

**Symposium on
“Complexity & Computation in the Natural Sciences”**

A Simple Approach to Study
Designs in Complex
Biochemical Pathways

SOMDATTA SINHA

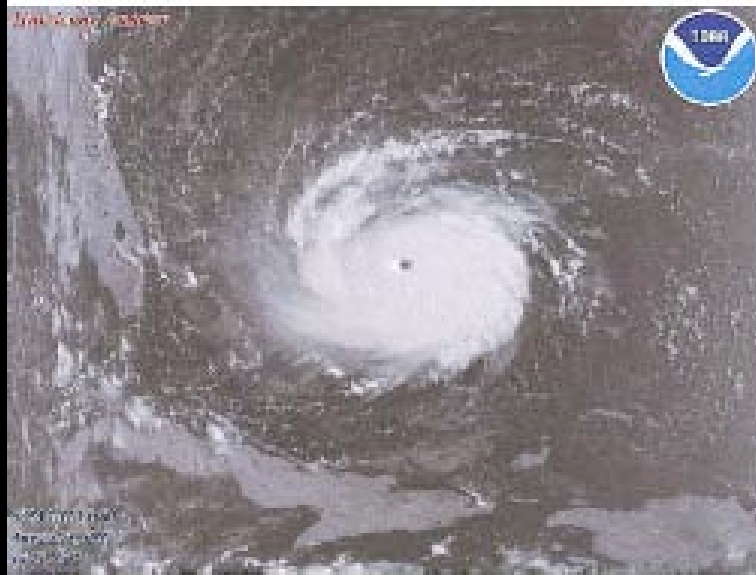
***Centre for Cellular and Molecular Biology (CSIR)
Hyderabad***



Indian Academy of Sciences,
74th Annual Meeting, New Delhi
Oct. 31 – Nov. 2, 2008

Indian Institute of Technology,
New Delhi, 31 October 2008

Hurricane approaching
Florida



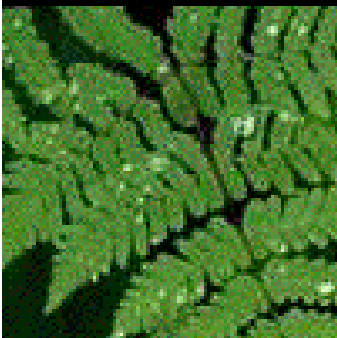
The Spiral Galaxy
M51



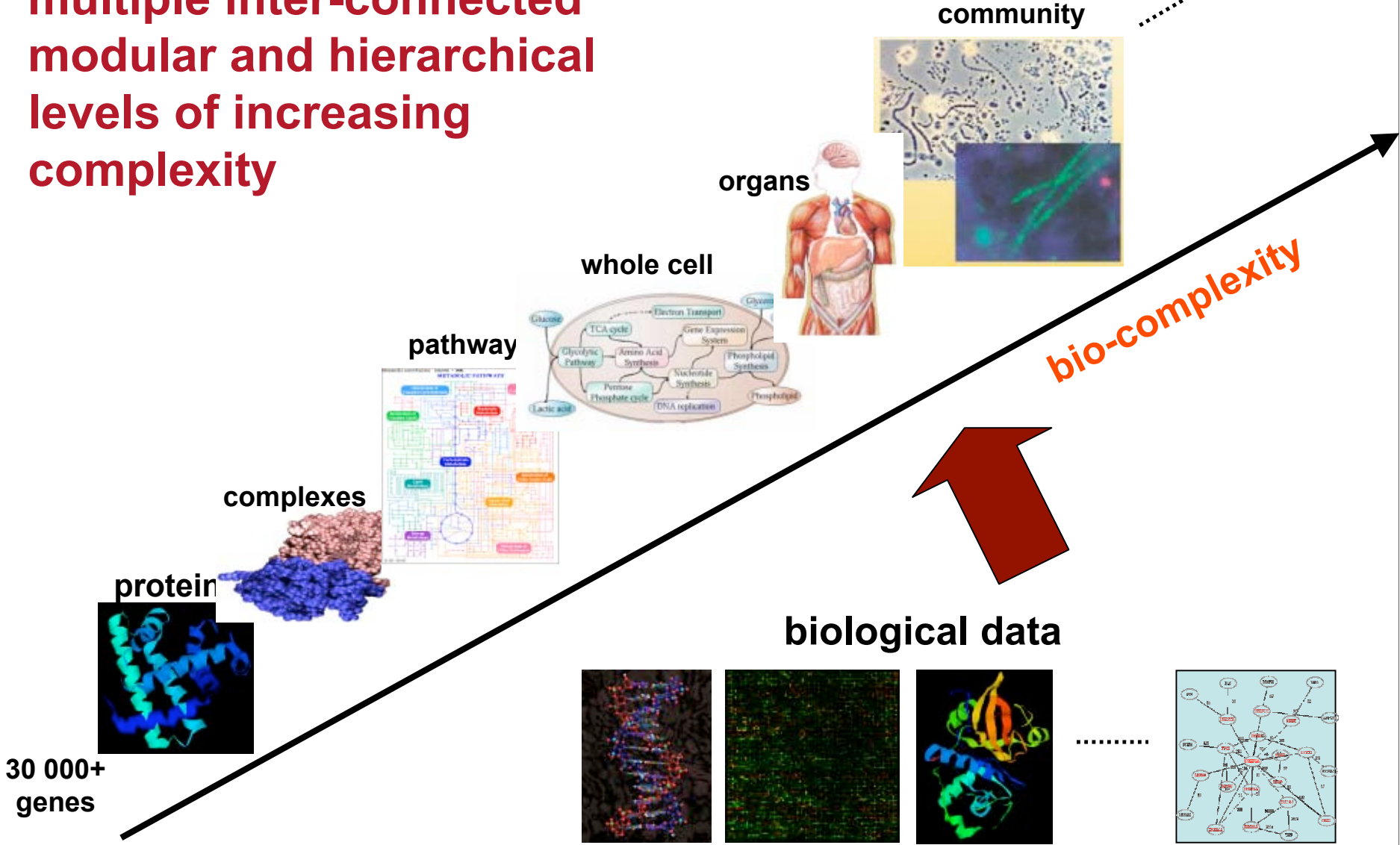
(Machta, Complexity, 2006: courtesy NOAA and NASA)

LIVING SYSTEMS ARE COMPLEX SYSTEMS

Complexity arises from selective and nonlinear interactions of functionally diverse components to produce a coherent structure/function

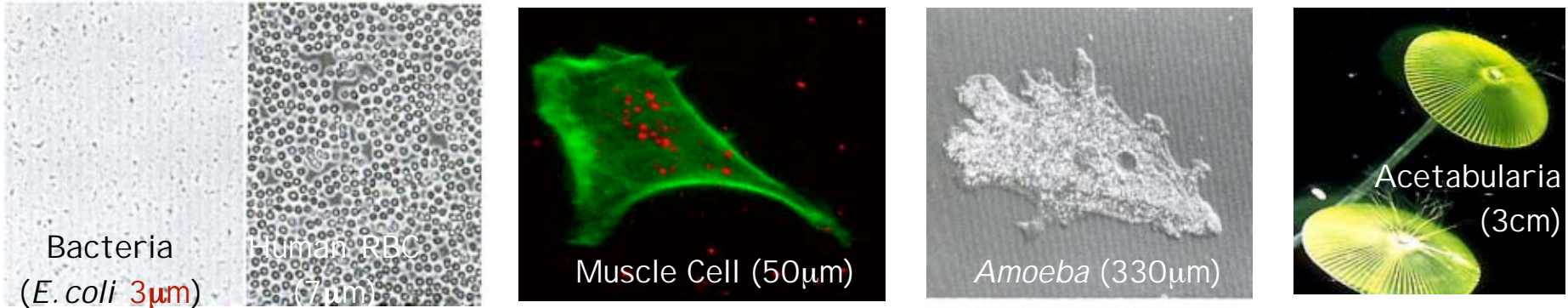


The genomic information and its regulation is organised at multiple inter-connected modular and hierarchical levels of increasing complexity



Living systems are made up of cells

– *single or multi-cellular*



Cell is the basic unit of life

Large cells -

nerve cells in giraffe's neck ~ 3 m
(9.7 ft) in length.

Smallest cell -

Mycoplasma ~ 10⁻⁷ cm diameter

Bacterium *E. coli* divide in 20 min

Yeast cell cycle - 90 -120 min

Rapidly dividing

mammalian cell cycle ~ 24 hours

Individual cells need to have mechanisms to monitor their environmental composition.

