Computational Systems Biology

. . . Biology X – Lecture 7 . . .

Bud Mishra

Professor of Computer Science, Mathematics, & Cell Biology

Systems Biology

Goal

- The goal of this subject is to understand, design and create large-scale computational system centered on the biology of
  - individual cells,
  - population of cells,
  - intra-cellular processes, and
  - realistic simulation, visualization and reasoning about these processes at multiple spatio-temporal scales.

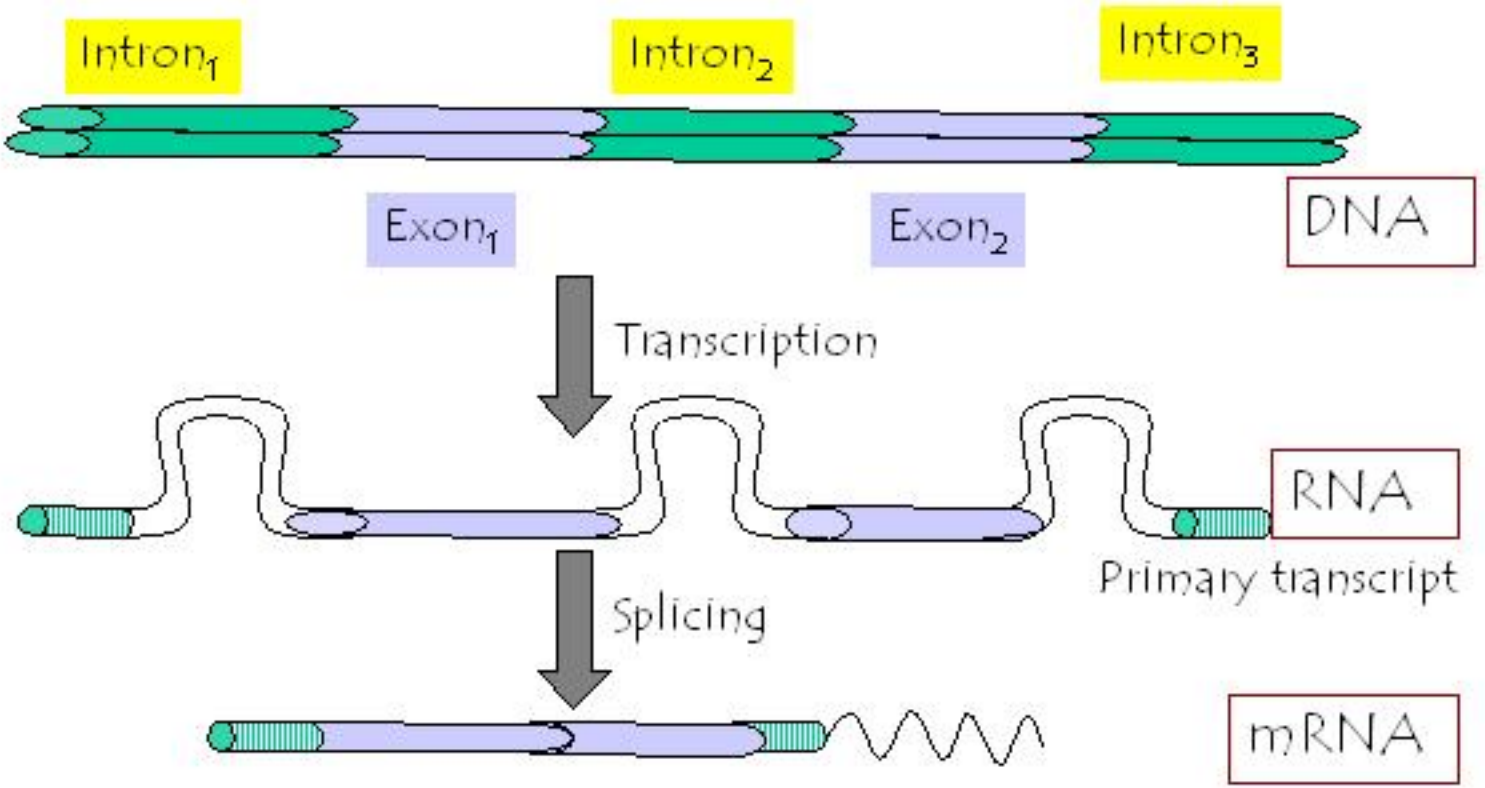
Why?

