{\overset{k_{14}}{\rightleftarrows}}$ *P2*·CII	$k_{14} = 0.00025\ M^{-1}sec^{-1}$	Selected to match CII half-life in Gottesman and Gottesman (1981)
P2·CII $\xrightarrow{k_{16}}$ *P2*	$k_{15} = 0.065\ sec^{-1}$	
	$k_{16} = 0.6\ sec^{-1}$	
CIII + *P2* $\underset{k_{18}}{\overset{k_{17}}{\rightleftarrows}}$ *P2*·CIII	$k_{17} = 0.01\ M^{-1}sec^{-1}$	Selected to match CIII protection of CII from degradation (Hoyt *et al.* 1982; Rattray *et al.* 1984) and CIII half-life (Kornitzer *et al.* 1991a,b)
P2·CIII $\xrightarrow{k_{19}}$ *P2*	$k_{18} = 0.01\ sec^{-1}$	
	$k_{19} = 0.001\ sec^{-1}$	

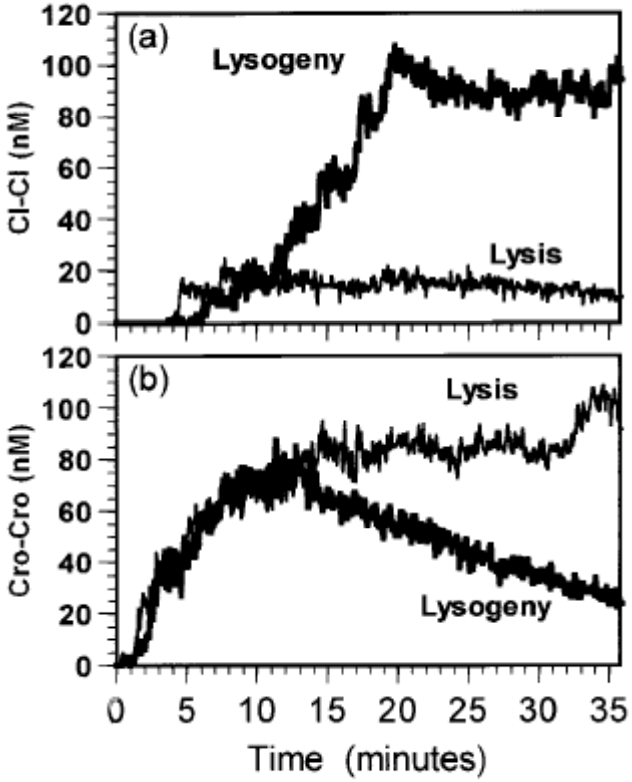
[a] The ( ) notation indicates degradation.

[b] The parameters on this and following lines for the CII/CIII proteases (here labeled *P1* and *P2*) are those corresponding to the "Full" curve in Figure 6a. The Hfl-related parameters below are adjusted to match half-lives of their targeted proteins and to match the percent lysogeny *vs.* API data in Kourilsky (1973) as described in the text.


a
Increasing activation
P_R promoter
number of molecules
10^3
10^2
10
1
S
Log[Cro_2]
Log[Cl_2]
b
P_{RM} promoter
Increasing activation
S
Log[Cl_2]
c
lytic case stabilizes in this region
lysogenic case
S
Log[Cro_2]
Log[Cl_2]
d
Average of lytic-fated cases
Average of lysogeny-fated cases
S
Log[Cl_2]


(a) all cells
Cro2
Cl2
(b) lytic subpopulation
Cro2
Cl2
(c) lysogenic subpopulation
Cl2
Cro2
Nanomolar
Time (minutes)
0
5
10
15
20
25
30
35
0
10
20
30
40
50
60
70
80
90
100


(a)
Lysogeny
Lysis
Free CII (nM)
(b)
Lysogeny
Lysis
Free CIII (nM)
(c)
Lysis
Lysogeny
% Proteolytic Activity
Time (minutes)


(a)
Lysogeny
Lysis
CI-CI (nM)
(b)
Lysis
Lysogeny
Cro-Cro (nM)
Time (minutes)

Insights Gained from Systems Biology Approach

- Study reveals how thermal fluctuations can be exploited by the regulatory circuit designs of developmental switches to produce different phenotypic outcomes
- Specific conclusions about role of termination sites on level of lysogeny (*in silico* mutations)
- Generic switch insights
- Robust yet random performance (hypothesis: dispersion in timing across population and not dispersion in outcome)

Circadian Rhythm

Circadian Rhythms

Circadian rhythms = self-sustained biological rhythms characterized by a *free-running period* of about 24h (*circa diem*)

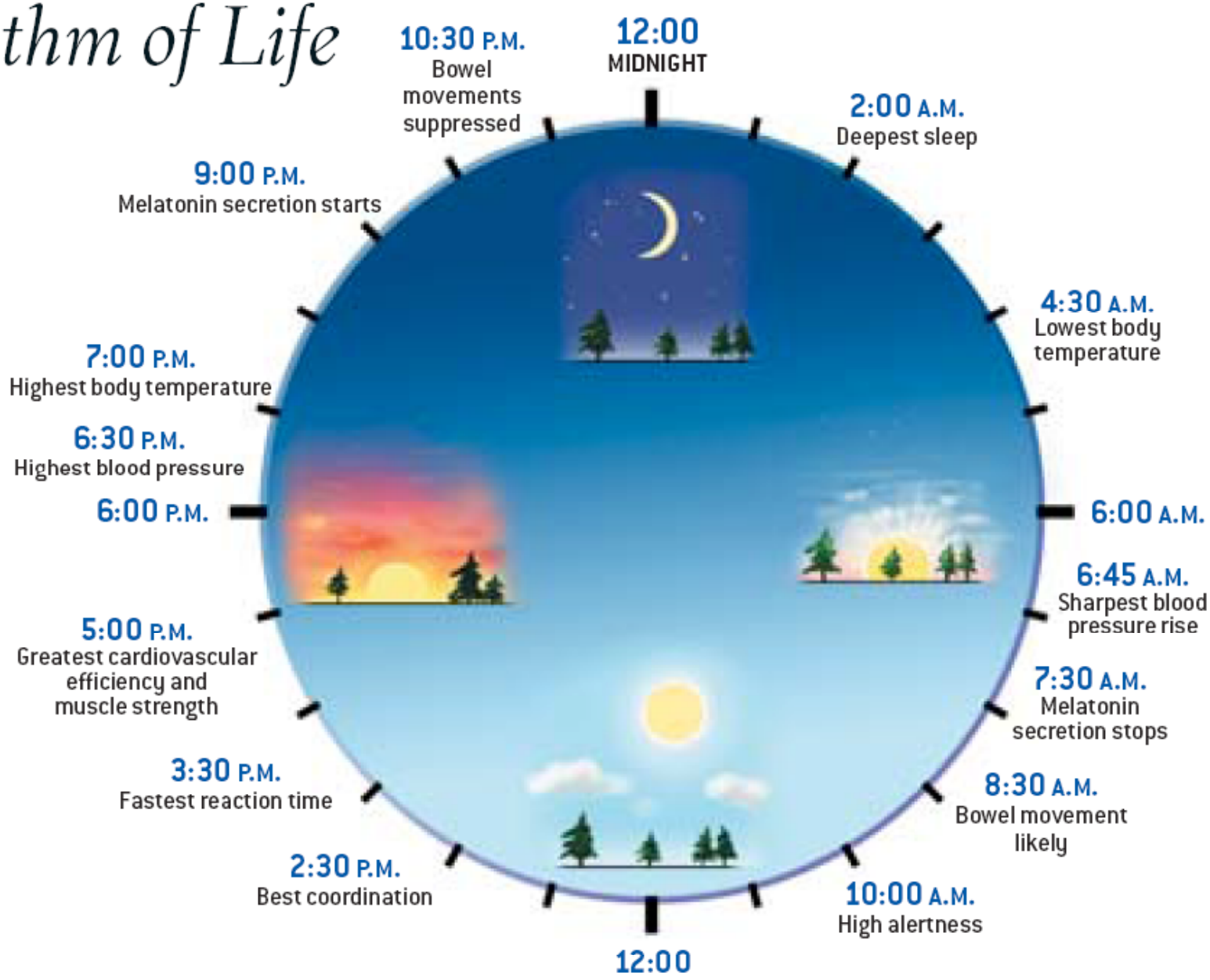
Circadian rhythms characteristics:

- General – bacteria, fungi, plants, flies, fish, mice, humans, etc.
- Entrainment by light-dark cycles (**zeitgeber**)
- Phase shifting by light pulses
- Temperature compensation

Circadian rhythms occur at the molecular level

The Rhythm of Life

THE CIRCADIAN CLOCK affects the daily rhythms of many physiological processes. The diagram at the right depicts the circadian patterns typical of someone who rises early in the morning, eats lunch around noon and sleeps at night. Although circadian rhythms tend to be synchronized with cycles of light and dark, other factors—such as ambient temperature, meal times, stress and exercise—can influence the timing as well. —*K.W.*

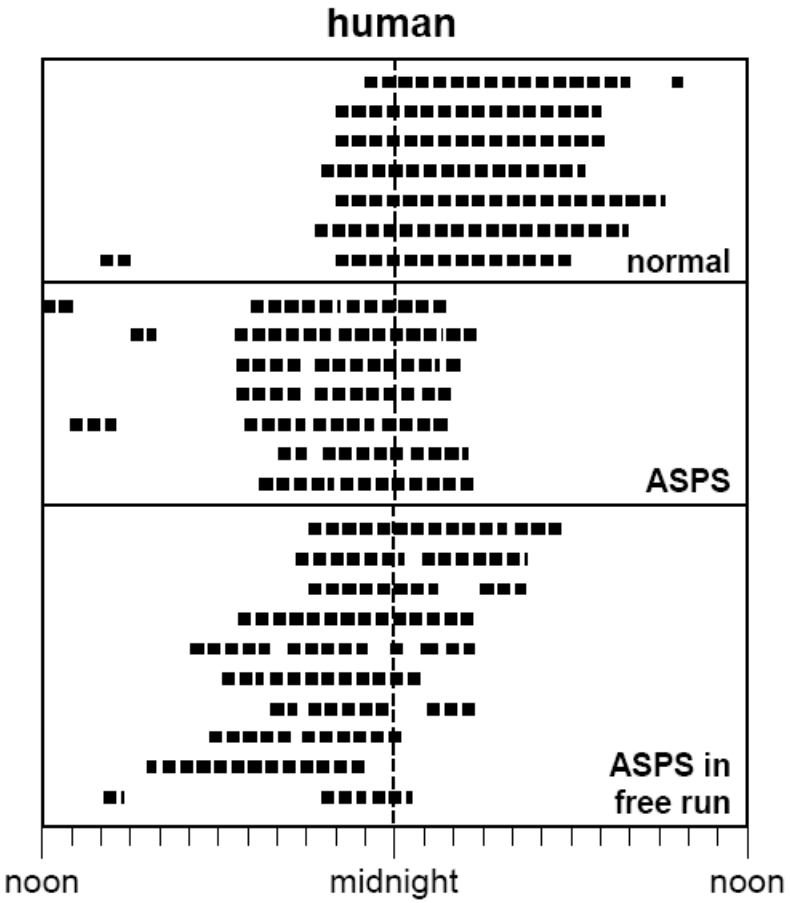


Source: The Body Clock Guide to Better Health, 2000

Circadian Rhythm and Gene Studies

[M. Rosbash, HHMI]

Connections to Sleep Disorder

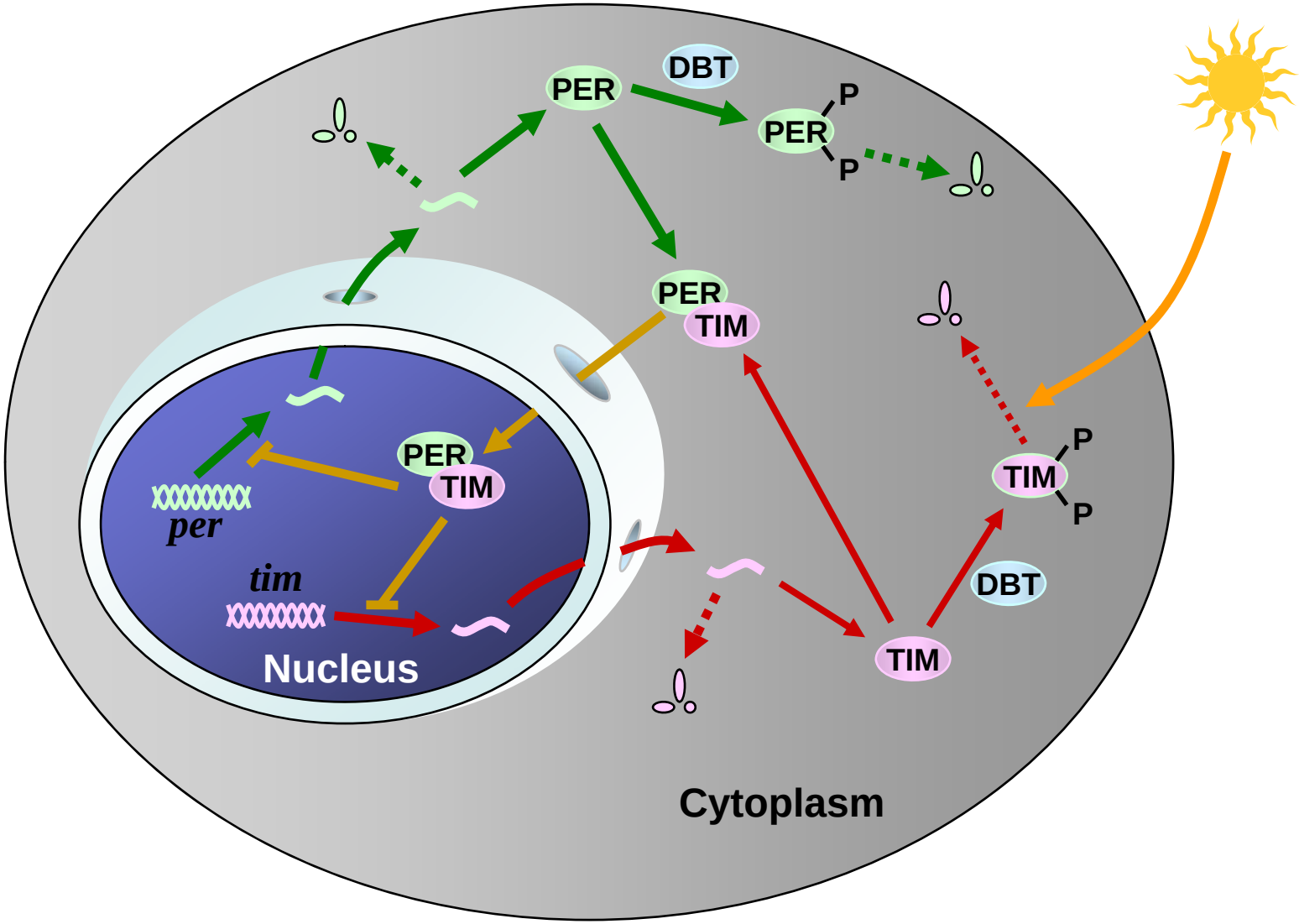


[Wagner-Smith & Kay, *Nat. Genet*., 2000]

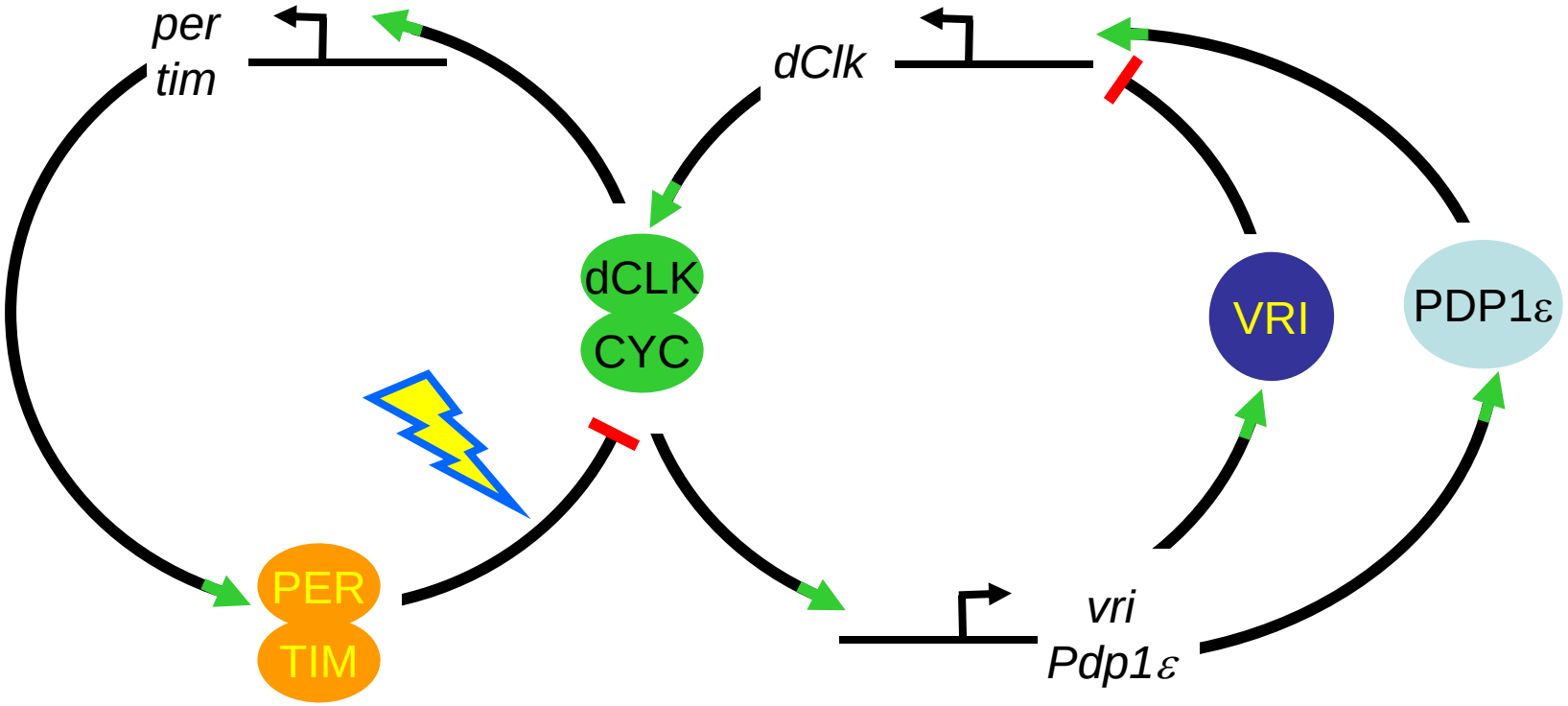
Circadian Gene Regulation

- Cellular circadian rhythmicity arises from a complex transcriptional feedback structure
- Several model systems have generated insight
  - *Drosophila*
  - *Neurospora*
  - Mouse
- Tremendously **robust** regulatory architecture
- Key structural elements
  - Autoregulatory transcriptional/translational negative feedback loop(s)
  - Positive feedback loop(s) between autoregulatory loops (clock, period/time)
  - Protein processing time delays (phosphorylation, dimerization, transport)

Drosophila Circadian Oscillator

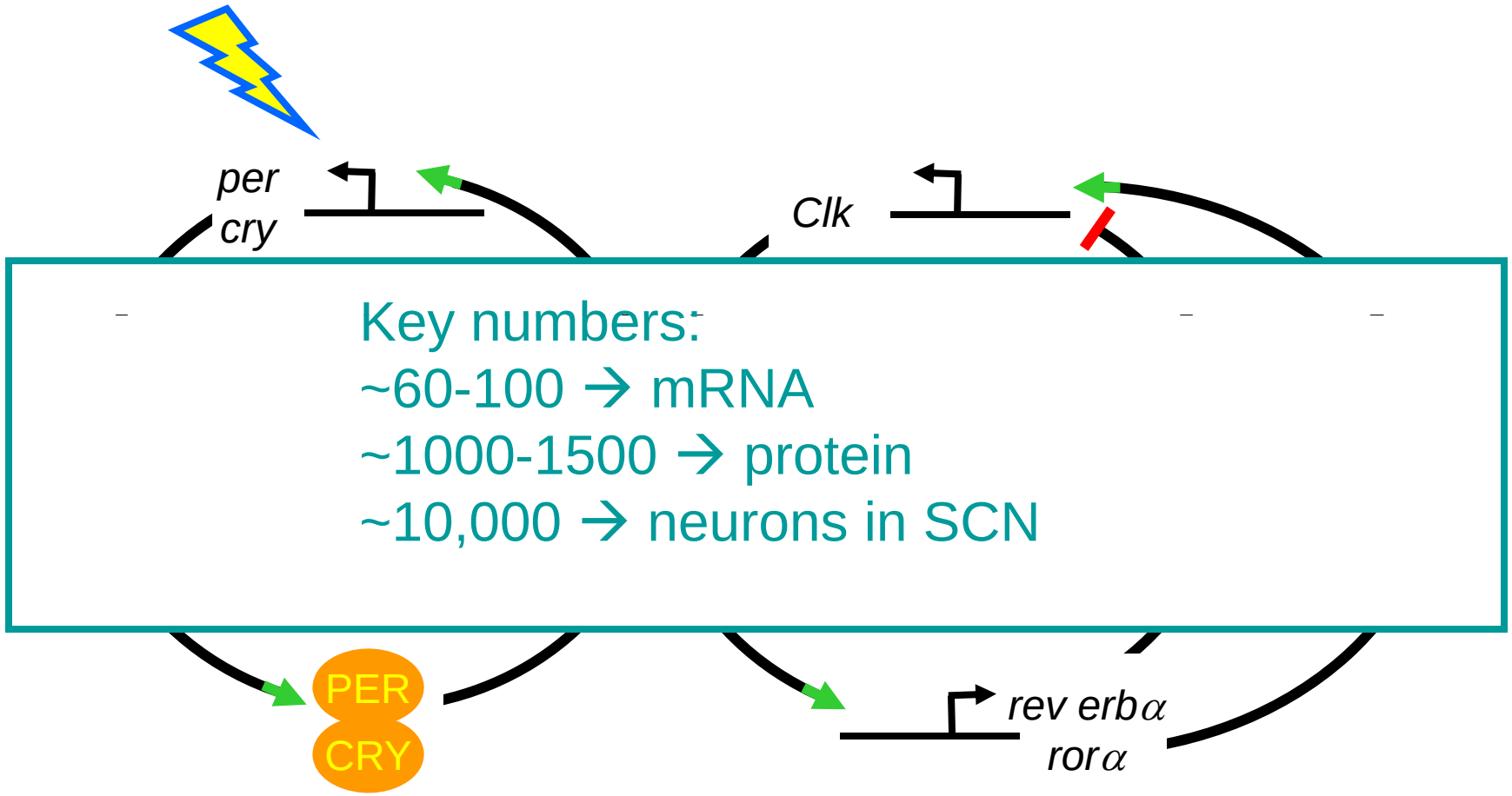


Circadian Rhythm Gene Network
[*Drosophila*]



adapted from [Cryan *et al.*, *Cell*, 2003]