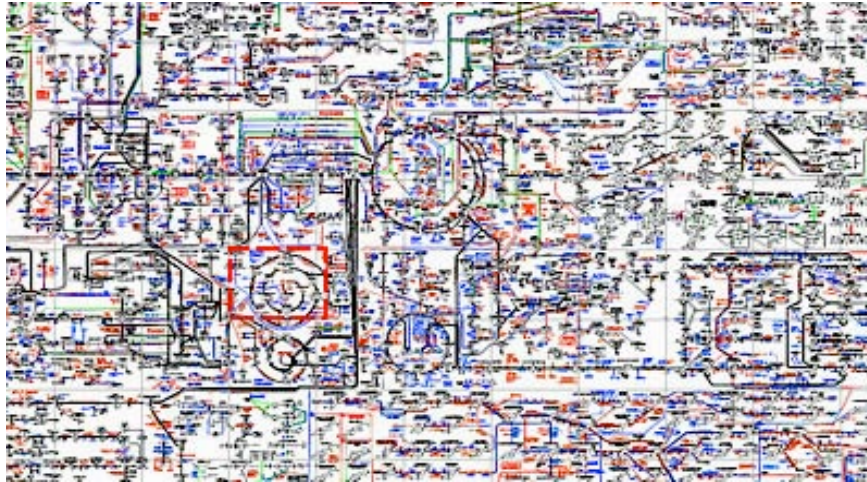
They need to discern and synchronize their responses according to variations in external and internal conditions.

To achieve this level of coordination, metabolites and chemical compounds are used by the cell as messages to know the composition of these environments.

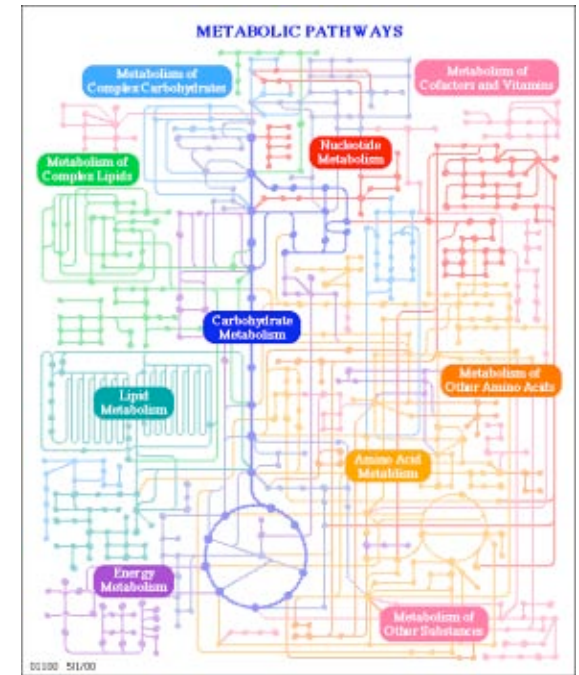
AND

it requires computation and information transfer across breadth and depth of processes at all levels of organisation

Cellular functions are controlled by networks of biochemical reactions



Intricate networks of inter-connected chemical reactions between molecular species in the cell.



Cellular behaviour is the emergent property of many biochemical reactions networked through feedback/feed-forward processes

Negative Feedback ensures stability and conservation of energy by desensitizing the system to perturbations - are naturally selected to be the most common form of regulation in pathways

Positive Feedback is potentially destabilizing, and primarily employed for excitable, amplification, and switching processes.

Complex network of biochemical reactions in cells
co-ordinate and control cellular functions

two interacting sets

GENETIC
REACTIONS

Gene induction,
repression,
replication,
transcription

**Gene & Transcription Factor
networks**

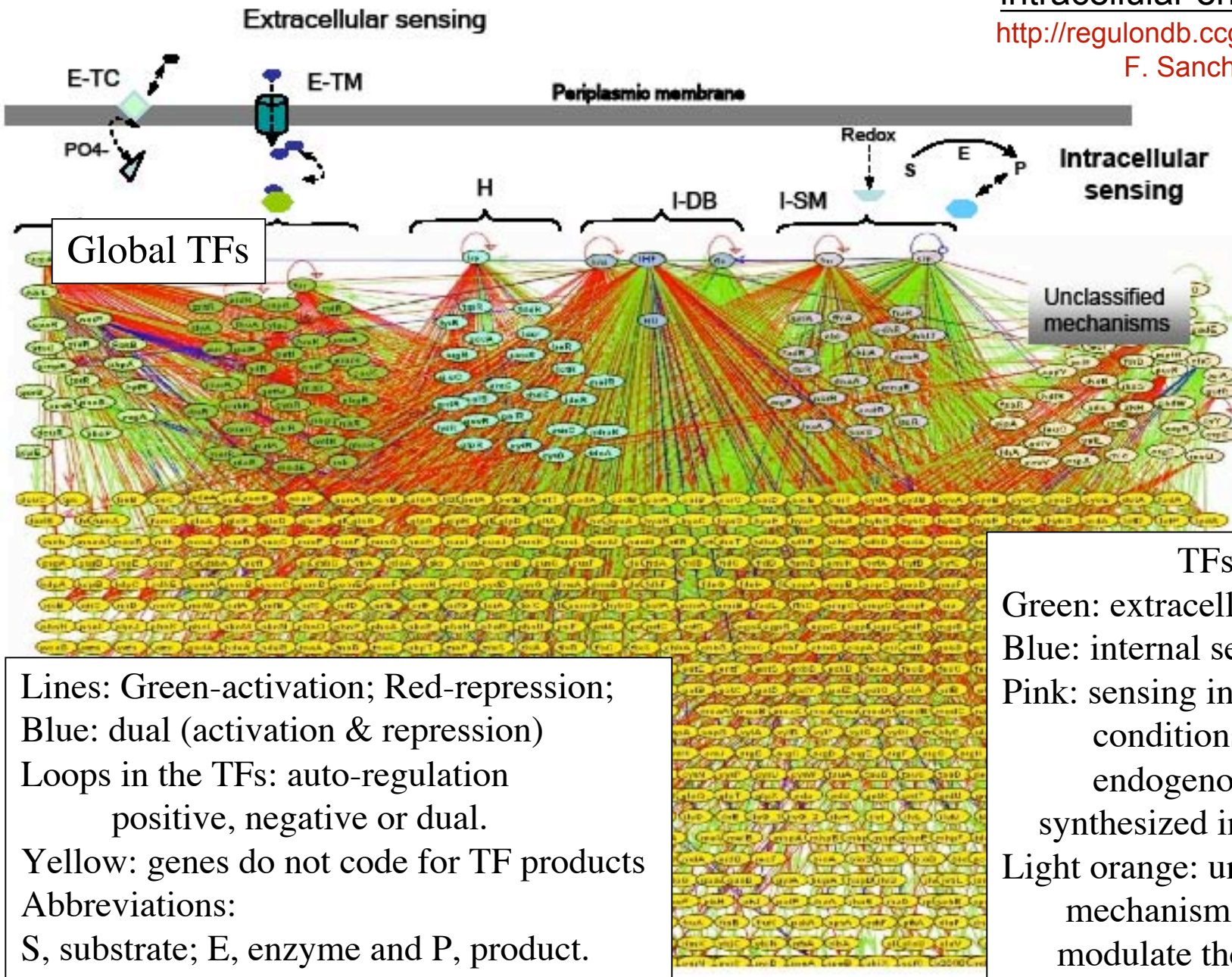
METABOLIC
REACTIONS

Conversion of substrate
molecules by enzymes,
enzyme inhibition or
activation

**Metabolic reaction
networks**

Escherichia coli transcriptional regulatory network for sensing the extra-cellular and intracellular environment