- Such a reasoning system, in the hands of a working biologist, can then be used to
  - gain insight into the underlying biology,
  - design refutable biological experiments, and
  - ultimately, discover intervention schemes to suitably modify the biological processes for therapeutic purposes.

Interrupted Genes:

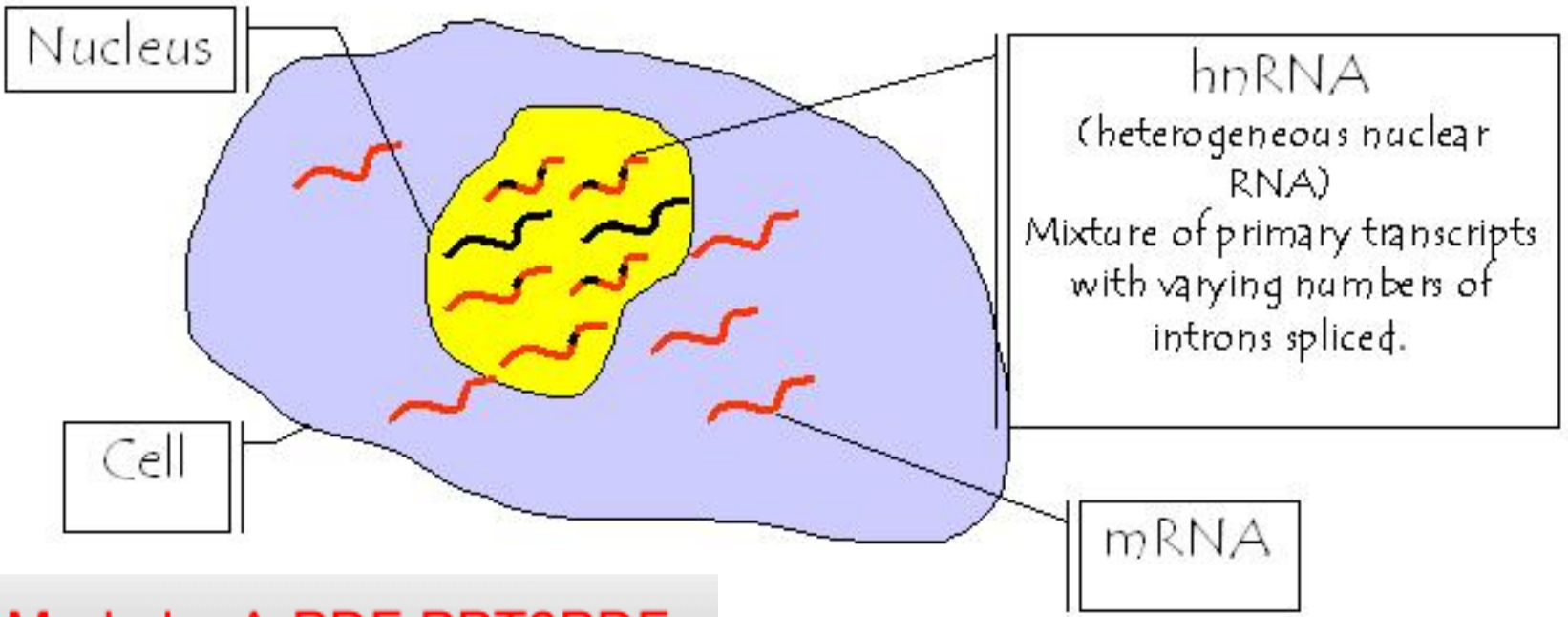
- An open reading frame (containing a gene) consists of
  - INTRONS: Intervening sequences ↦ Noncoding regions
  - EXONS: Protein coding regions
- Introns are abundant in eukaryotes and certain animal viruses.

