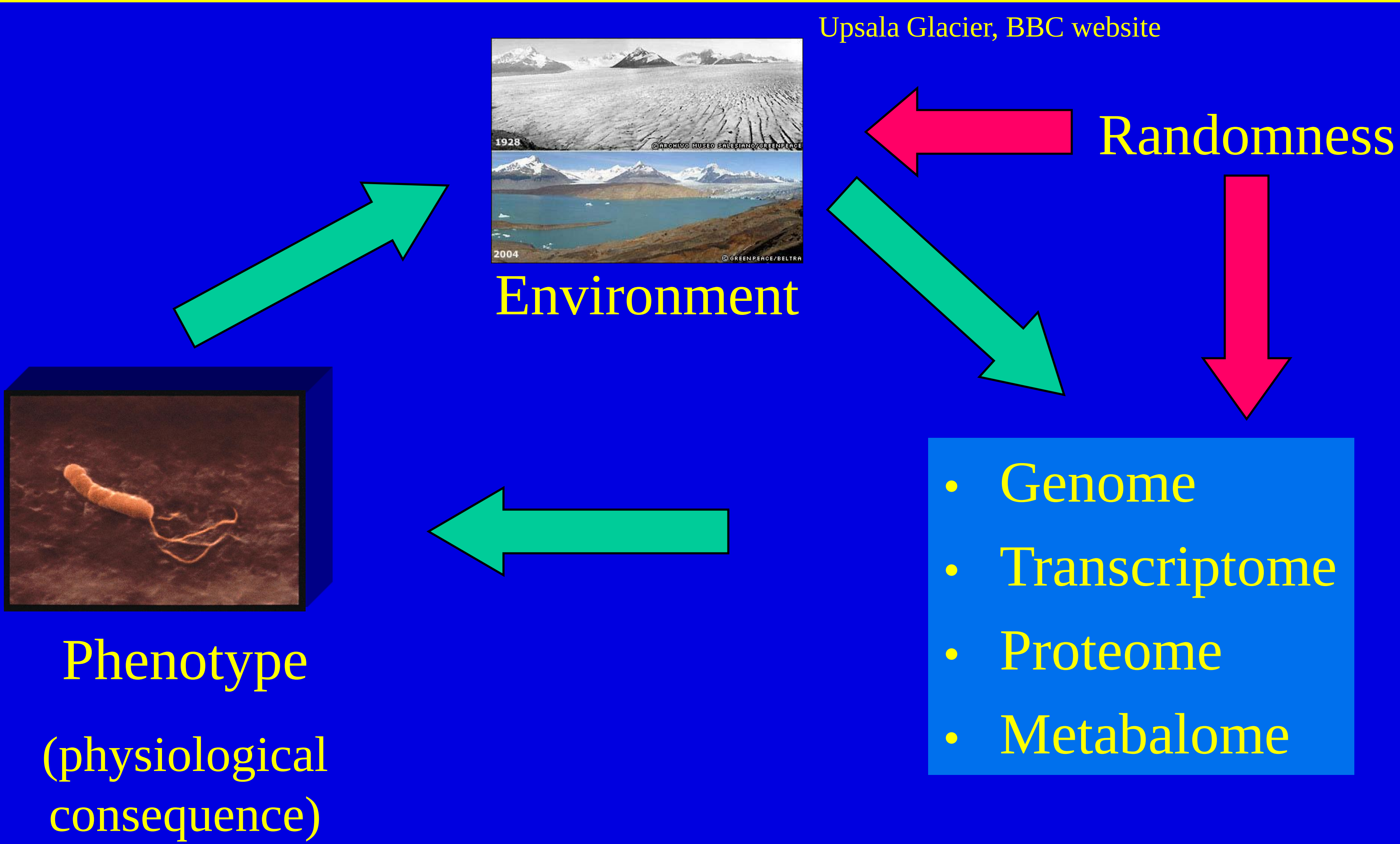
Design and analysis of multiple feedback loops using natural and synthetic genetic constructs

K V Venkatesh
Department of Chemical Engineering
IIT Bombay, Powai, Mumbai 400 076

Genotype to Phenotype: An Integrated Approach



Upsala Glacier, BBC website

Environment

Randomness

- Genome
- Transcriptome
- Proteome
- Metabalome

Phenotype

(physiological consequence)

Genotype to Phenotype: Modeling and Analysis

Modeling

Logical networks

ODE/PDE

Delayed ODE

Stochastic approach

Multiscale

Analysis

Data mining

Estimation theory

Nonlinear systems theory

Feedback control theory

Sensitivity

Key: Identify design principles

Simplicity in Biology

Alon, *Nature* **446**:497 (2007)

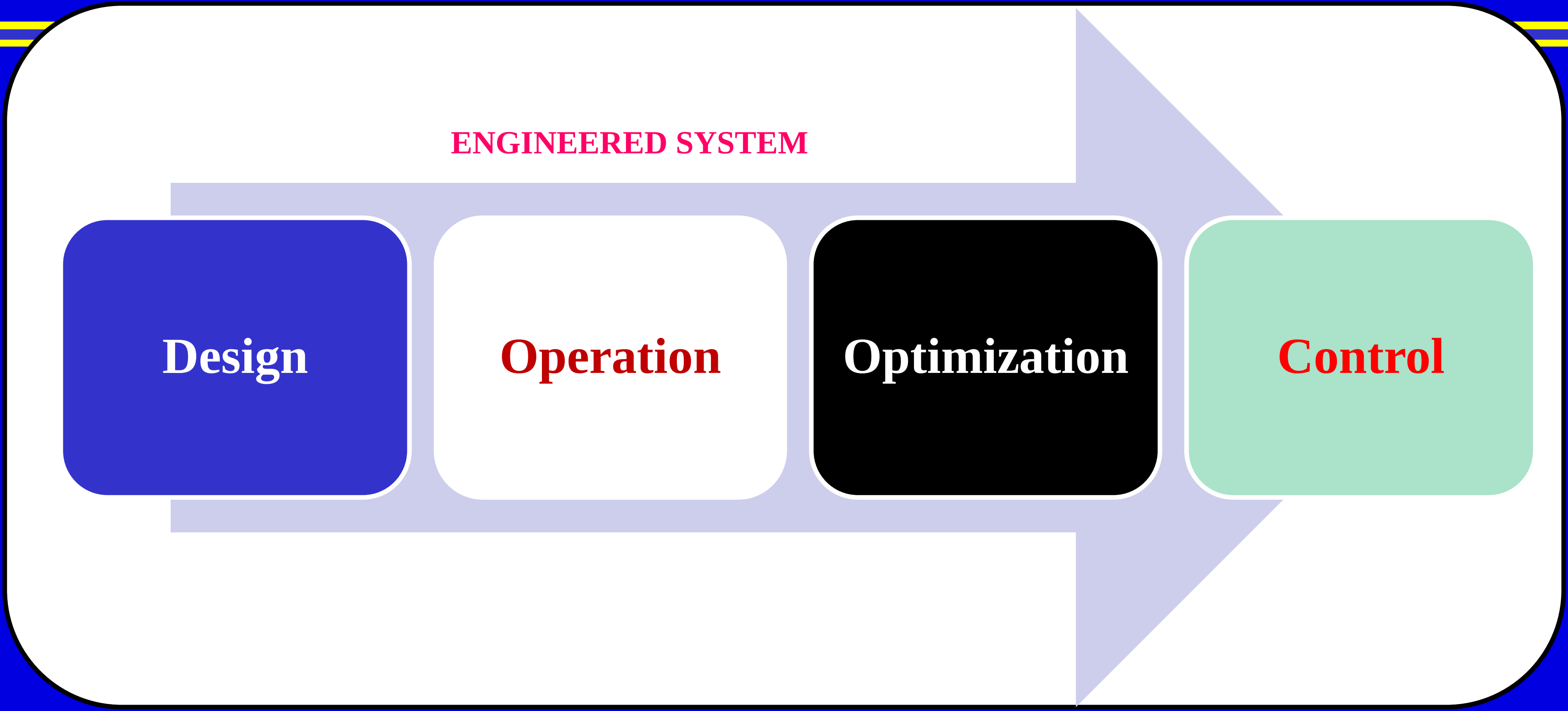
- Diversity in genes NOT in motifs of regulatory networks
- These networks are robust (yet fragile)
- Combination of motifs yield new dynamical properties
- Network motifs conserved across organisms (animal, plant)

Analog Motifs in Natural Systems

— Freeman, *Nature*, **408**, 2000

- PI control (plasma calcium homeostasis in mammals)
- Negative feedback (autorepression)
- Positive feedback (growth in cell development)
- Negative/positive feedback (prolong weak signals)
- Feedforward (heat shock response)
- Cascades (insulin signaling pathways)

Engineered Versus Natural System



ENGINEERED SYSTEM

Design

Operation

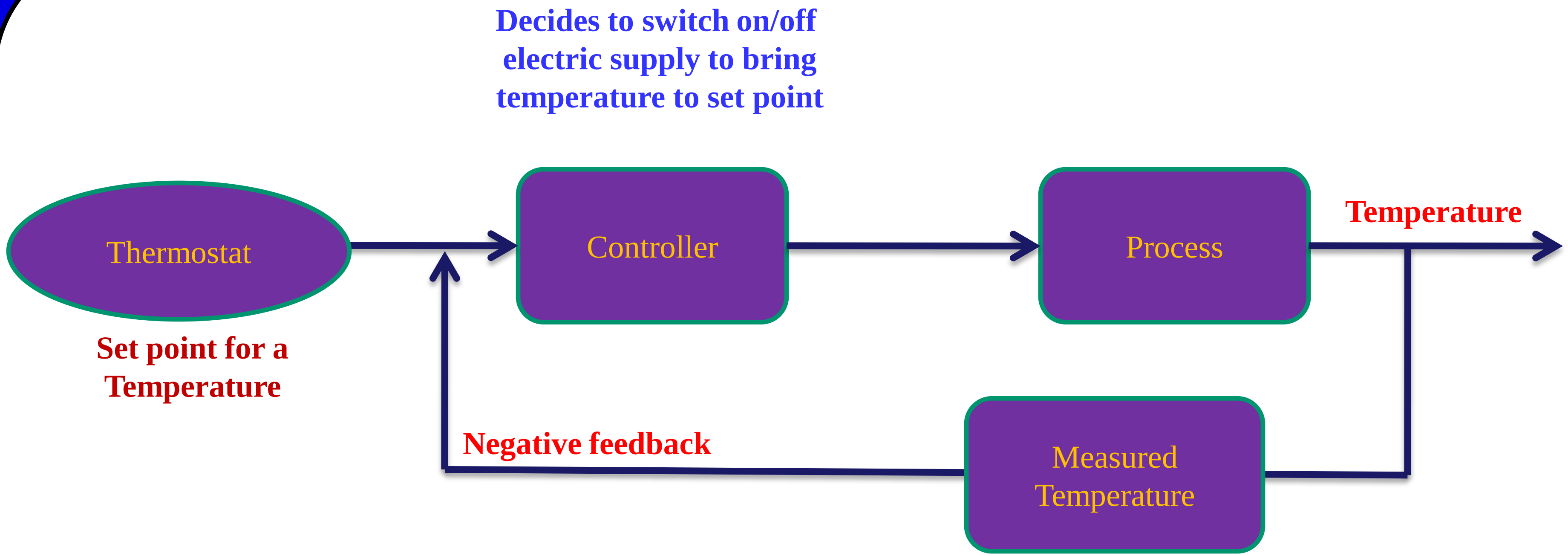
Optimization

Control

Engineered system: **bottom-up design** with known functionality of components

Natural system: **top down design** with unknown inherent property of various motifs

Engineered Systems : Room Heater



Decides to switch on/off electric supply to bring temperature to set point

Thermostat

Set point for a Temperature

Controller

Process

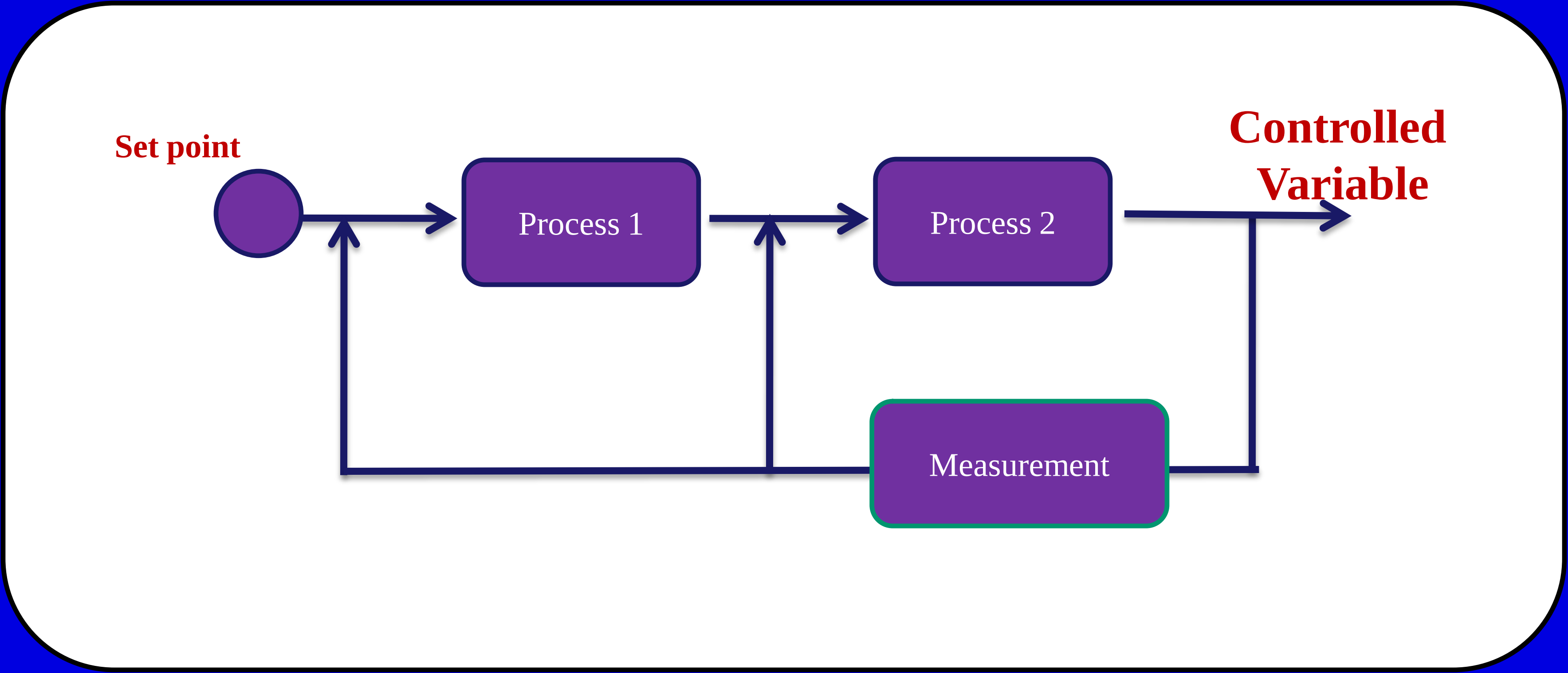
Temperature

Measured Temperature

Negative feedback

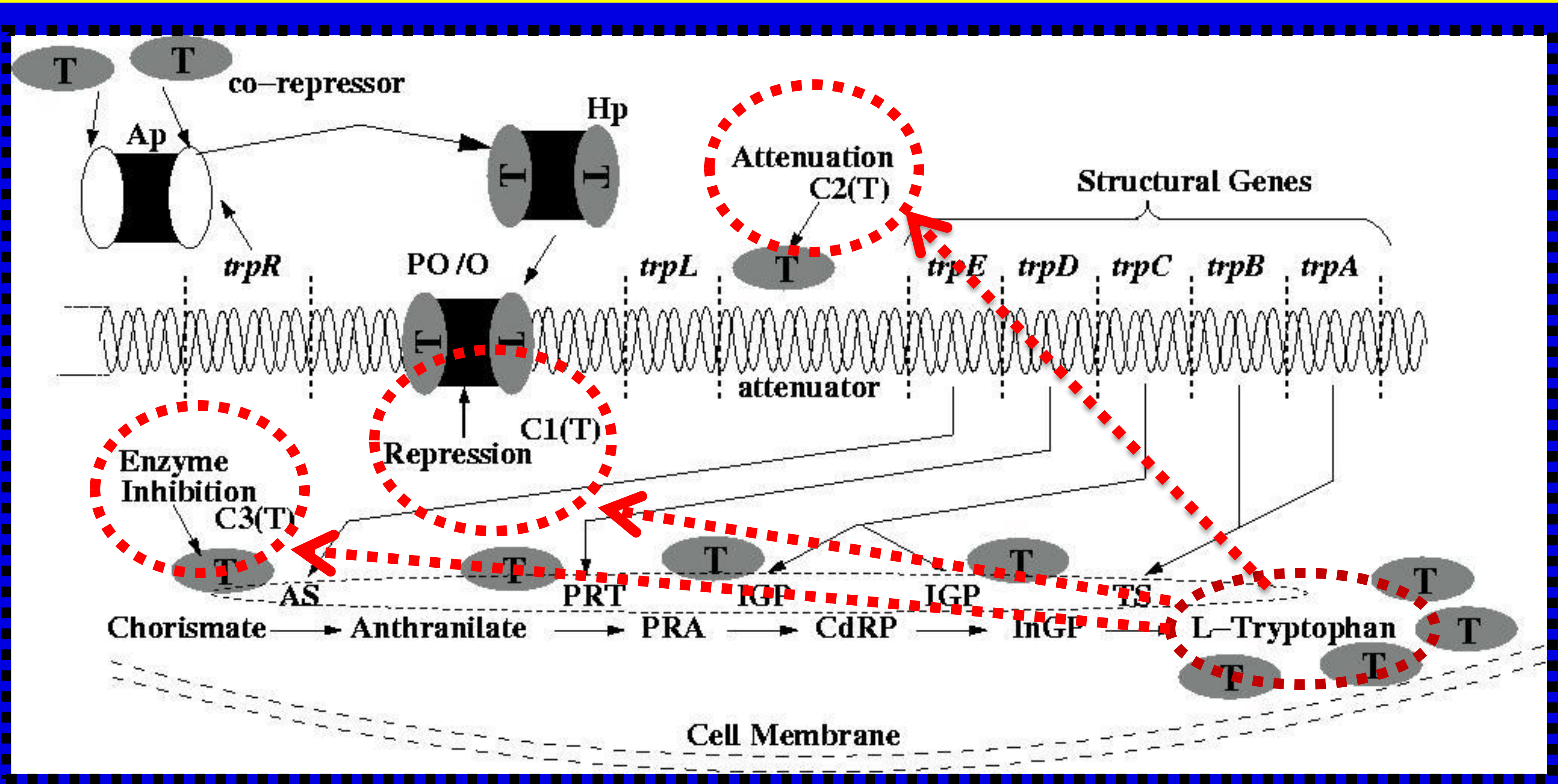
SINGLE INPUT SINGLE OUTPUT (SISO)

Multiple Input Multiple Output: a motif observed in Biological System



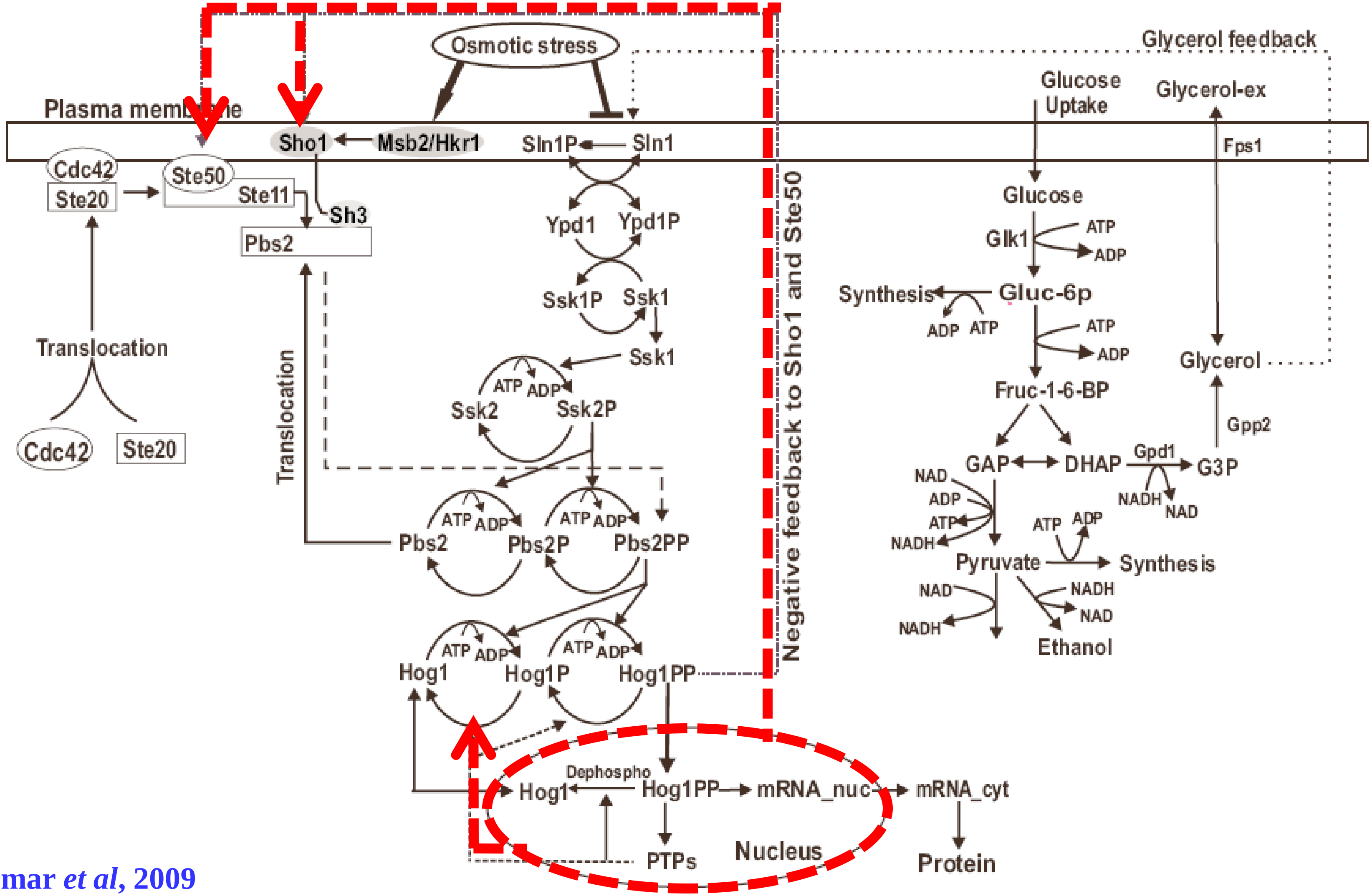
Single output is regulating the multiple upstream processes