<http://regulondb.ccg.unam.mx/>; by F. Sanchez and E. Diaz



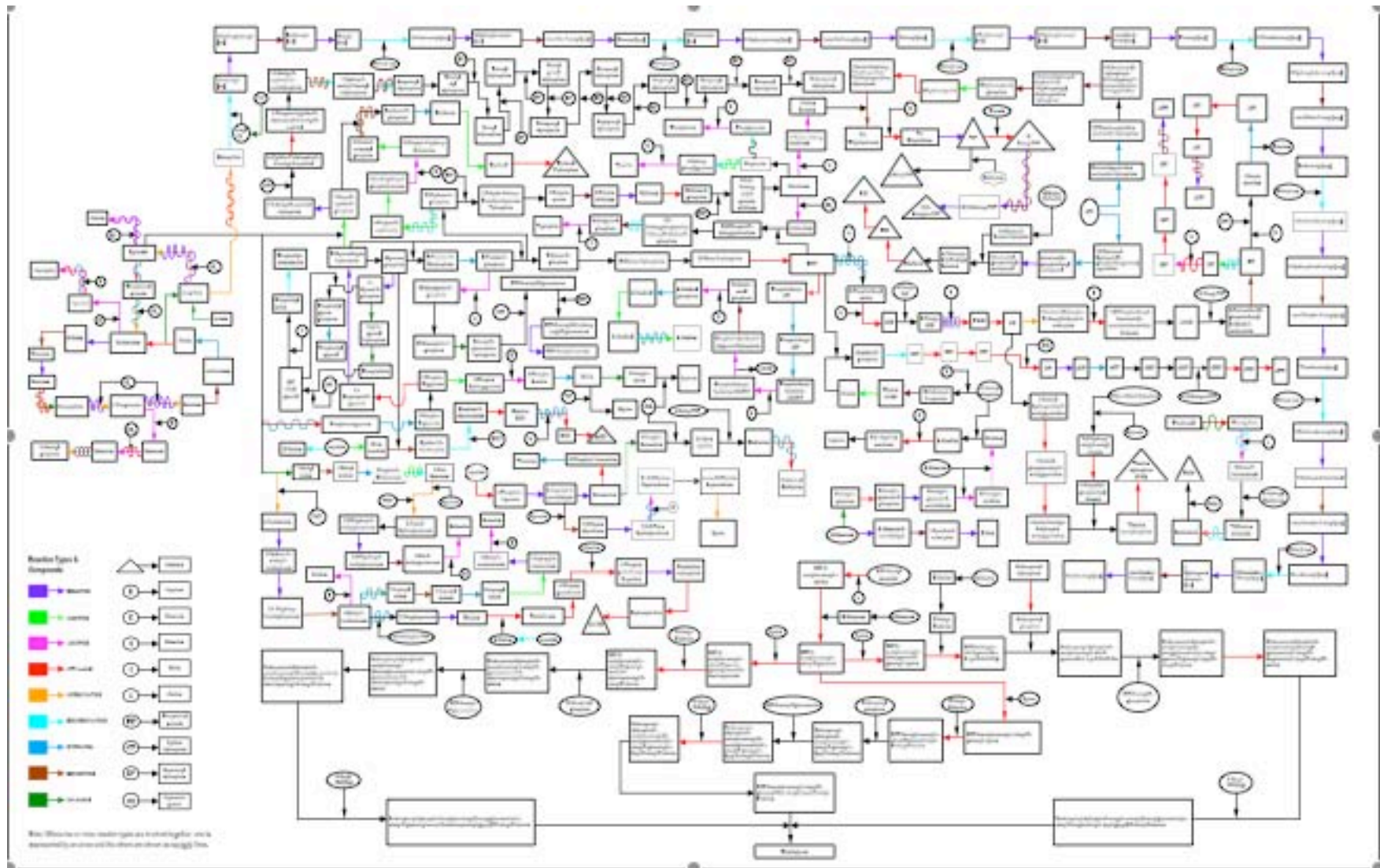
Lines: Green-activation; Red-repression;
 Blue: dual (activation & repression)
 Loops in the TFs: auto-regulation
 positive, negative or dual.
 Yellow: genes do not code for TF products
 Abbreviations:
 S, substrate; E, enzyme and P, product.

TFs:
 Green: extracellular sensing
 Blue: internal sensing
 Pink: sensing intracellular
 conditions using
 endogenous signals
 synthesized inside the cell
 Light orange: unknown
 mechanisms to
 modulate their activities.

Core Metabolic Network of reductive chemoautotrophs

(bacteria that fix carbon by the reductive TCA cycle)

(Minimal metabolome - 287 metabolites)



Srinivasan & Morowitz, Biological Bulletin, 2008

**How can one do predictive studies of
these complex information processing
units/pathways**

&

**understand the role of pathway designs in
their function ?**

“Multiple steps in every pathway”

&

“Regulation - Positive/Negative”

**The dynamical consequences of these designs can
be quite opposite in pathway functions.**

MODELLING BIOCHEMICAL PATHWAYS

Three complementary approaches

REVERSE ENGINEERING

Model existing pathways based on information derived from –

- **Genome sequences**
- **Protein sequences**
- **Biochemical & Genetic information**

FORWARD ENGINEERING

All designs that are not physically forbidden are realizable, but not all realizable designs are functionally effective

(in relation to context and constraints of the system and environment).

LARGE NETWORKS

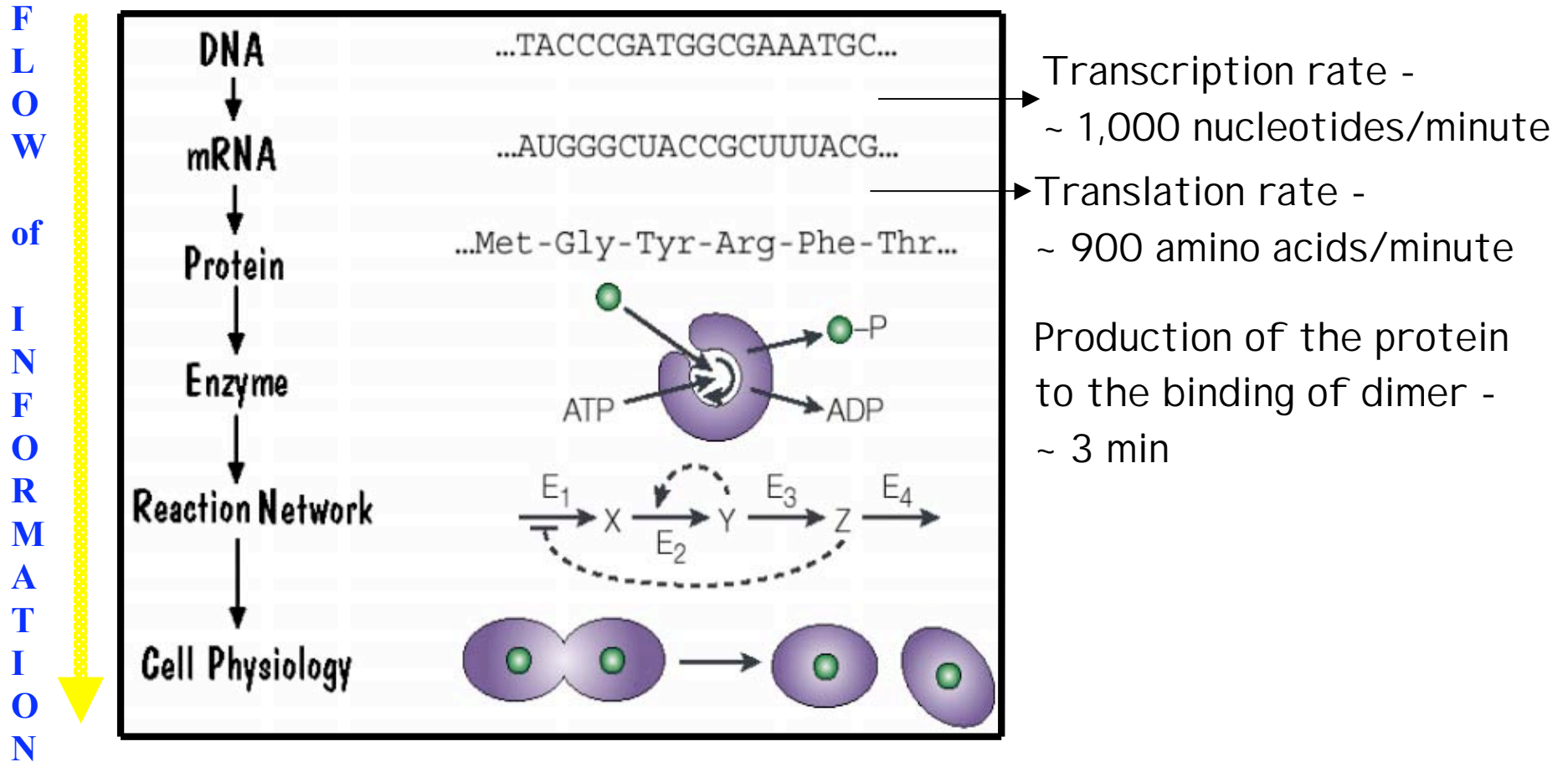
Construction & analysis of functionally related pathways from large scale gene expression and protein interaction data using network theory

→ *'Rational Network Design'*

Artificial genetic and enzymatic networks with specific properties constructed based on mathematical models

*Synthetic oscillatory circuit;
Toggle switch in bacteria;
Amplifiers of gene expression.*

THE INHERENT DELAY IN THE TRANSCRIPTION-TRANSLATION PROCESS



Starting from gene expression to cellular function involves a sequence of reactions over a period of time - **a common feature**

Question

- ❖ A generic feature in all intracellular biochemical processes is the time required to complete the whole sequence of reactions to yield any observable quantity - **widespread presence of time delay** in biological functions.
- ❖ Theoretically **time delay is known to be a source of instability**, and has been attributed to lead to oscillations or transient dynamics in several biological functions.
- ❖ **Negative feedback loops**, common in biochemical pathways, are known to **provide stability**, and withstand considerable variations and random perturbations of biochemical parameters.
(Savageau, 1974; Becskei and Serrano, 2000).

**Interaction of these two opposing factors
- instability and homeostasis -
are common features in intracellular processes**

Effect of these divergent forces in the dynamics of gene expression?