Interrupted Genes:



Interrupted Genes:

- Introns can occur between individual codons or within a single codon

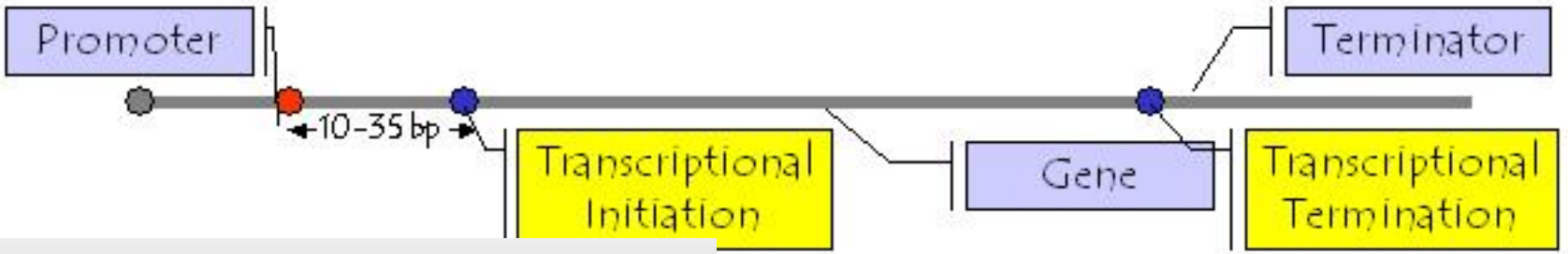


Some Genes...

Gene Product	Organism	Exon Length	#Introns	Intron Length
Adenoshine deaminase	Human	1500	11	30,000
Apolipoprotein B	Human	14,000	28	29,000
Erythropoietin	Human	582	4	1562
Thyroglobulin	Human	8500	≧40	100,000
α-interferon	Human	600	0	0
Fibroin	Silk Worm	18,000	1	970
Phaseolin	French Bean	1263	5	515

Regulation of Gene Expns

- Motifs (short DNA sequences) that regulate transcription
  - Promoter
  - Terminator
- Motifs that modulate transcription
  - Repressor
  - Activator
  - Antiterminator



Promoters

- pol I (RNA polymerase I)
  - Transcribes ribosomal RNA genes 100 ~ 1000 bp in front of the gene
- pol II (RNA polymerase II)
  - Transcribes genes encoding polypeptides
  - Complex and variable regulatory regions
- pol III (RNA polymerase III)
  - Transcribes transfer RNA and other small RNAs
  - Both up and down stream

Motifs

- Each motif is a binding site for a specific protein
- Transcription Factor:
  - Transcription factors (specific to a cell/environmental conditions) bind to regulatory regions and facilitate
    - Assembly of RNA polymerase into a transcriptional complex
    - Activation of a transcriptional complex.
- Termination Factor:
  - Assembly of proteins for termination and modification of the end of the RNA
- Epigenetic Changes
  - Methylation of the cytosine in the 5' region
  - Structural changes in chromatin

Organization of Genetic Info

- Bacterial Genome:
  - Genes are closely spaced along the DNA.
  - The sequences of genes may overlap.
  - Related genes (encoding enzymes whose functions are part of the same pathway or whose activities are related) are linked as a single transcription unit.

Organization of Genetic Info

- Eukaryotic Genome:
  - Genes are separated by long stretches of noncoding DNA sequences.
  - Multiple genes in a single transcription unit is extremely rare.
  - Multiple chromosomes – Linear
  - Chloroplasts and mitochondria – Circular
  - Genes appearing on the same chromosome are syntenic.

Gene Locations

Genes	chromosomes
α-globin cluster	16
β-globin cluster	11
Immunoglobulin	
κ (light chain)	2
λ (light chain)	22
Heavy Chain	14
Pseudogenes	9,32,15,18
Growth Hormone gene cluster	17

Genes	chromosomes
Insulin	11
Galactokinase	11
Viral oncogene homologues	
C-sis	22
C-mos	8
C-Ha-Ras-1	11
C-myb	6
Interferons	
α & β cluster	9
γ	12

Eukaryotic Genome

- Multiple copies of the same gene
  - Solve "supply problem"
  - There are several hundred ribosomal RNA genes I mammals
- Pseudogenes
  - Nonfunctional copies of genes...(Deletions or alterations in the DNA sequence)
  - Number of pseudo genes for a particular gene varies greatly...Different from one organism to another.