Tryptophan in *E. coli* (bacteria)



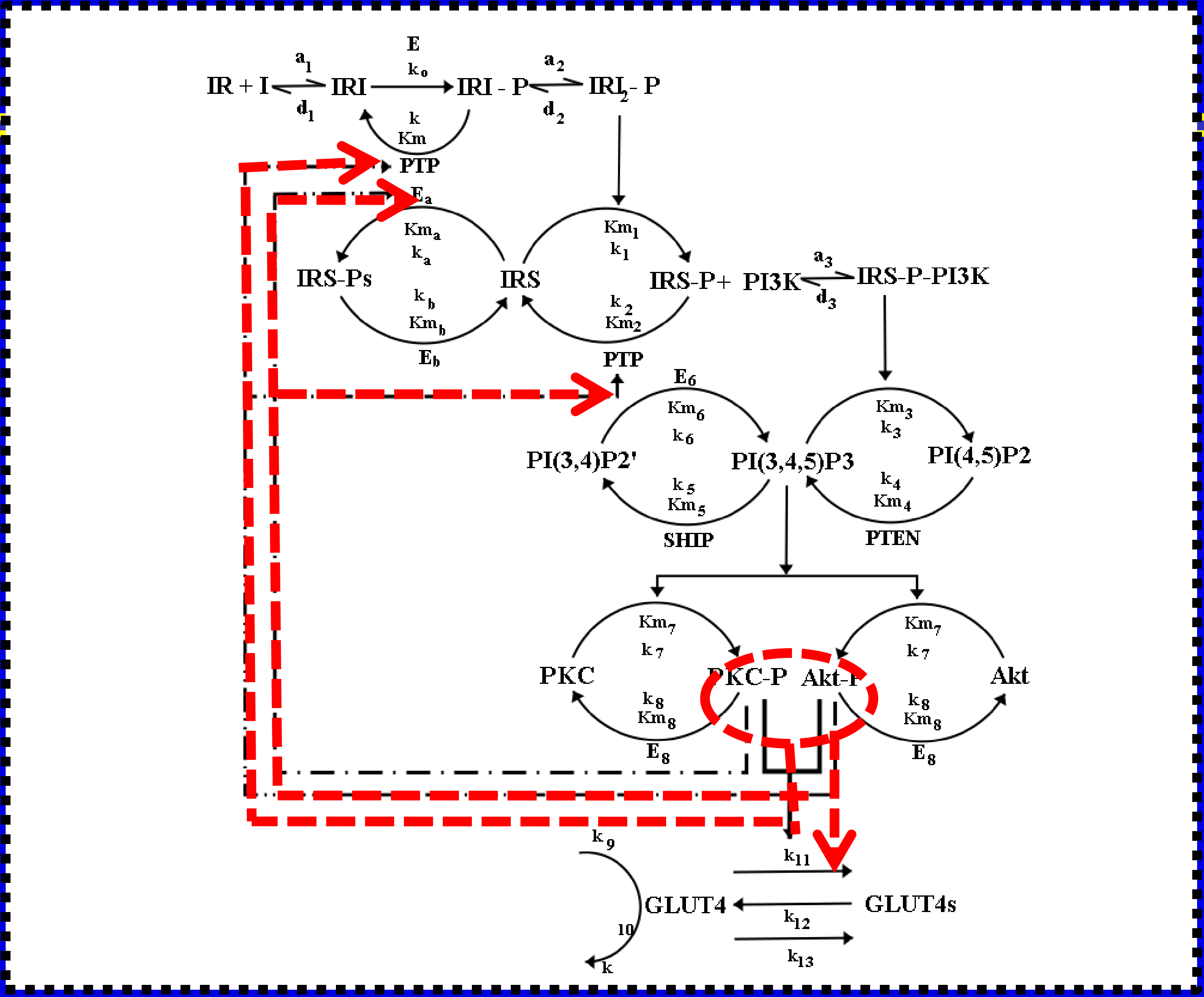
Ref. Venkatesh K V *et al*, 2004

Osmotic Stress Pathway in Yeast



Ref. Parmar *et al*, 2009

Insulin Signaling Pathway in Mammals



Are multiple feedbacks in biological systems redundant or overkill?

Ref. Freeman, 2000

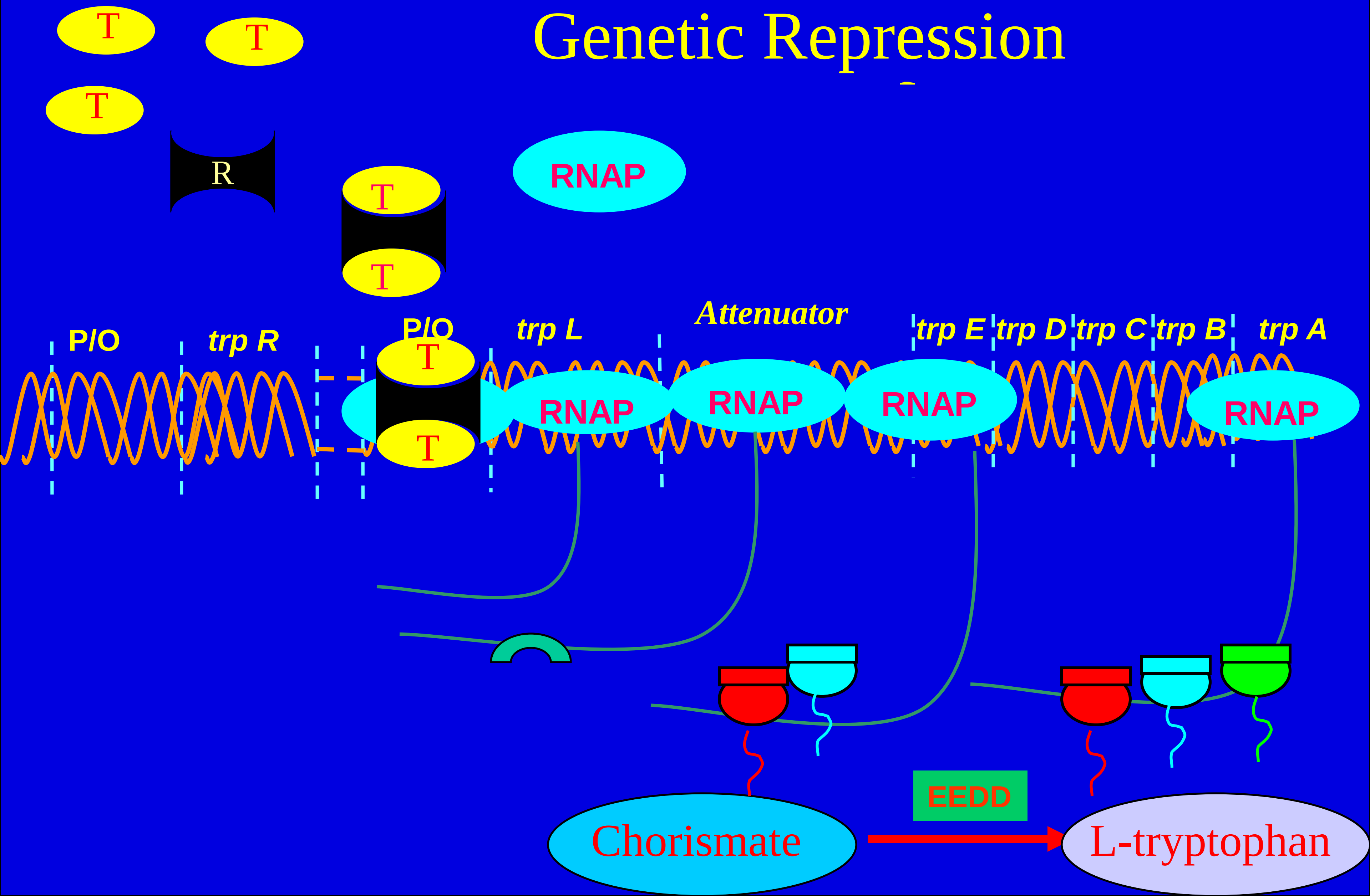
Systems Analysis of the Tryptophan System in *Escherichia coli*

Modeling and Analysis of Tryptophan System in *Escherichia coli*

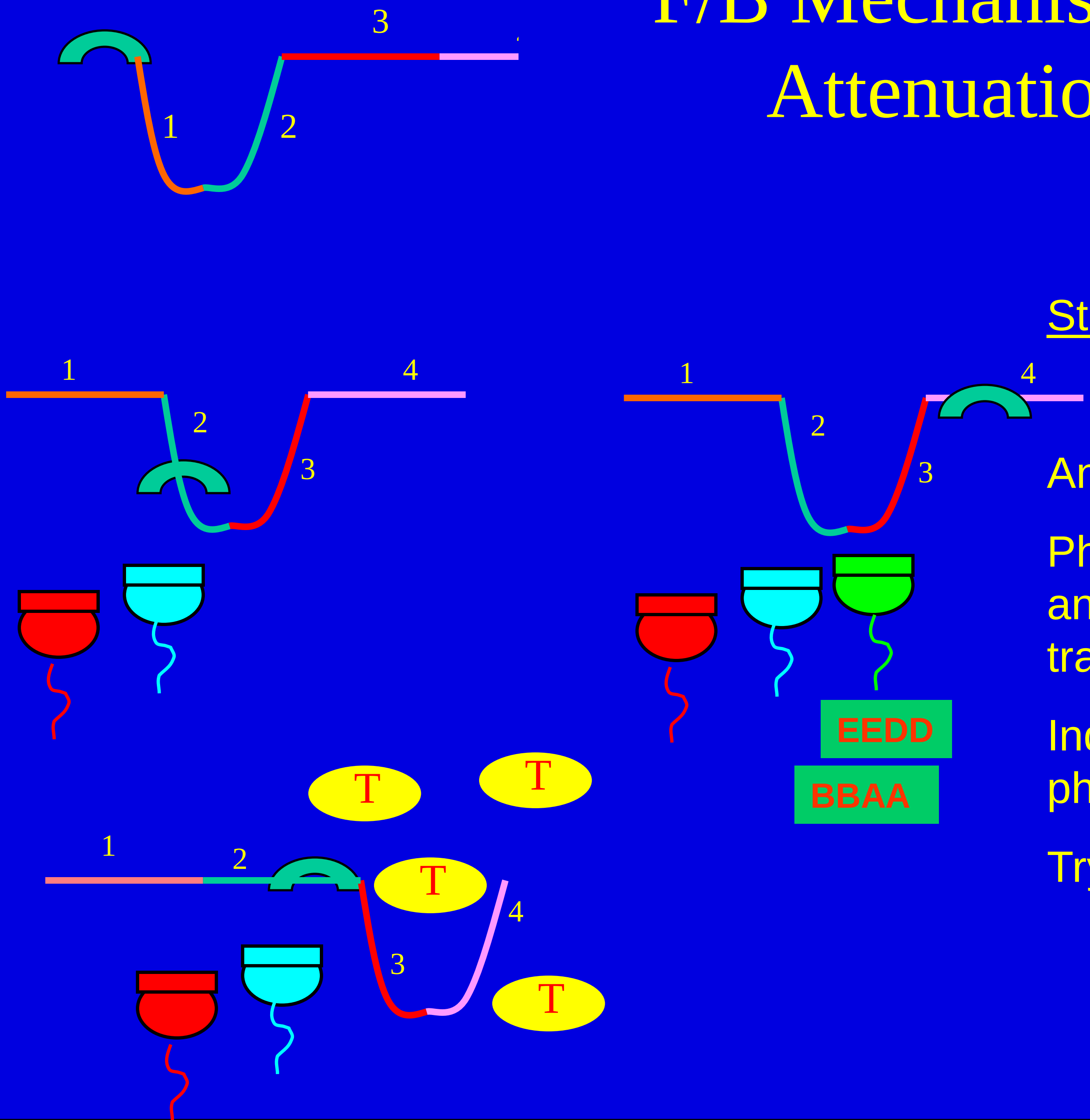
— *Yanofsky and co-workers, 1972, 1984, 1987, 2000; Xie et al., 2003*

- Goal: make tryptophan if none available in medium
- stop making tryptophan if available from medium
- Multiple feedback loop motif for autoregulation
- Widely occurring motif in biological systems (parallel cascade)
  - HOG pathway activation during osmotic shock (Hohmann, 2002)
  - Insulin signalling pathway (Sedaghat et al., 2002)
  - p53 regulation in cell cycle and apoptosis (Kohn, 1999)
  - circadian rhythms

F/B Mechanism I:
Genetic Repression



F/B Mechanism II: Attenuation



Structural Enzymes

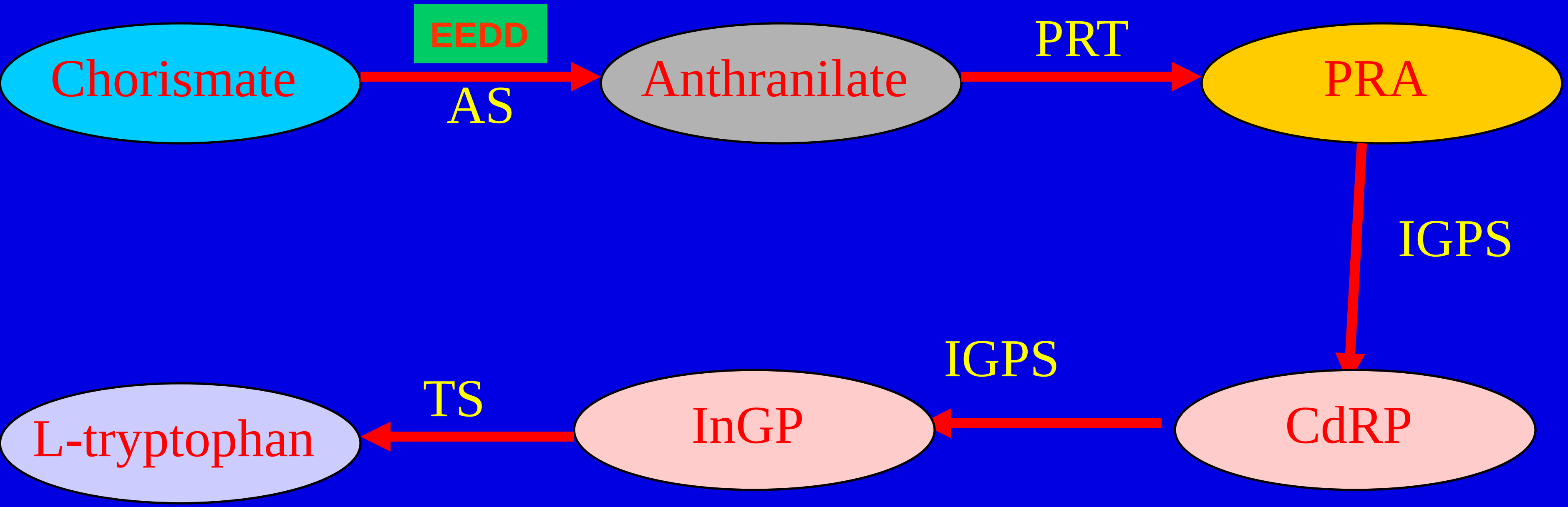
Anthranilate synthase

Phosphoribosyl anthranilate transferase

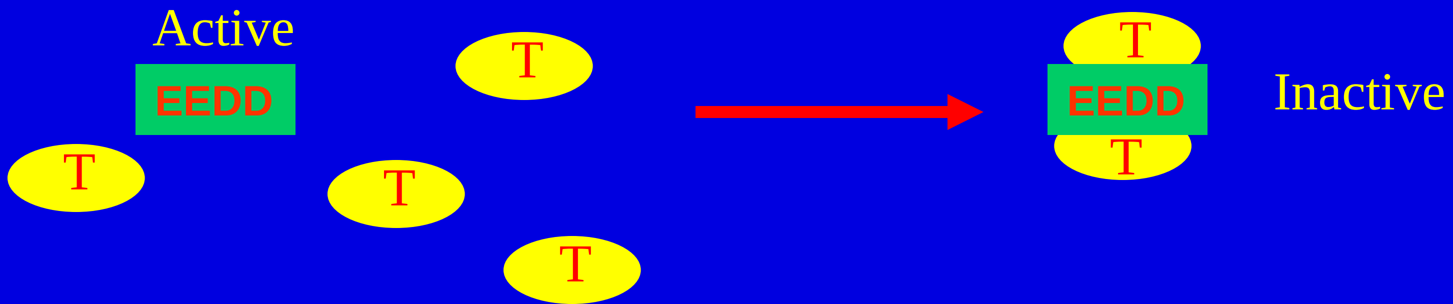
Indole glycerol phosphate synthase

Tryptophan synthase

F/B Mechanism III: Enzyme Inhibition



Chorismate → (EEDD, AS) → Anthranilate → (PRT) → PRA → (IGPS) → CdRP → (IGPS) → InGP → (TS) → L-tryptophan



Active

EEDD

T

T

T

T

→

T

EEDD

T

Inactive

Models

Bliss et al. (1982): repression and inhibition, time delays

Sinha (1988): detailed repression, tryptophan consumption constant

Sen and Liw (1990): non-constant tryptophan consumption

Santillan Mackey (2001): attenuation is modeled

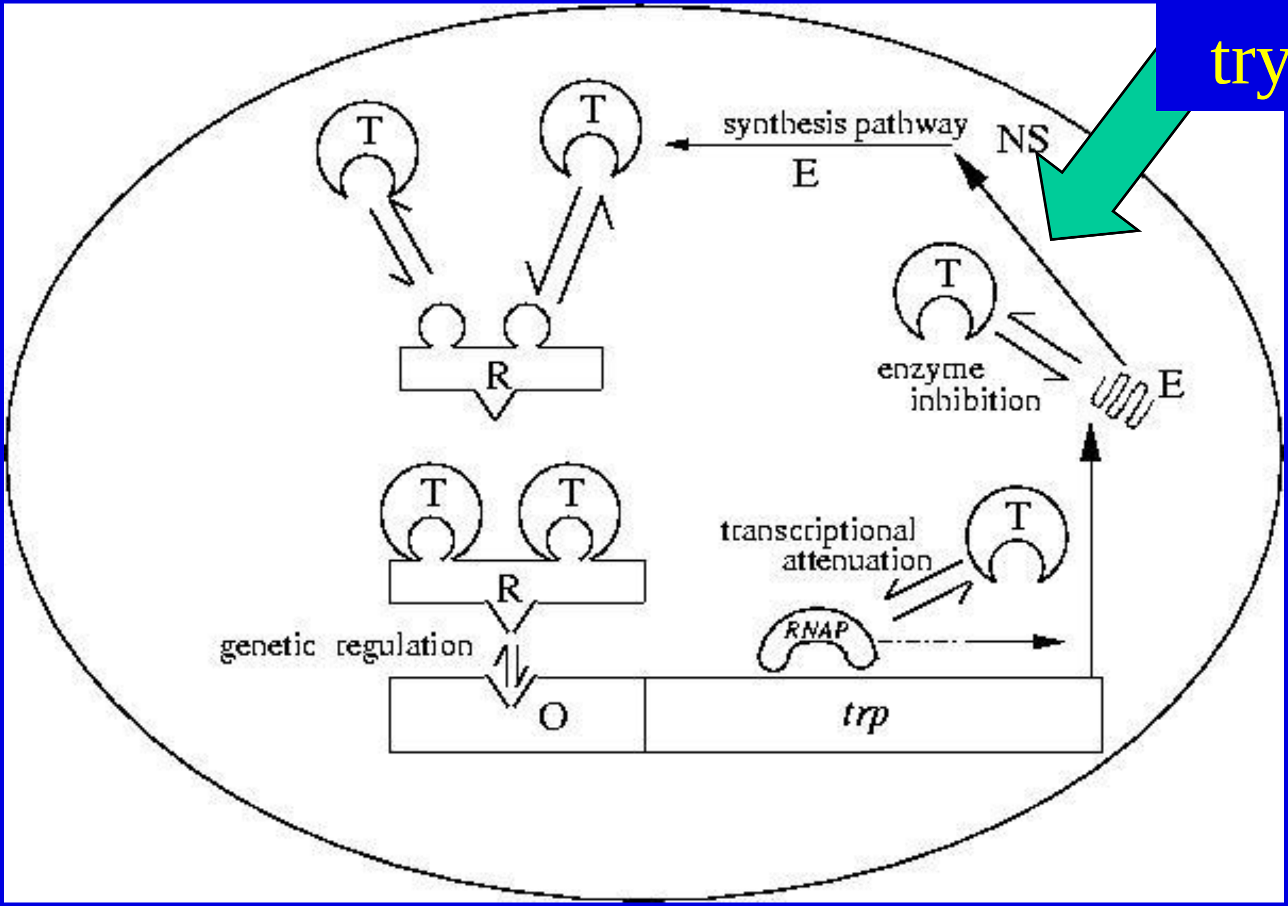
Xiu et al. (2002): repressor autoregulation dynamics