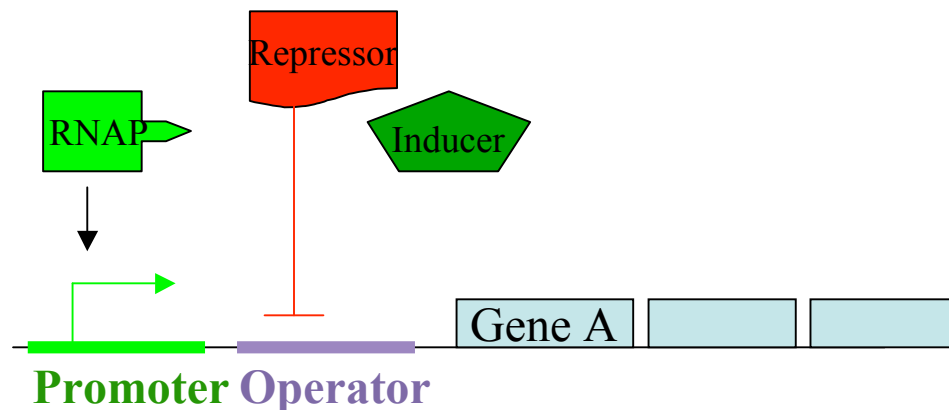
Forward Engineering of gene circuits

(Rational network design)

Construction of desired network *with specific properties predicted from mathematical models* using knowledge from biochemistry, molecular biology, and genetics.

Boolean/Logical Circuits in Biology :

Organisms take decisions based on input signals and give a binary (0/1) response in some cases.



Jacob & Monod Model of the prokaryotic operon (1961)

“It is obvious from analysis of these [bacterial genetic regulatory] mechanisms that their known elements could be connected into a wide variety of ‘circuits’ endowed with any desired degree of stability”

Genetic Circuit Engineering Paradigm

Design - Simulate - Implement & Test

A basic assumption underlying such 'synthetic' biology

The properties of individual genetic components can be used to understand and quantitatively predict circuit-level behaviors.

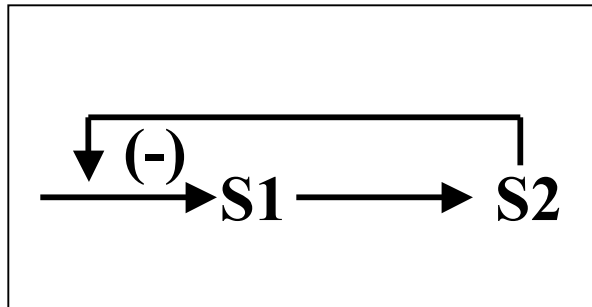
“ **Rational Network Design** ” can -

- a) engineer new cellular behaviour, and
- b) improve understanding of naturally occurring networks.

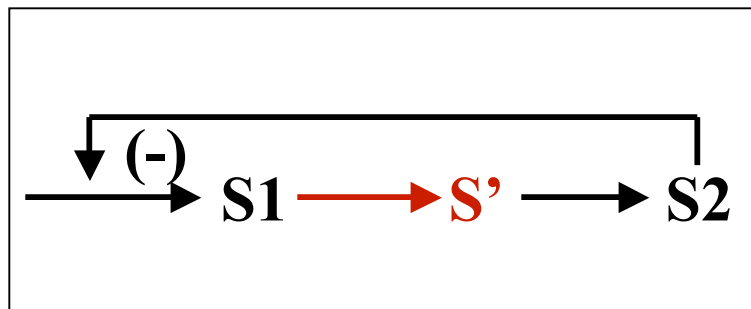
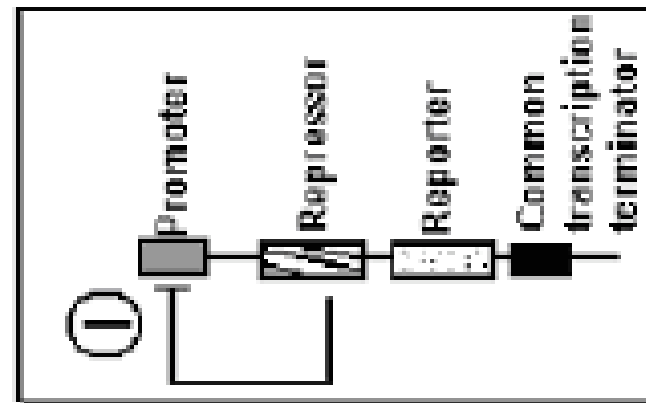
Approaches

- v Designed simple negatively auto-regulated transcriptional modules consisting of a basic regulator and transcriptional repressor - with and without delay in repression - and their controls.
Constructed gene circuits in *E. coli*.
- v Developed mathematical models of a simple negative feedback pathway based on the design of the negatively auto-regulated gene circuit -
deterministic and stochastic
- v Compared the gene expression dynamics with theoretical predictions.

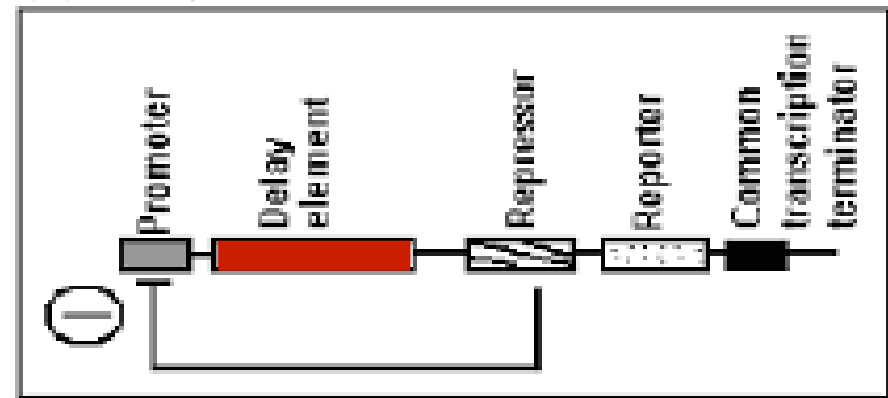
Designing negatively auto-regulated gene circuits - *with and without delay in repression*



Basic Circuit



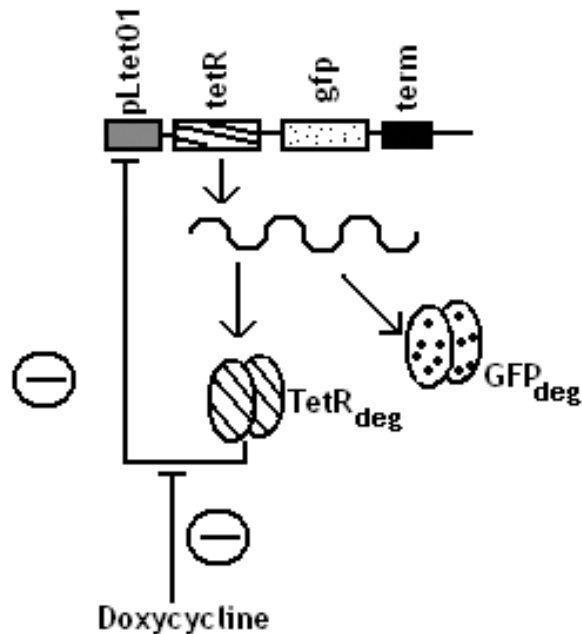
Delay Circuit



The presence of one or more genes (“Delay element”) increase the length of the transcript, thereby introducing a delay in establishment of negative feedback by the repressor in **Delay circuit** compared to the **Basic circuit**.

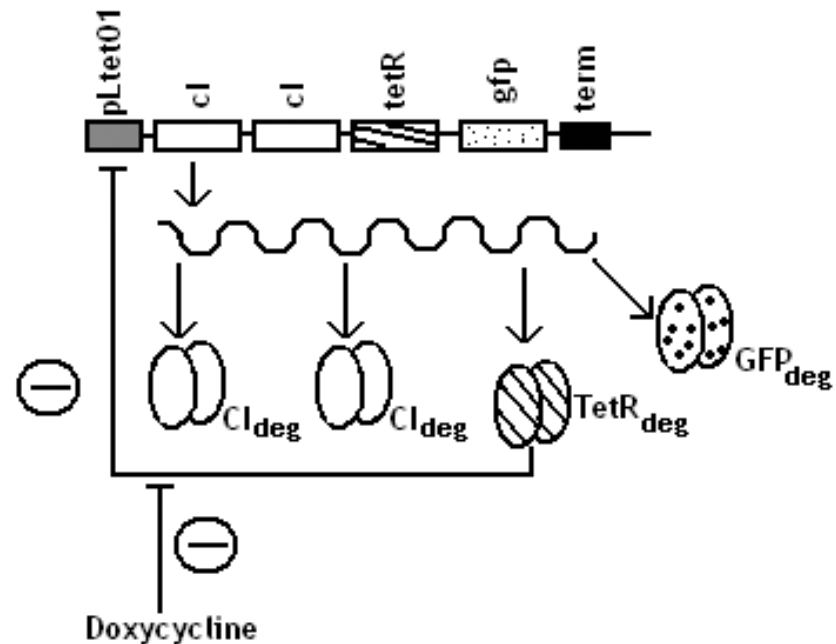
Design of negatively auto-regulated gene circuits

(a) Basic Circuit (TG)



TG: *tetR* gene and reporter gene (*gfp*) after the promoter-operator unit (*pLtet-01*).