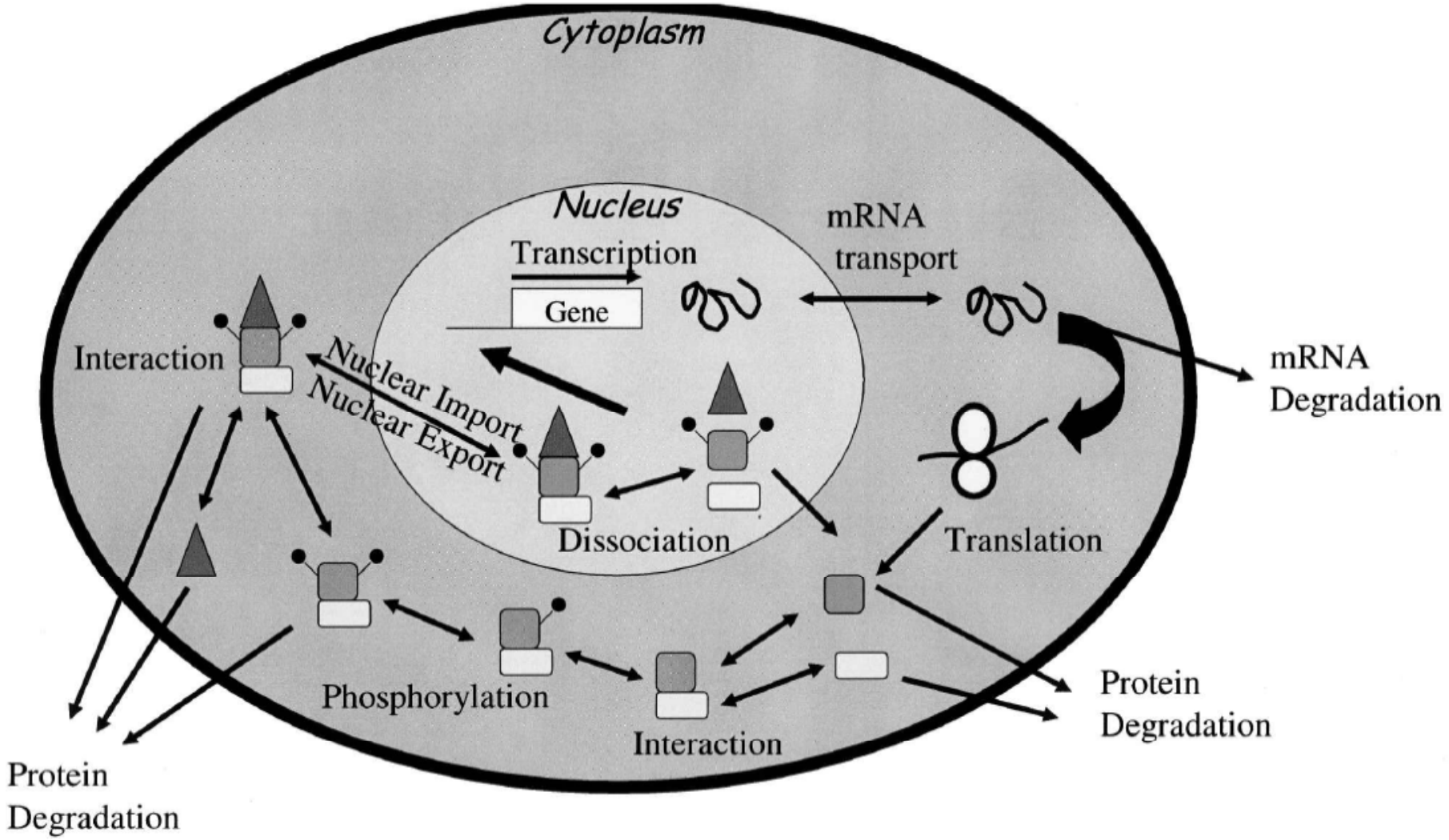
Circadian Rhythm Gene Network
[*Mouse*]



adapted from [Cryan *et al.*, *Cell*, 2003]

Generic Model

[Forger et al., 2003]



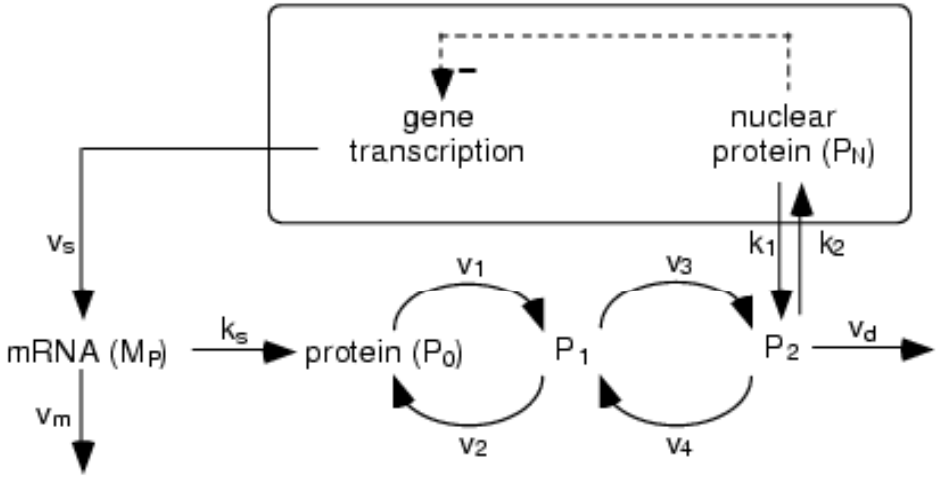
Model Validation Criteria

[Forger et al., 2003]

Criterion	*Expected model performance*
1. Period	Molecular oscillations occur with a free-running period (tau) of approximately 24 h.
2. Phase	(a) Molecular oscillations occur with appropriate phase relationships to each other in free-running conditions (appropriate = consistent with experimental data or plausible).
	(b) Molecular oscillations occur with appropriate phase relationships to each other and to the light-dark cycle.
3. Entrainment	(a) Input repeated at 24-h intervals results in 24-h periodicity of the molecular oscillations.
	(b) The molecular basis for the input to influence molecular oscillations should be based on experimental data.
4. Phase response	(a) Single stimuli lead to alterations in the phase of molecular oscillation.
	(b) The response to a stimulus depends on the phase at which it is administered.
	(c) The molecular basis for the input to influence molecular oscillations should be based on experimental data.
5. Mutations	Mutations affecting the level or activity of circadian-relevant genes *in vivo* should produce similar effects on oscillations *in silico*.

5-State Model

[Goldbeter, 1996; Gonze *et al.*, 2002]



5 states, 18 parameters

$$\frac{dM_P}{dt} = v_s \frac{K_I^n}{K_I^n + P_N^n} - v_m \frac{M_P}{K_m + M_P}$$

$$\frac{dP_0}{dt} = k_s M_P - v_1 \frac{P_0}{K_1 + P_0} + v_2 \frac{P_1}{K_2 + P_1}$$

$$\frac{dP_1}{dt} = v_1 \frac{P_0}{K_1 + P_0} - v_2 \frac{P_1}{K_2 + P_1} - v_3 \frac{P_1}{K_3 + P_1} + v_4 \frac{P_2}{K_4 + P_2}$$

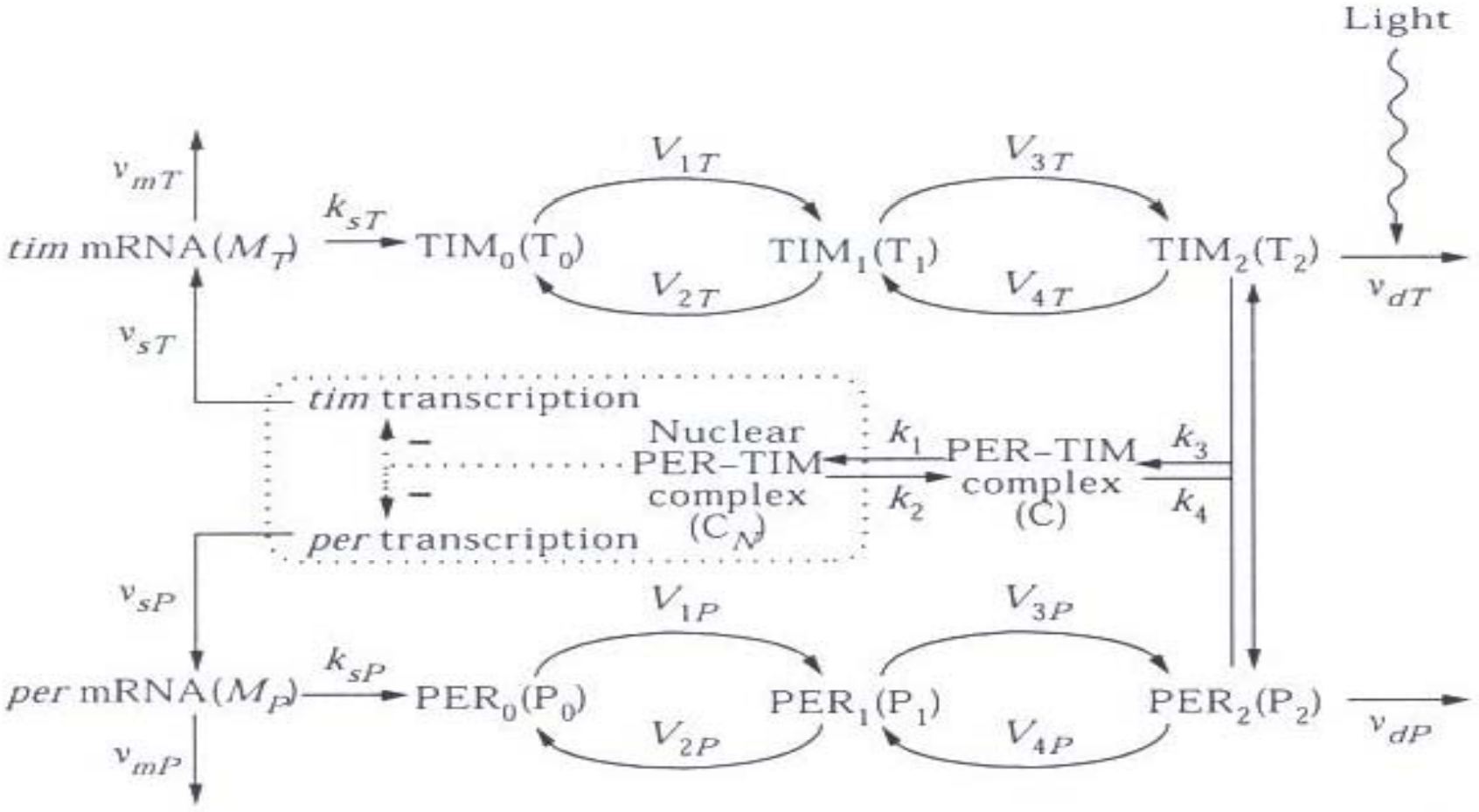
$$\frac{dP_2}{dt} = v_3 \frac{P_1}{K_3 + P_1} - v_4 \frac{P_2}{K_4 + P_2} - v_d \frac{P_2}{K_d + P_2} - k_1 P_2 + k_2 P_N$$

$$\frac{dP_2}{dt} = k_1 P_2 - k_2 P_N$$

10-state Model

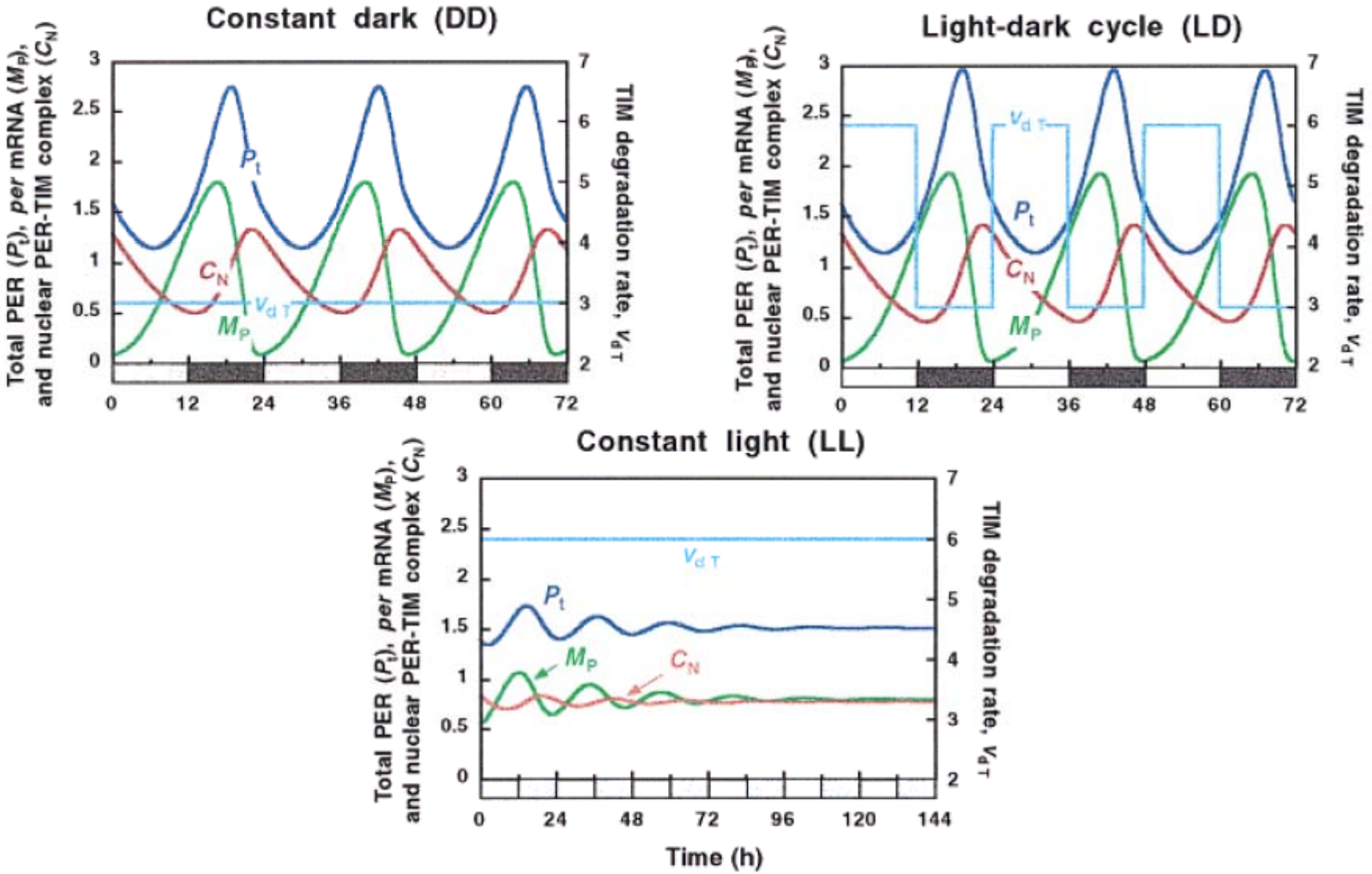
[Gonze *et al*., 2002]

10 states, 38 parameters

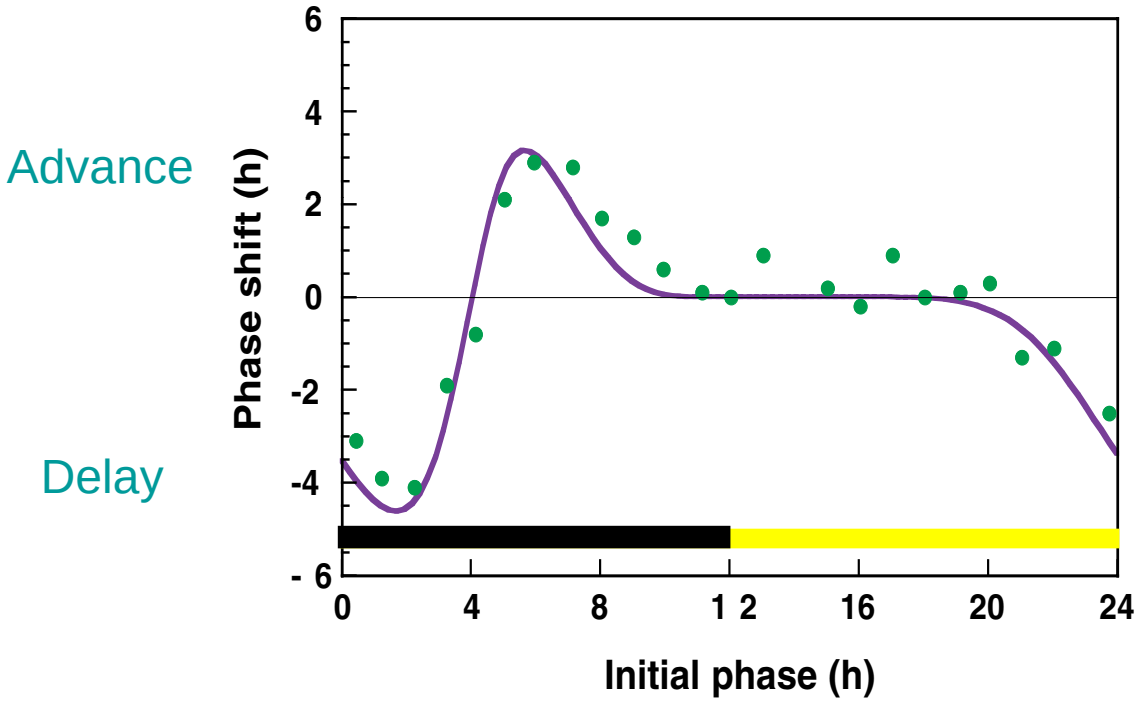


Entrainment Behavior

[Leloup & Goldbeter]