Bhartiya et al. (2003): model simplifications, attenuation not modeled

Ruhela et al. (2004): attenuation modeled

A Systems-Relevant Model for Tryptophan System

— Bhartiya, Rawool, Venkatesh, Eur. J. Biochem, **270,** 2003
— Ruhela, Bhartiya and Venkatesh, FEBS Letters, **563**, 2004



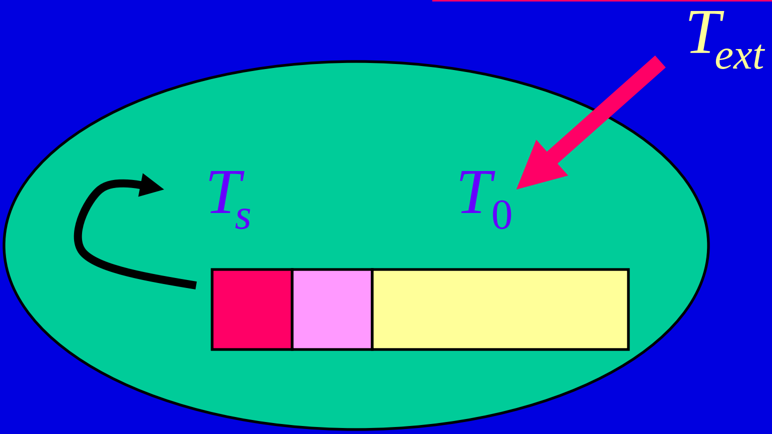
External tryptophan

Tryptophan System Model

$$T_0 = \overset{\overset{100\mu\text{M}}{\nearrow}}{T_{0,max}} \frac{T_{ext}}{T_{ext}\left(1+\frac{T_t}{f}\right)+e}$$

Total tryptophan

$$T_t = T_s + T_0$$

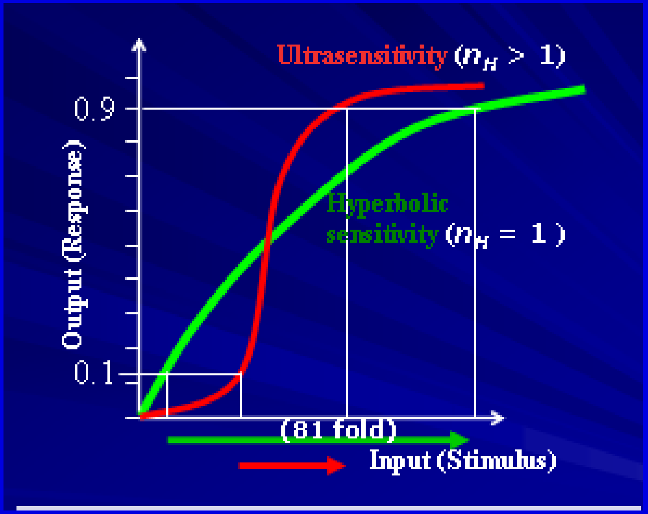


Operon Activation

$$\frac{d}{dt}O_F = k_1 O_t C_1(T_t) - k_{d,1}O_F - \mu O_F$$

Feedback Mechanism I: Genetic Repression

$$C_1(T_t) = \frac{K_1^{1.92}}{K_1^{1.92} + T_t^{1.92}}$$



Tryptophan System Model

Transcription

$$\frac{d}{dt}mRNA = k_2 O_R C_2(T_t) - k_{d,2} mRNA - \mu mRNA$$

Feedback Mechanism II: Attenuation

$$C_2(T_t) = \frac{K_1^{1.72}}{K_1^{1.72} + T_t^{1.72}}$$

Tryptophan System Model

Translation

$$\frac{d}{dt}E = k_3 mRNA - \mu E$$

Synthesis

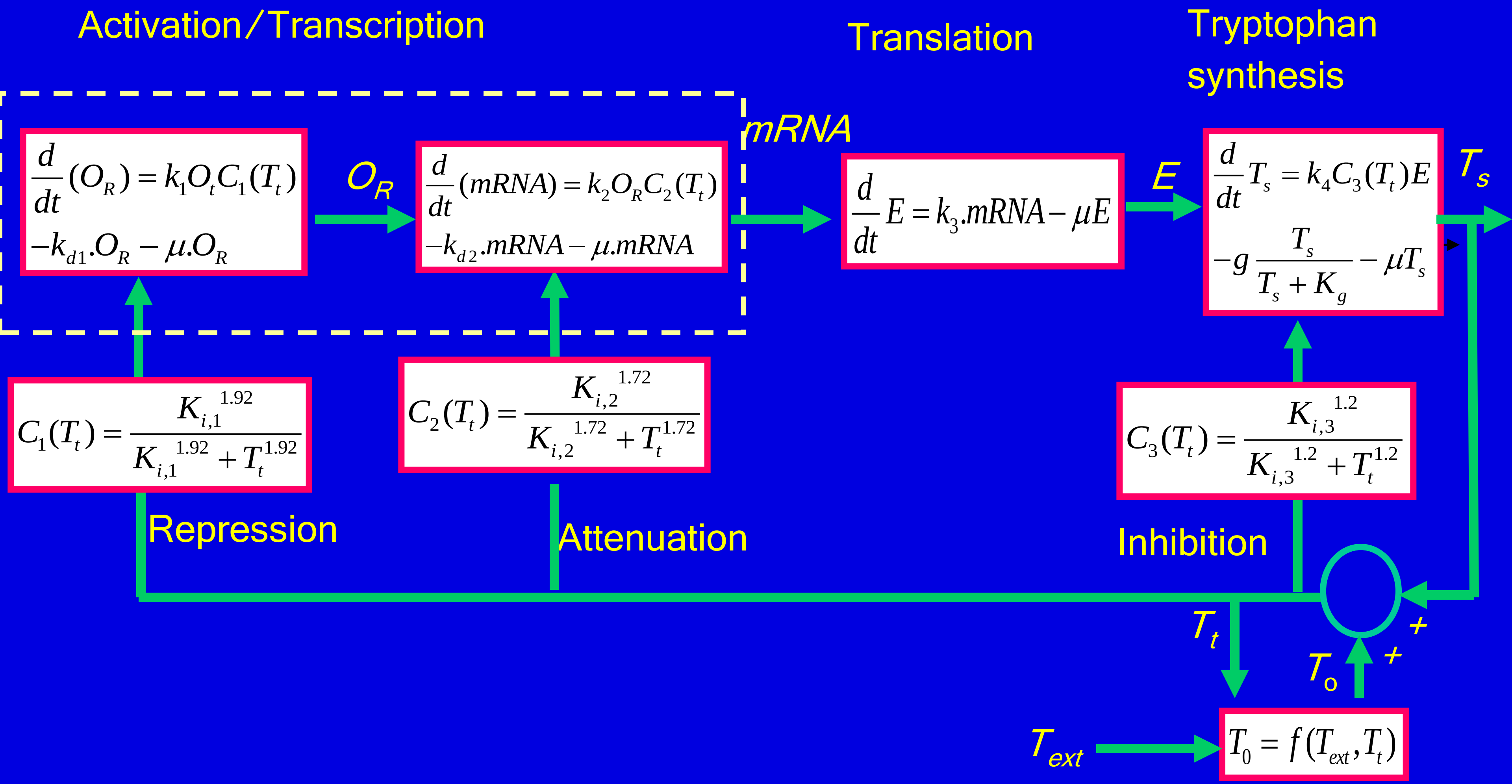
$$\frac{d}{dt}T_s = k_4 C_3(T_t)E - g\frac{T_s}{k_g + T_s} - \mu T_s$$

Feedback Mechanism
III: Enzyme Inhibition

$$C_3(T_t) = \frac{K_1^{1.2}}{K_1^{1.2} + T_t^{1.2}}$$

- Enables delineation of process and regulator
- Does not use delay differential equations

Tryptophan System in *Escherichia coli*: Regulator and Process — Venkatesh, Bhartiya and Ruhela, FEBS Letters, **563,** 2004



Activation/Transcription

$$\frac{d}{dt}(O_R) = k_1 O_t C_1(T_t) - k_{d1}.O_R - \mu.O_R$$

O_R

$$\frac{d}{dt}(mRNA) = k_2 O_R C_2(T_t) - k_{d2}.mRNA - \mu.mRNA$$

$mRNA$

Translation

$$\frac{d}{dt}E = k_3.mRNA - \mu E$$

E

Tryptophan synthesis

$$\frac{d}{dt}T_s = k_4 C_3(T_t)E - g\frac{T_s}{T_s + K_g} - \mu T_s$$

T_s

$$C_1(T_t) = \frac{K_{i,1}^{1.92}}{K_{i,1}^{1.92} + T_t^{1.92}}$$

Repression

$$C_2(T_t) = \frac{K_{i,2}^{1.72}}{K_{i,2}^{1.72} + T_t^{1.72}}$$

Attenuation

$$C_3(T_t) = \frac{K_{i,3}^{1.2}}{K_{i,3}^{1.2} + T_t^{1.2}}$$

Inhibition

T_t

T_o

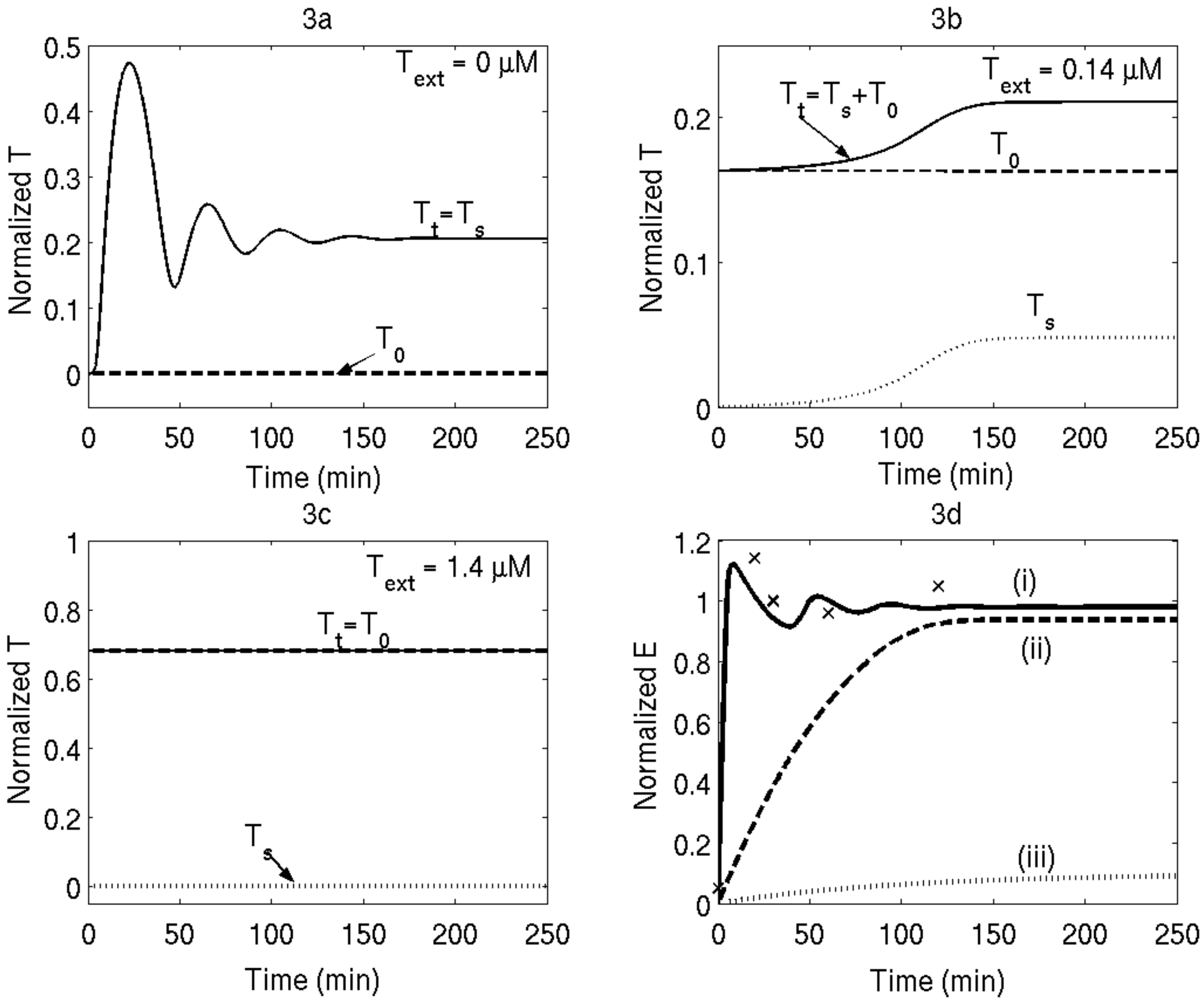
+

+

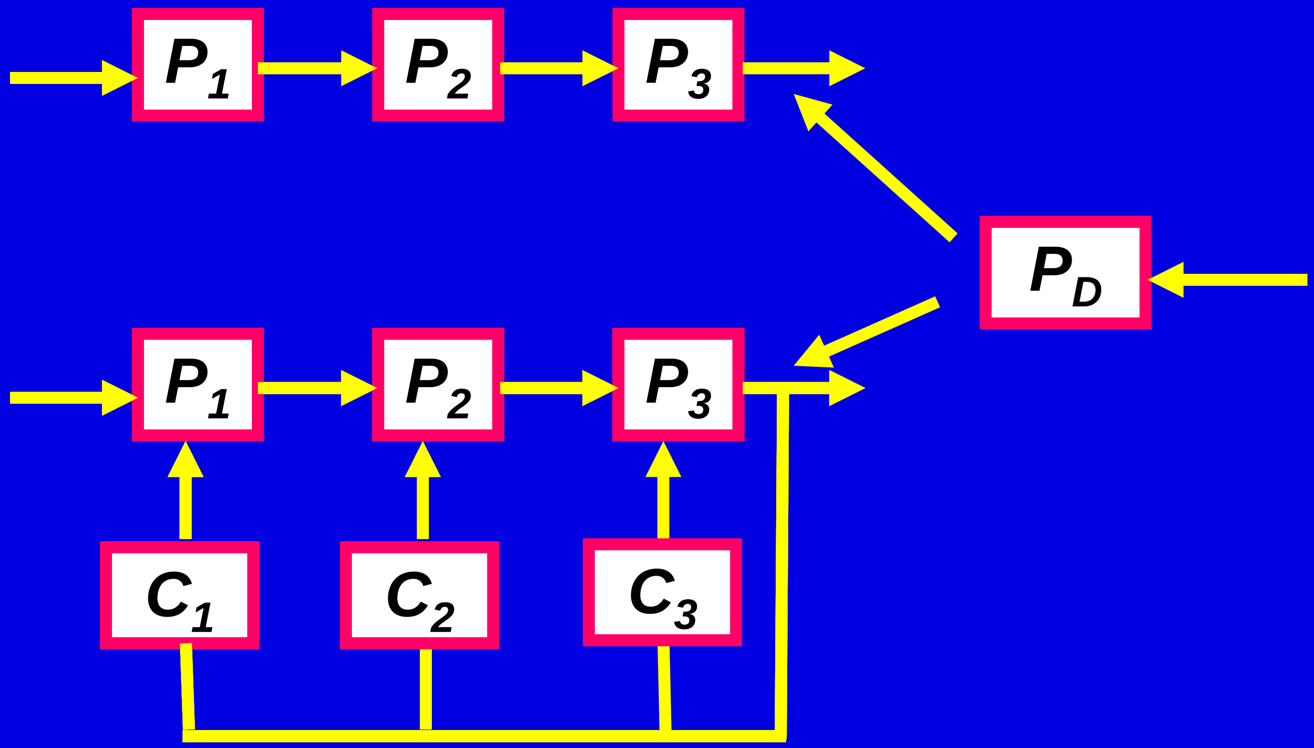
T_{ext}

$$T_0 = f(T_{ext}, T_t)$$

Model Simulation and Validation



Network Structure: Multiple feedback loops for regulation of Processes-in-series



Open Loop

Closed Loop

$C_1C_2C_3$ active : Triple feedback loop

Results

Nominal performance

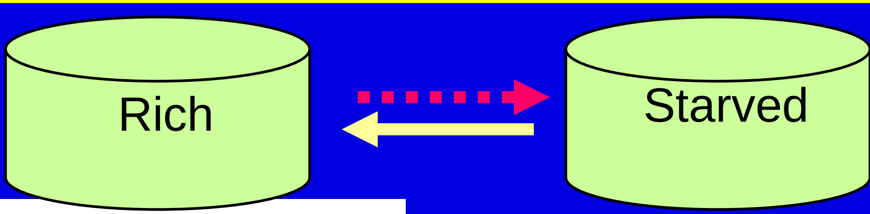
Robust performance

Frequency Response Analysis

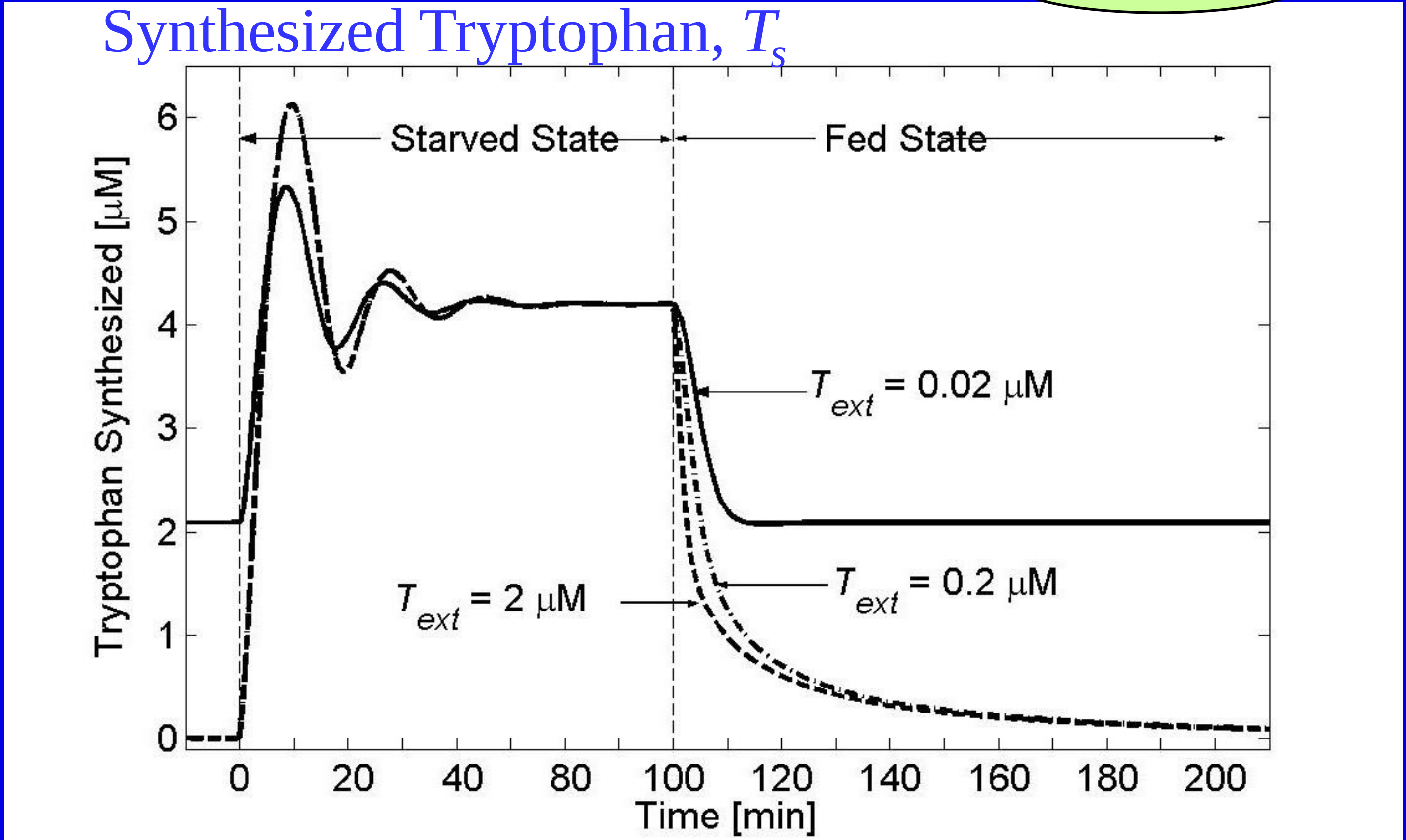
System Response Under Different Nutritional Levels