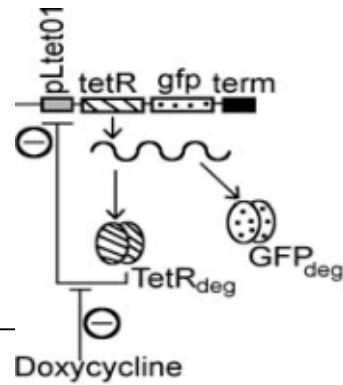
(b) Delay (C2TG)



C2TG: two copies of *cl* gene from λ phage before the repressor gene - production of the repressor is delayed.

Control Delay circuit (TC2G): Position of repressor same as in TG, but position of reporter is as in C2TG. This is identical to the Delay circuit (C2TG) in length, number of cistrons, and position of the Reporter gene, except for the position of the repressor, TetR.

Deterministic model



$$\frac{dm}{dt} = \beta_1 g - \alpha_1 m$$

$$\frac{dp}{dt} = \beta_2 m_{(t-\tau_1)} - \alpha_2 p$$

$$\frac{df}{dt} = \beta_2 m_{(t-\tau_2)} - \alpha_3 f$$

$$\frac{dg}{dt} = k_2 (g_t - g) - k_1 gp$$

m = mRNA

p = repressor

f = reporter

g = free promoters

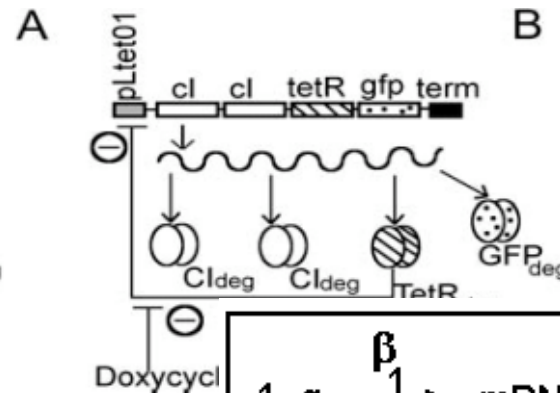
g_t = total no. of promoters;

$\alpha_1, \alpha_2, \alpha_3$ = degradation rates

β_1, β_2 = transcription & translation rates

K_1, k_2 = promoter-repressor complex reaction rate

τ_1, τ_2 = time delays on the production of **p** and **f** from the common mRNA;



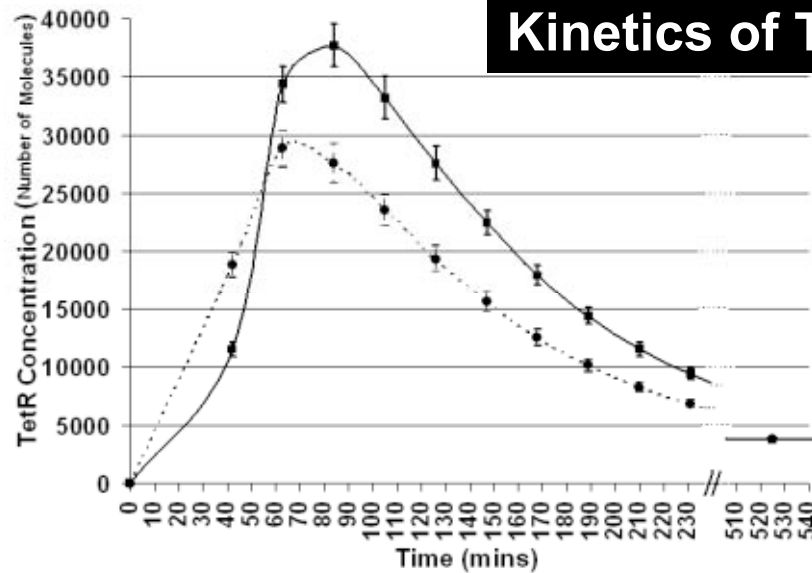
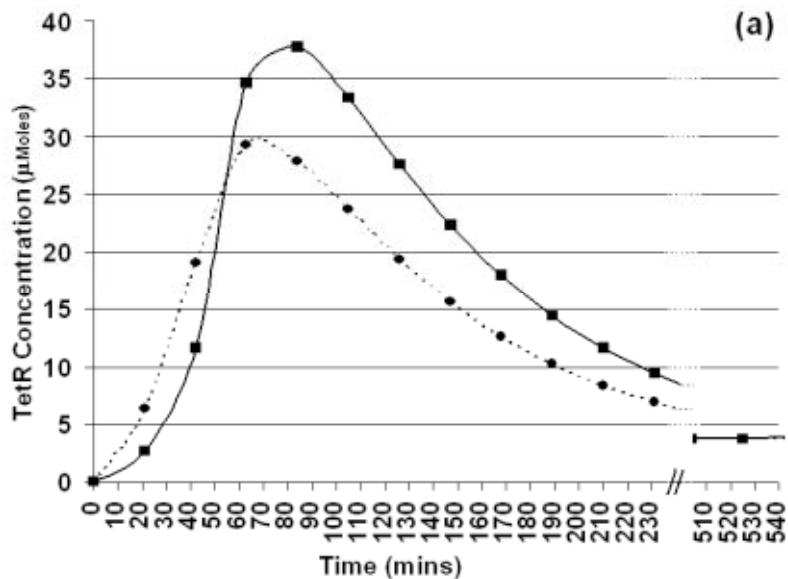
Stochastic model molecular reactions

(Gillespie, 1977;
Bratsun et al., 2005)

1. $g \xrightarrow{\beta_1} \text{mRNA}$ Transcription
2. $\text{mRNA} \xrightarrow{\beta_2} \text{TetR}$ Translation
3. $\text{mRNA} \xrightarrow{\beta_2} \text{GFP}$ Translation
4. $g + \text{TetR} \xrightarrow{k_1} [g\text{TetR}]$ Repressor binding
5. $[g\text{TetR}] \xrightarrow{k_2} g + \text{TetR}$ Repressor dissociation
6. $\text{mRNA} \xrightarrow{\alpha_1} \phi$ Degradation
7. $\text{TetR} \xrightarrow{\alpha_2} \phi$ Degradation
8. $\text{GFP} \xrightarrow{\alpha_3} \phi$ Degradation

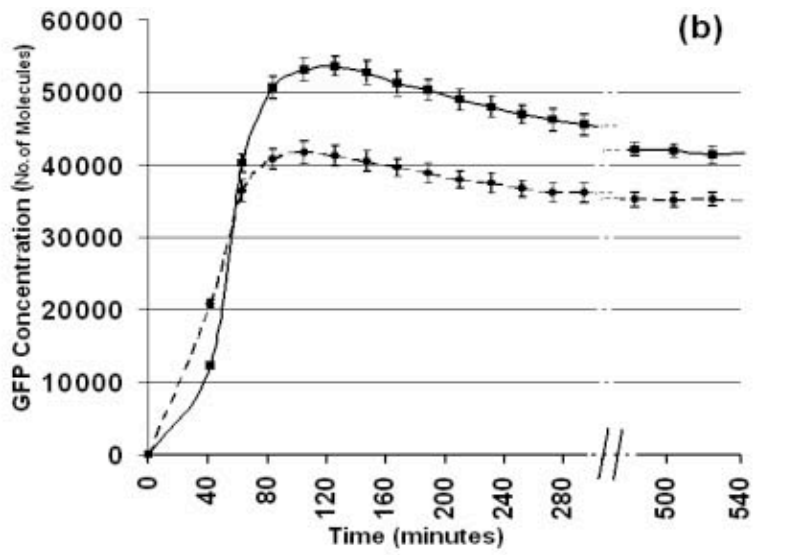
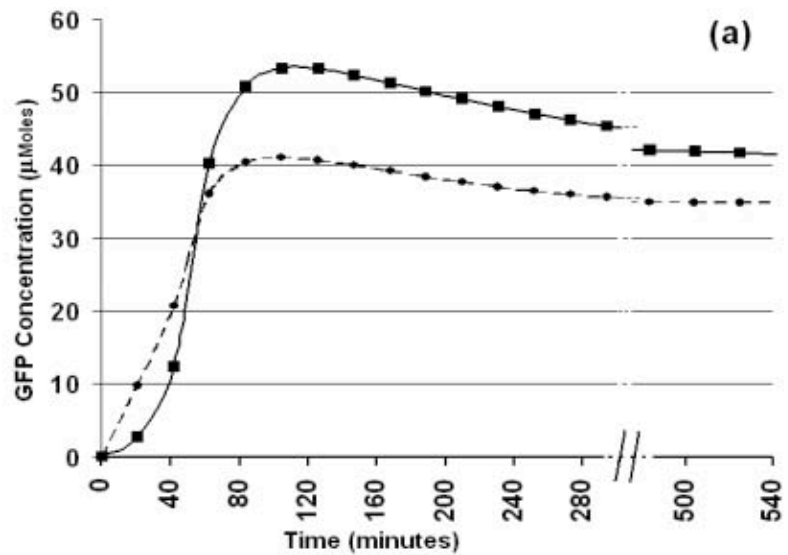
ϕ - degradation products of mRNA, TetR and GFP

Deterministic model



Kinetics of TetR

Kinetics of GFP



(average of 100 simulations).