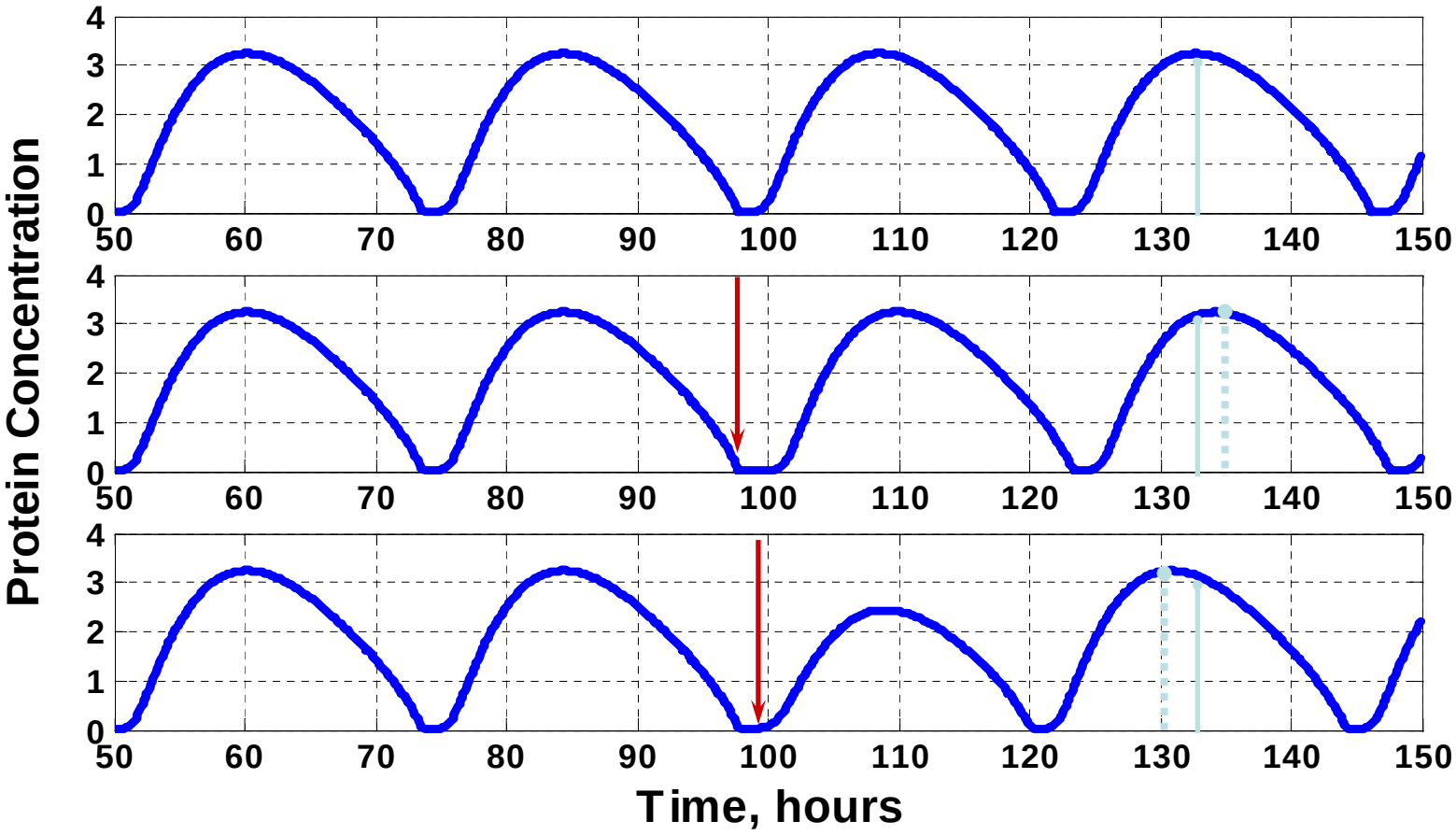
Light as a Zeitgeber



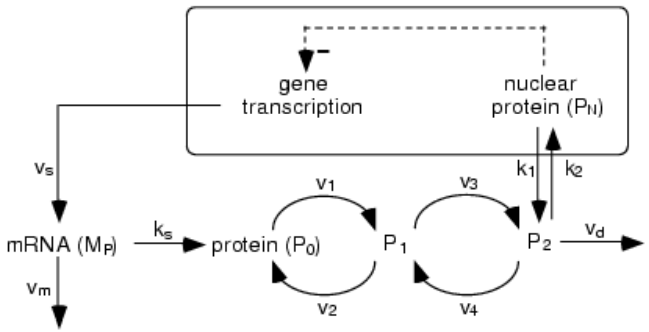
- Experimental data [Hall & Rosbash, 1987]
- Theoretical phase response curve [LeLoup]

Influence of Light Pulses

[Bagheri et al., 2004]



Stochastic Model - Simple



$$\frac{dM_P}{dt} = \nu_s \frac{K_I^n}{K_I^n + P_N^n} - \nu_m \frac{M_P}{K_m + M_P}$$

$$\frac{dP_0}{dt} = k_s M_P - \nu_1 \frac{P_0}{K_1 + P_0} + \nu_2 \frac{P_1}{K_2 + P_1}$$

$$\frac{dP_1}{dt} = \nu_1 \frac{P_0}{K_1 + P_0} - \nu_2 \frac{P_1}{K_2 + P_1} - \nu_3 \frac{P_1}{K_3 + P_1} + \nu_4 \frac{P_2}{K_4 + P_2}$$

$$\frac{dP_2}{dt} = \nu_3 \frac{P_1}{K_3 + P_1} - \nu_4 \frac{P_2}{K_4 + P_2} - \nu_d \frac{P_2}{K_d + P_2} - k_1 P_2 + k_2 P_N$$

$$\frac{dP_2}{dt} = k_1 P_2 - k_2 P_N$$

1	G$\longrightarrow$MP+G	$w_1 = (\upsilon_s \Omega) \frac{(K_I \Omega)^n}{(K_I \Omega)^n + P_N^n}$	$M_P \longrightarrow M_P + 1$
2	MP$\longrightarrow$	$w_3 = (\upsilon_m \Omega) \frac{M_P}{(K_m \Omega) + M_P}$	$M_P \longrightarrow M_P - 1$
3	MP$\longrightarrow$P0+MP	$w_2 = k_s M_P$	$P_0 \longrightarrow P_0 + 1$
4	P0$\longrightarrow$P1	$w_4 = (\upsilon_1 \Omega) \frac{P_0}{(K_1 \Omega) + P_0}$	$P_0 \longrightarrow P_0 - 1$ $P_1 \longrightarrow P_1 + 1$
5	P1$\longrightarrow$P0	$w_5 = (\upsilon_2 \Omega) \frac{P_1}{(K_2 \Omega) + P_1}$	$P_0 \longrightarrow P_0 + 1$ $P_1 \longrightarrow P_1 - 1$
6	P1$\longrightarrow$P2	$w_6 = (\upsilon_3 \Omega) \frac{P_1}{(K_3 \Omega) + P_1}$	$P_1 \longrightarrow P_1 - 1$ $P_2 \longrightarrow P_2 + 1$
7	P2$\longrightarrow$P1	$w_7 = (\upsilon_4 \Omega) \frac{P_2}{(K_4 \Omega) + P_2}$	$P_1 \longrightarrow P_1 + 1$ $P_2 \longrightarrow P_2 - 1$
8	P2$\longrightarrow$	$w_8 = (\upsilon_d \Omega) \frac{P_2}{(K_d \Omega) + P_2}$	$P_2 \longrightarrow P_2 - 1$
9	P2$\longrightarrow$PN	$w_9 = k_1 P_2$	$P_2 \longrightarrow P_2 - 1$ $P_N \longrightarrow P_N + 1$
10	PN$\longrightarrow$P2	$w_{10} = k_2 P_N$	$P_2 \longrightarrow P_2 + 1$ $P_N \longrightarrow P_N - 1$

Stochastic Model - Detailed

Reaction number	Reaction step	Probability of reaction step
1a	$G + P_N \xrightarrow{a_1} GP_N$	$w_1 = a_1 \times G \times P_N/\Omega$
1b	$GP_N \xrightarrow{d_1} G + P_N$	$w_2 = d_1 \times GP_N$
1c	$GP_N + P_N \xrightarrow{a_2} GP_{N2}$	$w_3 = a_2 \times GP_N \times P_N/\Omega$
1d	$GP_{N2} \xrightarrow{d_2} GP_N + P_N$	$w_4 = d_2 \times GP_{N2}$
1e	$GP_{N2} + P_N \xrightarrow{a_3} GP_{N3}$	$w_5 = a_3 \times GP_{N2} \times P_N/\Omega$
1f	$GP_{N3} \xrightarrow{d_3} GP_{N2} + P_N$	$w_6 = d_3 \times GP_{N3}$
1g	$GP_{N3} + P_N \xrightarrow{a_4} GP_{N4}$	$w_7 = a_4 \times GP_{N3} \times P_N/\Omega$
1h	$GP_{N4} \xrightarrow{d_4} GP_{N3} + P_N$	$w_8 = d_4 \times GP_{N4}$
1i	$[G, GP_N, GP_{N2}, GP_{N3}] \xrightarrow{v_s} M_P$	$w_9 = v_s \times (G + GP_N + GP_{N2} + GP_{N3})$
2a	$M_P + E_m \xrightarrow{k_{m1}} C_m$	$w_{10} = k_{m1} \times M_P \times E_m/\Omega$
2b	$C_m \xrightarrow{k_{m2}} M_P + E_m$	$w_{11} = k_{m2} \times C_m$
2c	$C_m \xrightarrow{k_{m3}} E_m$	$w_{12} = k_{m3} \times C_m$
3	$M_P \xrightarrow{k_s} M_P + P_0$	$w_{13} = k_s \times M_P$
4a	$P_0 + E_1 \xrightarrow{k_{11}} C_1$	$w_{14} = k_{11} \times P_0 \times E_1/\Omega$
4b	$C_1 \xrightarrow{k_{12}} P_0 + E_1$	$w_{15} = k_{12} \times C_1$
4c	$C_1 \xrightarrow{k_{13}} P_1 + E_1$	$w_{16} = k_{13} \times C_1$
5a	$P_1 + E_2 \xrightarrow{k_{21}} C_2$	$w_{17} = k_{21} \times P_1 \times E_2/\Omega$
5b	$C_2 \xrightarrow{k_{22}} P_1 + E_2$	$w_{18} = k_{22} \times C_2$
5c	$C_2 \xrightarrow{k_{23}} P_0 + E_2$	$w_{19} = k_{23} \times C_2$
6a	$P_1 + E_3 \xrightarrow{k_{31}} C_3$	$w_{20} = k_{31} \times P_1 \times E_3/\Omega$
6b	$C_3 \xrightarrow{k_{32}} P_1 + E_3$	$w_{21} = k_{32} \times C_3$
6c	$C_3 \xrightarrow{k_{33}} P_2 + E_3$	$w_{22} = k_{33} \times C_3$
7a	$P_2 + E_4 \xrightarrow{k_{41}} C_4$	$w_{23} = k_{41} \times P_2 \times E_4/\Omega$
7b	$C_4 \xrightarrow{k_{42}} P_2 + E_4$	$w_{24} = k_{42} \times C_4$
7c	$C_4 \xrightarrow{k_{43}} P_1 + E_4$	$w_{25} = k_{43} \times C_4$
8a	$P_2 + E_d \xrightarrow{k_{d1}} C_d$	$w_{26} = k_{d1} \times P_2 \times E_d/\Omega$
8b	$C_d \xrightarrow{k_{d2}} P_2 + E_d$	$w_{27} = k_{d2} \times C_d$
8c	$C_d \xrightarrow{k_{d3}} E_d$	$w_{28} = k_{d3} \times C_d$
9	$P_2 \xrightarrow{k_1} P_N$	$w_{29} = k_1 \times P_2$
10	$P_N \xrightarrow{k_2} P_2$	$w_{30} = k_2 \times P_N$

Enzymatic Degradation of mRNA

$$\frac{dM_P}{dt} = v_s \frac{K_I^n}{K_I^n + P_N^n} - v_m \frac{M_P}{K_m + M_P}$$

$v_m = 0.3$ nMh^{-1}
$K_m = 0.2$ nM

$$\text{Mp} \longrightarrow \qquad w_3 = (v_m \Omega) \frac{M_P}{(K_m \Omega) + M_P} \qquad M_P \longrightarrow M_P - 1$$

$v_m = 0.3$ mol h^{-1}
$K_m = 0.2$ mol

$$\text{Mp} + \text{Em} \xrightarrow{k_{m1}} \text{Cm} \qquad w_{10} = k_{m1} \times M_P \times E_m / \Omega$$

$$\text{Cm} \xrightarrow{k_{m2}} \text{Mp} + \text{Em} \qquad w_{11} = k_{m2} \times C_m$$

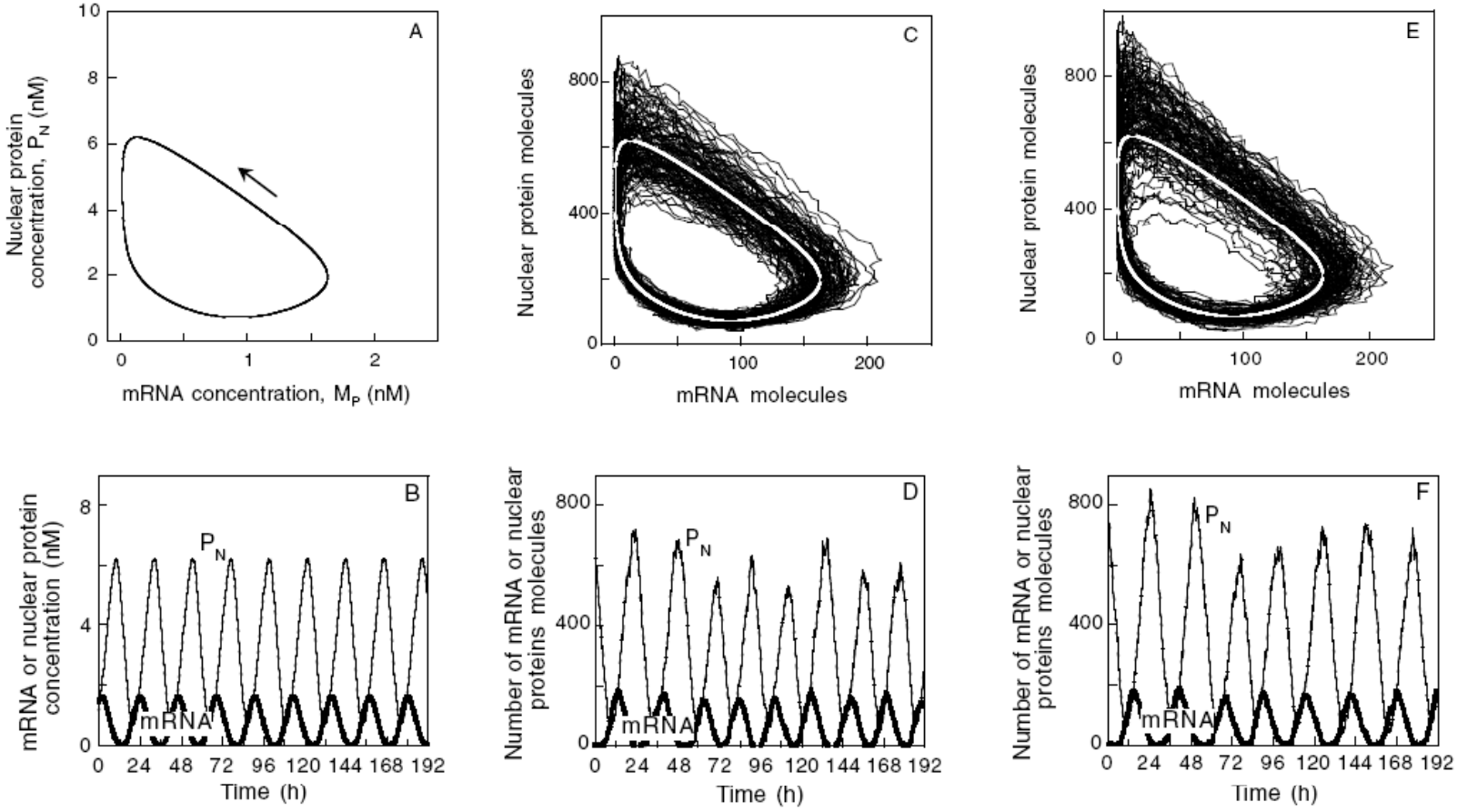
$$\text{Cm} \xrightarrow{k_{m3}} \text{Em} \qquad w_{12} = k_{m3} \times C_m$$

$k_{m1} = 165$ mol^{-1} h^{-1},
$k_{m2} = 30$ h^{-1}, $k_{m3} = 3$ h^{-1},
$E_{mtot} = E_m + C_m = (0.1 \times \Omega)$ mol

Stochastic Behavior

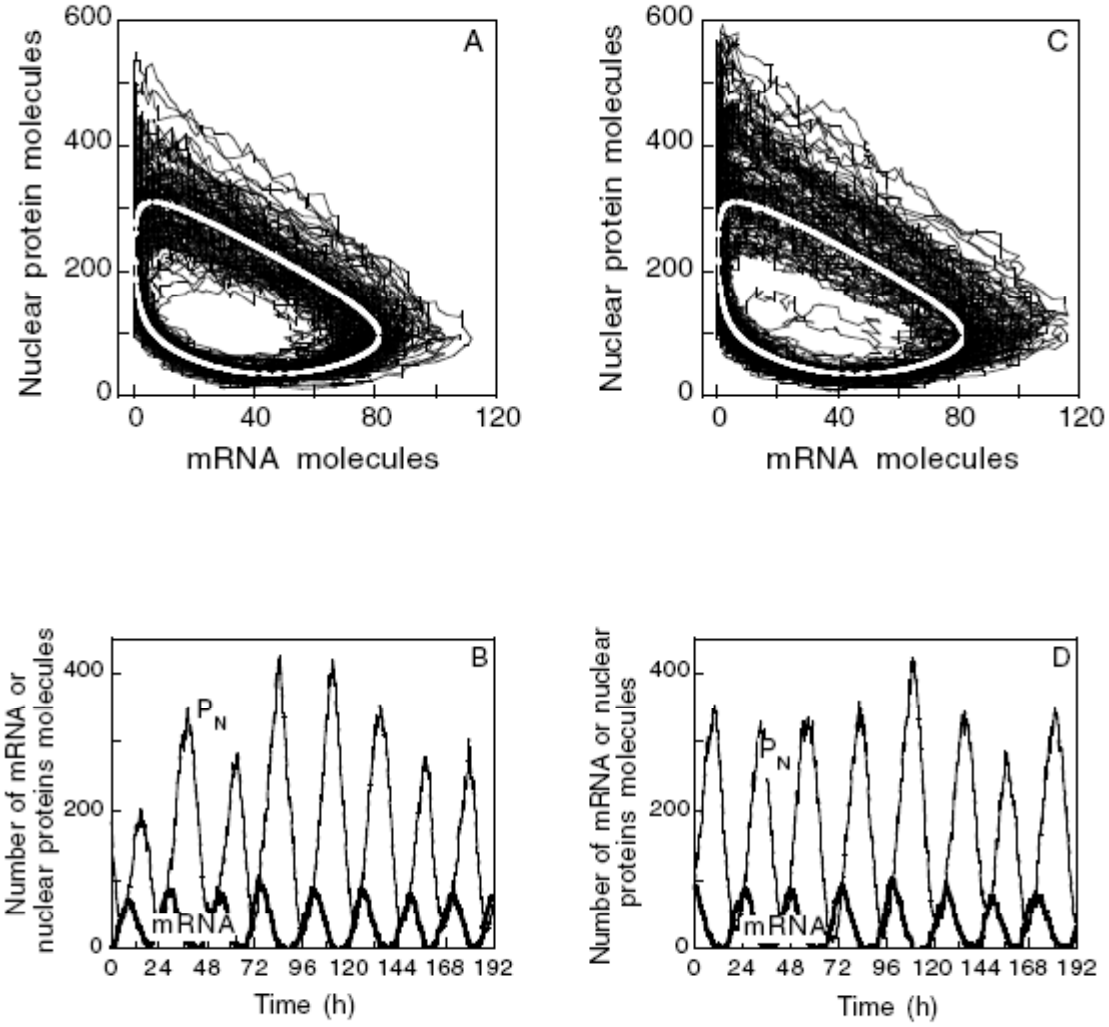
[Ω=100]

Deterministic **Simple Stoch** **Complex Stoch**



Stochastic Behavior

[Ω=50]

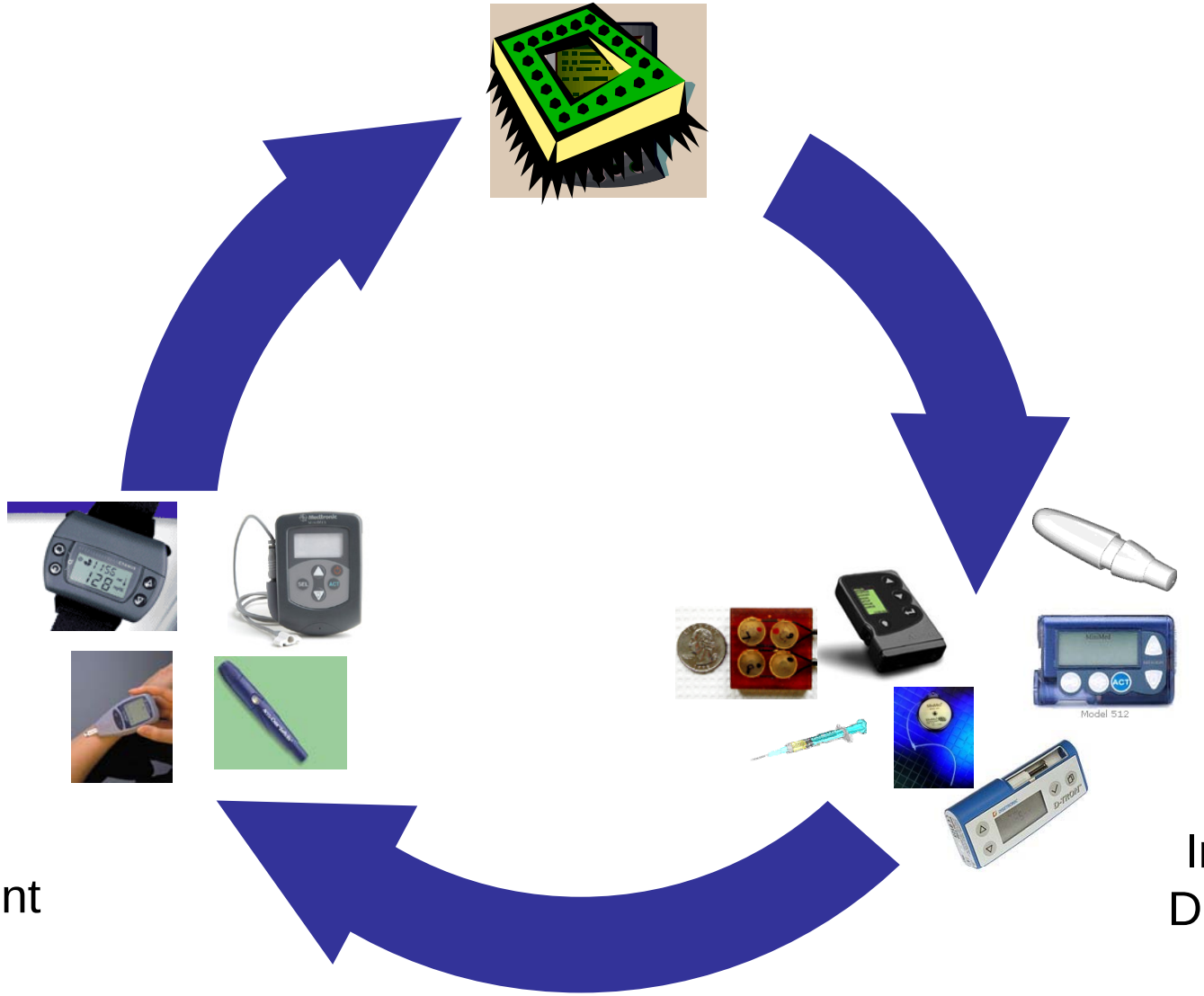


Implications from Systems Biology Studies

- Robustness characteristics of feedback architecture under stochastic uncertainty
- Underlying design principles
- Nature of entrainment, and systems characterization
- Possible therapeutic ramifications (mutants, etc.)
- General biological oscillator insights