-Chaudhary, Bhartiya and Venkatesh, IET Systems Biology, **1,** 2007

Studies with Nonlinear Model


Rich

Starved


Synthesized Tryptophan, T_s

- Homeostasis

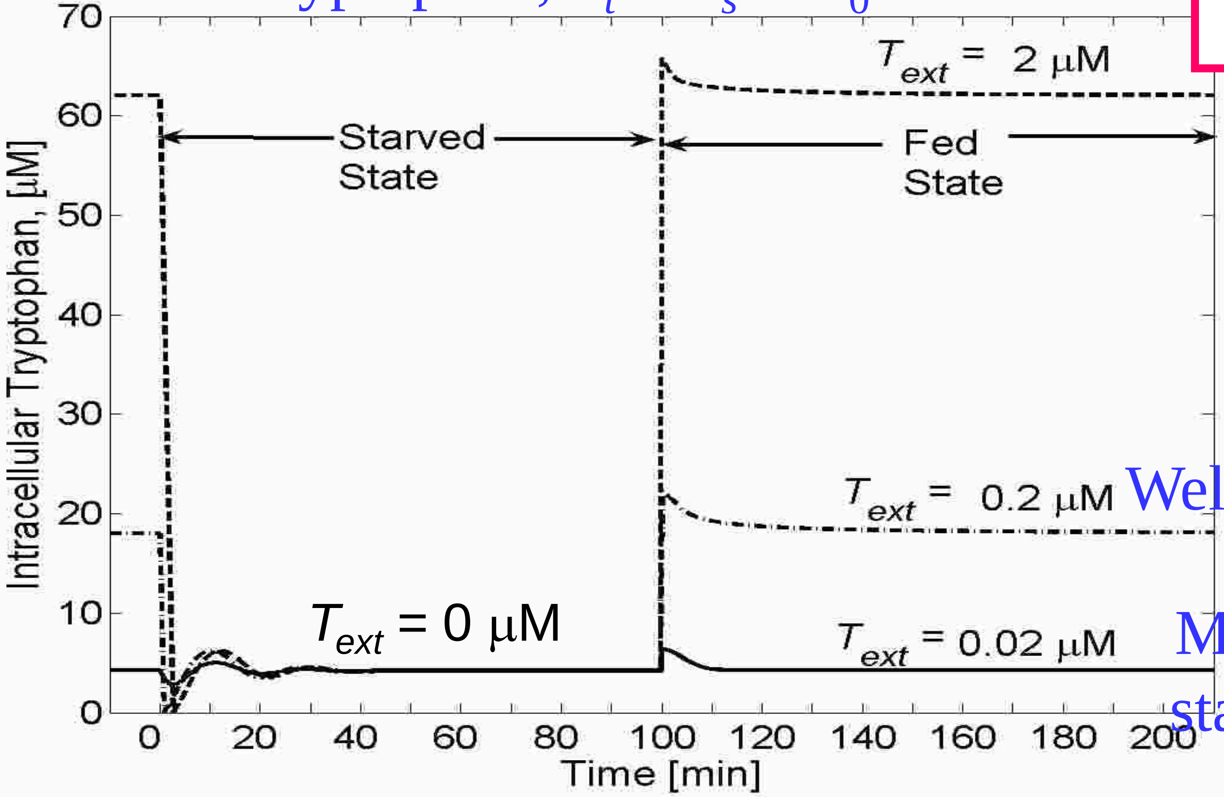
System Response Under Different Nutritional Levels

-Chaudhary, Bhartiya and Venkatesh, IET Systems Biology, **1,** 2007

Studies with Nonlinear Model

$$T_0 = \underset{100\mu\text{M}}{\underset{\nearrow}{T_{0,max}}}\,\frac{T_{ext}}{T_{ext}\left(1+\frac{T_t}{f}\right)+e}$$

Total Tryptophan, $T_t = T_s + T_0$



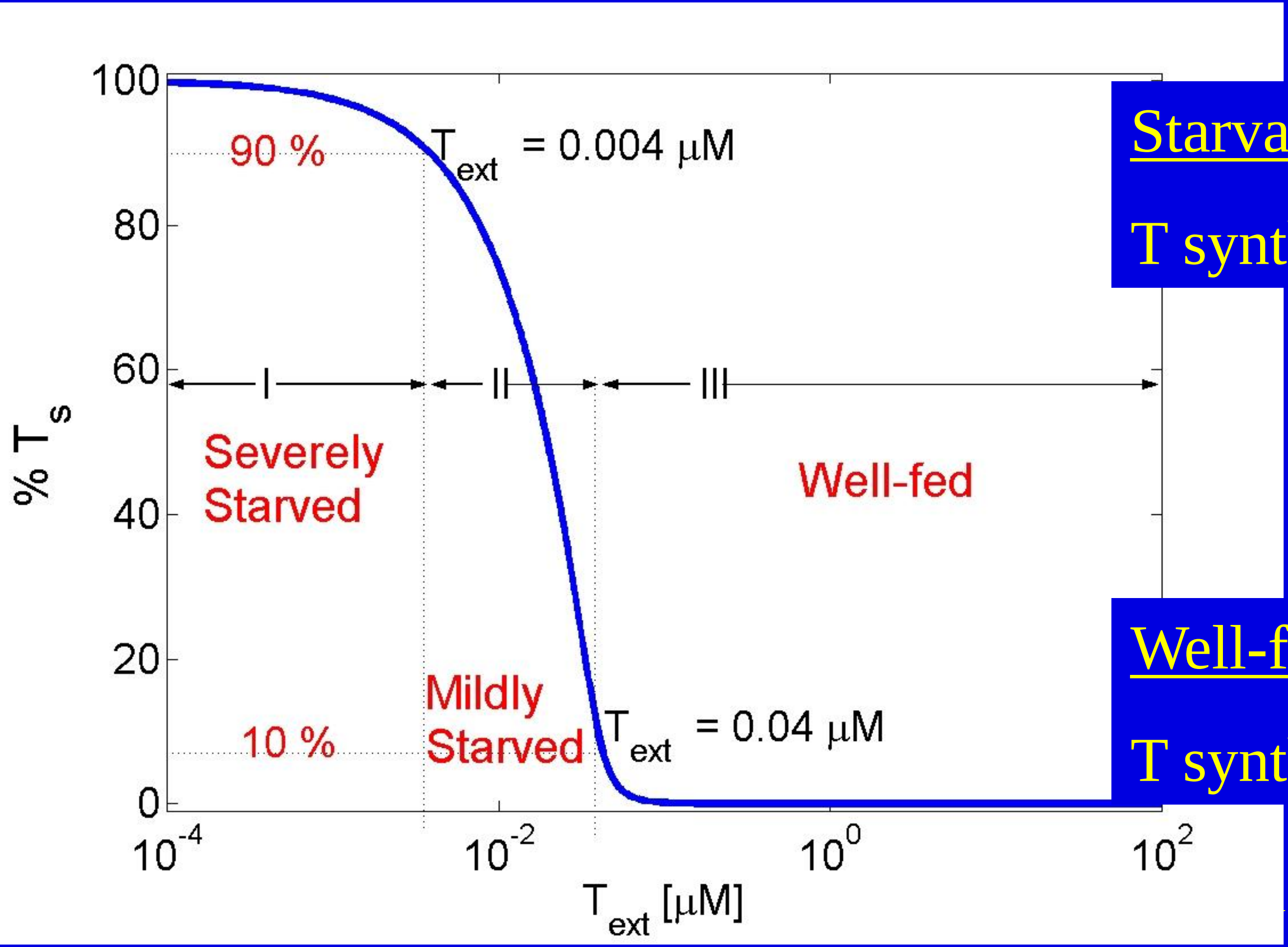
Well fed

Mildly starved

- Output disturbance

Characterization of Nutritional Status

Studies with Nonlinear Model



Starvation: $T_{ext} < 0.004$ μM

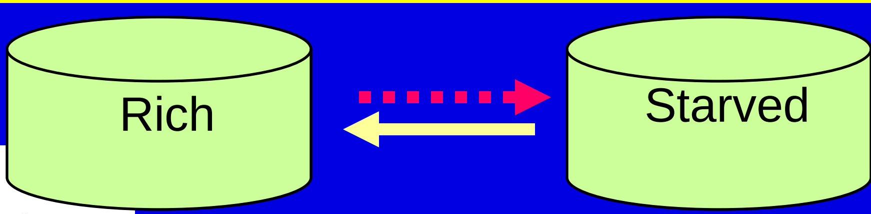
T synthesized > 90% T total

Well-fed: $T_{ext} > 0.04$ μM

T synthesized < 10% T total

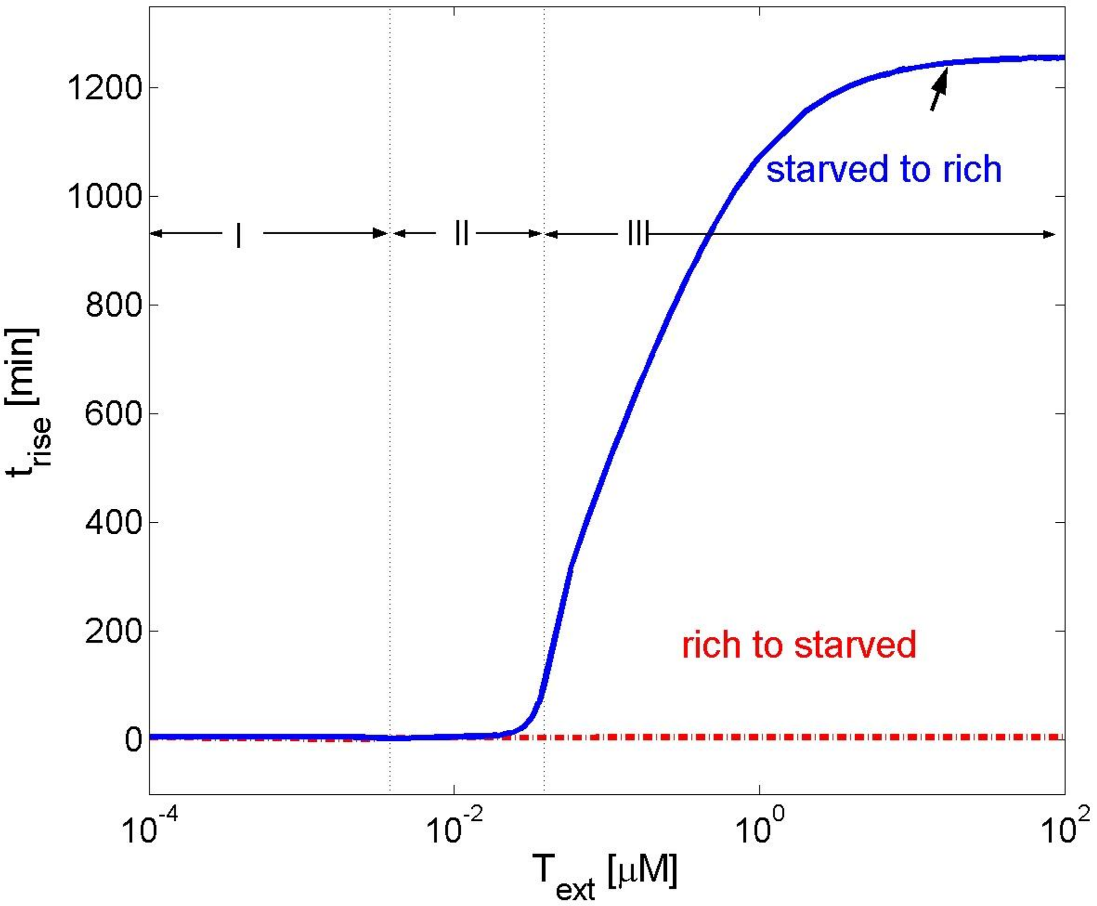
Network Goal: Robust rise time necessary for survival during starvation

Studies with Nonlinear Model



Rich

Starved



Nominal Performance

- Rapid tryptophan synthesis in severely to mildly starved conditions
- Under starvation, rise time of 5 minutes regardless of initial state
- Under well-fed conditions, sluggish shut-off of synthesis
- Identified three regions of nutrition

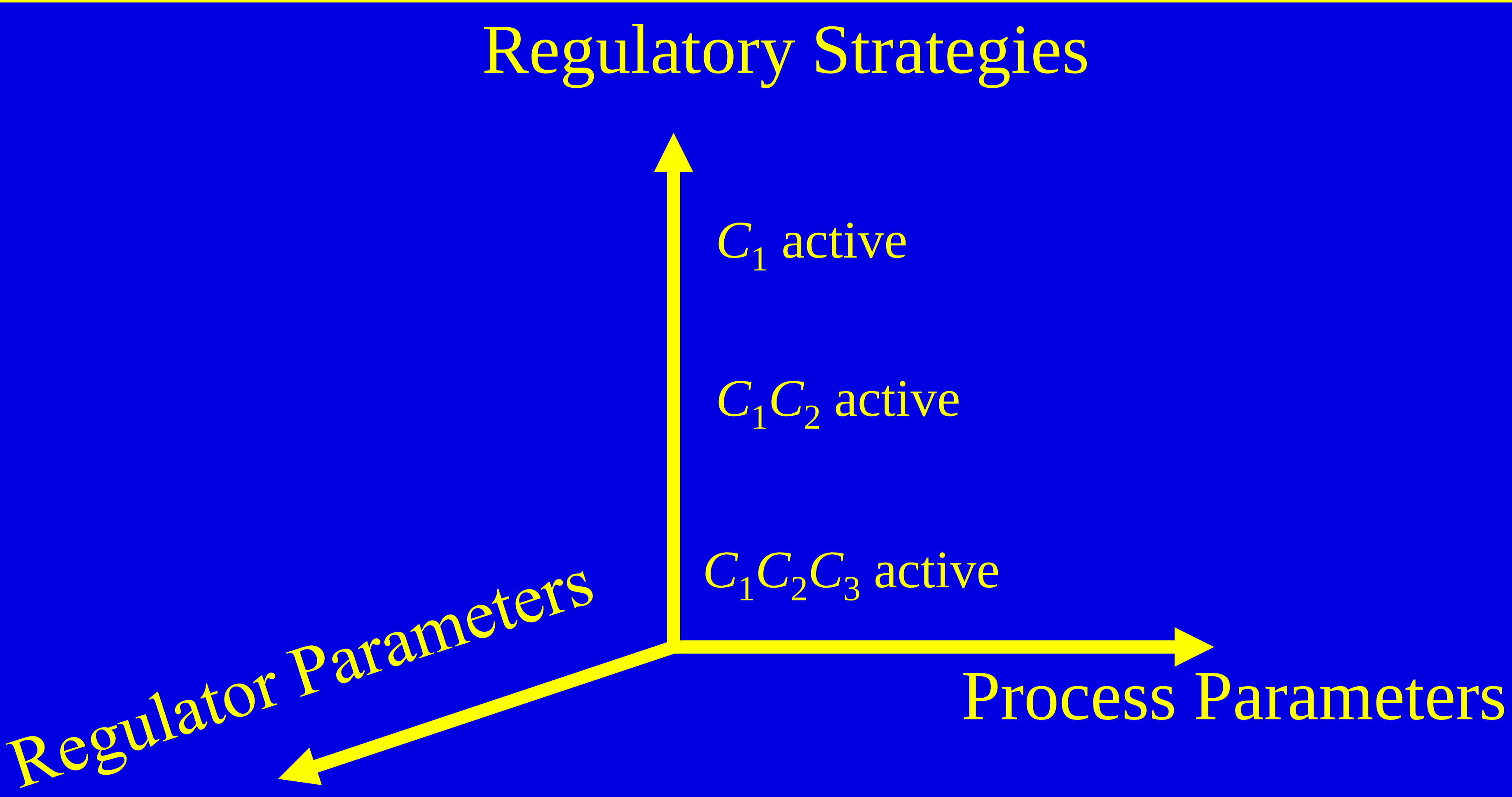
Results

Nominal performance

Robust performance

Frequency Response Analysis

Perturbations



Regulatory Strategies

C_1 active

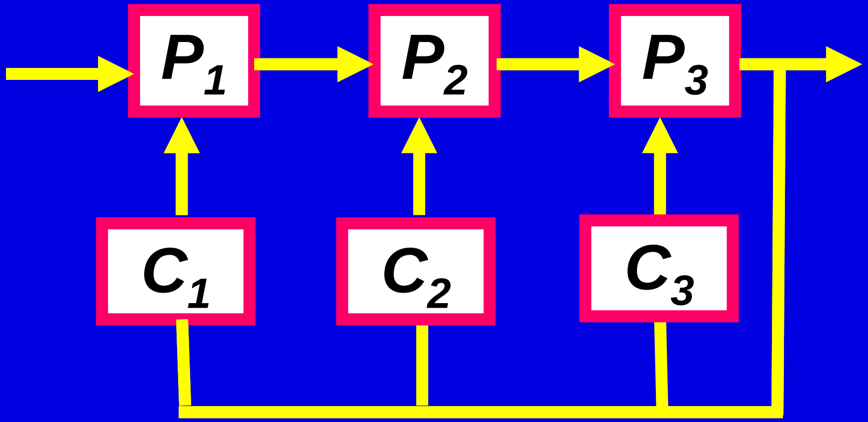
$C_1 C_2$ active

$C_1 C_2 C_3$ active

Regulator Parameters

Process Parameters

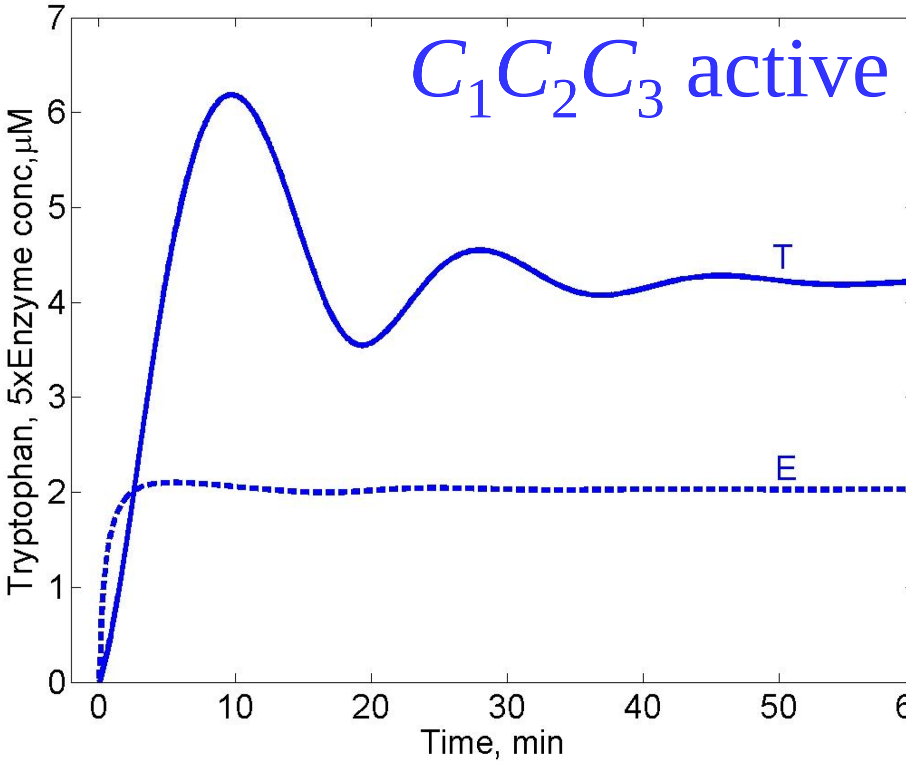
$C_1C_2C_3$ active v/s C_1 active Design