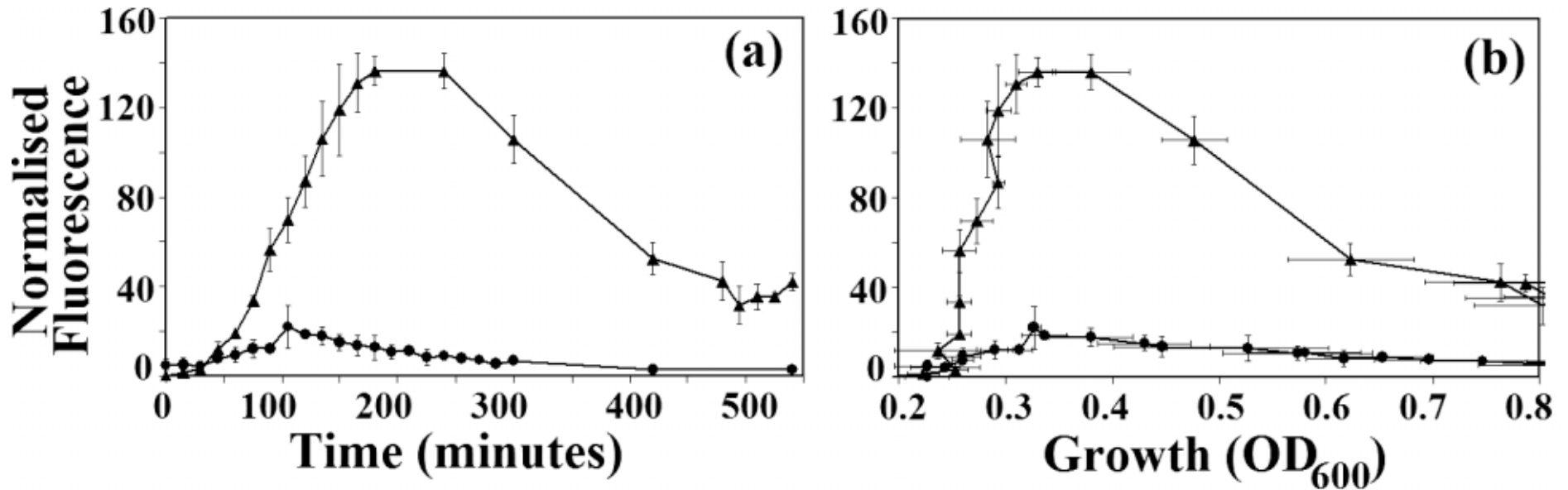
Stochastic model

Basic circuit (circles with dashed lines)
Delay circuit (squares with solid lines)

The circuit dynamics is stable in all conditions considered, but with increasing delay time, the system shows damped oscillations, and the steady state is reached with progressively larger excursion in the phase plane.

Our theoretical analysis predicts that the negatively auto-regulated pathway, as represented in these models, can show a transient overshoot in gene expression and protein production due to the delay in the kinetics of the repression process.

Experimental kinetics of GFP



Basic (*circles*) and the Delay (*triangles*) circuits upon induction (25 ng/ml) in four independent experiments.

(a) Normalised fluorescence versus time (minute);

(b) Normalised fluorescence versus growth (OD_{600}).

Error bars (one standard deviation) are for both fluorescence and growth.

Delay circuit shows a large overshoot in gene expression.

**Gene expression in population of *E. coli* cells
with the gene circuits.**

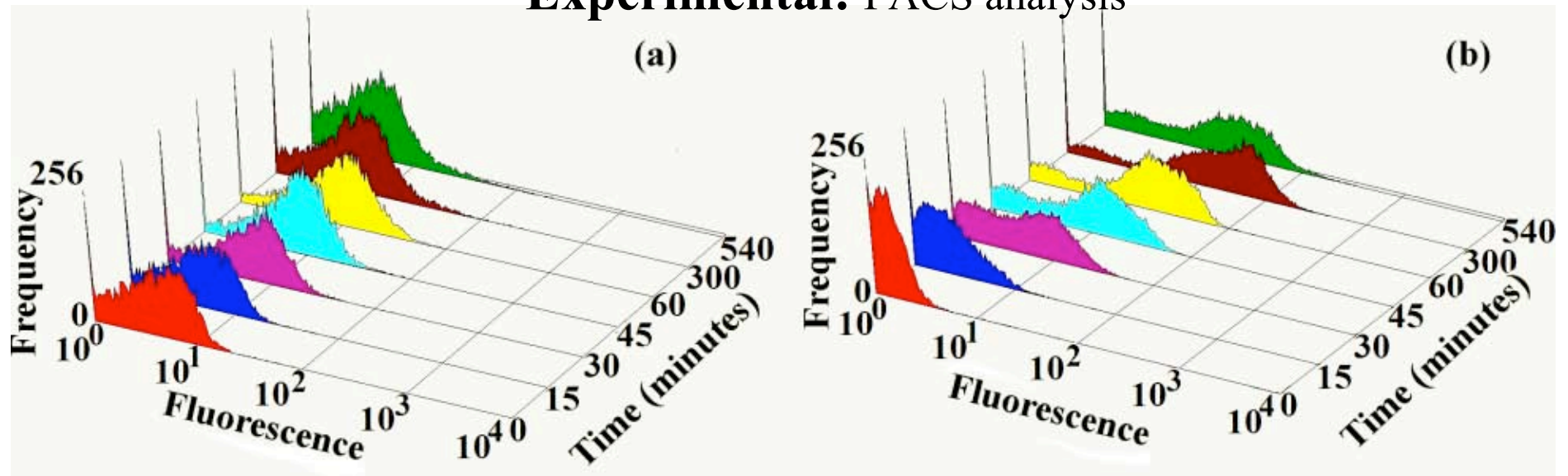
Model prediction

and

Experimental measurements.

Intra-population heterogeneity in gene expression

Experimental: FACS analysis



Frequency distributions of GFP in cell populations:

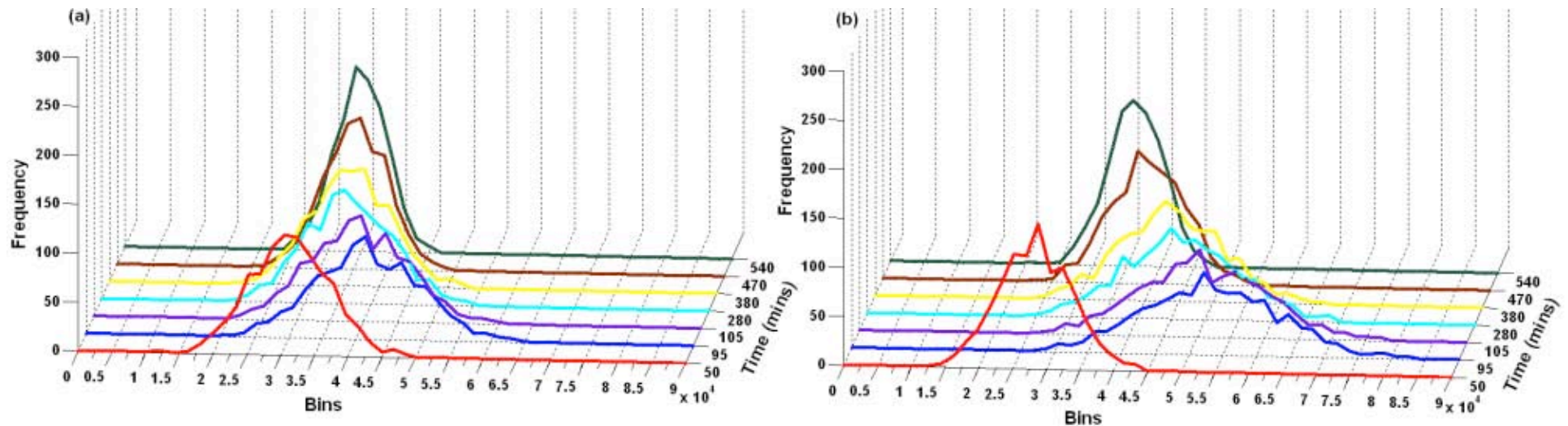
(a) Basic (TG), and (b) Delay (C2TG) circuits at different time intervals.

The population of cells with Delay circuit shows –

- (i) A significantly higher fluorescence in time, which later return to lower levels;
- (ii) Presence of bimodality in fluorescence distribution; and,
- (iii) Broader distribution of fluorescence in cell population - larger heterogeneity in gene expression among the individual cells within a population


Intra-population heterogeneity in gene expression

Theoretical: GFP expression at different time points in model 1000 cells with both circuits having plasmid copy number variation (50 ± 10 , normally distributed)



Frequency distributions of GFP molecules in cell populations:
(a) Basic (TG), and (b) Delay (C2TG) circuits at different time intervals.