Biomedical Control

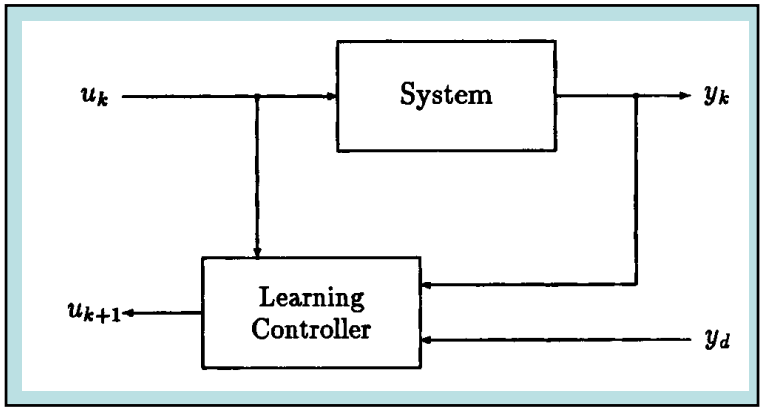
The Glucose – Insulin “Loop”



Episodic Measurements: Run-to-Run Control

Run-to-Run Control Preliminaries

- Iterative Learning Control (ILC)
  - arose from repetitively operated systems
    - antenna servomechanism
  - also useful for switching between inputs
    - robot actions



Learning Control Scheme

- Basic algorithm:

$$e_k = y_d(t) - y_k(t)$$
$$u_{k+1} = u_k + \Gamma \dot{e}_k$$

- For LTI plant ([A,B,C,D]), convergence:

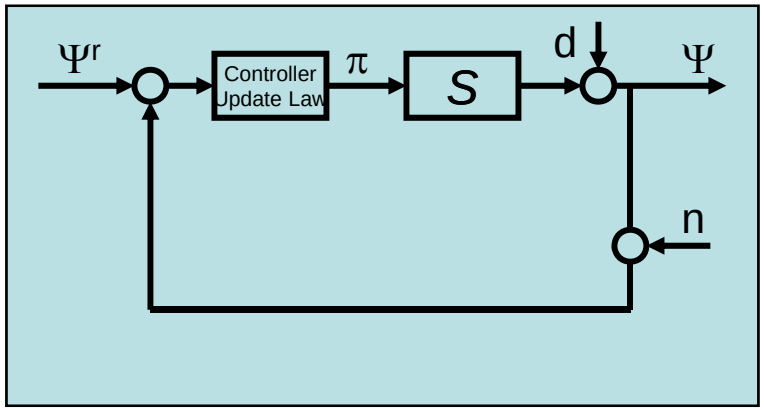
$$\left\| I - CB\Gamma \right\|_i < 1$$
$$\lim_{k \to \infty} y_k(t) \to y_d(t)$$

- Algorithm generalization:

$$u_{k+1} = u_k + \Phi e_k + \Gamma \dot{e}_k + \Psi \int e_k dt$$

Optimization-based r2r Preliminaries

- Optimization-based r2r
  - model-based framework
    - gradient-based update between iterations
  - model-free framework
    - terminal constraint handling



Optimization Problems

- Nominal Problem:

$$\min_{u(t)} \quad J = \varphi(x(t_f))$$
$$\text{s.t.} \quad \dot{x} = f(x,u); \quad x(0) = x_0$$
$$S(x,u) \le 0, \quad \mathrm{T}(x(t_f)) \le 0$$

- Robust Optimization:

$$\min_{u(t)} \quad J = \varphi(x(t_f))$$
$$\text{s.t.} \quad \dot{x} = F(x,\theta,u) + d; \quad x(0) = x_0$$
$$S(x,u) \le 0, \quad \mathrm{T}(x(t_f)) \le 0$$

Optimization Problems (cont'd)

- Measurement-based Optimization (MBO):

$$\min_{u^k(t)} \quad J^k = \varphi(x^k(t_f))$$

$$\text{s.t.} \quad \dot{x}^k = f(x^k,\theta) + g(x^k,\theta)u^k + d^k$$

$$x^k(0) = x_0$$

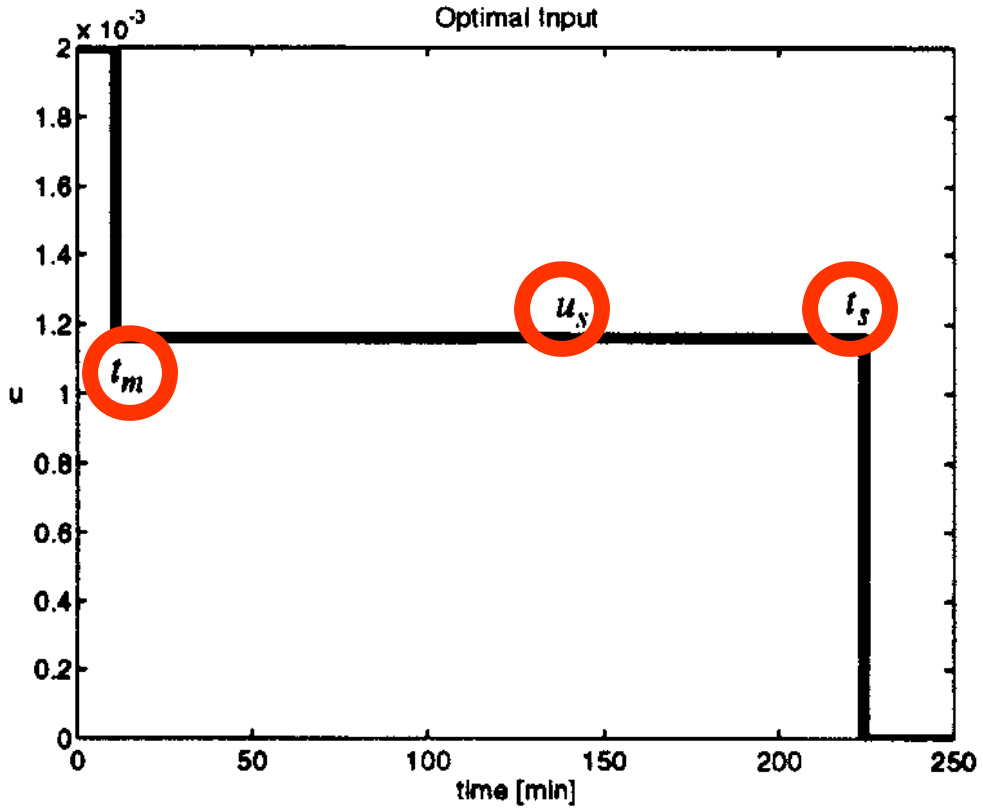
$$y^k = h(x^k,\theta) + v^k$$

$$S(x^k,u^k) \le 0$$

$$\mathrm{T}(x^k(t_f)) \le 0$$

$$\text{given} \quad y^j(i) \quad \forall i, \forall j \le (k-1)$$

Parameterization of Input Vector



Classification of Measurement-based Optimization Approaches

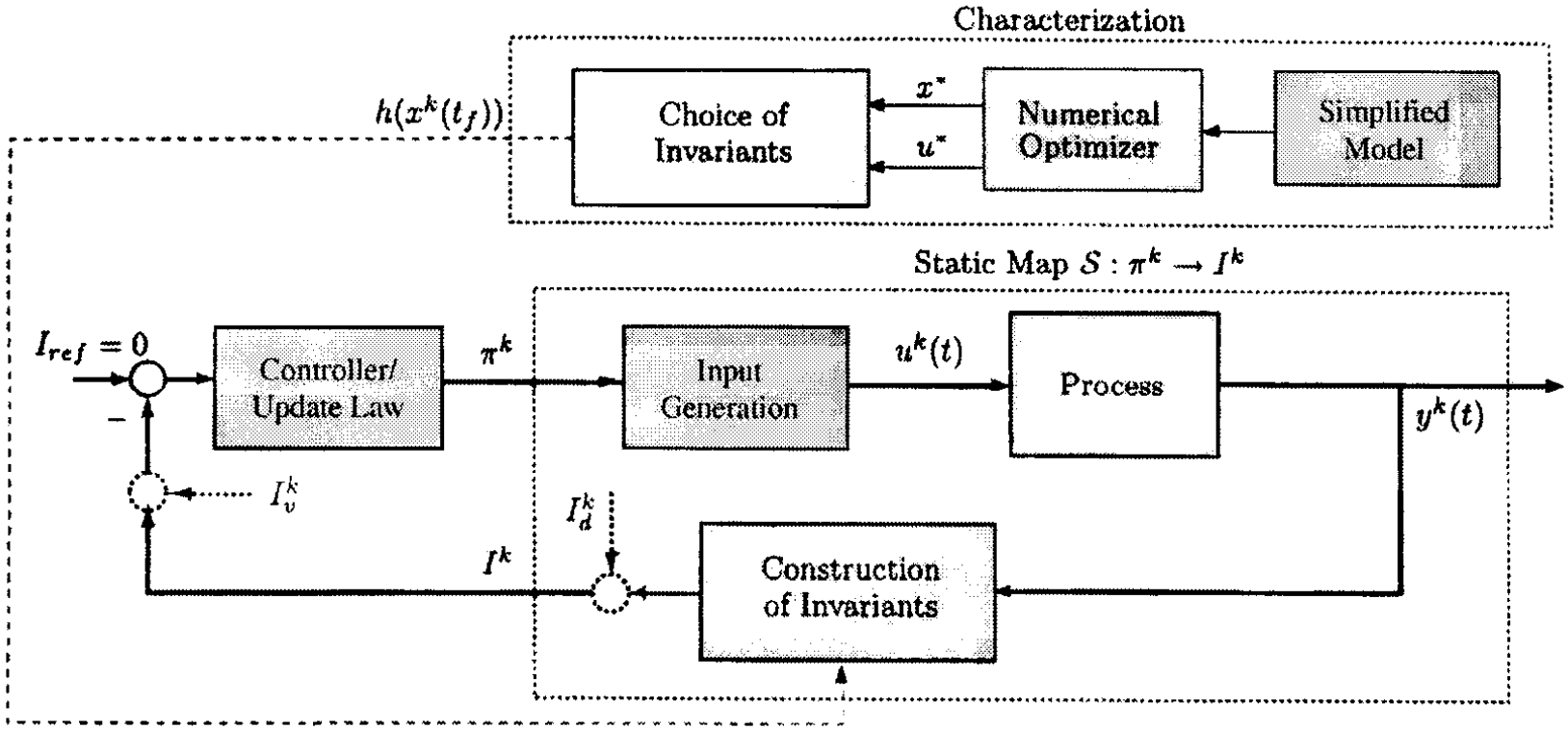
- Fixed model – MBO (repeated)
  - *accuracy of model*
- Refined mode – MBO (repeated)
  - *persistency of excitation*
- Evolutionary Approaches
  - *curse of dimensionality*
- Reference Tracking
  - *what to track for optimality*

Application Summary

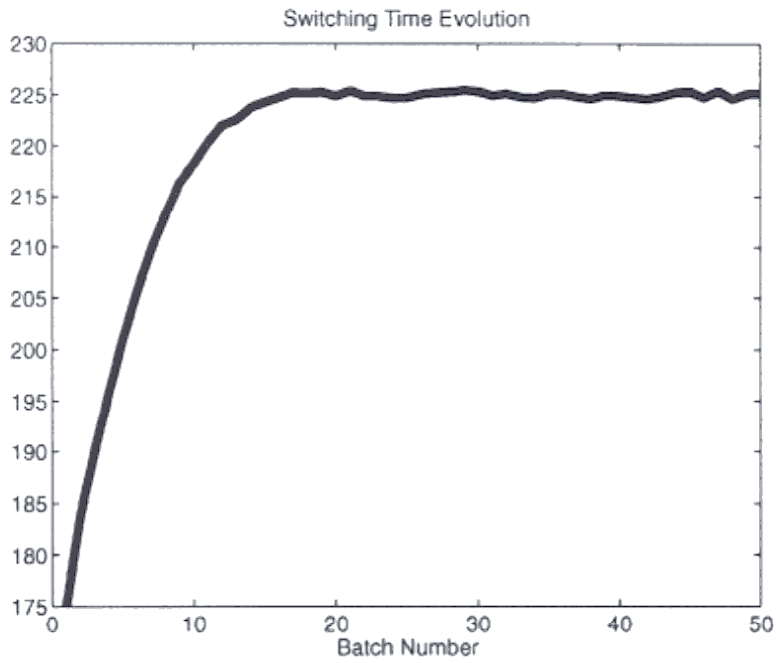
Simulated Chemical Reactor

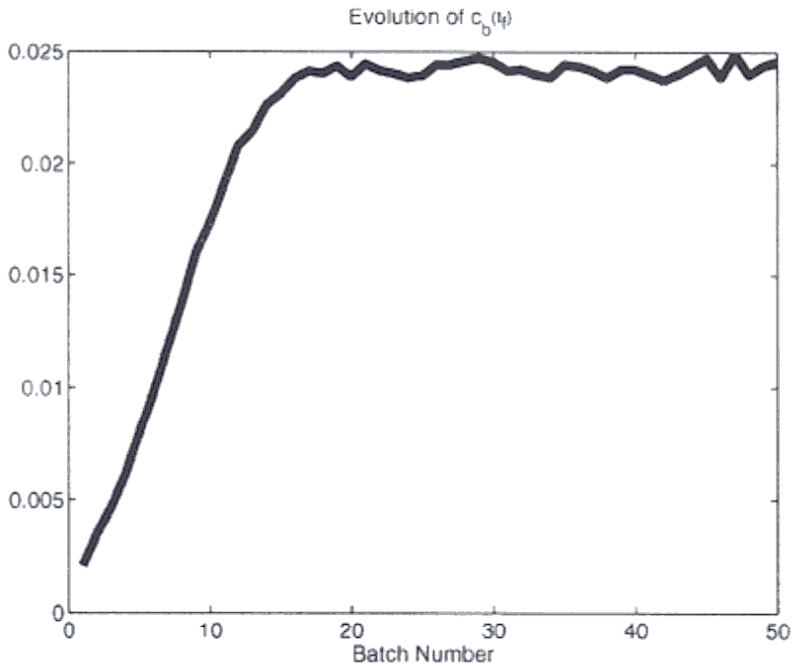
- (Srinivasan et al., 2001)
- Optimize productivity of complex reaction
- Approach:
  - Feed rate is actuator, parameterized by 3-D
  - No model – simple gain matrix
  - Optimization-based approach
  - Complex constraints
- Effective convergence in ~10-20 batches

Algorithm Architecture



Results


Switching Time Evolution
230
225
220
215
210
205
200
195
190
185
180
175
0
10
20
30
40
50
Batch Number


Evolution of $c_b(t_f)$
0.025
0.02
0.015
0.01
0.005
0
0
10
20
30
40
50
Batch Number

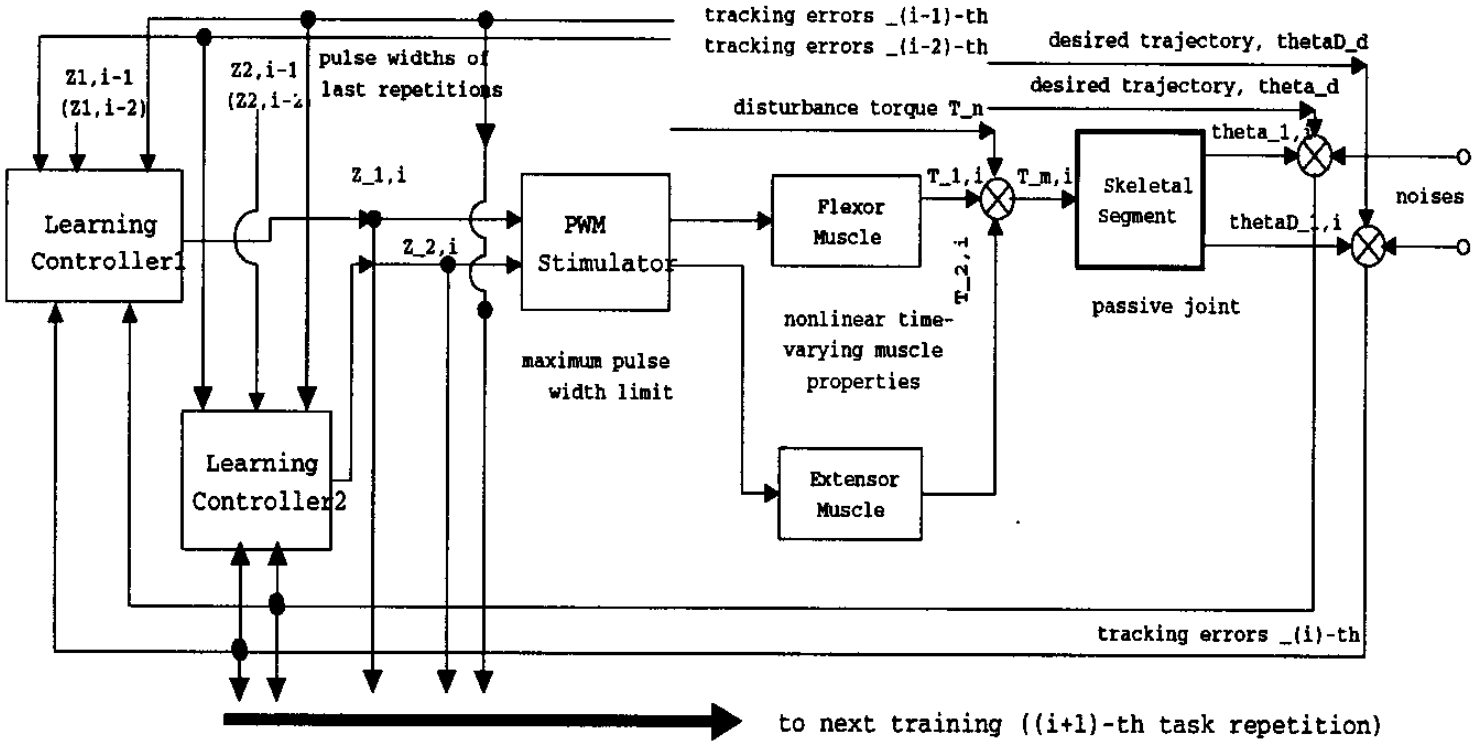
Experimental Chemical Reactor

- Lee & Lee in (Bien & Xu, 1998)
- Reactor temperature recipe (5000 sec)
  - (charge,heat-up,reaction,cooling,discharge)
- Approach:
  - Jacket temperature profile is actuator
  - Inaccurate ARX model employed
  - Combined iterative and feedback algorithm
- Effective convergence in ~6 batches

Functional Neuromuscular Stimulation (FNS)