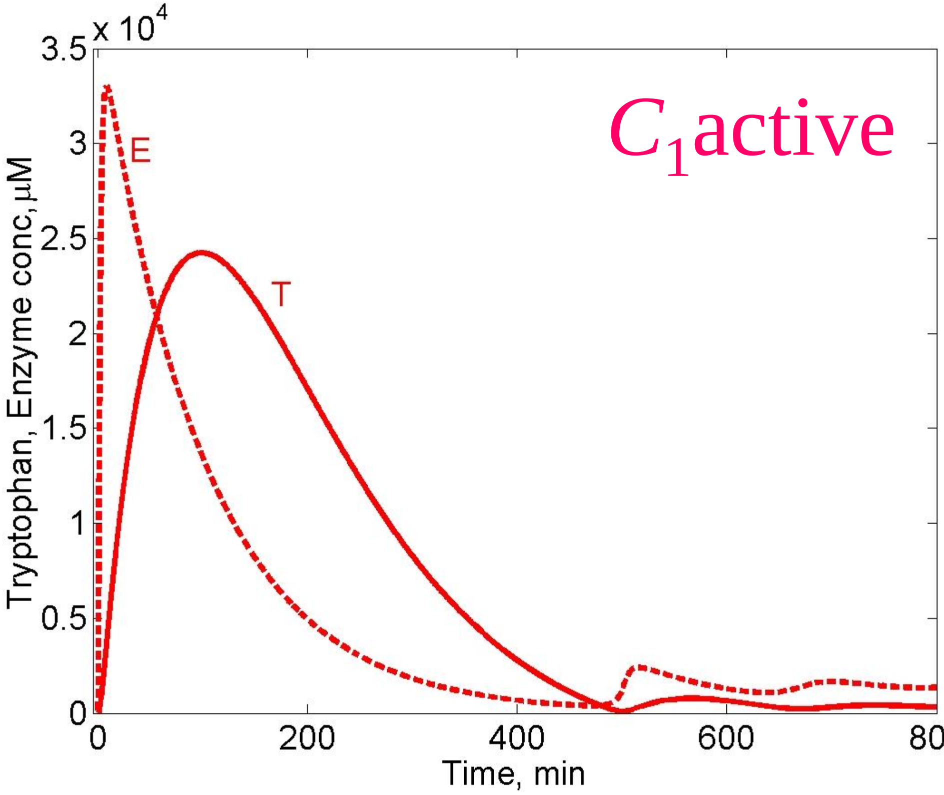


$C_1C_2C_3$ active : Triple feedback loop

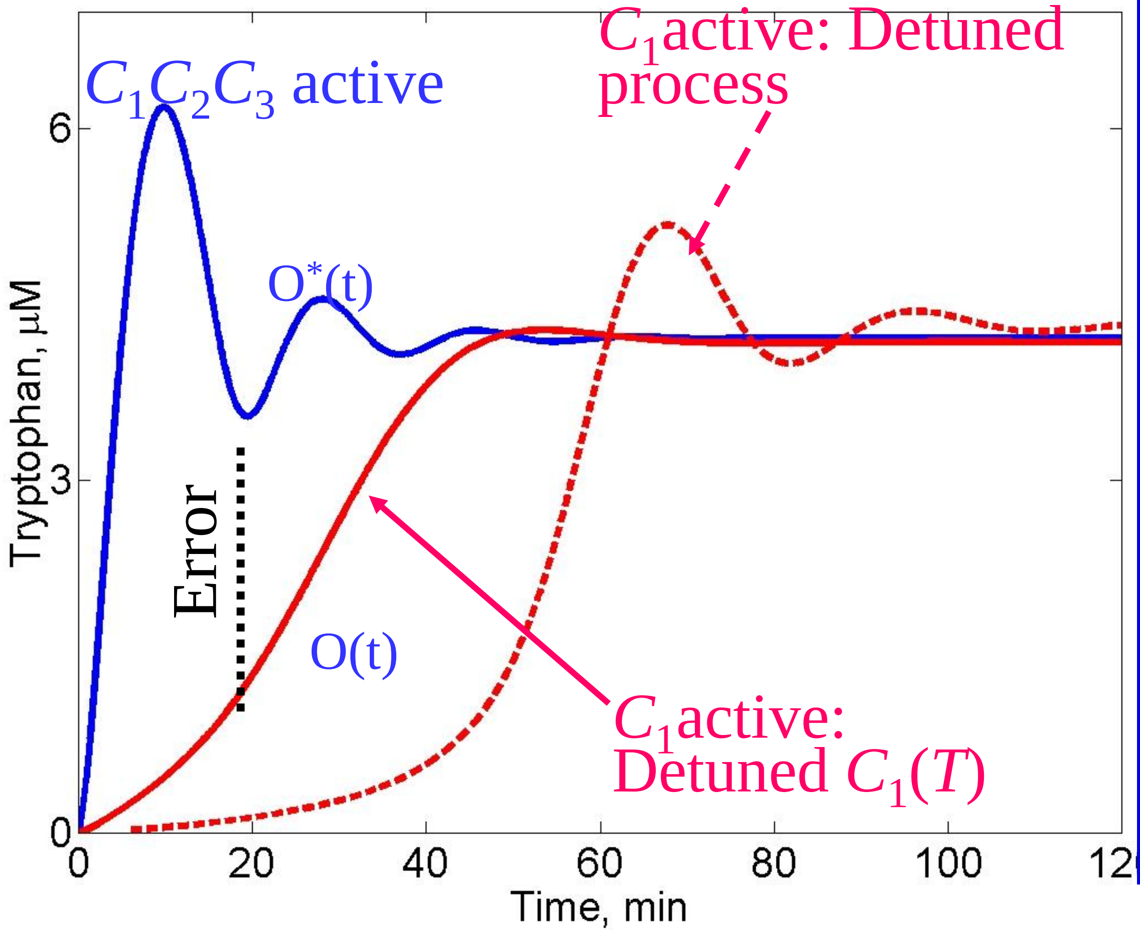
Are Multiple feedbacks loops a regulatory overkill?
(Freeman, Nature, 2003).

Starvation: $C_1C_2C_3$ active v/s C_1 active design





Starvation: Improve C_1 active mutant performance by retuning



- Retuning of single loop not sufficient to yield performance as in multiple loop design
- Multiple feedback architecture is key to meet physiological needs
- Rise time = 5 min
- Settling time = 30 min

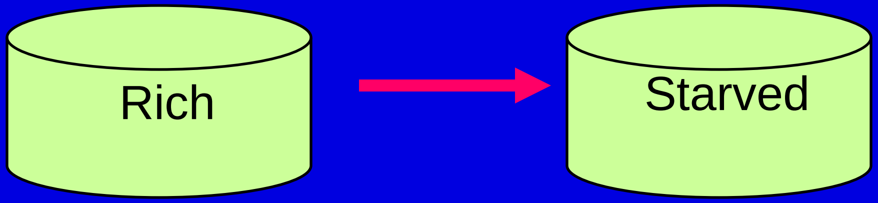
Robust Performance Metrics

Network Goals

- Rise time (time needed to first attain 5% of final value)
- Root mean square error (error relative to nominal performance)

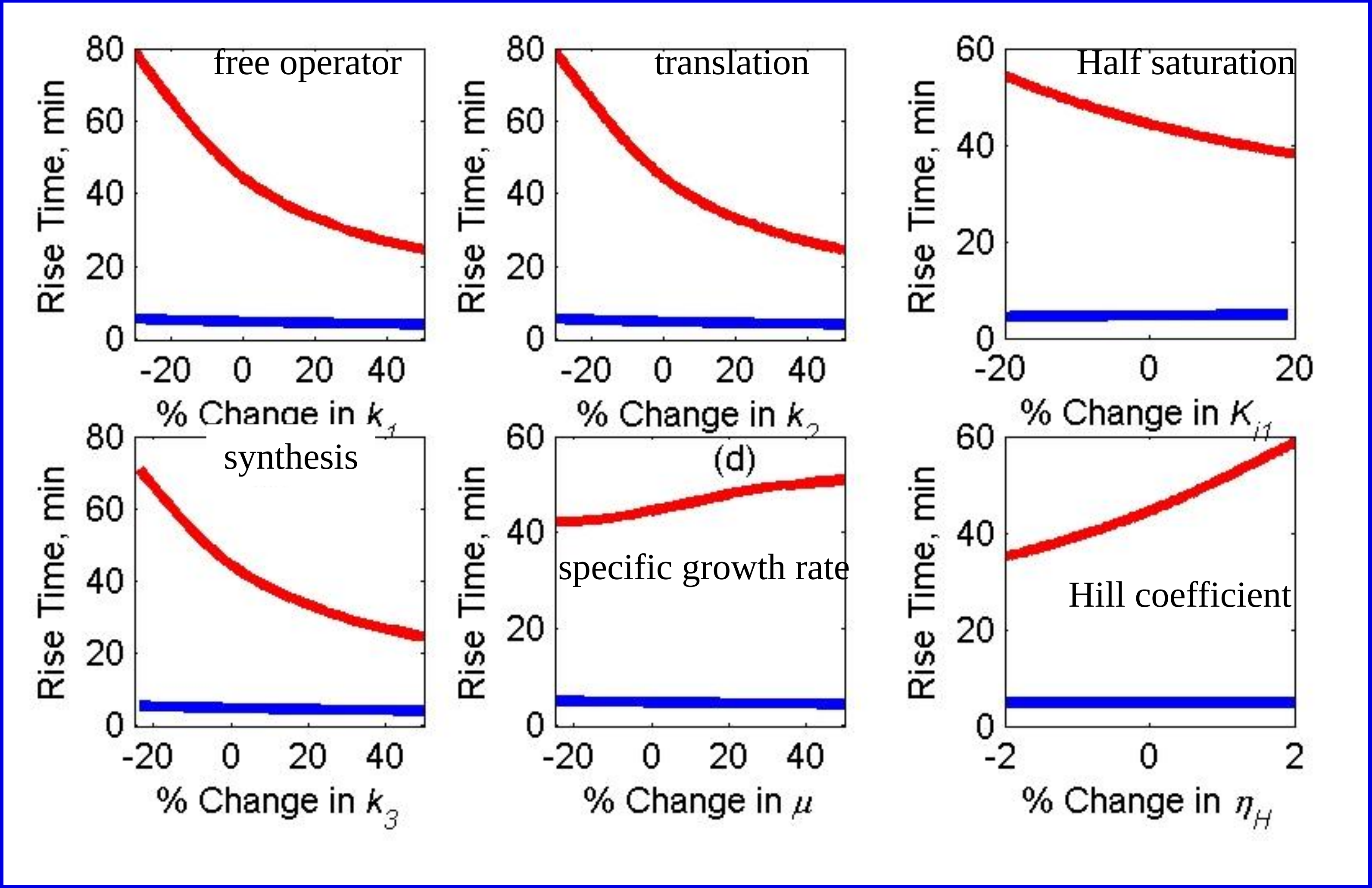
$$\boxed{I(p,s)=\frac{1}{\sqrt{t_f}}\sqrt{\int_0^{t_f}\left[O(t,p,s)-O^*(t,p^*,s^*)\right]^2 dt}}$$

- Perturb one parameter at a time (co-ordinate directions only)

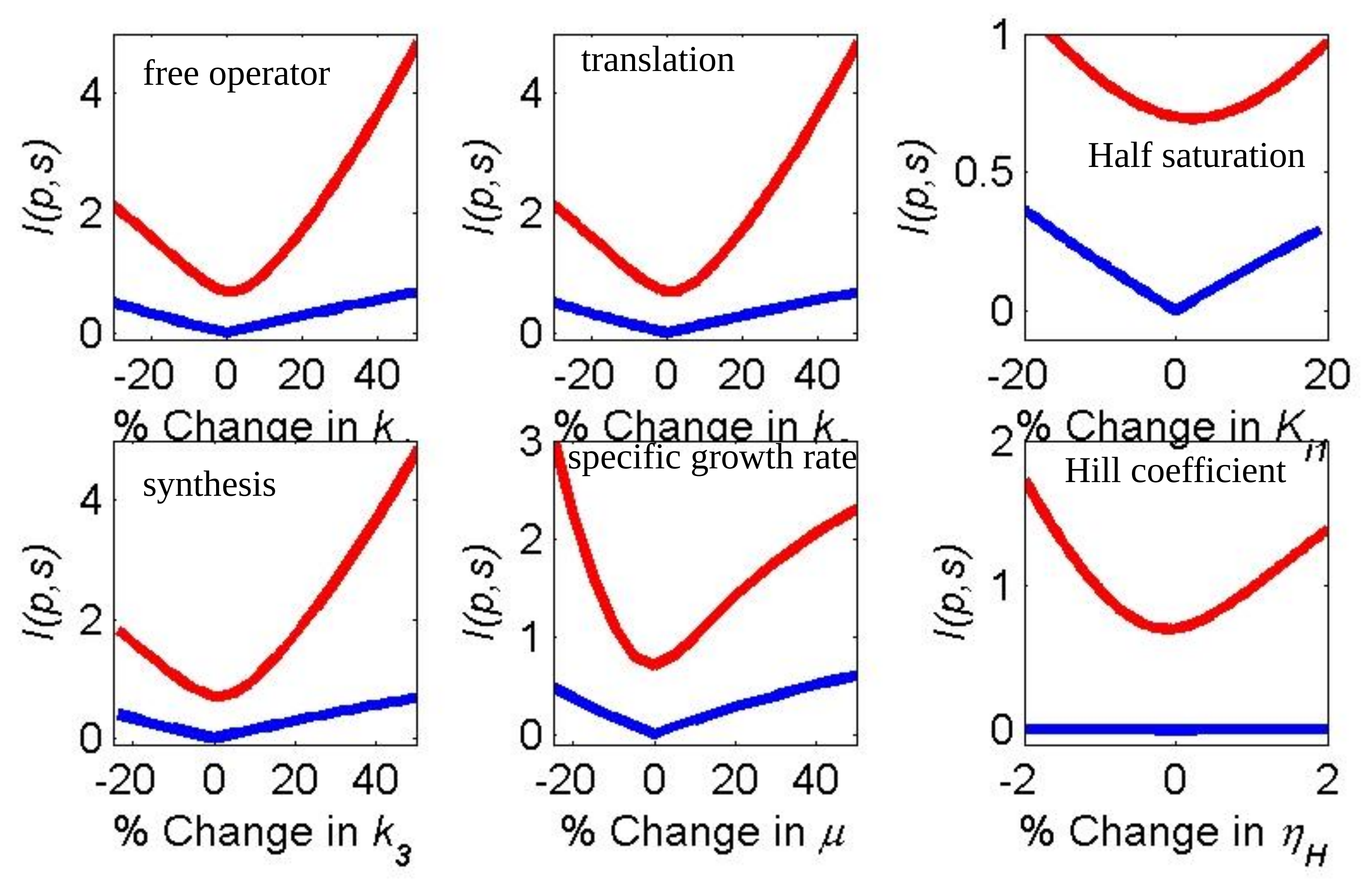


Robust Performance: $C_1C_2C_3$ active v/s C_1 active design; Metric: Rise time

Bhartiya, Chaudhary, Venkatesh and Doyle, Royal Society Interface, 2006



Robust Performance: $C_1C_2C_3$ active v/s C_1 active design; Metric: $I(p,s)$



Robust Performance: $C_1C_2C_3$ active v/s C_1 active design

- Multiple loop design yields superior dynamic performance
- Multiple loop design is robust thus making parameter values irrelevant (non-model based)
- Multiple loop design advantage for both *trp* physiological system as well as linearized system
- Robust to parameters yet fragile to structural mutations (HOT, RYF)

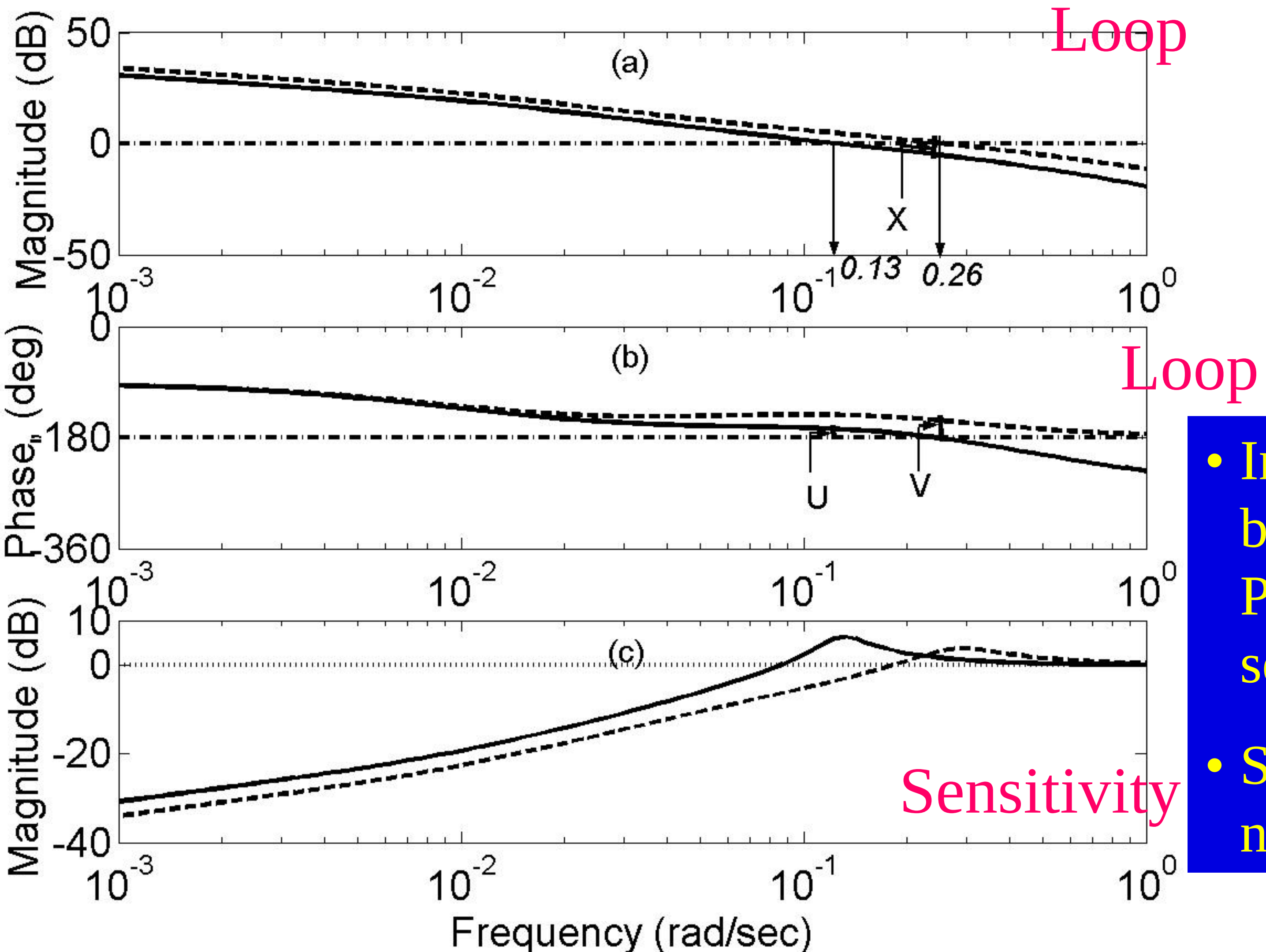
Results

Nominal performance

Robust performance

Frequency Response Analysis

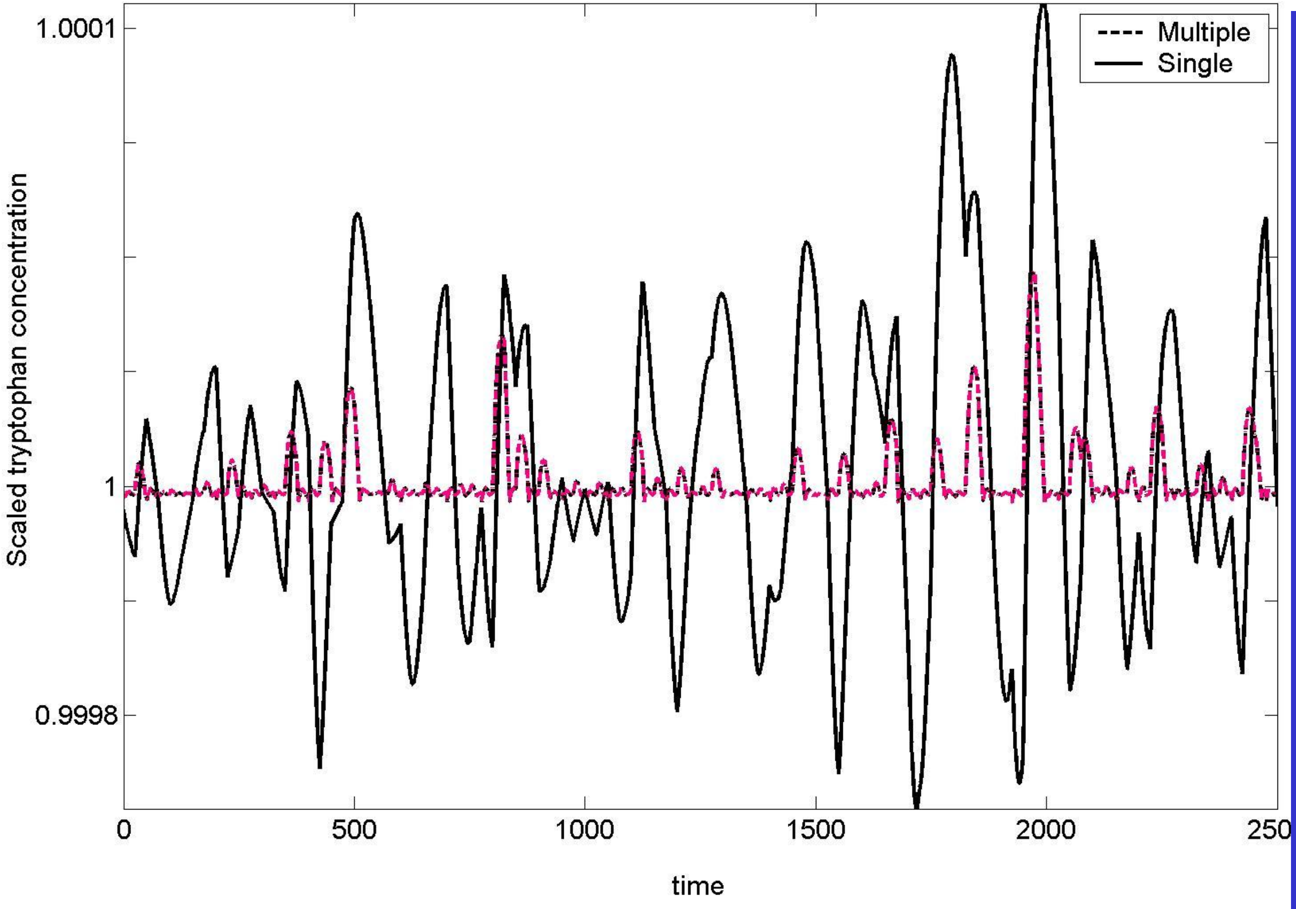
Frequency Response of Linearized *Trp* system $C_1C_2C_3$ active versus C_1 active mutant



--- multiple
— single

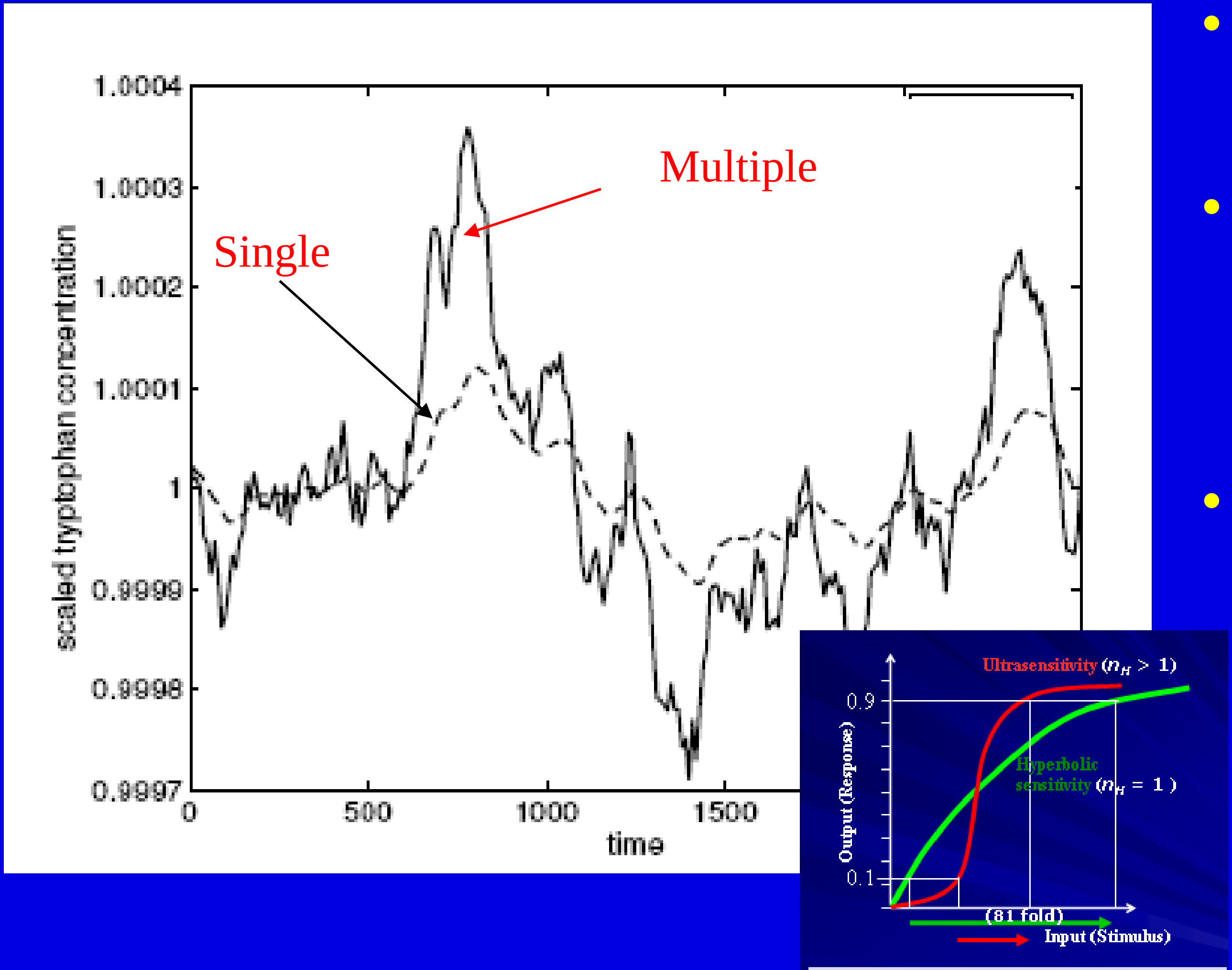
- Increased bandwidth, GM, PM, lower sensitivity peak
- Susceptibility to noise?