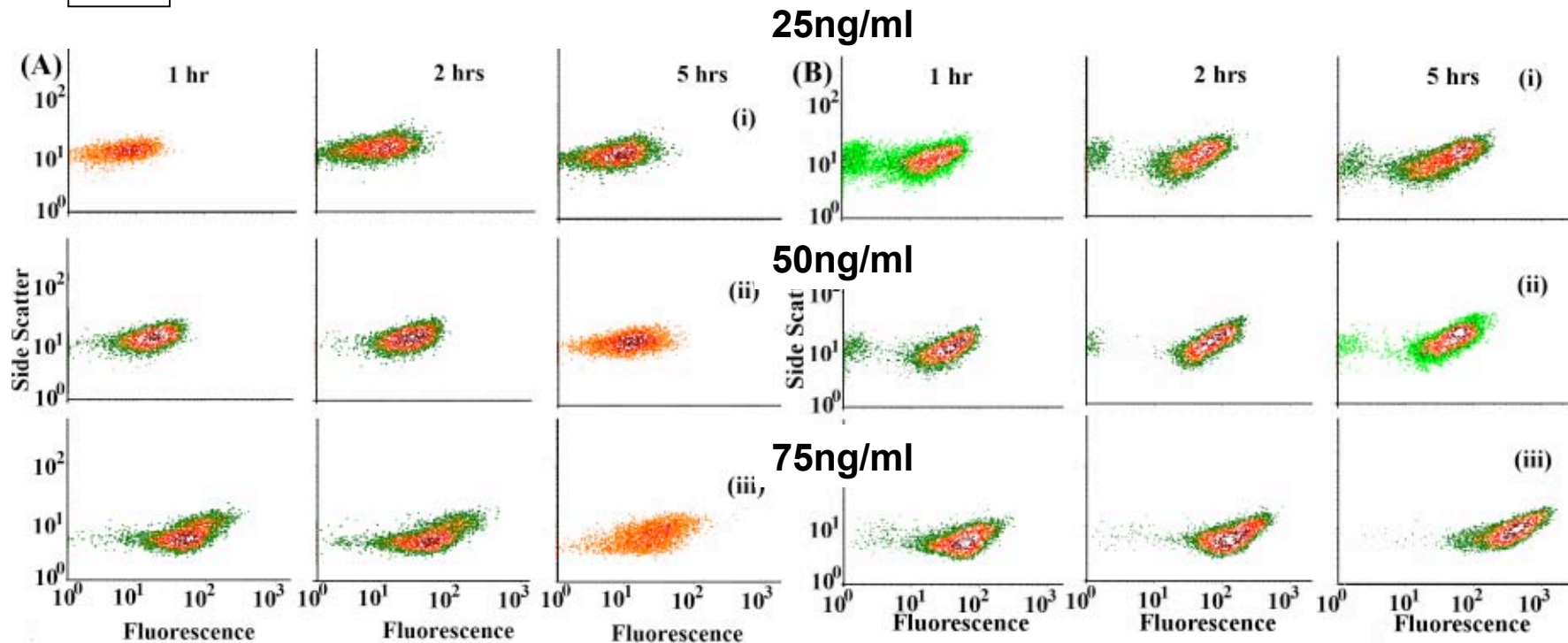
The population of cells with Delay circuit shows –

- (i) A significantly **higher fluorescence** in time, which later return to lower levels;
- (ii) **No bimodality** in fluorescence distribution; and, 
- (iii) Broader distribution of fluorescence in cell population - **larger heterogeneity** in gene expression among the individual cells within a population

Bimodality

TG

C2TG



Contour plots of GFP fluorescence distribution in cell populations.

At 1, 2 and 5 hrs after induction with different inducer concentrations - (i) 25ng/ml, (ii) 50 ng/ml, and (iii) 75ng/ml, of Doxycycline.

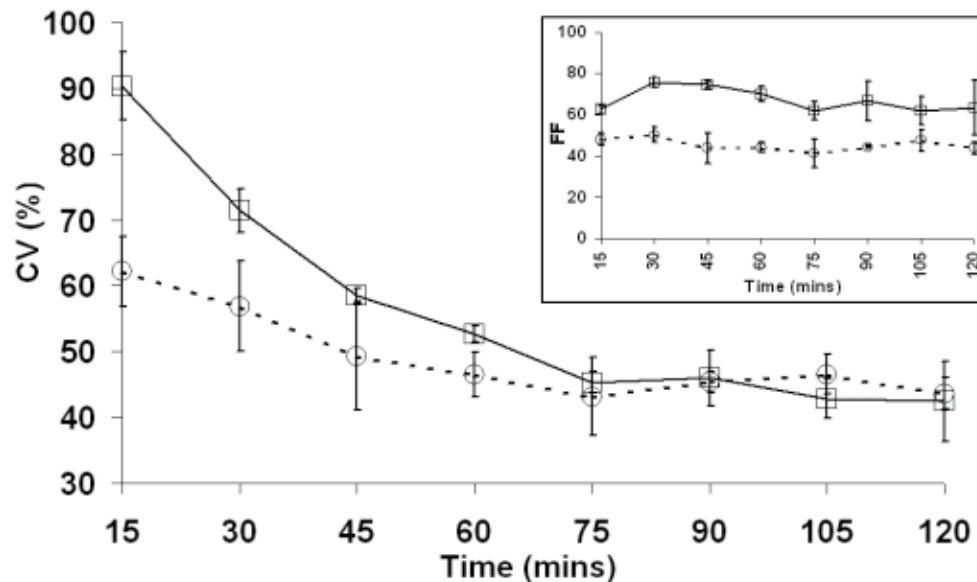
The presence of bimodality in Delay circuit cell populations, induced at 25ng/ml, is a consequence of, but not an inherent property of, the delay element in the circuit. Removal of this low-expressing fraction of cells by gating shows that C2TG continues to have a greater spread than TG.

Heterogeneity of gene expression in a population of cells

Common measures of comparing variability (noise) in a system -

Coefficient of Variation $CV = (\text{standard deviation}/\text{mean}) * 100$
and Fano Factor $FF = \text{variance}/\text{mean}$.

Experiments



Basic (TG - *dashed line + circle*)
Delay (C2TG - *solid line + square*)

Inset:

Changes in Fano Factor (FF) for both the circuits.

- During the time of the build-up of the overshoot (till 90 minutes):

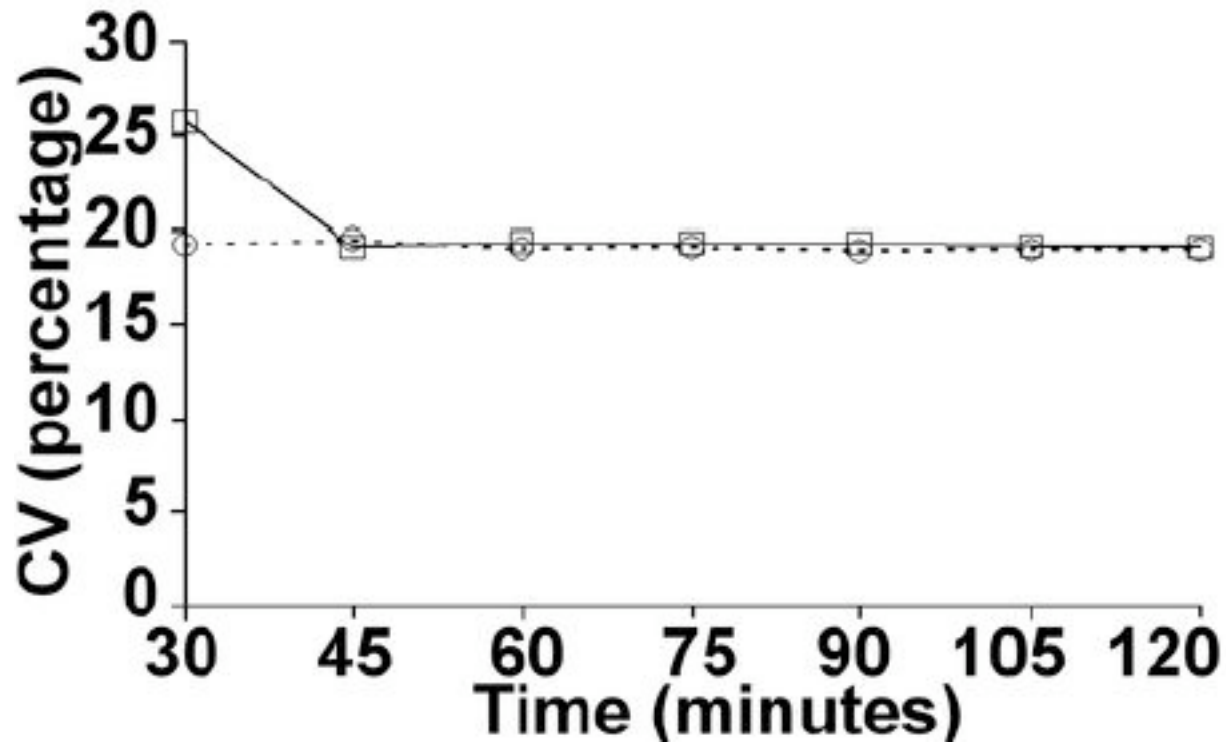
CV of the Delay population > Basic population

- The initial decreasing trend in CV in both the circuits indicates reduction in the intrinsic noise levels with time due to the establishment of the repression.
- Fano Factor shows continuing difference between the two circuits. Delay circuit exhibits greater variability compared to the Basic circuit.