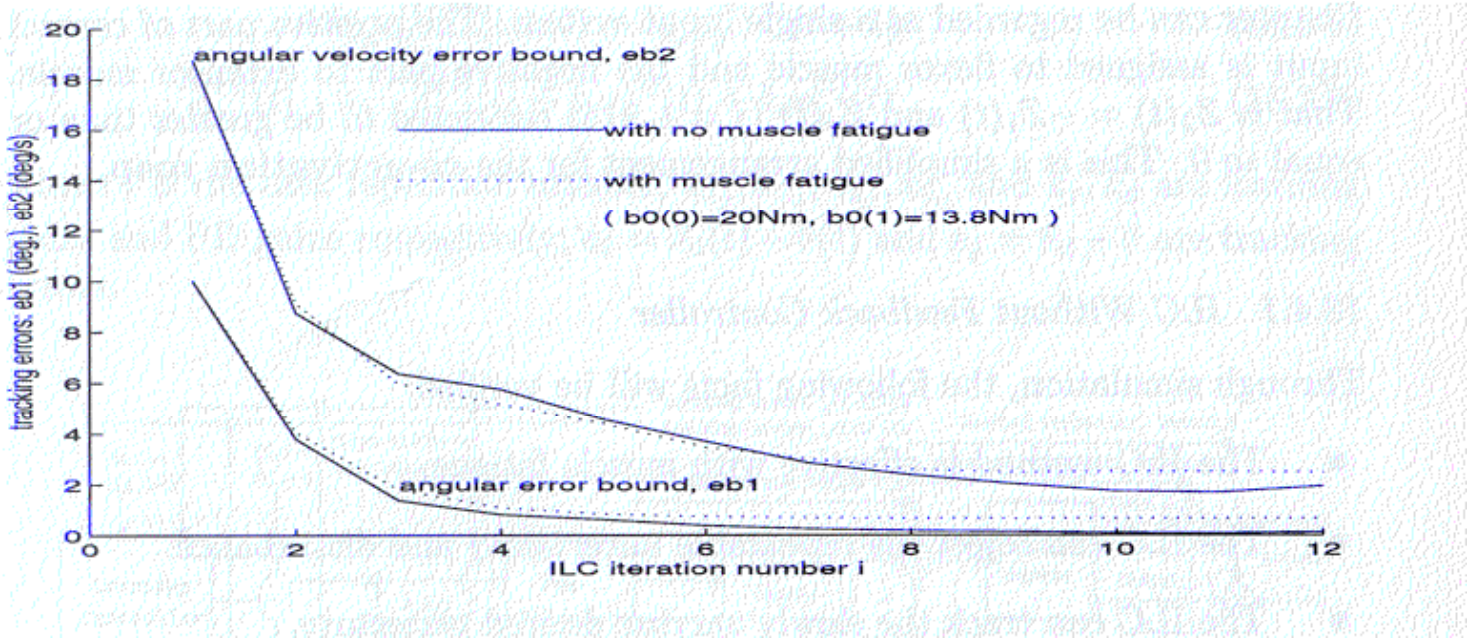
- Dou et al. in (Bien & Xu, 1998)
- FNS of limb no longer under voluntary control
- Challenges
  - customization to individual patients
  - adaptation from time-varying musculoskeletal system
  - robustness against exogenous disturbances
- Approach:
  - Simplified fundamental model employed
  - Track joint angle using pulse width (flexor, extensor)

Algorithm Architecture



Results

- ILC effective with muscle fatigue
- ILC rejects repeated uncertainty and disturbance
- ILC tracks slowly varying desired trajectory



Parameter Identification (ILI)

- (Chen & Wen, 1999)
- Aerodynamic drag coefficient modeling
- Approach:
  - ILC: given a reference trajectory and repeated data, refine the input profile
  - ILI: given I/O data (reference) and repeated data, refine uncertain coefficient
- Effective convergence in ~10-30 cycles

Robot Repeated Task

- (Moore, 1993)
- Two joint manipulator
- Approach:
  - simplified (nonlinear) fundamental model
  - torque is the actuator
  - adaptive gain adjustment
  - P-ILC algorithm
- Effective convergence in ~8 cycles

Insulin Injection Optimization

- (Doyle III et al., *IEEE EMBS Conf*., 2001)
- Timing/amount of (repeated) insulin injections
- Approach:
  - Gain model (implicit)
  - Gmax and Gmin are objectives
  - Fixed (decoupled) PI control structure
  - MBO framework
- Effective convergence in ~6-10 cycles

Observations – Current Patient Protocol

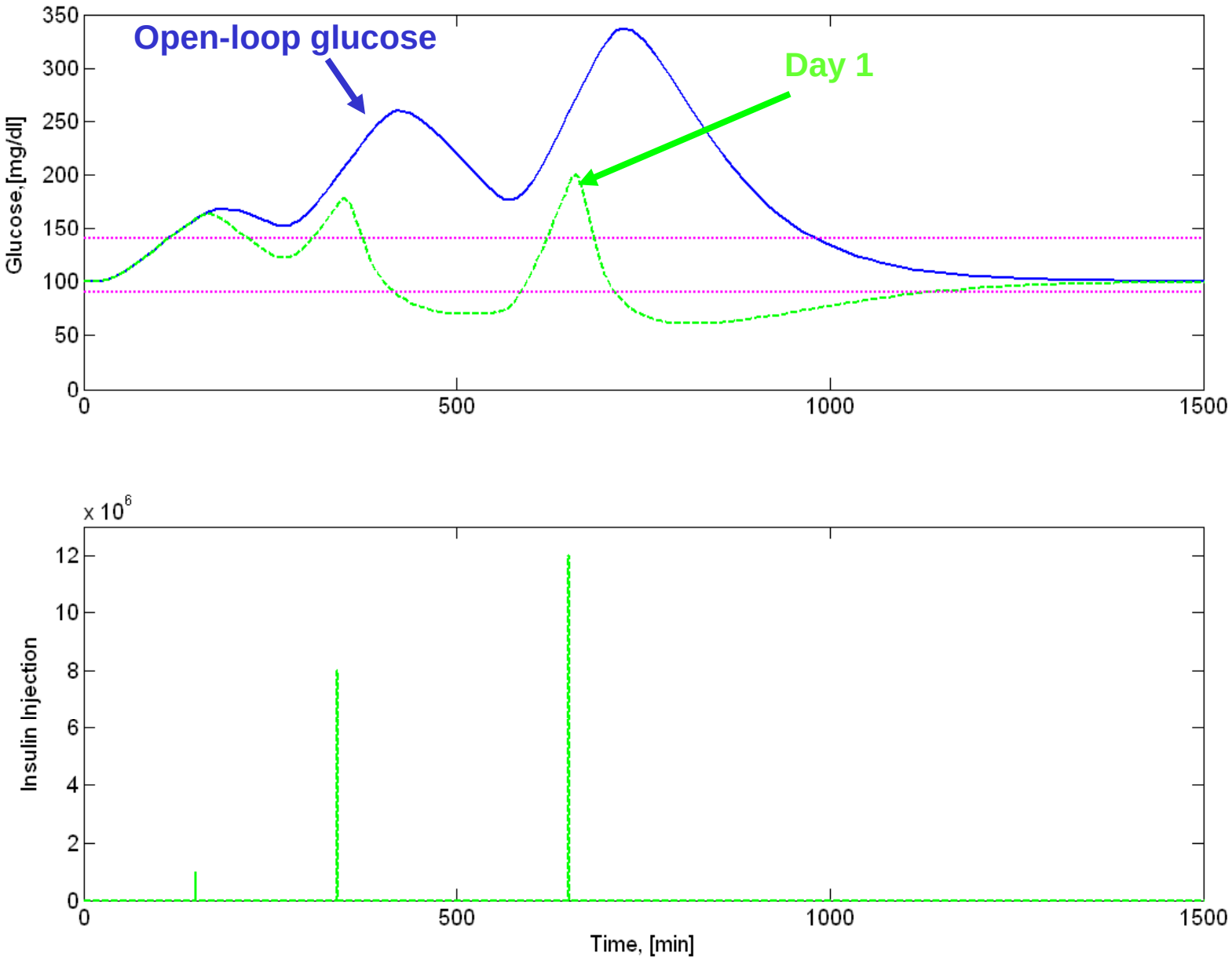
- Availability of periodic glucose measurement

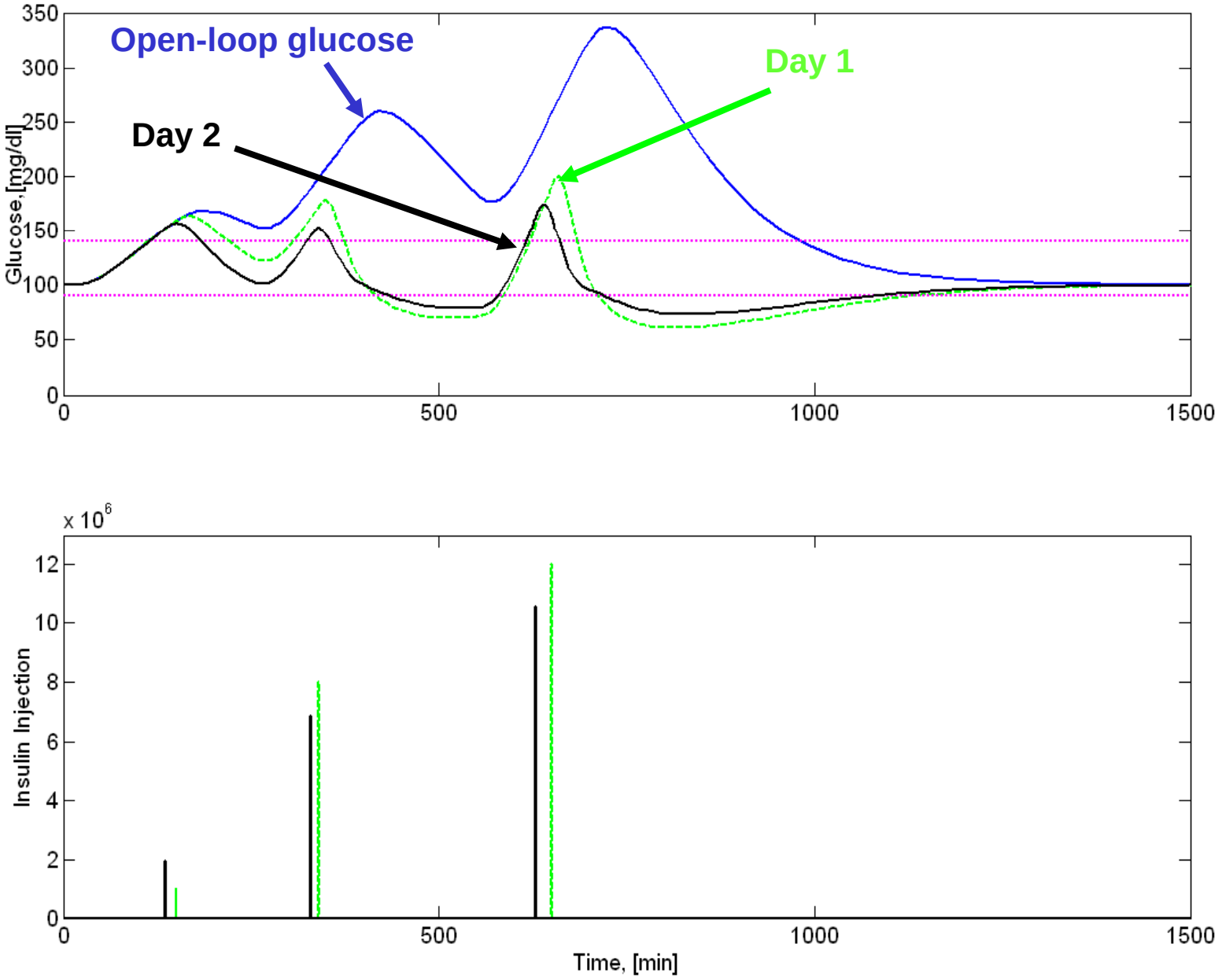
- Issues with accurate models for individual patients
- “Batch-like” = single meal or 24 hour cycle
- Few key variables
  - input: timing and size of insulin bolus
  - performance: maximum and minimum glucose values

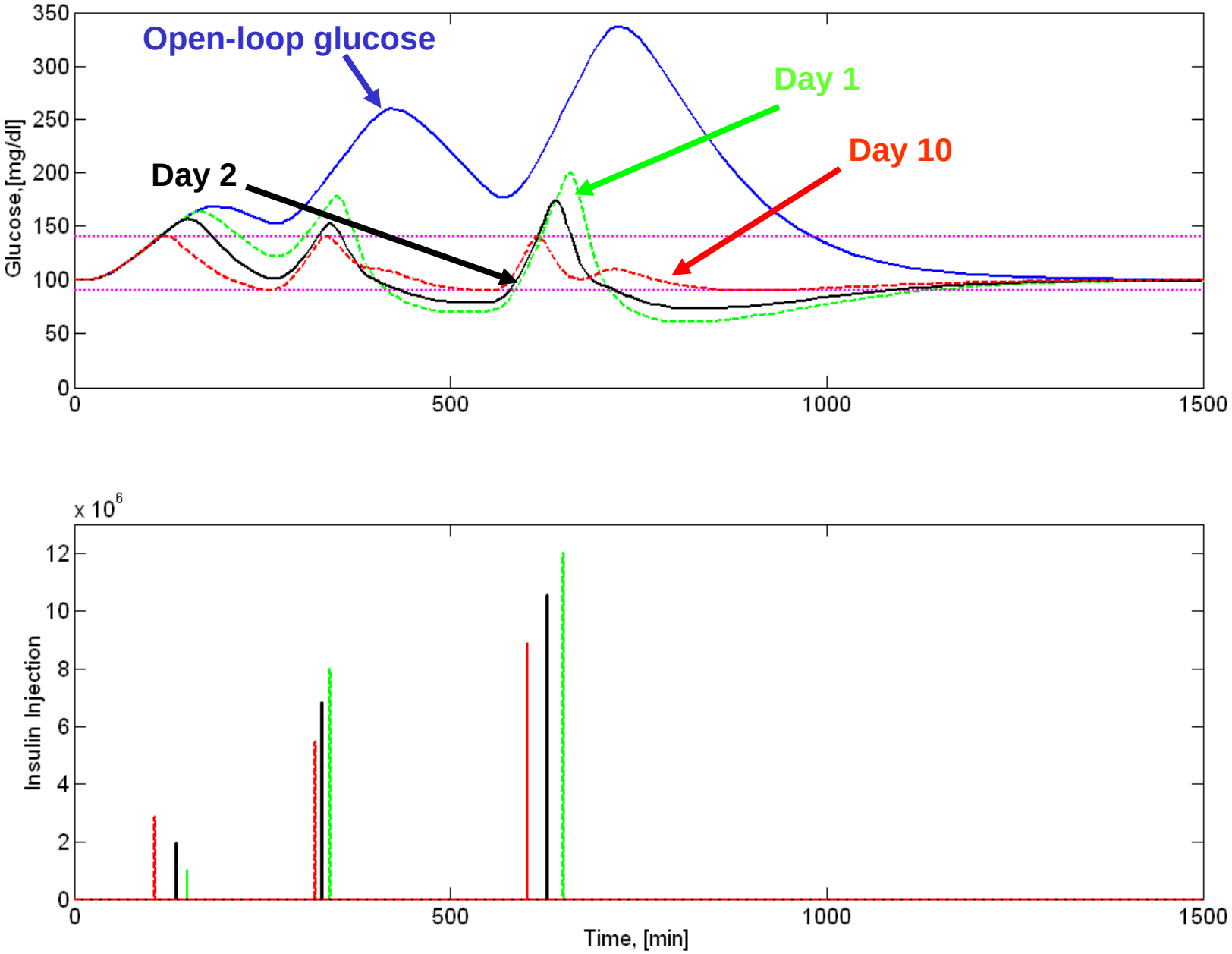
Run-to-Run Algorithm

$$T(k+1) = T(k) + K_T \min(0, G^r_{\max} - G_{\max}(k))$$
$$Q(k+1) = Q(k) + K_Q \max(0, G^r_{\min} - G_{\min}(k))$$

- Initial guesses T(1), Q(1)
- Reference values for G_{max} and G_{min}
- Can impose **hard** bounds on G_{max} and G_{min}
- Gains K_T and K_Q reflect compromise between speed and accuracy
- Straightforward generalization to 3-meal (24 hr) cycle


Open-loop glucose
Day 1
Glucose,[mg/dl]
350
300
250
200
150
100
50
0
0
500
1000
1500
x 10^6
Insulin Injection
12
10
8
6
4
2
0
0
500
1000
1500
Time, [min]

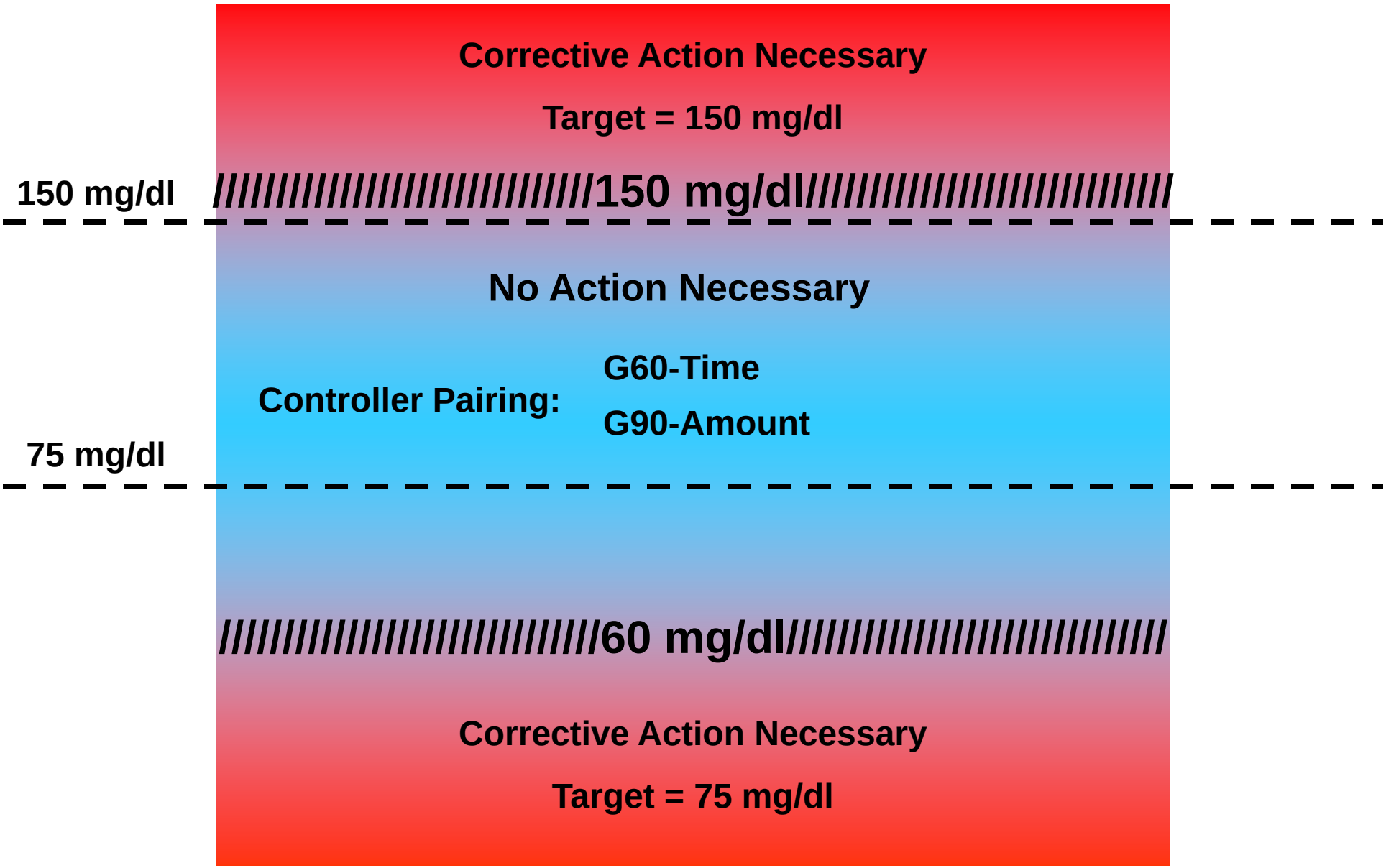

Open-loop glucose
Day 1
Day 2
Glucose, [mg/dl]
350
300
250
200
150
100
50
0
0
500
1000
1500
x 10^6
Insulin Injection
12
10
8
6
4
2
0
0
500
1000
1500
Time, [min]


Open-loop glucose
Day 1
Day 10
Day 2
Glucose,[mg/dl]
350
300
250
200
150
100
50
0
0
500
1000
1500
x 10^6
Insulin Injection
12
10
8
6
4
2
0
0
500
1000
1500
Time, [min]