Noise Simulation with in vivo regulators: Injected at end of transcription- A Langevin Approach



- •Fixed step Euler
- •Single realization
- •Multiple feedback mitigates noise better! (contrary to intuition)
- •What is the role of nonlinear regulators?

Simulation with Sub-Sensitive Regulation



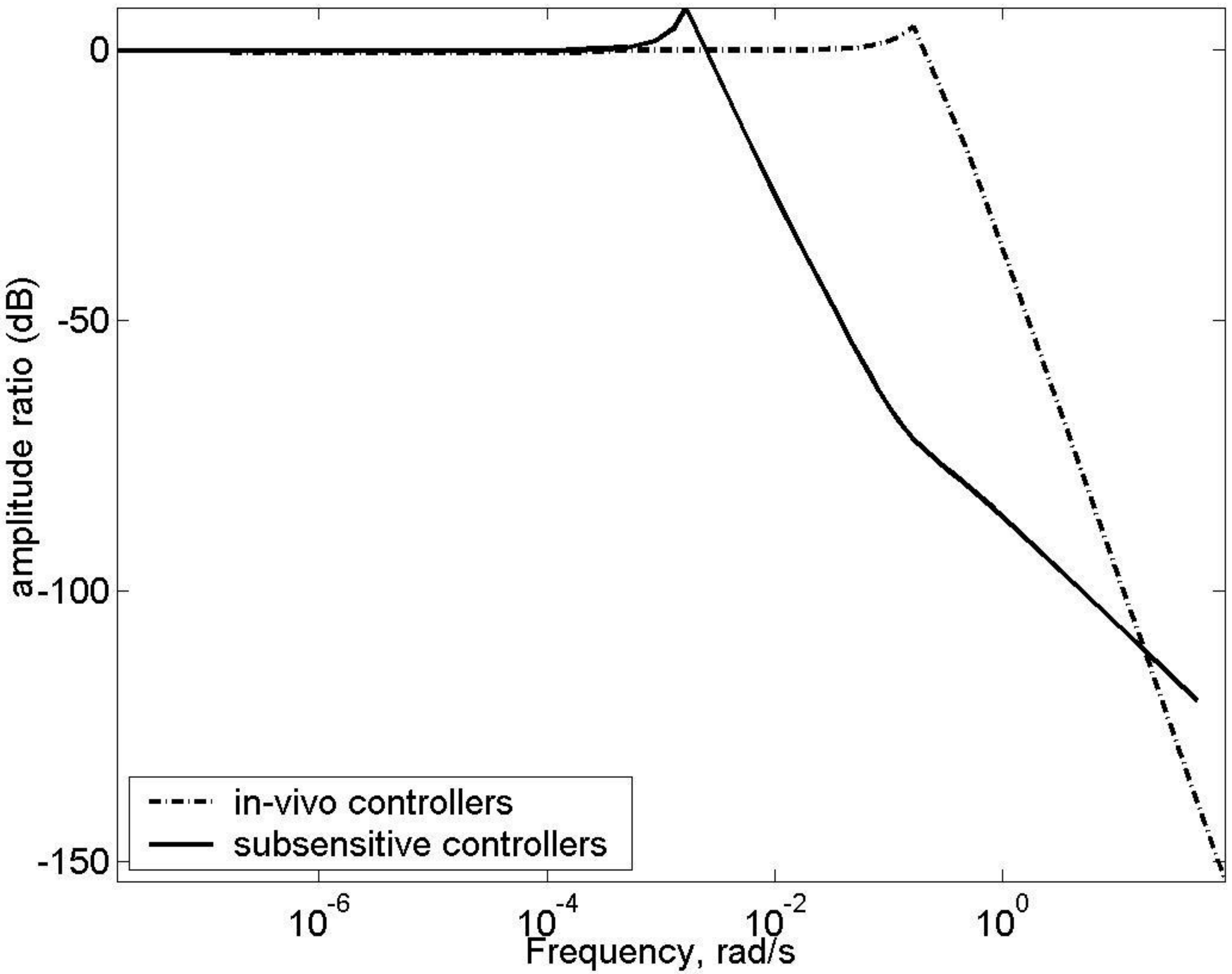
- η_H for $C_{1,}$ C_2 and C_3 is 0.5
- Multiple feedback is more noisy
- Observation: Is ultrasensitivity is responsible for noise mitigation?

Frequency Response of *Trp* System Using Describing Functions

- Quasi-linearisation: Approximation of a nonlinear system by a linear one, which depend on some properties of the input
- Describing Functions: quasi-linear approximating functions– describe the transfer characteristics of a nonlinearity
- The graphical method described by Gelb and Vander Velde (1968) used to plot frequency response
- The tryptophan system is quasi-linearised around steady state concentration T = 4.21 µM
- Since the Hill equation represents an asymmetric nonlinearity, we divide it into two regions

Role Of Ultrasensitivity in Multiple Loop Design

Bavdekar, Venkatesh and Bhartiya, AIChE Annual Meeting, Indianapolis, 2005



- Subsensitive Design: η_H for C_1, C_2 and C_3 is 0.5

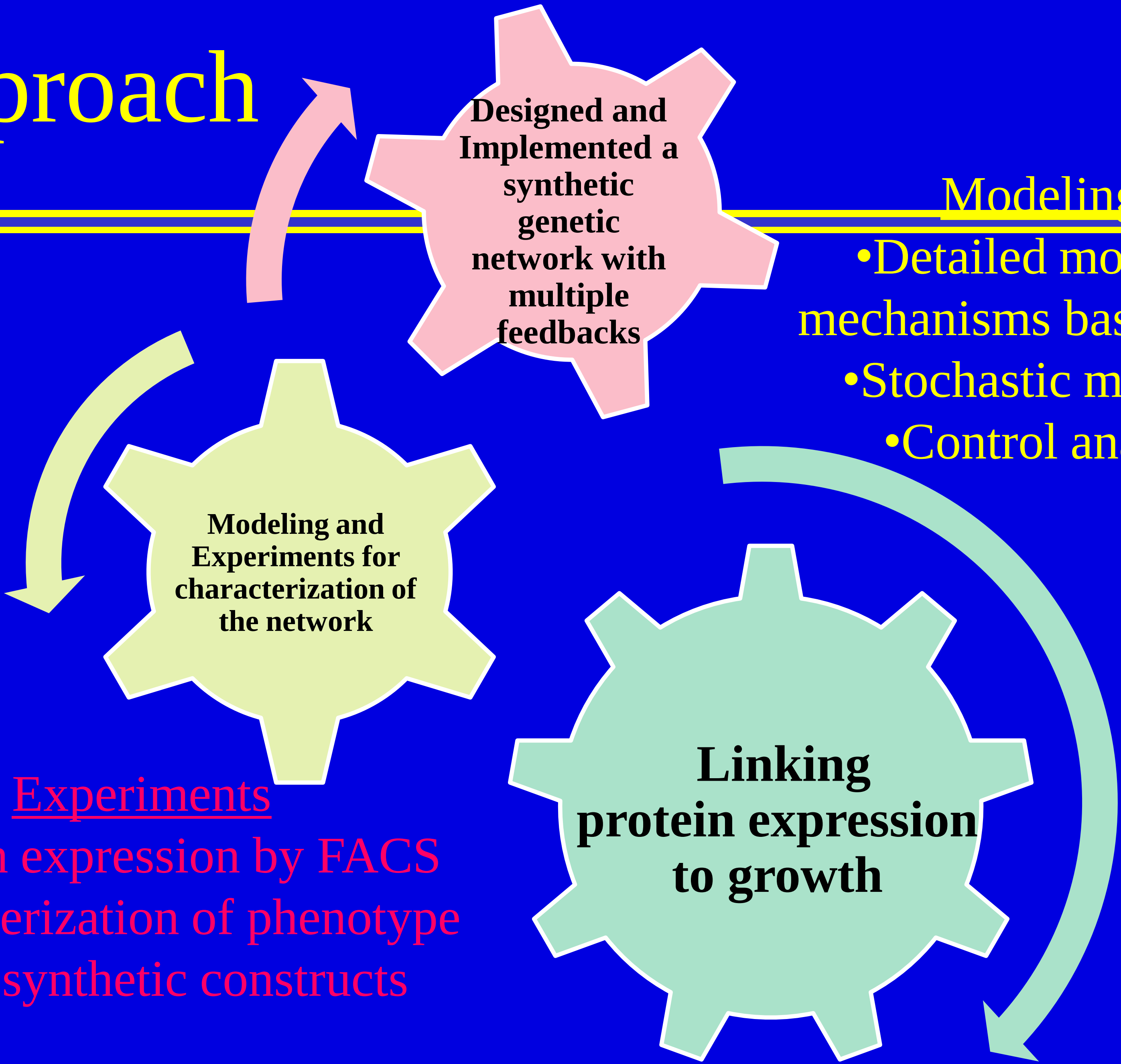
- Observation: ultrasensitivity results in higher roll-off as well as retain higher bandwidth

Conclusions

- Tryptophan System
  - Multiple feedback loops give bacterium a niche for survival during severe starvation
  - Nonlinear regulators counter the effect of fluctuations in nutritional environment
  - A prototype for analysis of naturally evolved systems

Implementation of multiple feed back loop strategy in a synthetic network

Approach



Designed and Implemented a synthetic genetic network with multiple feedbacks

Modeling and Experiments for characterization of the network

Linking protein expression to growth

Modeling –

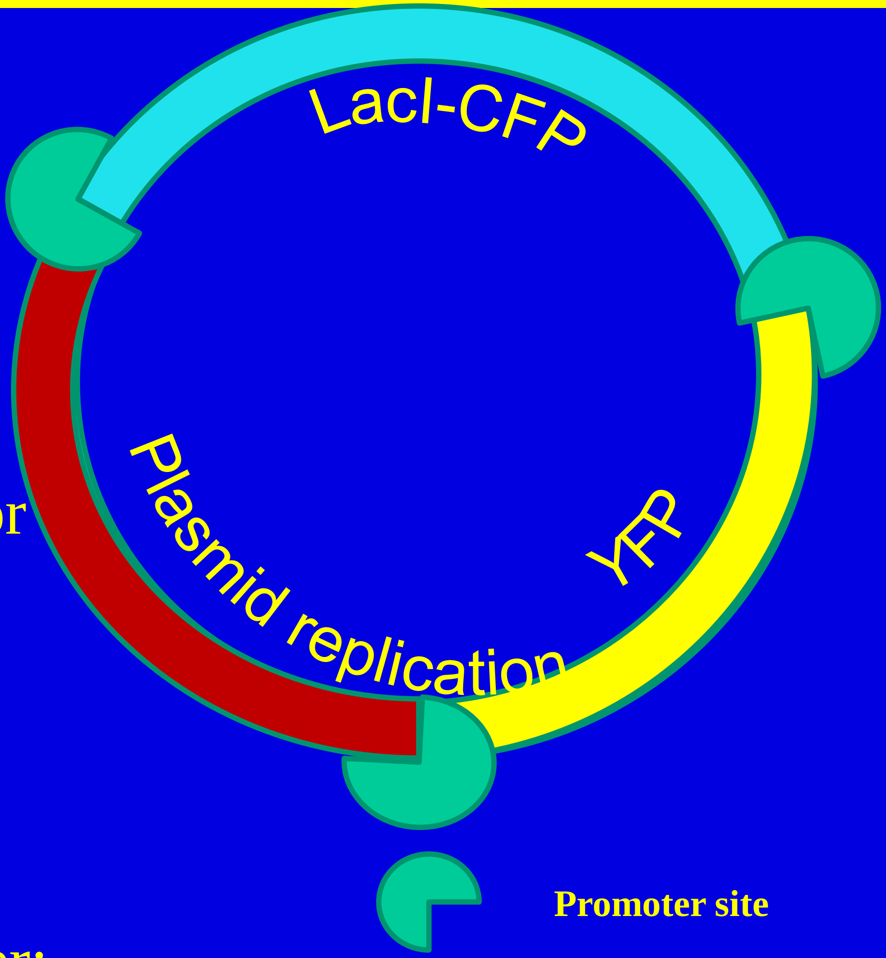
- Detailed molecular mechanisms based model
- Stochastic modeling
- Control analysis

Experiments

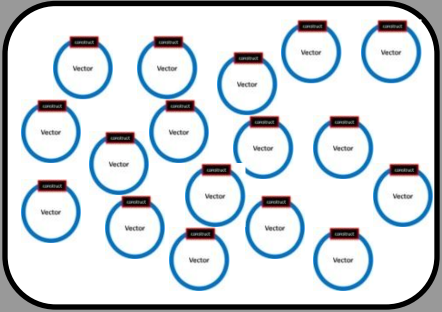
- Protein expression by FACS
- Characterization of phenotype in the synthetic constructs

Components of Synthetic Constructs

- Use of existing bio-bricks
- Four promoter sites used for the constructs: **pTet, pLac, pMB1** and **pLacOP** .
- **pMB1 and pLacOP** : promoters for plasmid replication.
- To characterize amount of LacI: **LacI-CFP fusion protein.**
- To characterize plasmid copy number: **YFP expression.**