Model Delay and Basic circuits -

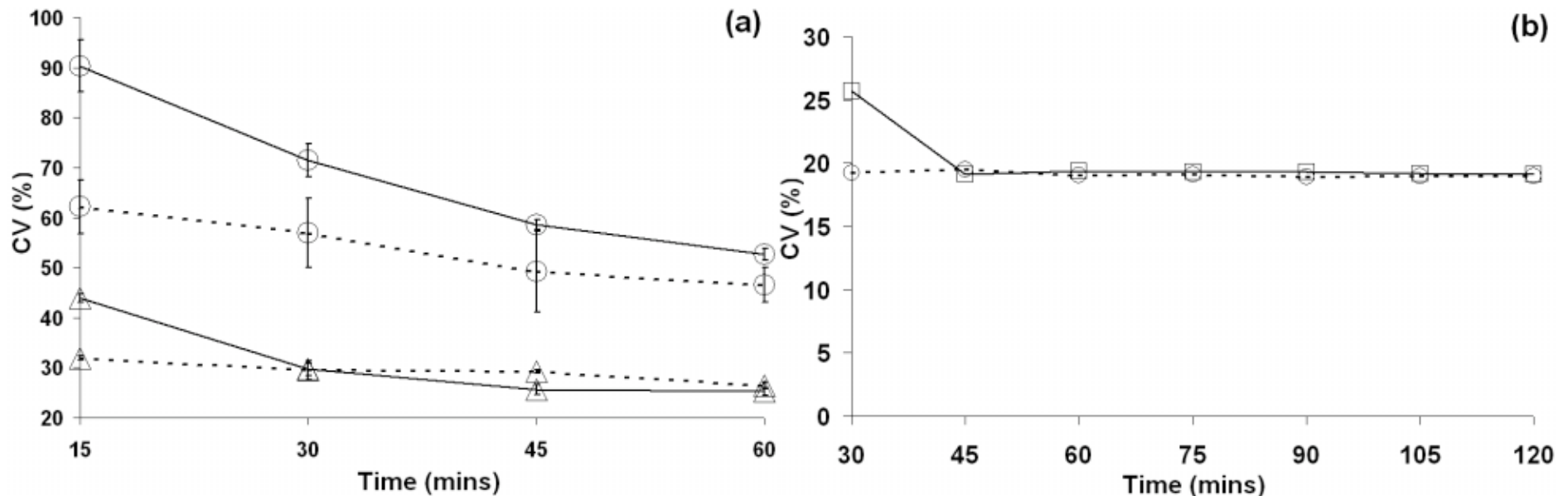
No significant difference in their CV over time except at an early time point.

Prediction is not consistent with the experimental results ?



This prediction is consistent with experimental results.

The experimental Delay circuit at 75ng/ml induction shows similar difference in CV as is seen in the model circuits.



Coefficient of Variation for both Basic (dashed lines) and Delay (solid lines) circuits:

- a) Experimental populations for inducer concentrations (25 ng/ml - circles, 75ng/ml - triangles) till 60 min from three experiments;
- b) Theoretical simulation (Basic: dashed line and Delay: solid line).

The hypothesis that delay in repression is the primary factor for inducing increased inter-cellular heterogeneity in gene expression in a population is shown theoretically and experimentally.

CONCLUSIONS

Robustness of the results

Experimental methodology used involved –

- 1) Population approach (Fluorimetry) – observations on $\sim 10^9$ cells
- 2) Cell level (FACS) – observation on $\sim 10^4$ cells

Theoretical model only highlighting the delay in repression

– calculations on $\sim 10^3$ cells

(no consideration of real factors, e.g., cell size changes, growth, folding delays of GFP, nonlinearities involved in degradation, etc)

**All three approaches show
“Overshoot” and “Heterogeneity” in gene expression**

CONCLUSIONS

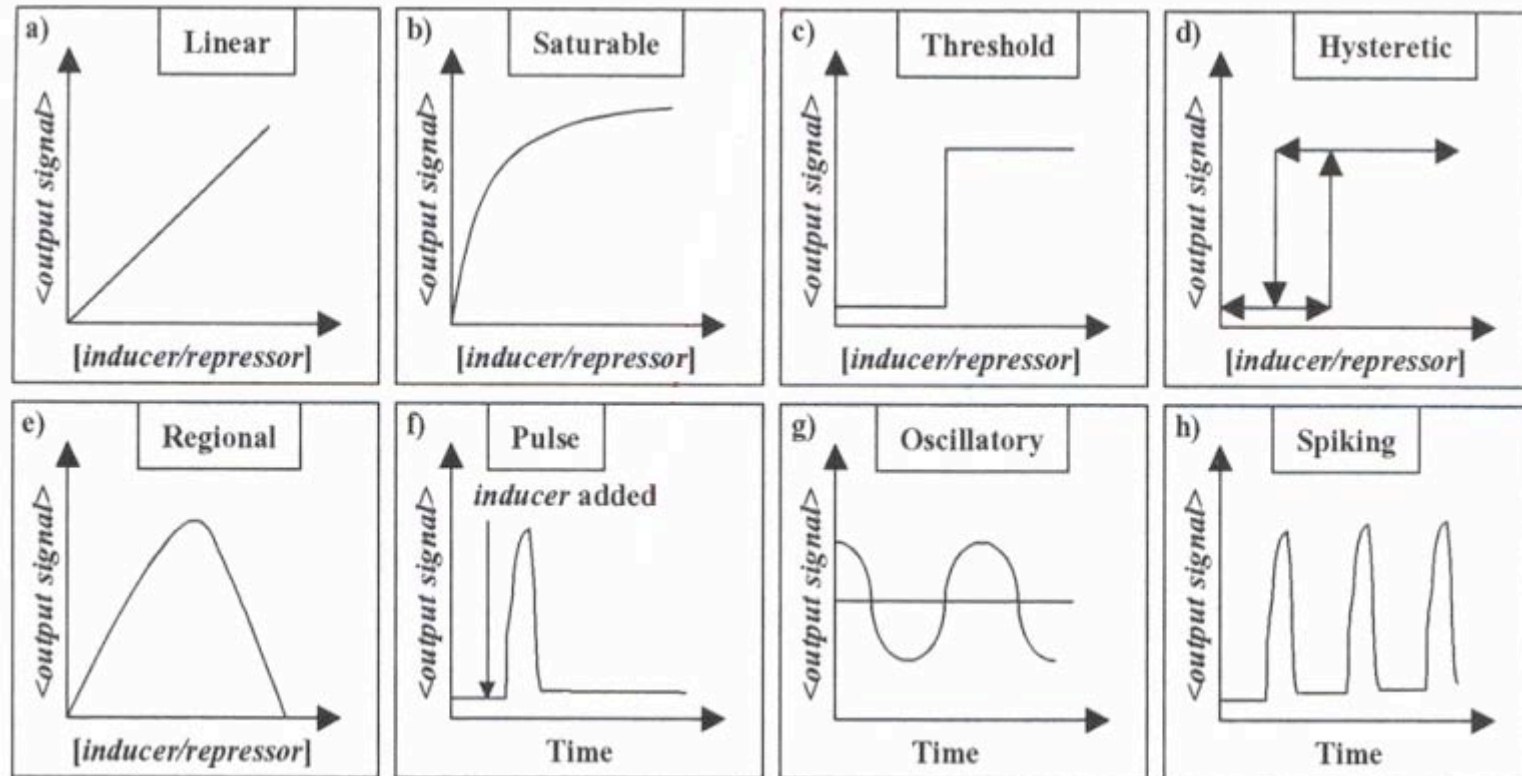
The generic origin of delay in biochemical pathways implies that there is a high likelihood that the two properties - **transient overshoot & generation of heterogeneity in gene expression in cell population** - play important roles in gene regulation.

- The overshoot allows for gene products being available in large amount for multi-step pathways to function,
- It can also act as a dominant source of large deterministic variability paving way to increase the phenotypic diversity in a population of cells before the negative regulation sets in.

Our theoretical and experimental results provide important clues and give possible rationale for delayed feedbacks to be such a generic feature in gene organisations in cells.

Development of Gene Circuits *The Circuit Engineering Vision*

Develop a standard library of interoperable “parts” that corresponds to various control functions (www.parts.mit.edu)



Develop integrated computational infrastructure for

Computer Aided Design (CAD) of genetic circuits

Simulation and dynamic analysis

Build increasingly complex genetic circuits using well-characterized parts

We have the "parts list"

How do these "parts" interact as a "whole", and how does this system function to create an organism?

**Ultimate goal is to link
behaviour of cells, organisms, and populations
to the information encoded in the genome**

"Systems Biology"

is about identifying, characterizing, and integrating the
parts-lists of complex biological systems
to find the underlying design and working principles of the
biological computational units

Acknowledgements

Dr. R. Maithreye

Dr. R. R. Sarkar

Dr. Veena K. Parnaik

Dr. Gopal Pande

Council for Scientific & Industrial Research (CSIR)

Department of Biotechnology