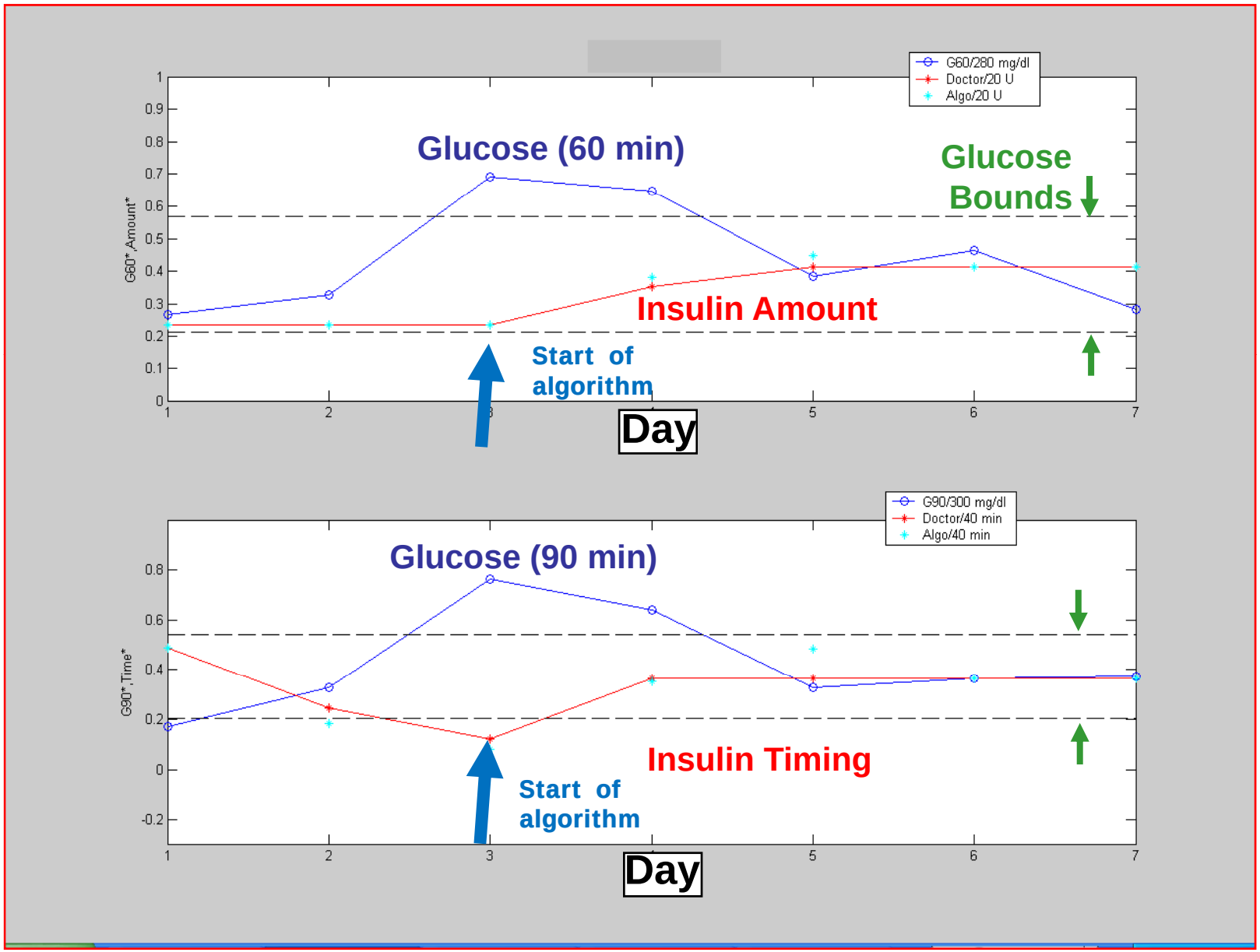
Preliminary Clinical Evaluation

- Summer 2003 @ Sansum Diabetes Research Institute
- 9 type I patients, pump users
- Separate phases for:
  - Bolus determination
  - Patient sensitivity identification
  - Single meal run-to-run studies
  - 3-meal run-to-run studies

Phase 6 (3 meal)



Preliminary Clinical Evaluation



Phase 6 Results

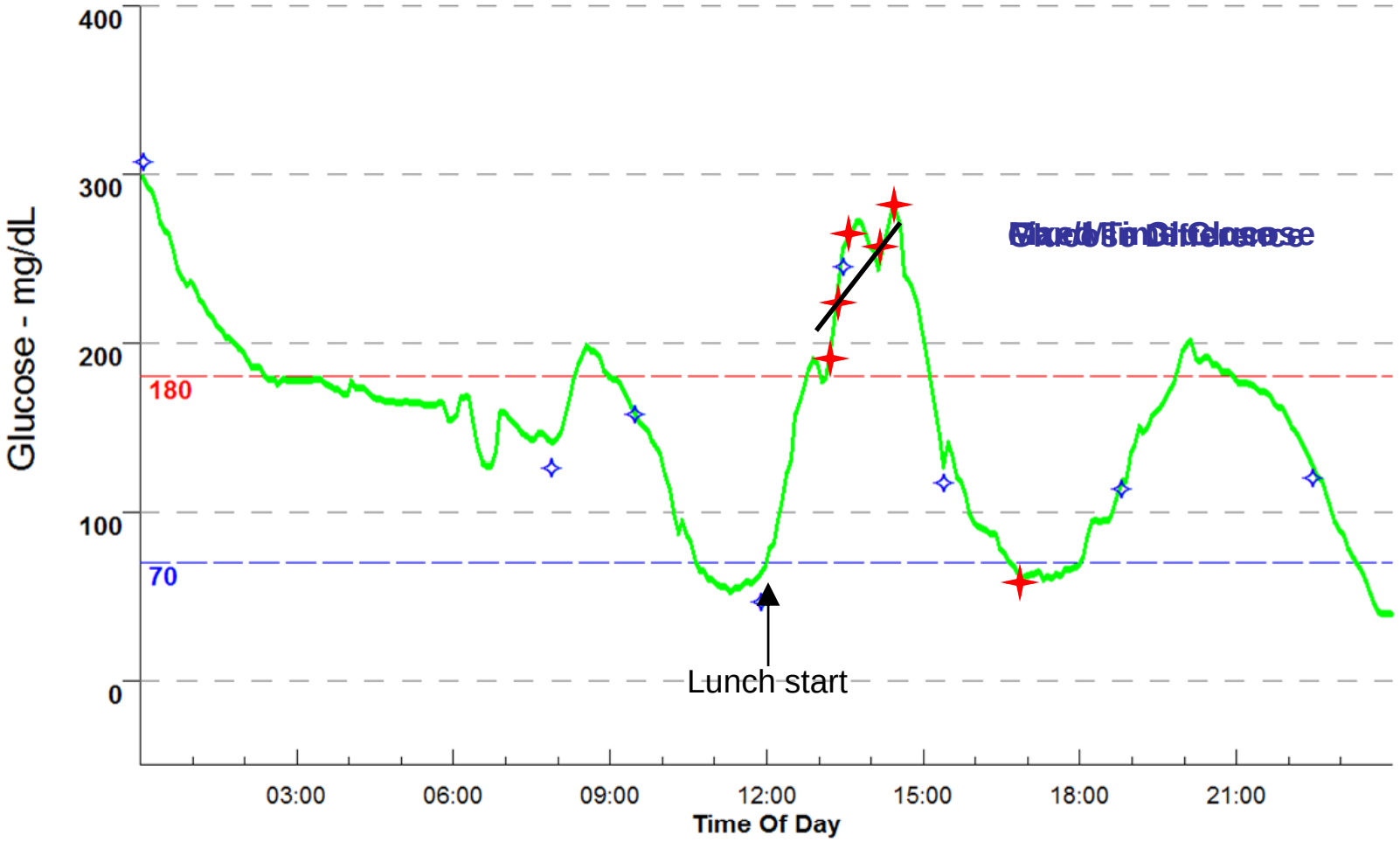
- **Class A (convergent, 3-4 days, clinically adequate range)** **[41%]**
- **Class B (always within range)** **[26%]**
- **Class C (divergent, incorrect sensitivity, mitigating circumstances)** **[33%]**

Dinner	Breakfast	Lunch
B	B	B
C	C	A
B	A	B
A	C	C
A	B	A
C	A	B
A	A	A
A	C	C
C	C	A

Summary - Observations

- By providing prepared meals to the patients, food chaos & variability was minimized.
- Due to the design of the trial, patients needed to record 3 BG measurements for each meal (9/day). If site problems or other events occurred, patients often checked more frequently as this was encouraged.
- In Phase V and Phase VI, the patients were under dosed by ~25% and were still able to keep G60 and G90 between 60 and 150 mg/dl suggesting that *once in good control, it is difficult to be bumped out of control*.
- As the trial progressed, the A1C values decreased significantly
- The impact of the bolus *timing* was unclear in terms of the effect on the glucose profile.

Comparison of Performance Metrics



New Algorithm Formulation

$$\nu_{k+1} = \nu_k + K\left(\psi^r - \psi_k\right)$$

$$\psi_k = \begin{bmatrix} G(T_{B_1}) - G(T_{B_2}) \\ G(T_{L_1}) - G(T_{L_2}) \\ G(T_{D_1}) - G(T_{D_2}) \end{bmatrix}$$

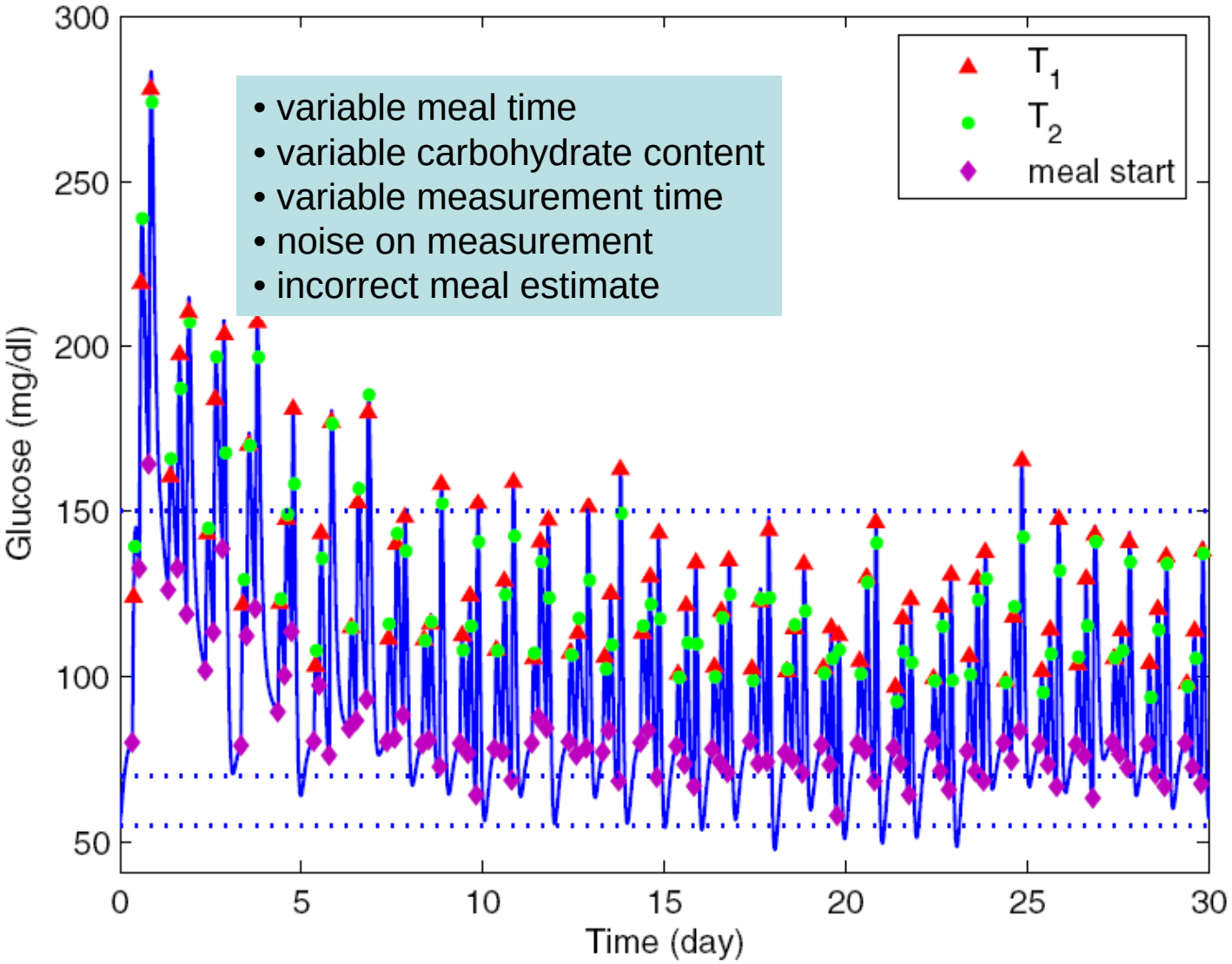
postprandial glucose difference

$$\nu_k = \begin{bmatrix} Q_B & Q_L & Q_D \end{bmatrix}^T$$

insulin meal bolus

- Only changing insulin dose, timing always fixed to the beginning of the meal
- Still require two post-meal measurements
  - First measurement 60-90 minutes after the start of the meal
  - Second measurement 30-60 minutes after the first
  - For each meal, denote these times as:
    T_{B1}, T_{B2}, T_{L1}, T_{L2}, T_{D1}, T_{D2}

Robustness of Algorithm



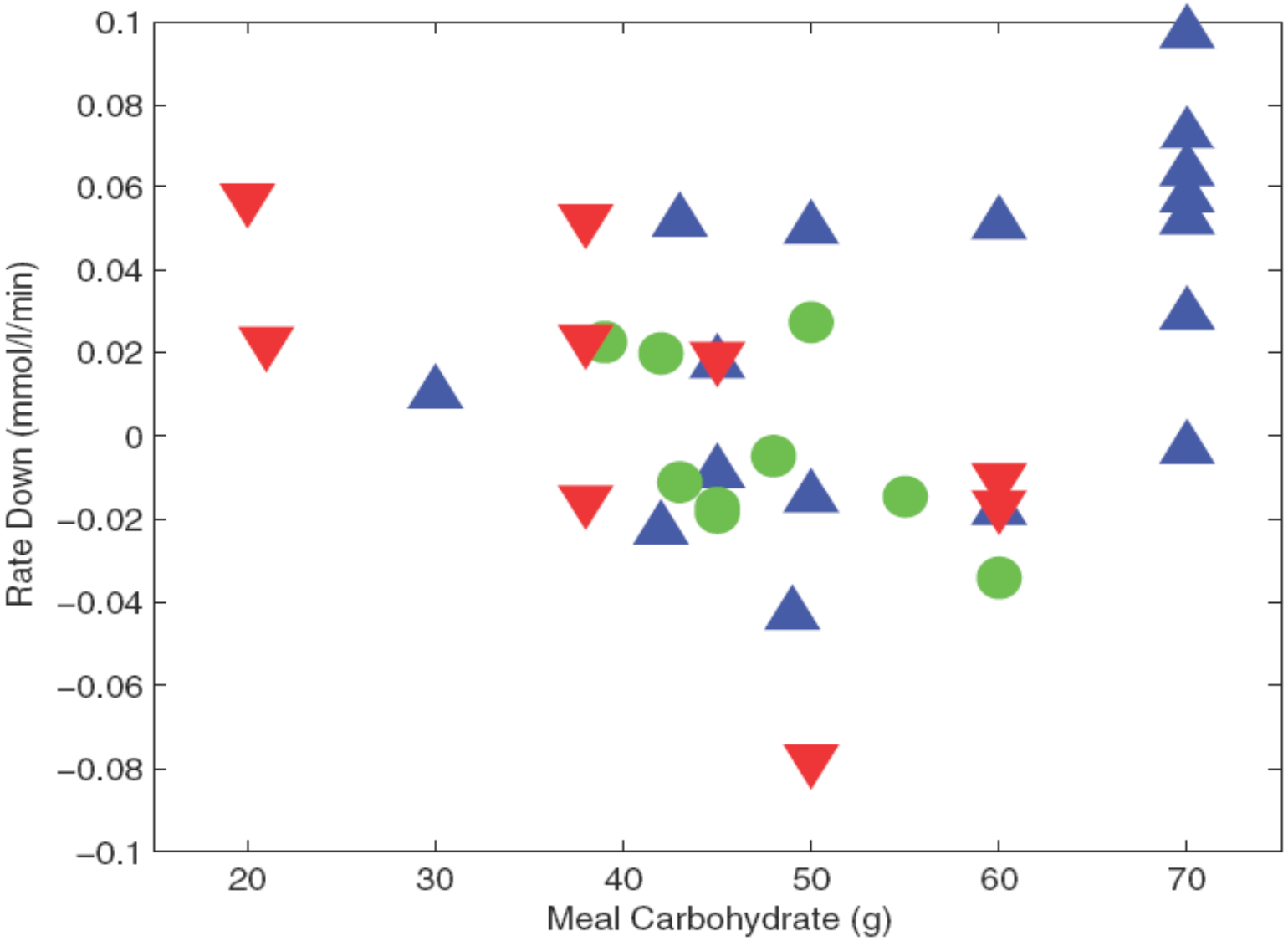
Clinical Evaluation of New Algorithm

- *11 subjects with type 1 diabetes & CSII pumps*

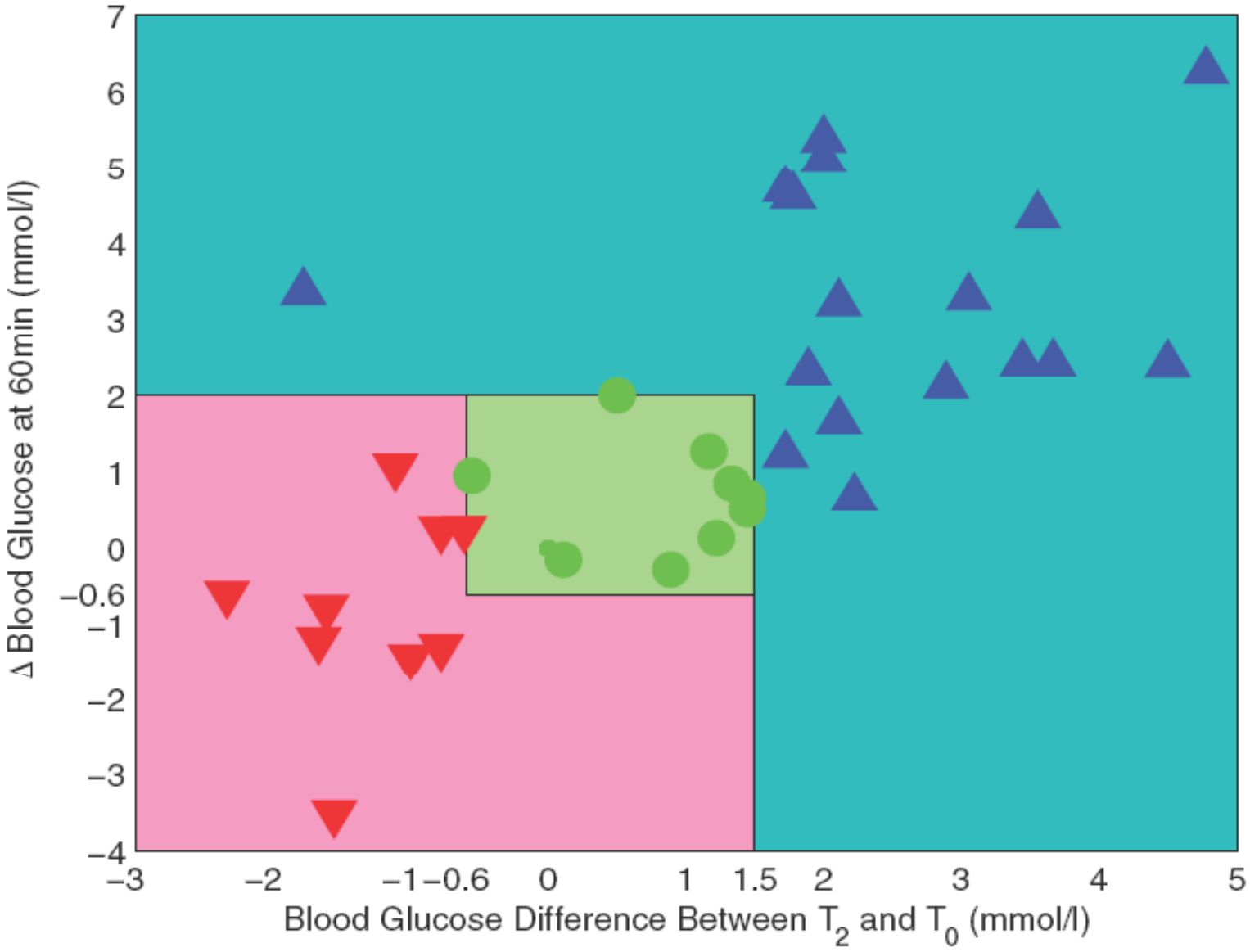
- **Phase 1**
  - Optimized basal rates
  - Brought out of control (1h post-prandial 170–200 mg/dl)
  - Lunch only
  - Carbohydrate content kept constant
  - Algorithm adjusted dosing over 2 weeks

- **Phase 2**
  - All three meals
  - Carbohydrate content varied
  - Algorithm adjusted dosing over 2–3 weeks