Genes in Eukaryotes

- A gene may appear exactly once
- It may be part of a family of repeated sequence . Members of a family may be clustered or dispersed.
- Members of a gene family may be related and functional (expressed at different times in development, or in different cells) or may be pseudo genes.
- **Chromosomal Morphology**:
  - *Nucleolar organizers* (genes for ribosomal RNA)
  - *Telomeric and Centromeric regions* (Tandemly

Genome Rearrangement

- Reshuffling of genes between homologous chromosomes via reciprocal crossing-over during both meiosis and mitosis.
- Gene synteny and linkages are usually preserved.
- Most rearrangements are random.
- Some rearrangements are normal processes altering gene expressions in an orderly and programmed manner.

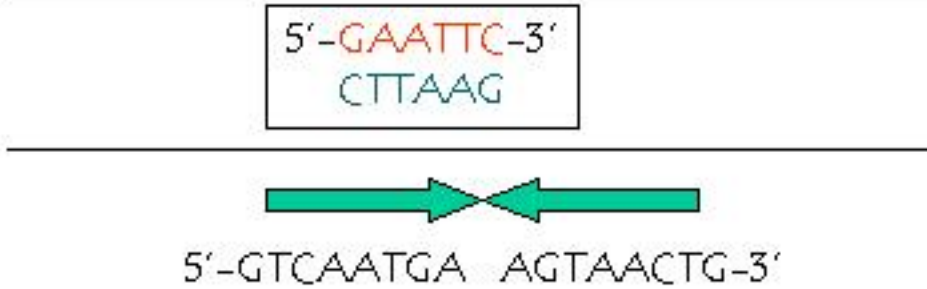
Repeat Structure

- Copy Number: $2 \sim 10^6$
- Direct Repeats "head-to-tail"
  - Tandem repeats or separated by other sequences
- Inverted Repeats "head-to-head"
  - Stem-and-loop structure
  - Hairpin structure
- Reverse Palindrome
- True Palindrome

Repeat Structure

- Tandem Direct Repeats
- Inverted Repeats

- Reverse Palindrome
- True Palindrome



Repeats within the Genome

- Gene Family
  - Genes and its cognate pseudogenes
- Satellite: Repeats made of noncoding units
  - Minisatellites: Tandem repeats…Mostly in centromeric regions
  - Satellite repeat units vary in length from 2 base pairs to several thousands.

Interspersed Repeats

- SINES: Short Interspersed Repeats
  - Each repeat unit is of length 100 – 500 bps
  - Processed pseudogenes derived from class III genes
  - Example: Alu repeats…dimeric head-to-tail repeats of 130 bp
- LINES: Long Interspersed Repeats
  - Each unit is of length > 6 Kb.

The Cell

- A cell is a small coalition of a set of genes held together in a set of chromosomes (and even perhaps unrelated extrachromosomal elements).
- They also have set of machinery made of proteins, enzymes, lipids and organelles taking part in a dynamic process of information processing.