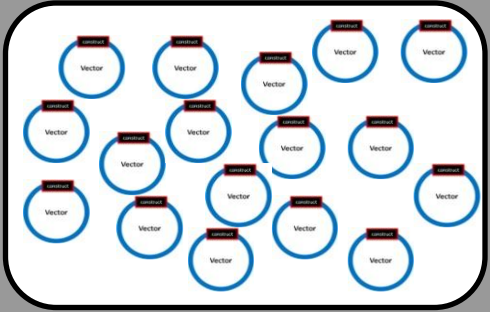


Characteristics of promoters used for Plasmid Replication

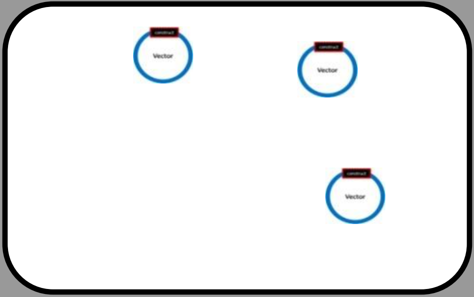


On addition of IPTG

Plasmid copy number does not change

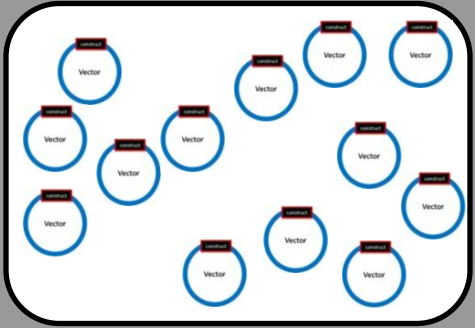


pMB1



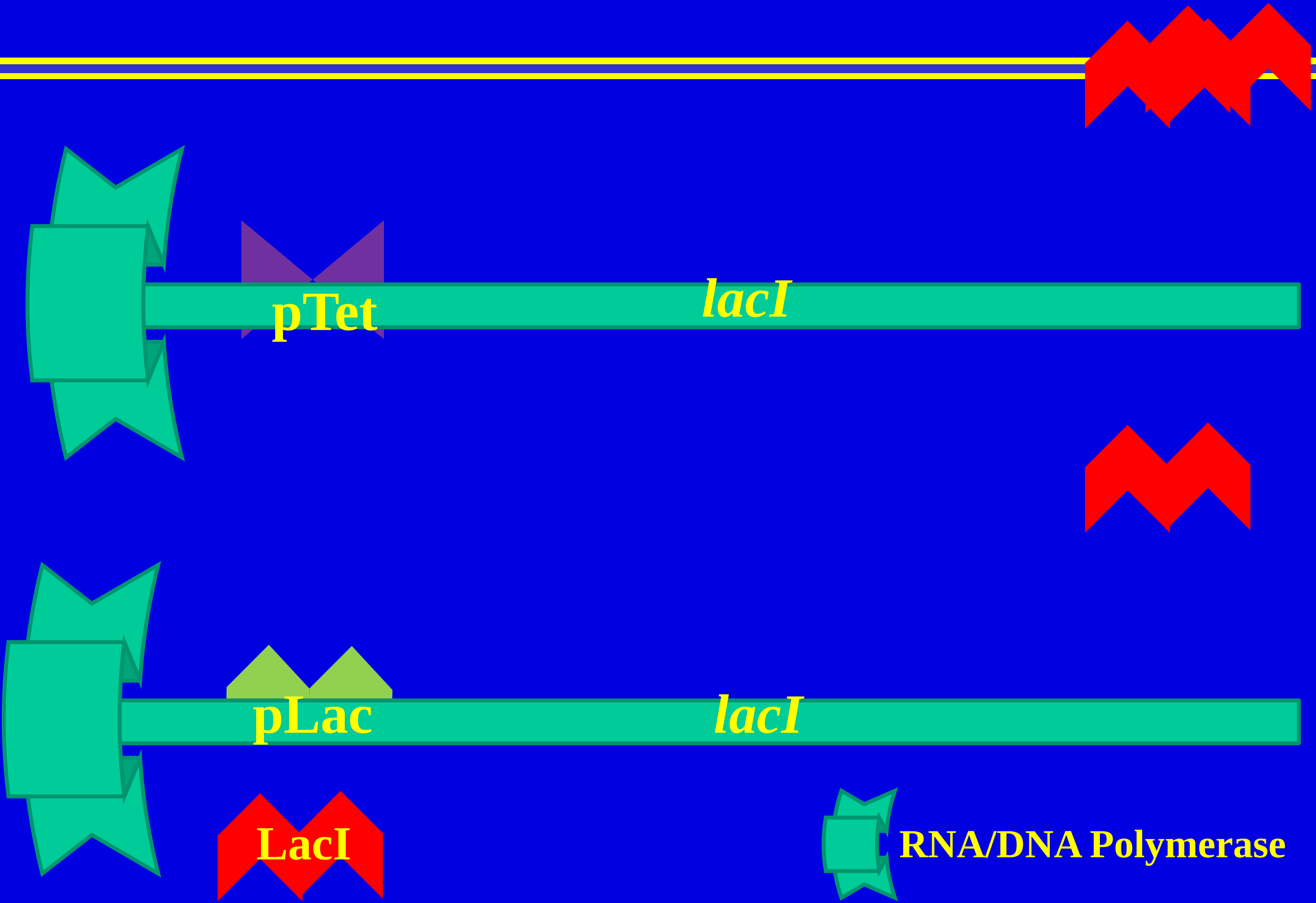
On addition of IPTG

Plasmid copy number increases



pLacOP

LacI regulation in pTet and pLac



pTet

lacI

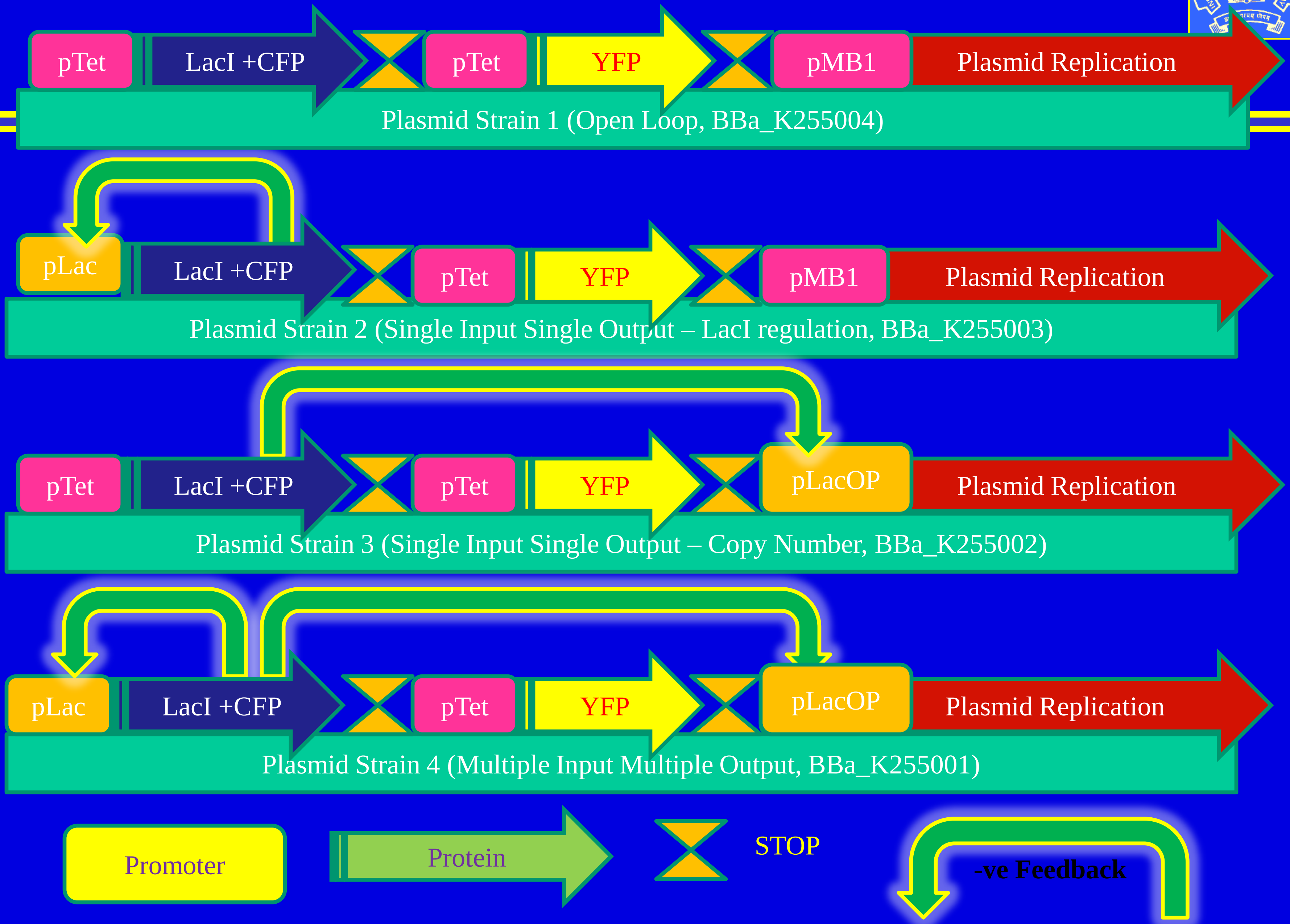
pLac

lacI

LacI

RNA/DNA Polymerase

Constructs



pTet | LacI +CFP | pTet | YFP | pMB1 | Plasmid Replication

Plasmid Strain 1 (Open Loop, BBa_K255004)

pLac | LacI +CFP | pTet | YFP | pMB1 | Plasmid Replication

Plasmid Strain 2 (Single Input Single Output – LacI regulation, BBa_K255003)

pTet | LacI +CFP | pTet | YFP | pLacOP | Plasmid Replication

Plasmid Strain 3 (Single Input Single Output – Copy Number, BBa_K255002)

pLac | LacI +CFP | pTet | YFP | pLacOP | Plasmid Replication

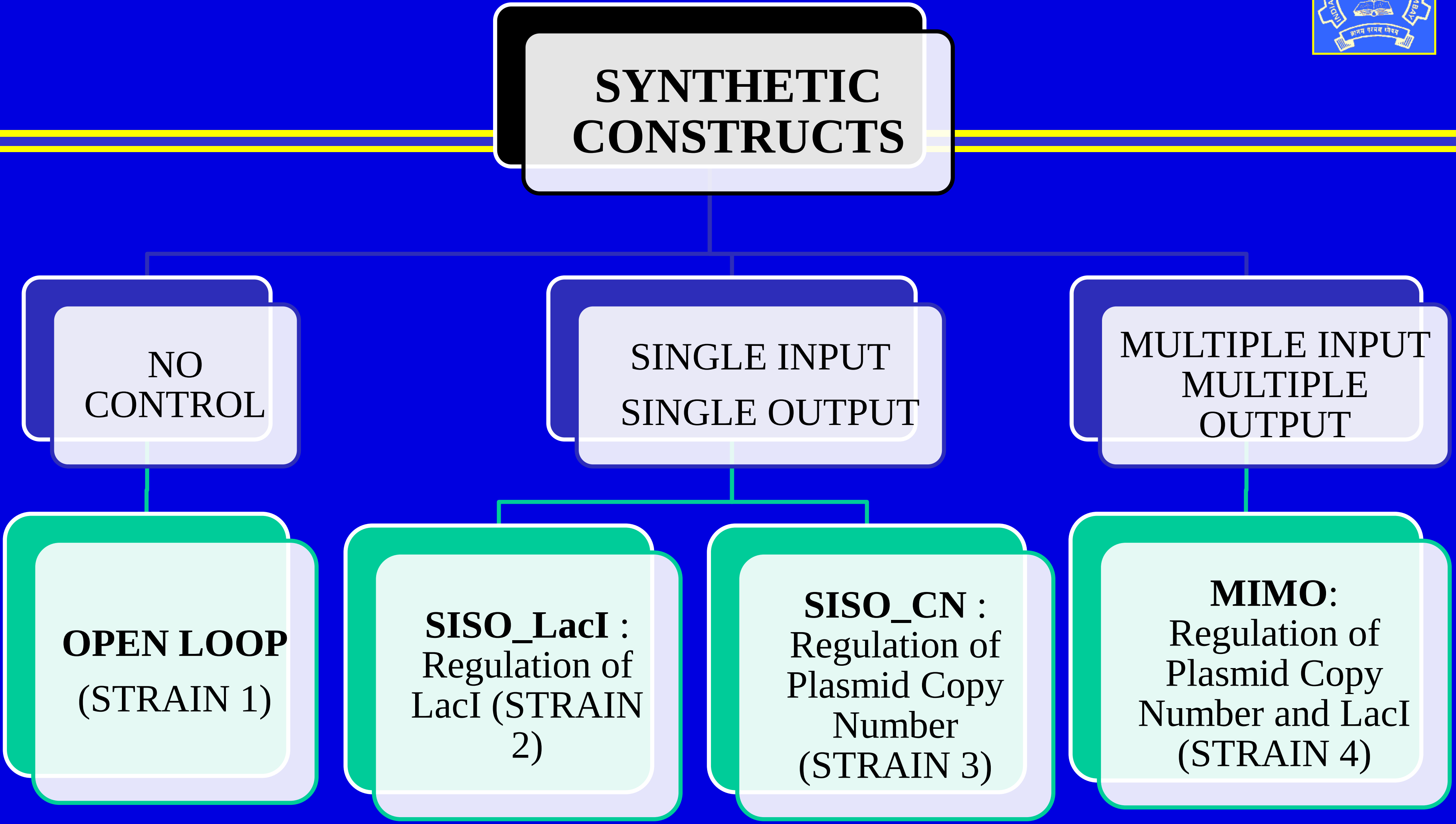
Plasmid Strain 4 (Multiple Input Multiple Output, BBa_K255001)

Promoter

Protein

STOP

-ve Feedback



SYNTHETIC CONSTRUCTS

NO CONTROL

OPEN LOOP (STRAIN 1)

SINGLE INPUT SINGLE OUTPUT

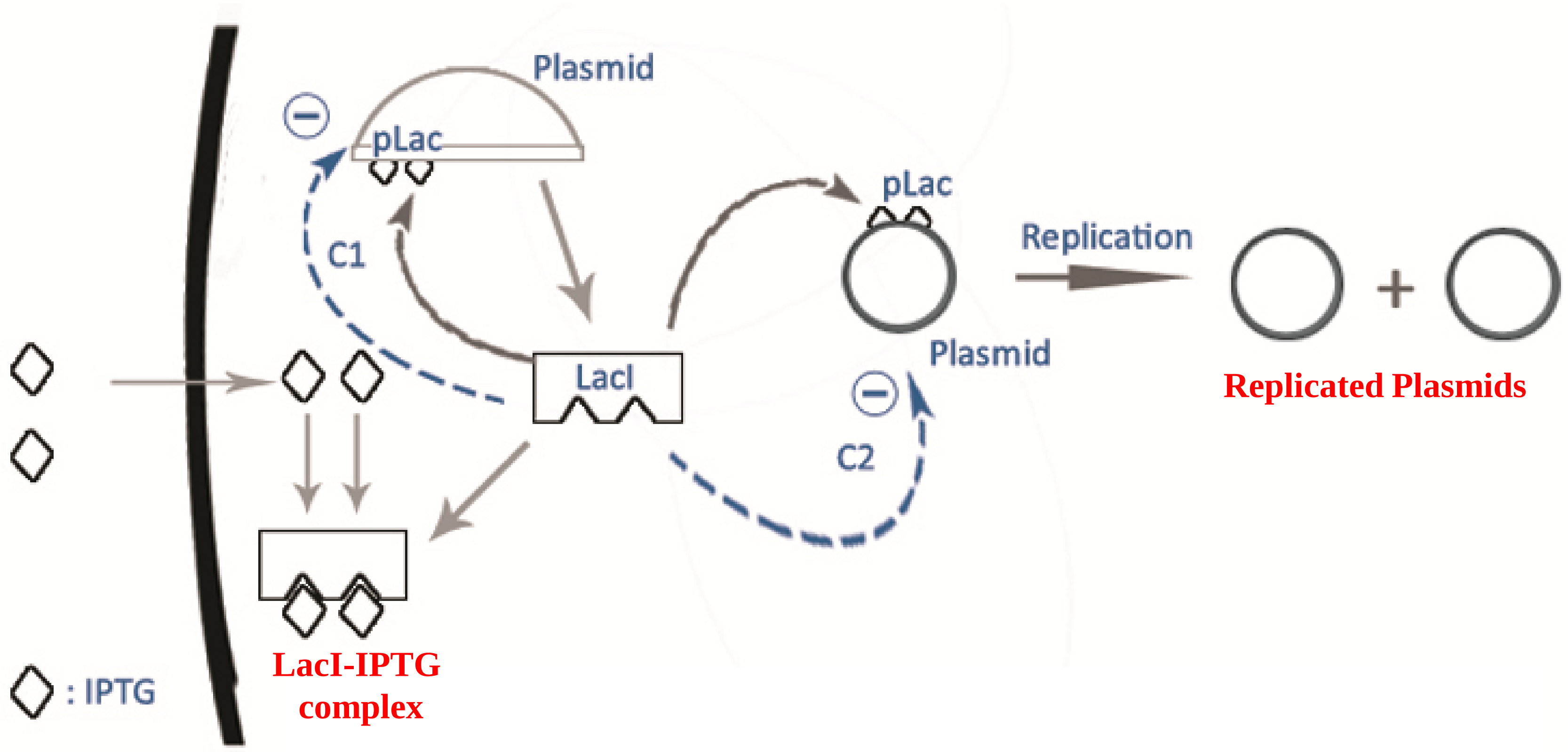
SISO_LacI : Regulation of LacI (STRAIN 2)

SISO_CN : Regulation of Plasmid Copy Number (STRAIN 3)

MULTIPLE INPUT MULTIPLE OUTPUT

MIMO: Regulation of Plasmid Copy Number and LacI (STRAIN 4)

Molecular Map of the Construct



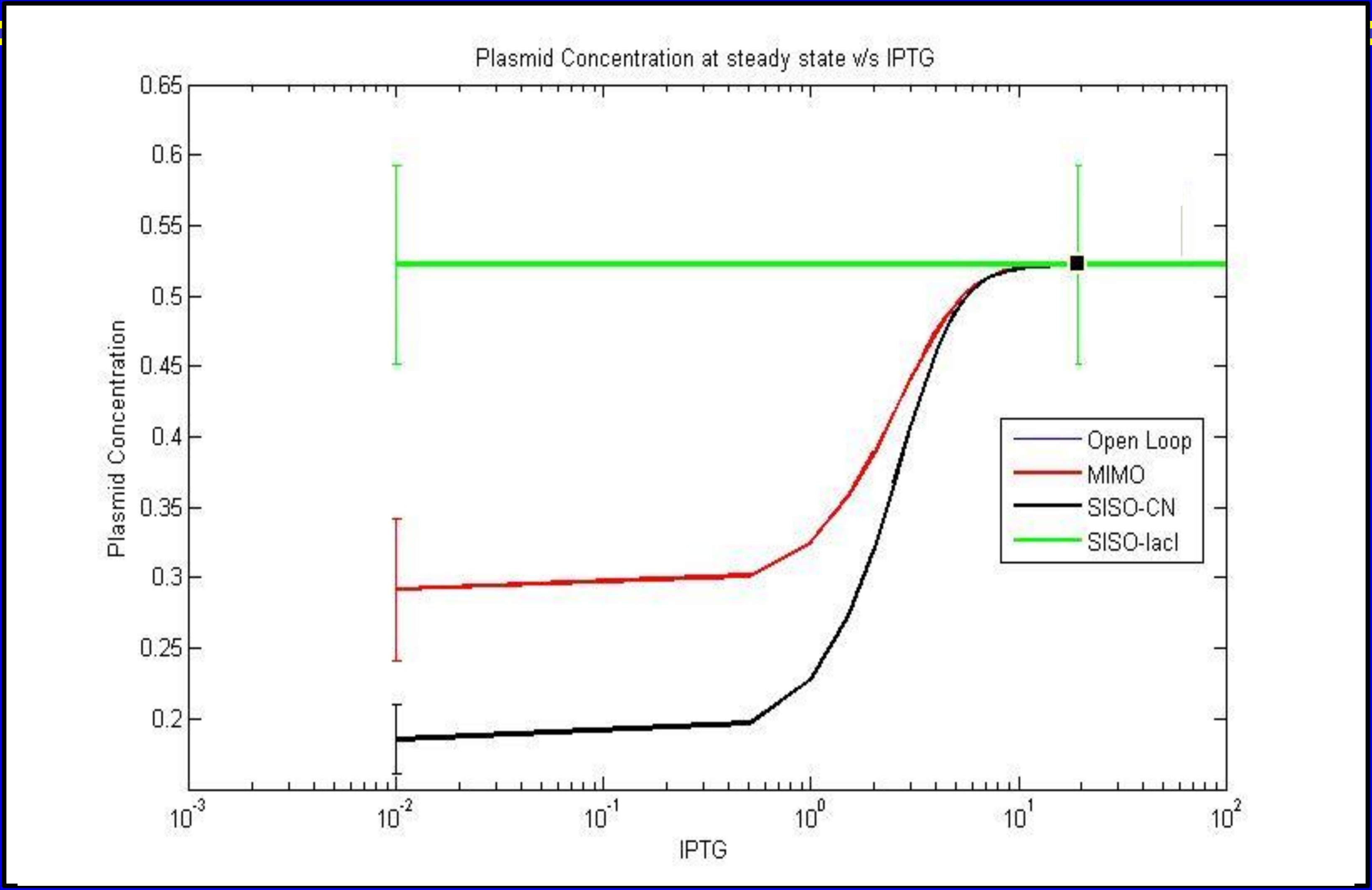
Replicated Plasmids

LacI-IPTG complex

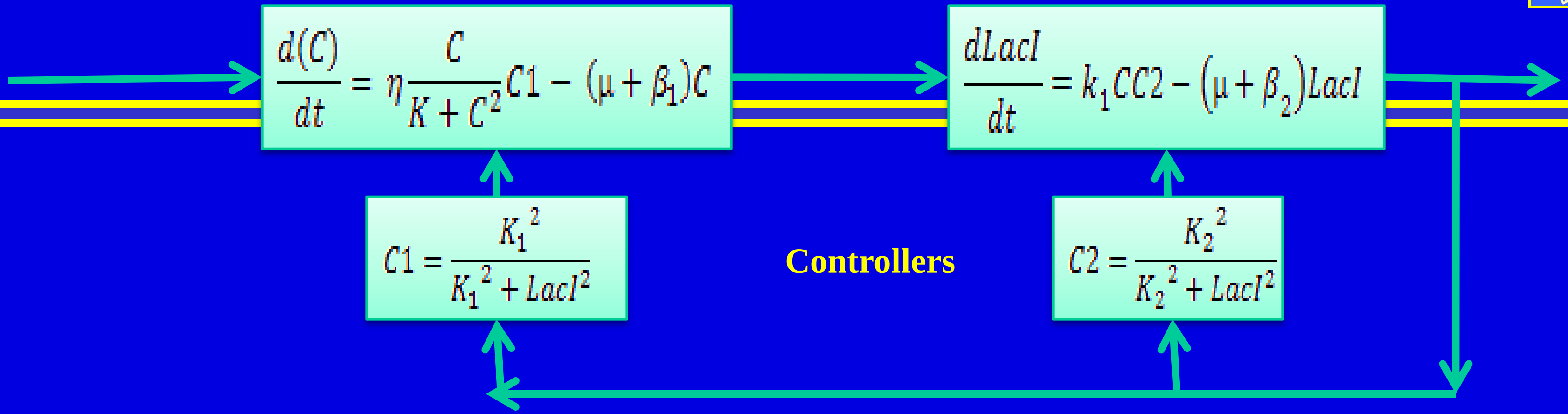
Modeling Methodologies

- Detailed Dynamic Modeling using all known molecular interactions
- Stochastic Analysis on a simplified model using Langevin approach
- Frequency response analysis on the linearised model

Prediction of Steady State Expression of YFP (Plasmid Copy Number)



Control Analysis to Characterize System Behavior



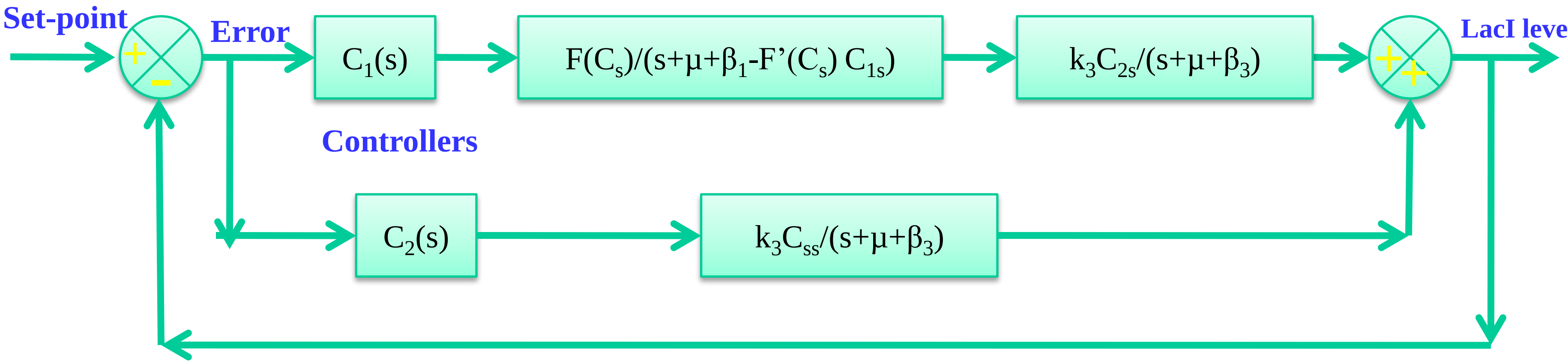
$$\frac{d(C)}{dt} = \eta \frac{C}{K + C^2} C1 - (\mu + \beta_1) C$$

$$\frac{dLacI}{dt} = k_1 C C2 - (\mu + \beta_2) LacI$$

$$C1 = \frac{K_1^{\,2}}{K_1^{\,2} + LacI^2}$$

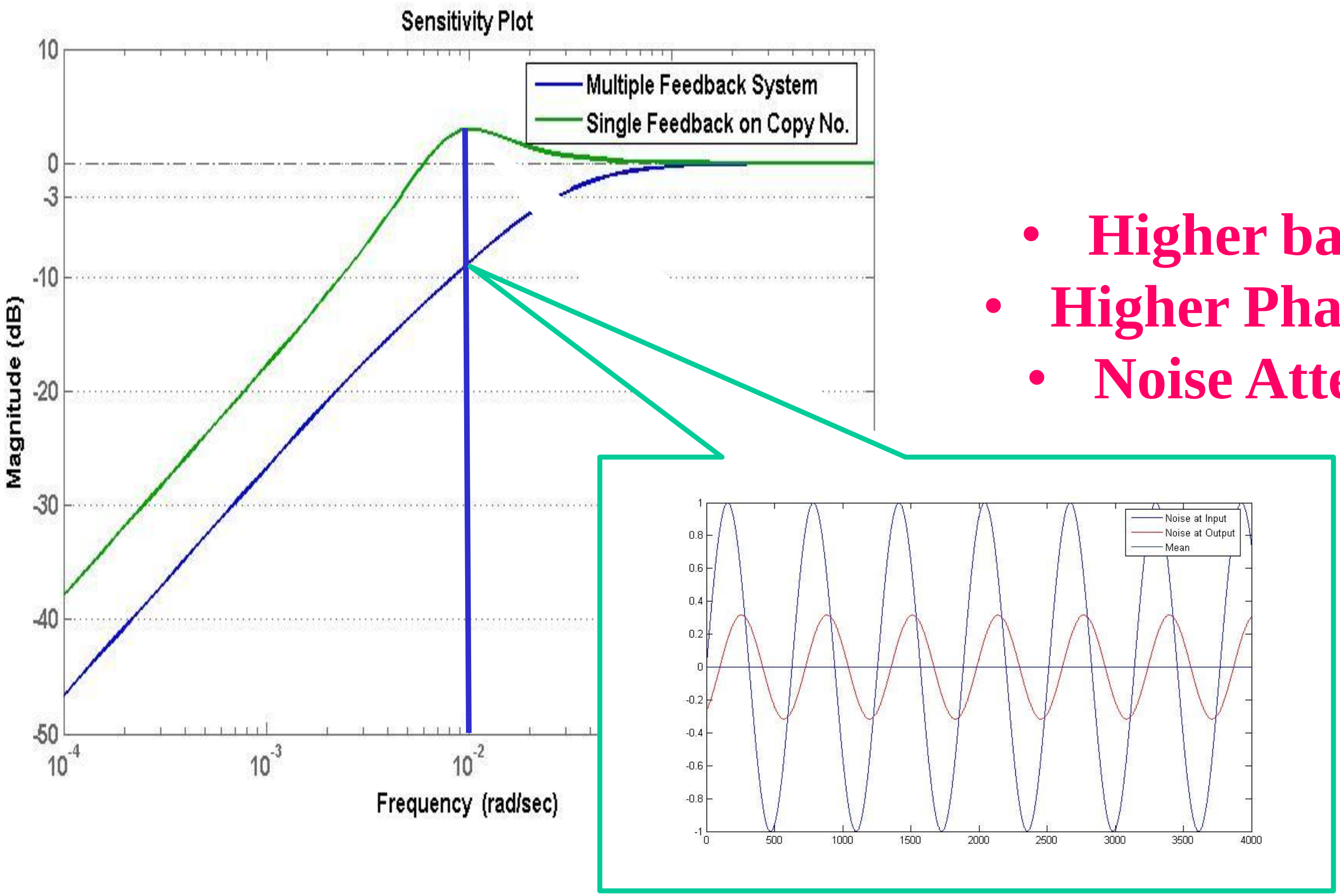
$$C2 = \frac{K_2^{\,2}}{K_2^{\,2} + LacI^2}$$