Challenge in Data Clustering



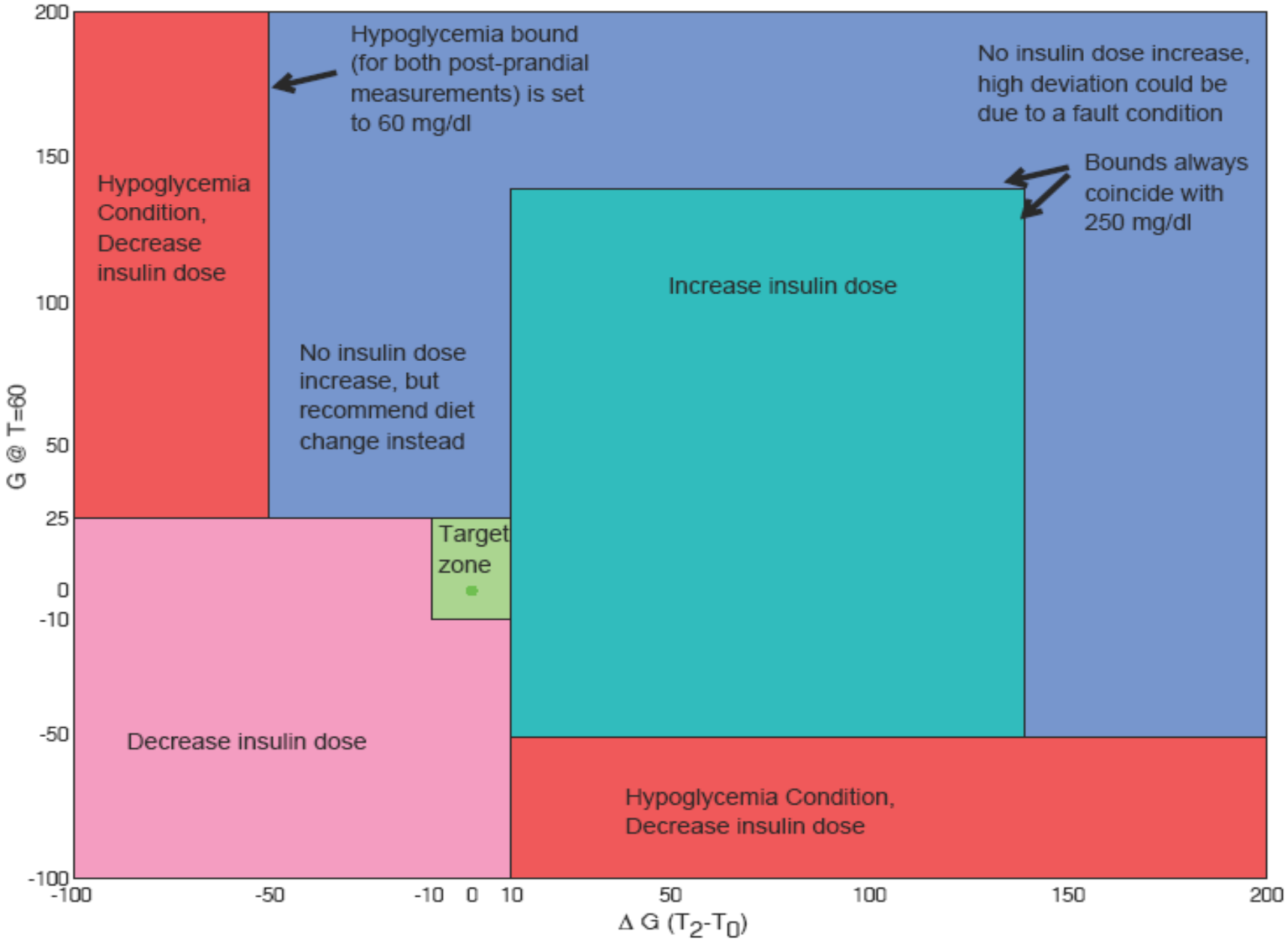
Medically-Inspired Performance Measure

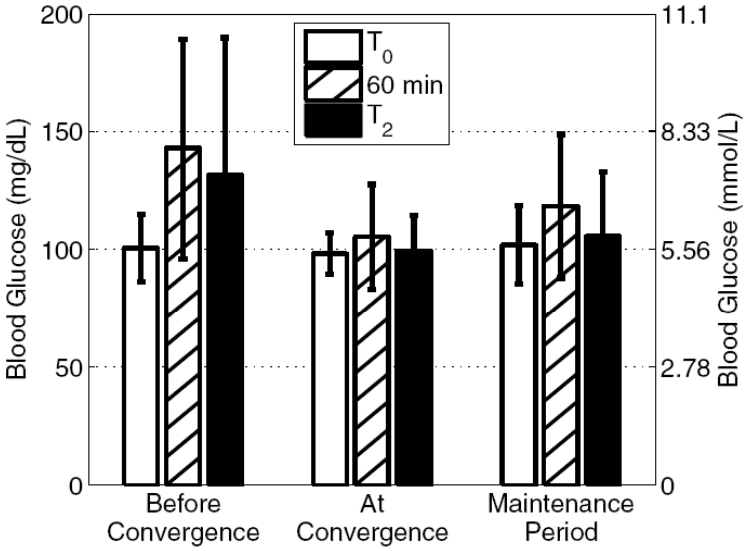


Phase 1 Results

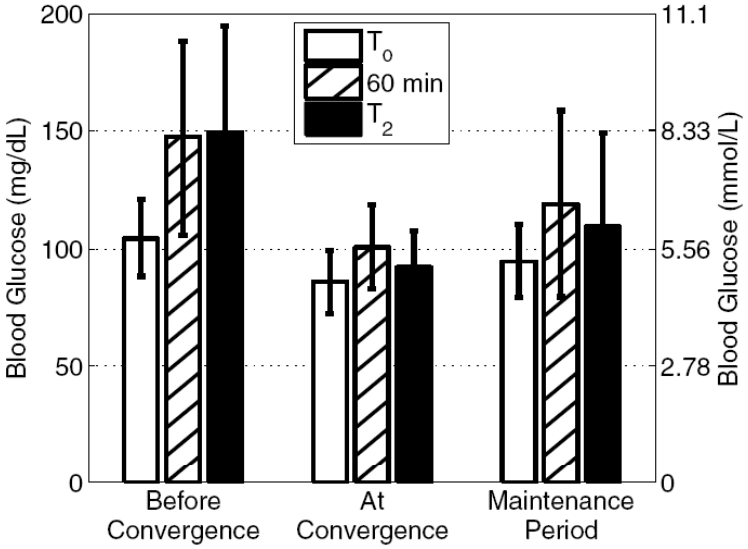
- Time to convergence: 5.4±3.6 days
- On the first day
  - pre-prandial BG: 101.7±22.4 mg/dl
  - 60 min post-prandial BG: 176.5±41.6 mg/dl
  - IC ratio: 1U to 14.15±3.95 g carbohydrate
- On convergence
  - pre-prandial BG: 94.7±23.9 mg/dl
  - 60 min post-prandial BG: 109.5±25.3 mg/dl
  - IC ratio: 1U to 9.47±2.27 g carbohydrate
- Over 118 meals, only two hypoglycemic (<55 mg/dl) events were reported (a 1.7% incidence rate), & none below 50 mg/dl

Modifications for Phase 2

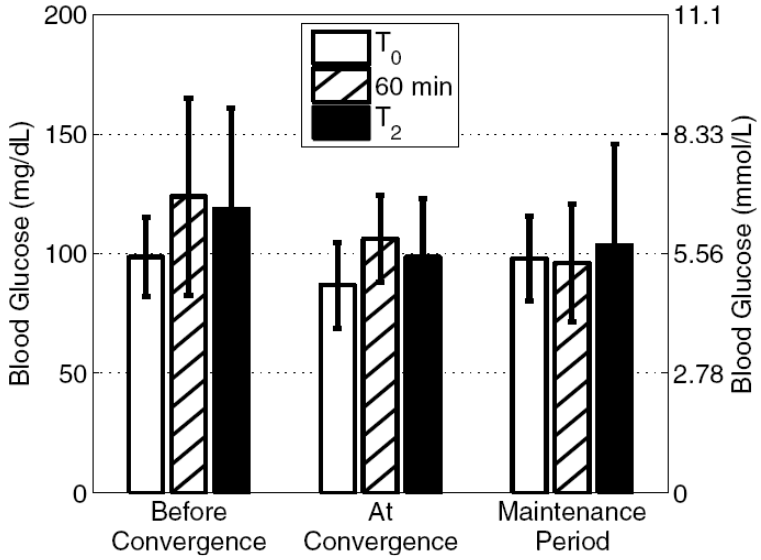




(a) Breakfast



(b) Lunch

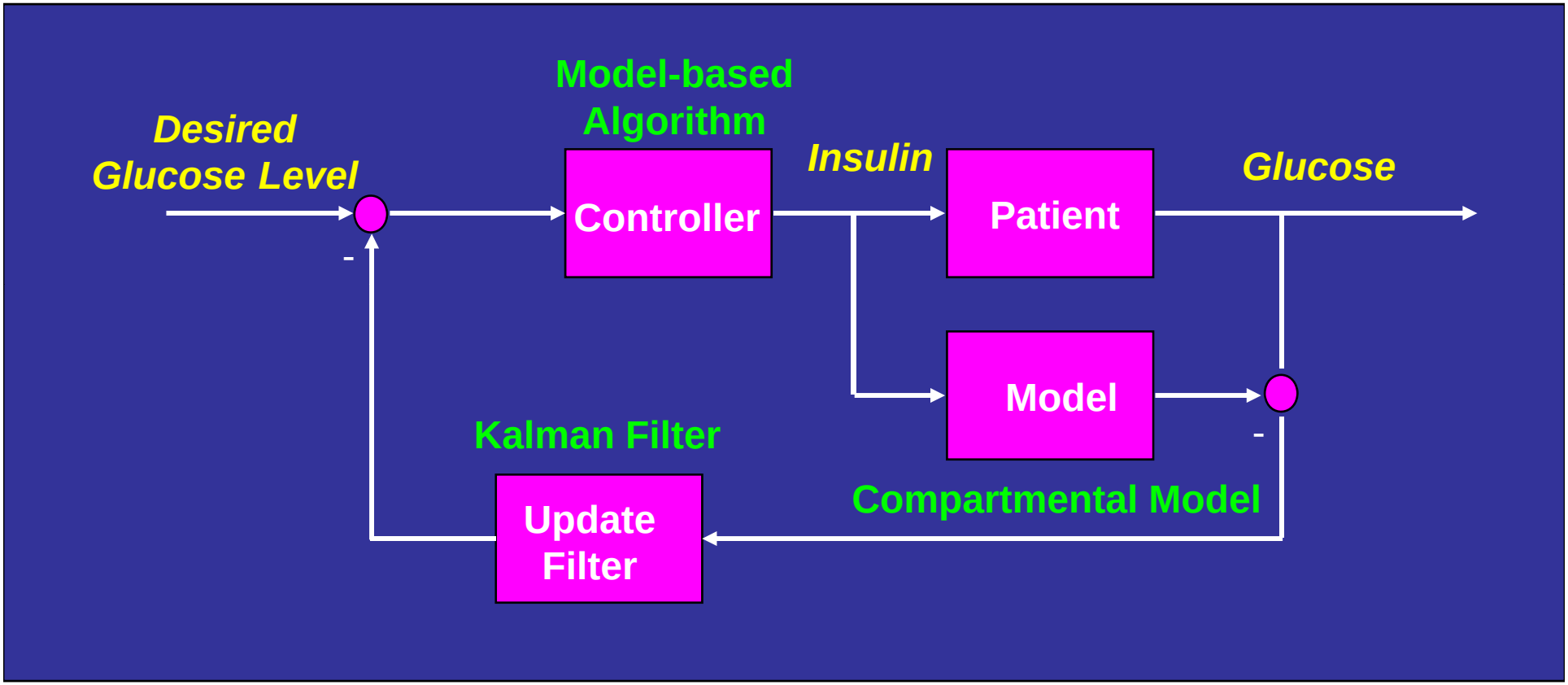


(c) Dinner

Closing the Loop

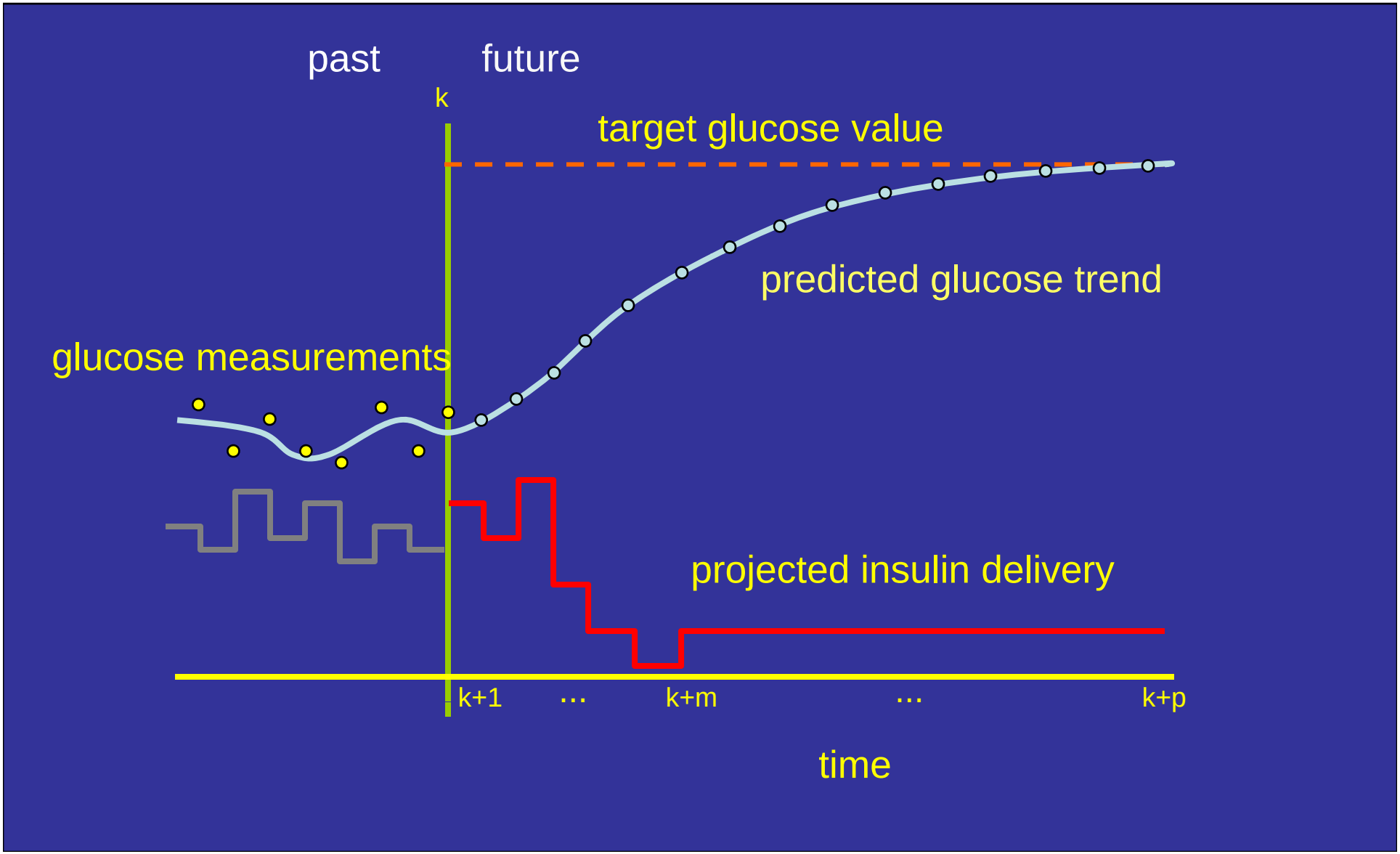
Model-Based Control Approach

[Parker, Doyle III, Peppas, *IEEE Trans. Biomed. Eng.*, 1999]



Key tenet of Robust Control Theory:
Model accuracy is directly tied to achievable performance

Moving Horizon Concept of MPC



MPC Components

- Reference Trajectory Specification
- Process Output Prediction (using Model)
- Control Action Sequence Computation
  (programming problem)
- Error Prediction Update (feedback)

Unconstrained MPC

- Recall: $$Y(k) = M^S Y(k-1) + S\Delta v(k-1)$$

- Model Prediction:

$$Y(k \mid k-1) = M^S Y(k-1 \mid k-1) + S\Delta v(k-1)$$

State estimate at time k-1

- Correction:

$$Y(k \mid k) = Y(k \mid k-1) + \mathbf{I}\left(\hat{y}(k) - y(k \mid k-1)\right)$$

measurement estimate

$$\mathbf{I} = \begin{bmatrix} I \\ \vdots \\ I \end{bmatrix} \Big\} n$$

Control Problem

$$\textbf{Objective :} \quad \min_{\Delta u(k),\ldots,\Delta u(k+m-1)} \sum_{\ell=1}^{p} \left\| y(k+\ell \mid k) - r(k+\ell) \right\|^2$$

$$\left(\|\bullet\| = \left(x^T x\right)^{1/2} \right)$$

- Suppose we want to control some outputs more tightly than others?
  - premultiply by: Γ_ℓ^y $\quad$ **for example :** $\Gamma_\ell^y = \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix} \ \forall \ell$
- Suppose we want to penalize manipulated variable moves?
  - premultiply by: Γ_ℓ^u

$$\min_{\Delta \mathbf{u(k)},\ldots,\Delta \mathbf{u(k+m-1)}} \sum_{\ell=1}^{p} \left\| \Gamma_\ell^y \left[y(k+\ell \mid k) - r(k+\ell) \right] \right\|^2 + \sum_{\ell=1}^{m} \left\| \Gamma_\ell^u \left[\Delta u(k+\ell-1) \right] \right\|^2$$

Vector notation:

$$\min_{\Delta U(k)} \left\{ \left\| \Gamma^y \left[Y_{pred}(k+1 \mid k) - R(k+1) \right] \right\|^2 + \left\| \Gamma^u \Delta U(k) \right\|^2 \right\}$$

$$\textbf{s.t.} \quad Y_{pred}(k+1 \mid k) = \overline{\overline{M}} Y_{pred}(k \mid k) + \overline{\overline{S}}^{\mathbf{u}} \Delta U(k) + \overline{S}^{\mathbf{d}} \Delta d(k)$$

Least Squares Formulation

$$\begin{bmatrix} \Gamma^y \overline{\overline{S^u}} \\ \Gamma^u \end{bmatrix} \Delta U(k) = \begin{bmatrix} \Gamma^y & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} R(k+1) - \overline{\overline{M}} Y_{pred}(k \mid k) \; - \overline{S}^{\mathbf{d}} \; \Delta d(k) \\ 0 \end{bmatrix}$$

$$\equiv \begin{bmatrix} \Gamma^y E_p(k+1 \mid k) \\ 0 \end{bmatrix}$$

$$E_p(k+1 \mid k) = \begin{bmatrix} e(k+1 \mid k) \\ \vdots \\ e(k+p \mid k) \end{bmatrix}$$

Overspecified problem (m>p):

$$\Delta U(k) = \left(\overline{\overline{S^u}}^T \Gamma^{y^T} \Gamma^y \overline{\overline{S^u}} + \Gamma^{u^T} \Gamma^u \right)^{-1} \overline{\overline{S^u}}^T \Gamma^{y^T} \Gamma^y E_p(k+1 \mid k)$$

Receding Horizon Implementation

$$\Delta u(k) = \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix} \left(\overline{\overline{S^u}}^T \Gamma^{y^T} \Gamma^y \overline{\overline{S^u}} + \Gamma^{u^T} \Gamma^u \right)^{-1} \overline{\overline{S^u}}^T \Gamma^{y^T} \Gamma^y E_p(k+1 \mid k)$$

Off-line computation (K_{MPC})

Unconstrained MPC Algorithm

- Do not vary manipulated inputs for n intervals
  (assume no disturbances - system is at equilibrium)
- Initialize - measure output $\hat{y}(0)$ $\quad Y(0\,|\,0) = \left[\hat{y}(0)^T, \ldots, \hat{y}(0)^T\right]^T$

  measure Δd(0), get new measurements $(\hat{y}(1), \Delta d(1))$
- Prediction: $Y(k\,|\,k-1) = M^S Y(k-1\,|\,k-1) + S^u \Delta u(k-1) + S^d \Delta d(k-1)$
- Correction: $Y(k\,|\,k) = Y(k\,|\,k-1) + \mathbf{I}\mathbf{K}_\mathbf{F}\left(\hat{y}(k) - y(k\,|\,k-1)\right)$
- Compute reference trajectory error

$$E_p(k+1\,|\,k) = R(k+1) - \overline{\overline{M}} Y(k\,|\,k) - \overline{S}^{\mathbf{d}}\ \Delta d(k)$$

- Control computation $\quad \Delta u(k) = K_{MPC} E_p(k+1\,|\,k)$
- Obtain measurement $\quad (\hat{y}(k+1), \Delta d(k+1)),\ \ \mathbf{set}\ k = k+1,\ \ \mathbf{goto}\ \ \mathbf{3}$

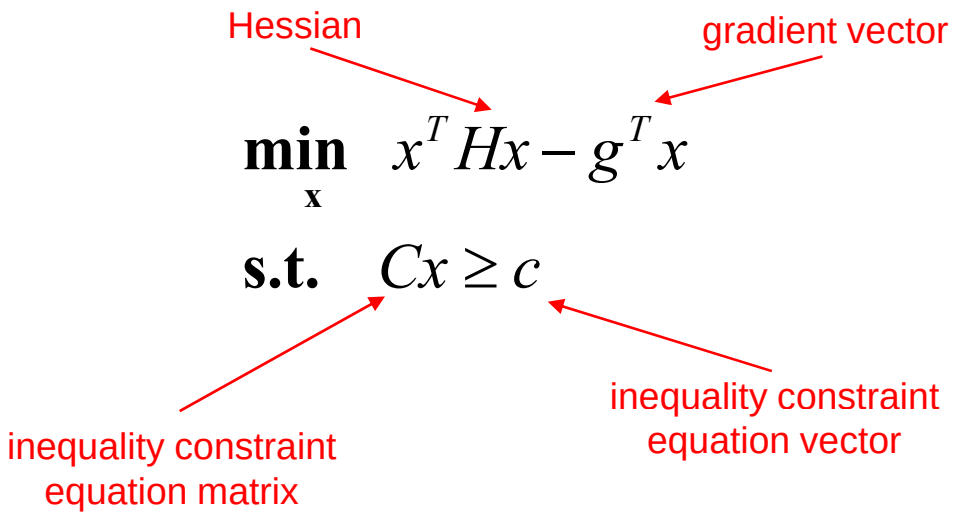
Tuning MPC

- Horizons (m,p)
- Penalty weights $\left(\Gamma^y, \Gamma^u\right)$
- Filters $\left(f_1, \ldots, f_{n_y}\right)$
  - unmeasured disturbance
  - setpoint filter:

$$\overline{R}(k \mid k-1) = \overline{\overline{M}}\,\overline{R}(k-1 \mid k-1)$$
$$\overline{R}(k \mid k) = \overline{R}(k \mid k-1) + \mathbf{IK}_{\mathbf{F}}''\left(R(k) - \overline{R}(k \mid k-1)\right)$$
$$\mathbf{IK}_{\mathbf{F}}'' = \mathbf{diag}\left\{f_1'', \ldots, f_{n_y}''\right\}$$

Constrained MPC

- General Structure of QP



- Several *robust* and *reliable* QP solvers available

MPC Constraints

Input Magnitude

$$\begin{bmatrix} \begin{bmatrix} I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ I & \dots & \ddots & 0 \\ I & I & \cdots & I \end{bmatrix} \\ \begin{bmatrix} -I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ -I & \dots & \ddots & 0 \\ -I & -I & \cdots & -I \end{bmatrix} \end{bmatrix} \Delta U(k) \geq \begin{bmatrix} u(k-1) - \bar{u}(k) \\ \vdots \\ u(k-1) - \bar{u}(k+m-1) \\ \underline{u}(k) - u(k-1) \\ \vdots \\ \underline{u}(k+m-1) - u(k-1) \end{bmatrix}$$

Input Rate

$$\begin{bmatrix} -I \\ I \end{bmatrix} \Delta U(k) \geq \begin{bmatrix} -\Delta\bar{u}(k) \\ \vdots \\ -\Delta\bar{u}(k+m-1) \\ -\Delta\bar{u}(k) \\ \vdots \\ -\Delta\bar{u}(k+m-1) \end{bmatrix}$$

Output Magnitude

$$\begin{bmatrix} -\bar{\bar{S}}^{u} \\ \bar{\bar{S}}^{u} \end{bmatrix} \Delta U(k) \geq \begin{bmatrix} \bar{\bar{M}} Y(k \mid k) + \bar{S}^{\mathbf{d}} \, \Delta d(k) - \begin{bmatrix} \bar{y}(k+1) \\ \vdots \\ \bar{y}(k+p) \end{bmatrix} \\ -\bar{\bar{M}} Y(k \mid k) - \bar{S}^{\mathbf{d}} \, \Delta d(k) + \begin{bmatrix} \underline{y}(k+1) \\ \vdots \\ \underline{y}(k+p) \end{bmatrix} \end{bmatrix}$$

Constrained Formulation

$$\min_{\Delta U(k)} \; \Delta U(k)^T H^u \Delta U(k) - G(k+1|k)^T \Delta U(k)$$

$$\text{s.t.} \quad C^u \Delta U(k) \geq c^u(k+1|k)$$

$$H^u = \overline{\overline{S^u}}^T \Gamma^{y^T} \Gamma^y \overline{\overline{S^u}} + \Gamma^{u^T} \Gamma^u$$

$$G^u = \overline{\overline{S^u}}^T \Gamma^{y^T} \Gamma^y \overline{\overline{S^u}} E_p(k+1|k)$$

Important Observations

- An unconstrained model predictive controller can be recast as a classical controller (PID, etc.)
  - First-order dynamic model = PI
  - Second-order dynamic model = PID
  - More complex models yield more complex controllers

- MPC is a very general framework and represents the state of the art in a number of commercial sectors
(refining, chemicals, aerospace, etc.)

- **Strategy: *understanding* influences control design**

MPC Approach for Biomedical Control

- Many characteristics and requirements in common with industrial process control
- Successful drug delivery (clinical) studies
  - atracurium (Linkens and Mahfouf, 1995)
  - sodium nitroprusside (Kwok et al., 1997)
  - sodium nitroprusside & dopamine (Rao et al., 1999)
  - anesthesia (Gentilini et al., 2001)
- Present studies – computer patient models
  - Bergman Model (Bergman et al., 1981)
  - Sorensen Model (Sorensen & Colton, 1985)
  - AIDA Model (Lehmann & Deutsch, 1992)

Insulin Delivery – Algorithmic Details

- Solves on-line optimization problem
- Controller objective function:

$$\min_{\Delta U(k)} \left\| \Gamma_y \, Y(k+1 \mid k) - R(k+1 \mid k) \right\|_2^2 + \left\| \Gamma_u \, \Delta U(k) \right\|_2^2$$

Glucose Tracking

Insulin Penalty

- Controller tuning
  - » move horizon, *m*, and prediction horizon, *p*
  - » setpoint tracking (Γ_y), move suppression (Γ_u) weighting
- Constraints:

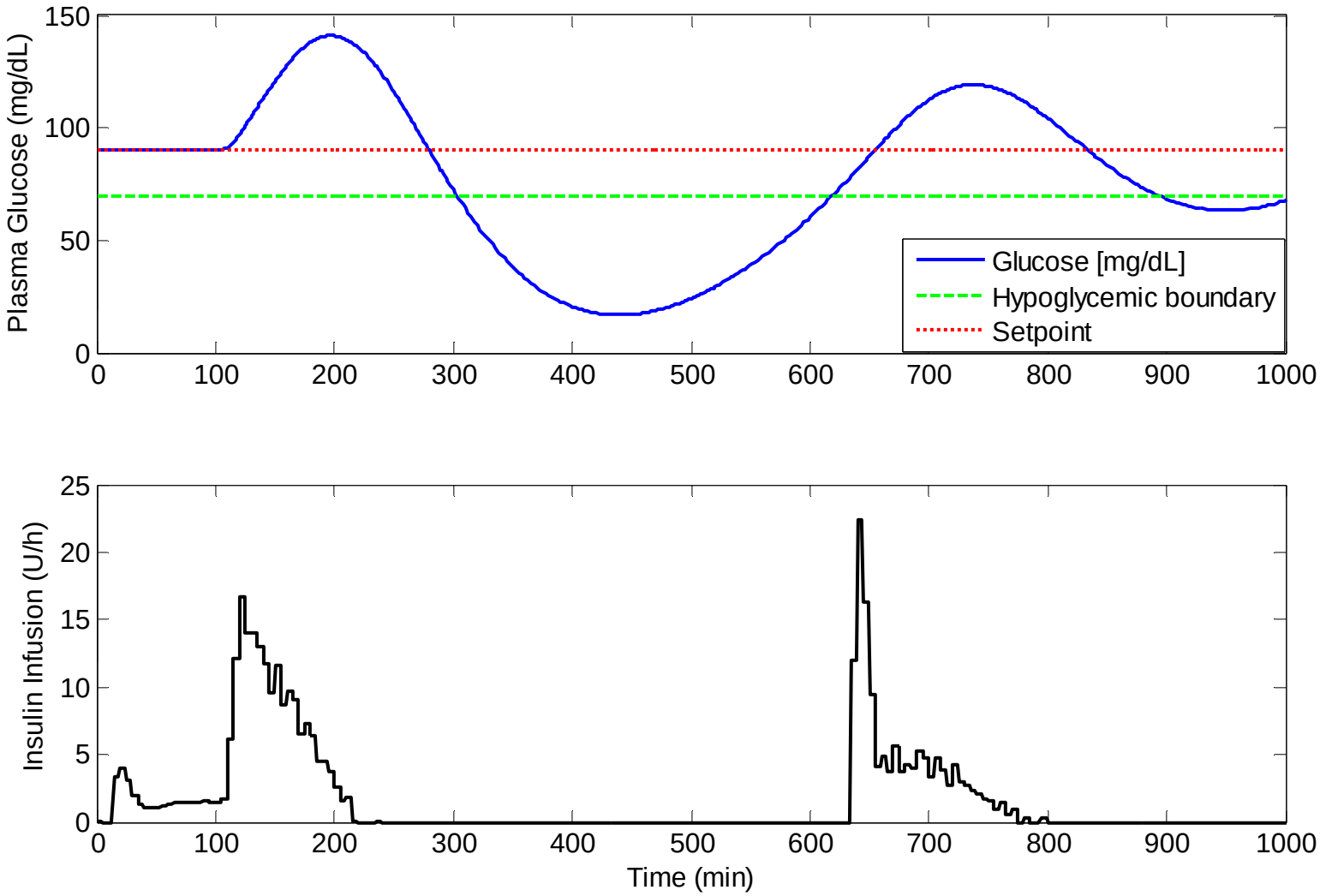
$$0\, mU/min \le U(k) \le 66.25\, mU/min$$

$$\left|\Delta U(k)\right| \le 16.5\, mU/min$$

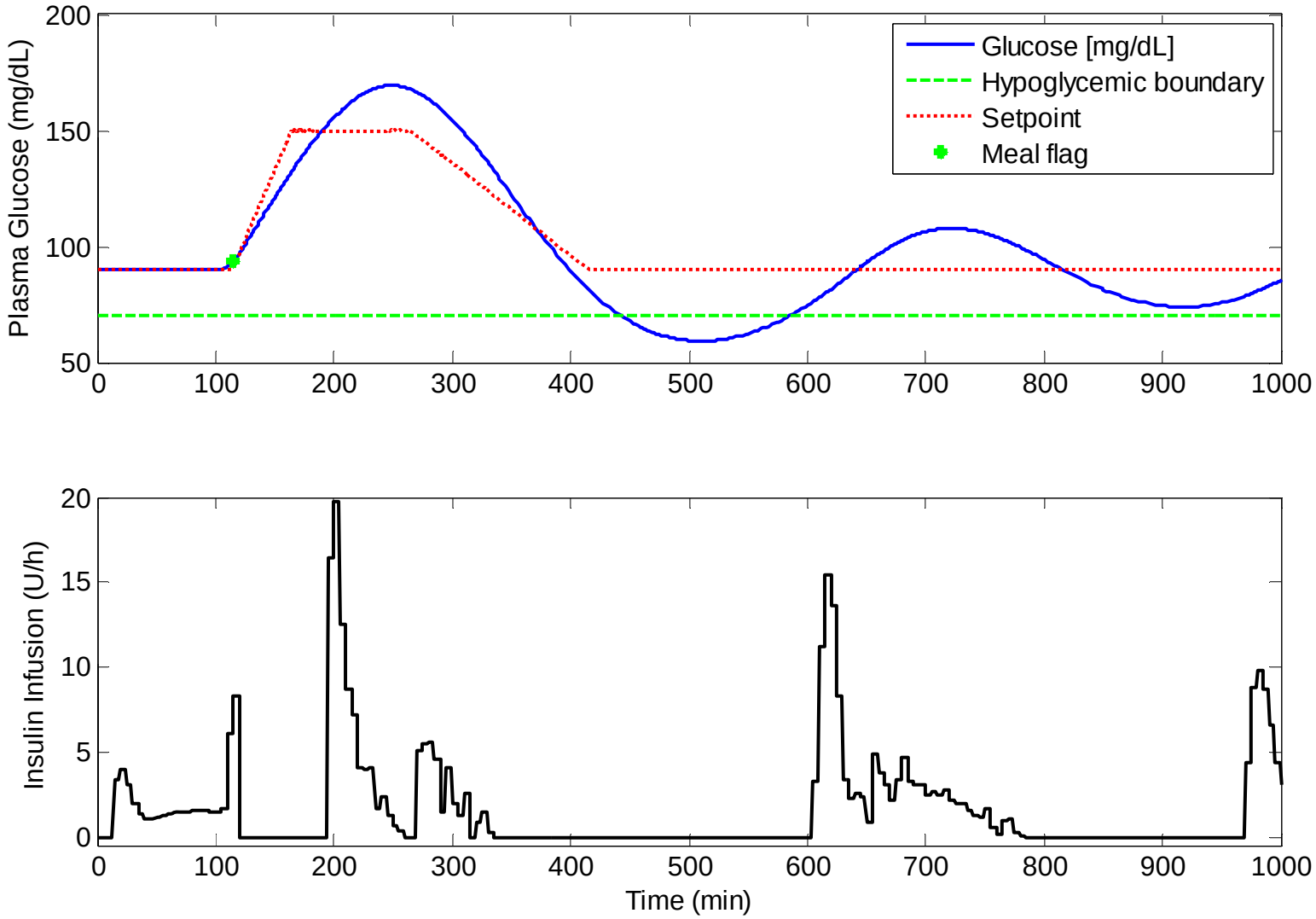
$$G_{\min}(k) \ge 60\ mg/dl$$

Pump Limitations

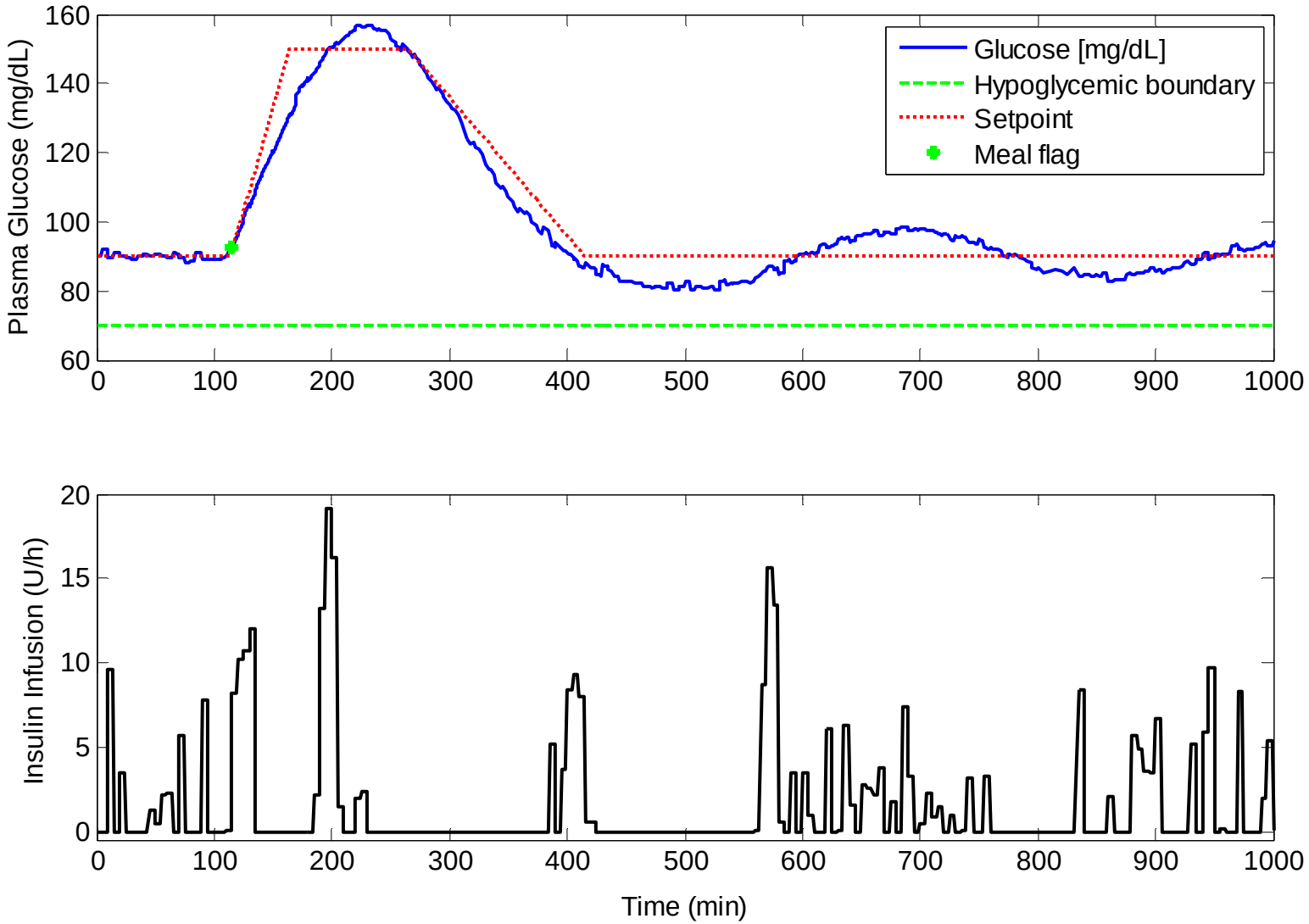
Safety



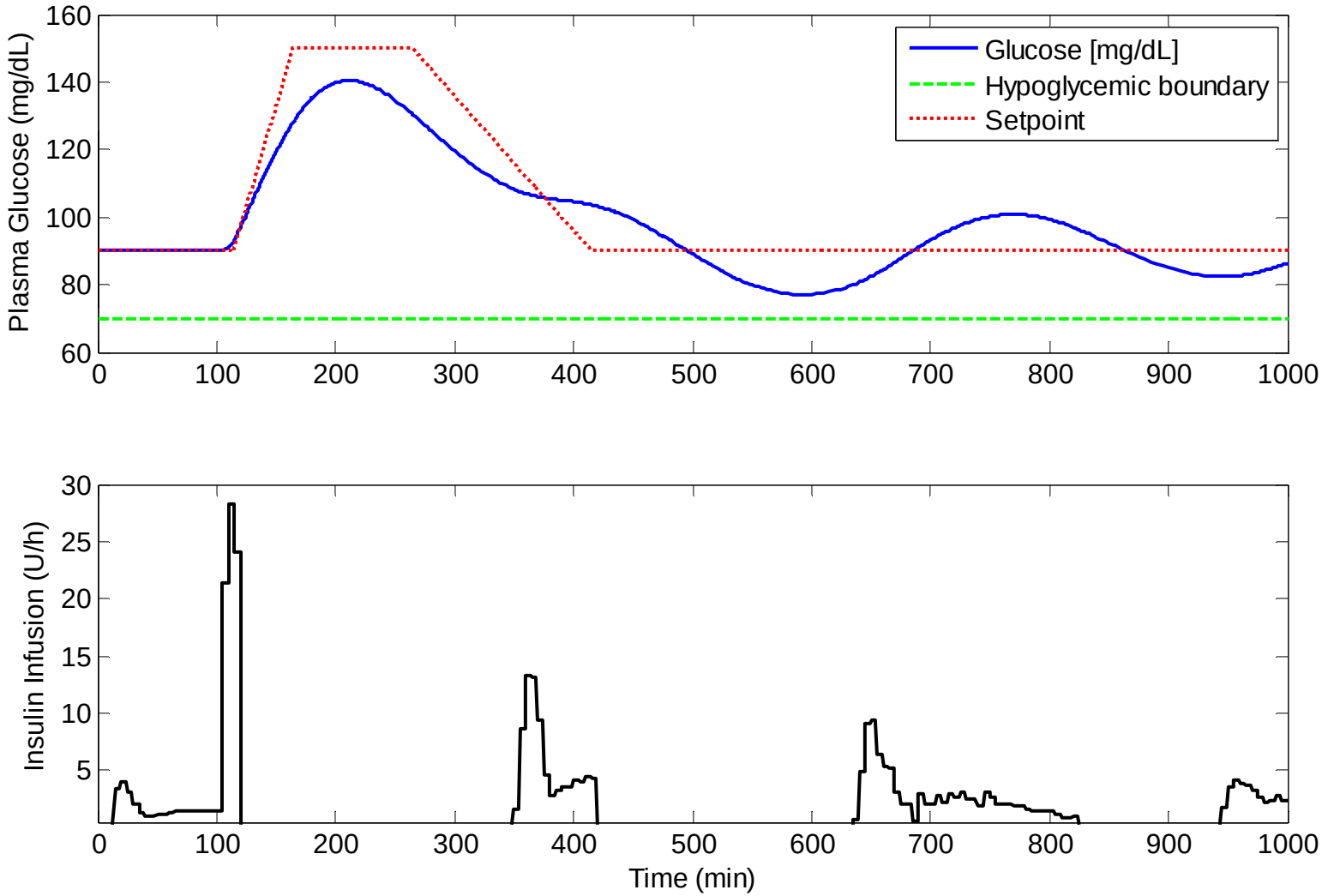
Simple MPC with no meal detection. 15 units of insulin were used to cover the meal (lower figure in black), which led to a severe hypoglycemia. The setpoint denoted by the red dotted line, plasma glucose with blue line, hypoglycemic boundary (70 mg/dL) with green dashed line and controller moves with the black line in the lower plot



MPC with meal detection and variable reference. 11.5 units of insulin were used to cover the meal, which led to a mild hypoglycemia. The setpoint denoted by the red dotted line, plasma glucose with blue line, hypoglycemic boundary (70 mg/dL) with green dashed line, meal flag point with the green circle and controller moves with the black line in the lower plot.



MPC with meal detection, variable reference and estimated glucose absorption profile and process noise of ± 3 mg/dL. 11 units of insulin were used to cover the meal. The setpoint denoted by the red dotted line, plasma glucose with blue line, hypoglycemic boundary (70 mg/dL) with green dashed line, meal flag point with the green circle and controller moves with the black line in the lower plot.



MPC with announced meal, variable reference and estimated glucose absorption profile. 10 units of insulin were used to cover the meal. The setpoint denoted by the red dotted line, plasma glucose with blue line, hypoglycemic boundary (70 mg/dL) with green dashed line, meal flag point with the green circle and controller moves with the black line in the lower plot.