Block diagram for the LacI system



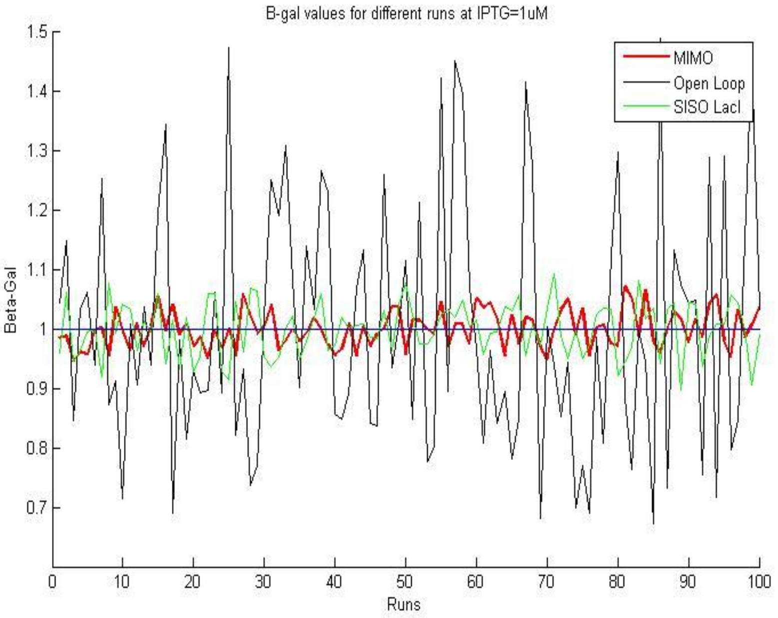
Block diagram for the Linearised LacI system

Frequency Response Analysis

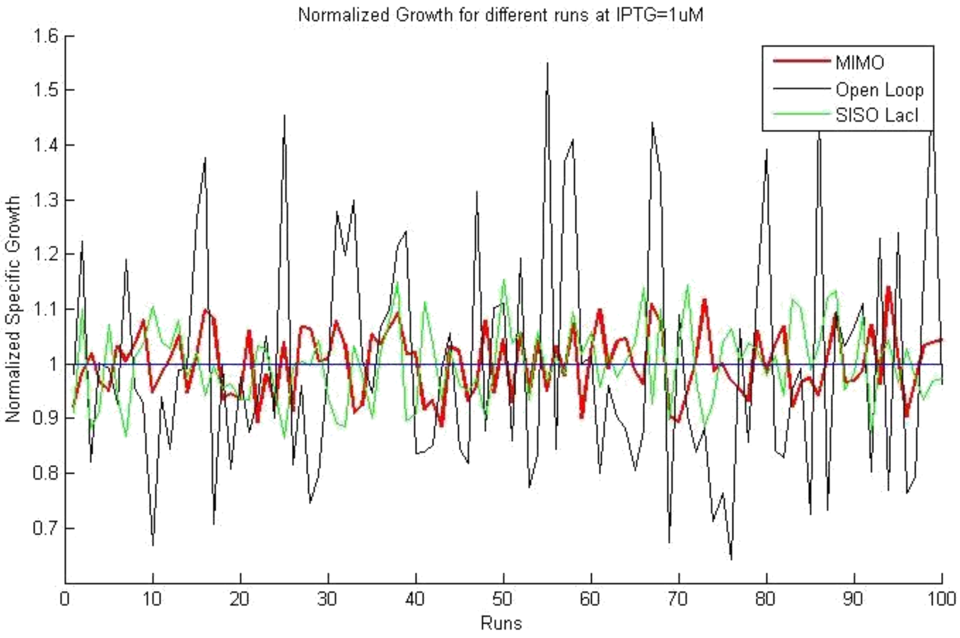


- **Higher bandwidth**
- **Higher Phase margin**
- **Noise Attenuation**

Stochastic Modeling on Growth Rate



NORMALIZED β-gal EXPRESSION



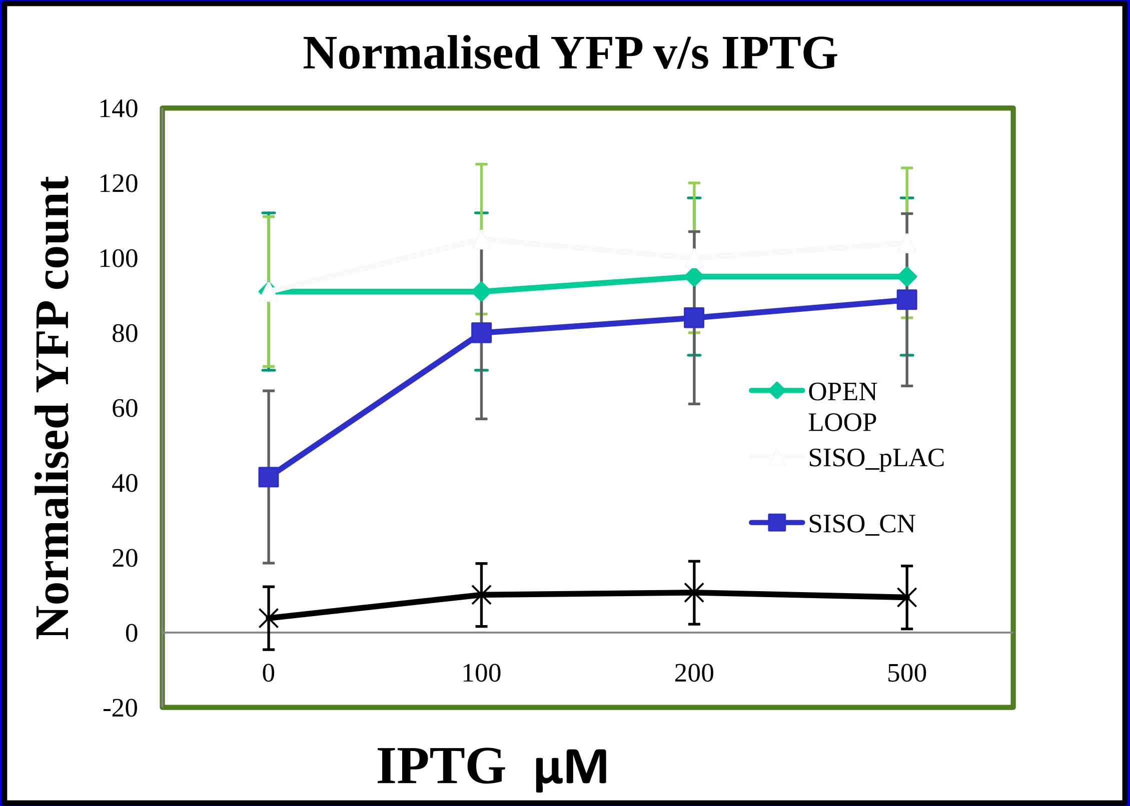
SPECIFIC GROWTH RATE

For perturbation of the kinetic parameters around the mean value, we see MIMO has the least variance compared to open loop or a single feedback system

Experimental Validation

- Experiments with various IPTG concentrations were conducted.
- Protein expression measured as YFP using FACS to quantify plasmid copy number.
- Mean and Variance obtained from the distribution.

Experimental YFP expression (characterizing Plasmid Copy Number)



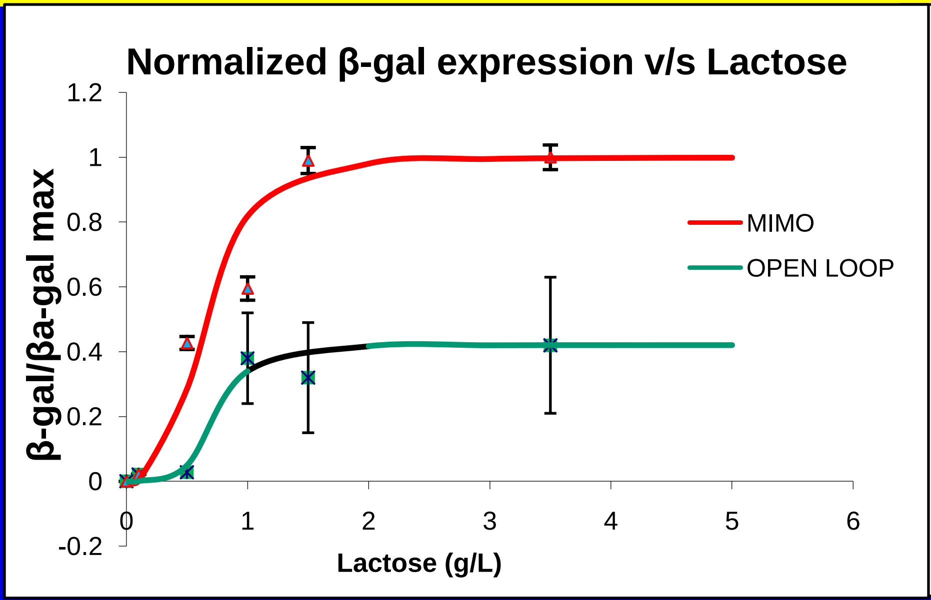
- **Open Loop and SISO_LacI:** No increase in YFP with inducer
- **SISO_CN and MIMO:** expression increase with inducer

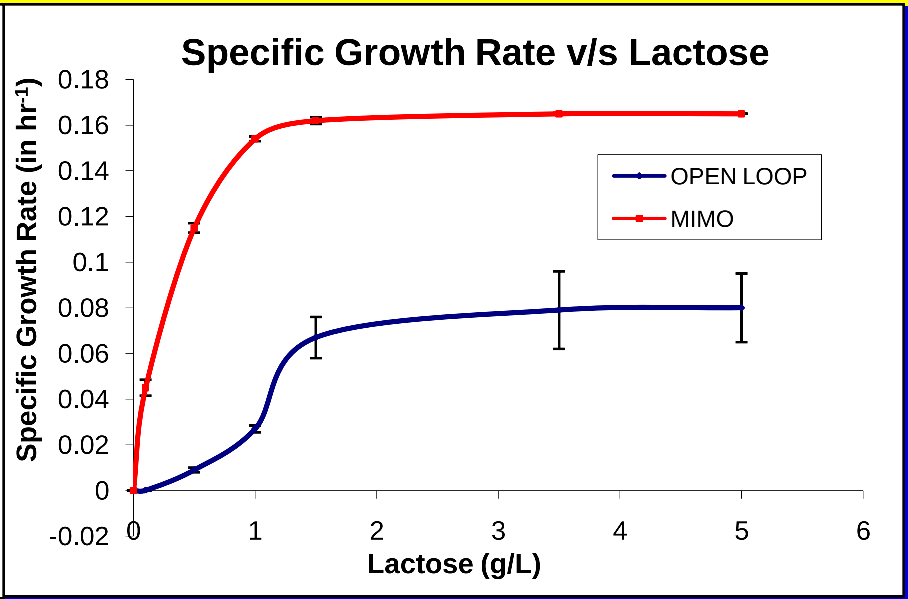
Higher variance in open loop

Characterization of LacI expression

- An indirect measure of LacI was obtained by measuring β-galactosidase from the ***lacZ*** of the host.
- Further the growth rate of the four transformants were also enumerated.

Experimental Results





Noise in protein expression propagates to growth

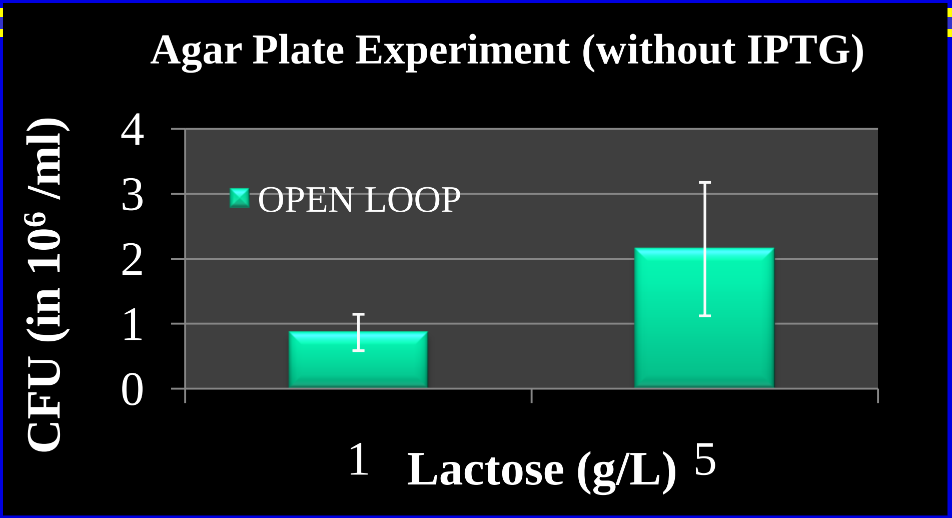
The variance in specific growth rate is less compared to that observed in protein expression.

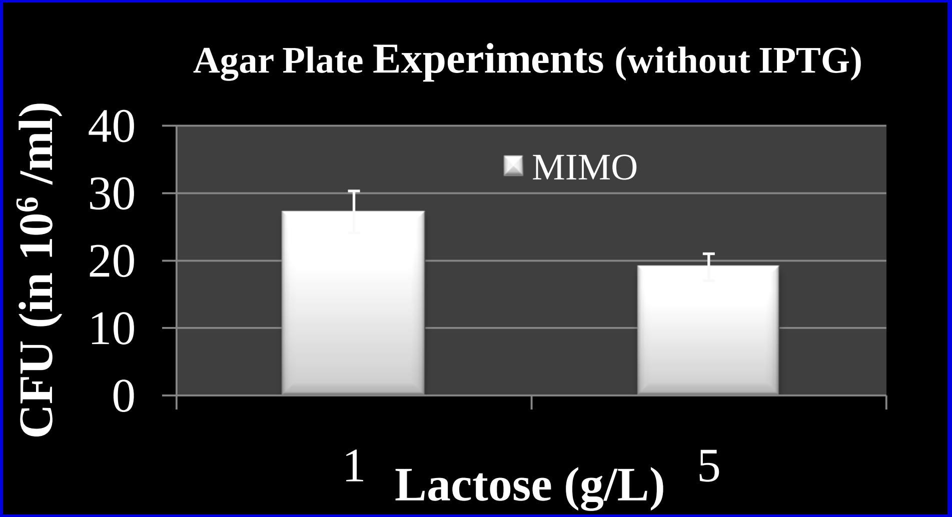
Agar Plate Experiments

Strains were grown on agar plate with different lactose concentrations.

Colony Forming Units in the agar plates were counted.

Variance in
Open Loop is 40 % and
MIMO is 10%.

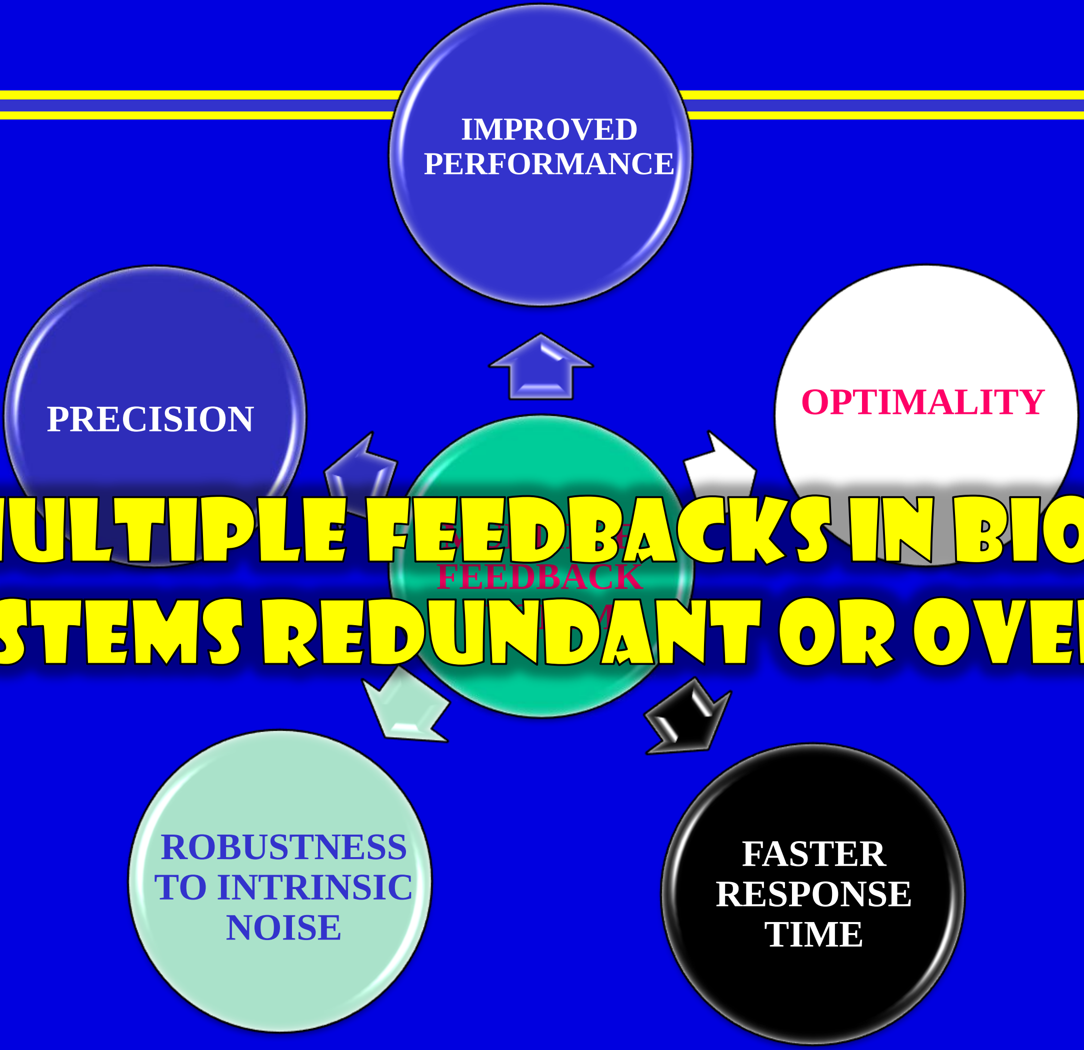




Recapitulating…

- Robustness in protein expression which leads to low variance in specific growth rate.
- The noise in protein expression is filtered leading to a decrease in the variance in growth rate. This may be due to metabolism and division process.
- The transformants with the synthetic network yields distinct phenotypic response.

DO MULTIPLE FEEDBACKS IN BIOLOGICAL SYSTEMS REDUNDANT OR OVERKILL?



IMPROVED PERFORMANCE

PRECISION

OPTIMALITY

[illegible] FEEDBACK [illegible]

ROBUSTNESS TO INTRINSIC NOISE

FASTER RESPONSE TIME

Collaborators / Students

- Prof. Sharad Bhratiya (IITB)
- Dr. Mukund Thattai (NCBS)
- Nikhil Chaudhary (PhD, IITB)
- iGEM team 2009
- Pushkar Malakar (PhD, IITB)

Thank you!!