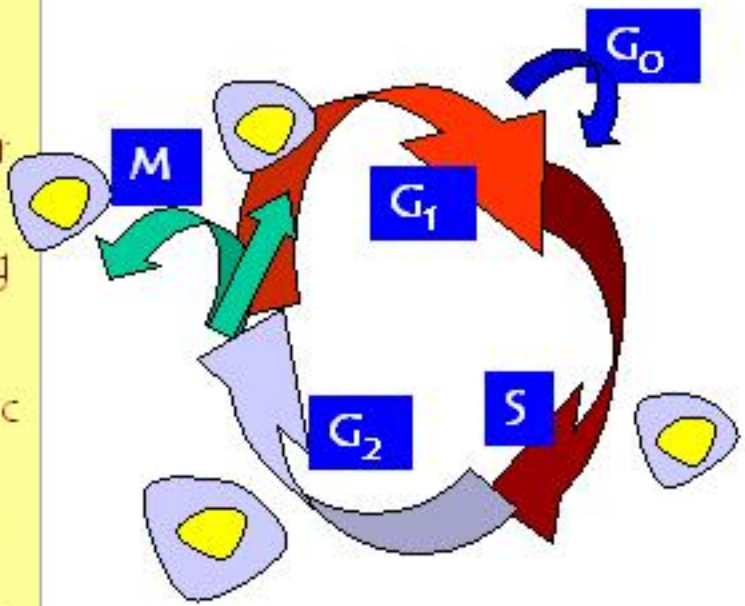
The Cell Cycle

The dynamics of cell:

- The cell cycle) the set of events that occur within a cell between its birth by mitosis and its division into daughter cells again by mitosis
  - interphase period when DNA is synthesized and
  - mitotic phase
    - The cell division by mitosis (into 2 daughter cells) and meiosis (into 4 gametes from germ-line cells);
    - Working of the machinery within the cell---mainly the ones involving replication of DNA, transcription of DNA into RNA and translation of RNA into protein.

The Cell Cycle:

- In growing cells, the four phases proceed successively, taking from 10-20 hrs.
- Interphase: comprises the G_1, S, and G_2 phases. DNA is synthesized in S and other cellular macromolecules are synthesized throughout interphase, roughly doubling cell's mass.
- During G_2 the cell is prepared for mitotic (M) phase when the genetic material is evenly proportioned and the cell divides.
- Nondividing cells exit the normal cycle, entering the quiesecent G_0 state.

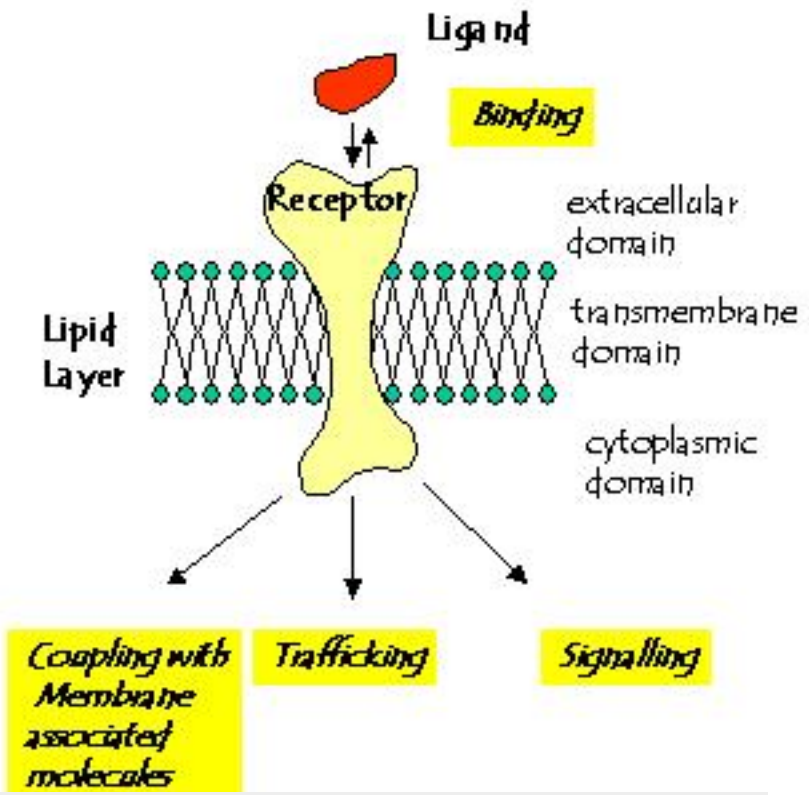


Differentiation & Suicide

- Cellular dynamics controls how a cell changes (or differentiates) to carry out a specialized functions
  - Structural or morphological changes (muscles, neural, skin..)
  - Immune systems: Many cell types come together in organized tissues designed to let the body distinguish self from non-self.
- Programmed Cell Death/Apoptosis:
  - Condensation of the nucleus.
  - Fragmentation of the DNA.
  - Morphological changes followed by consumption by macrophages.

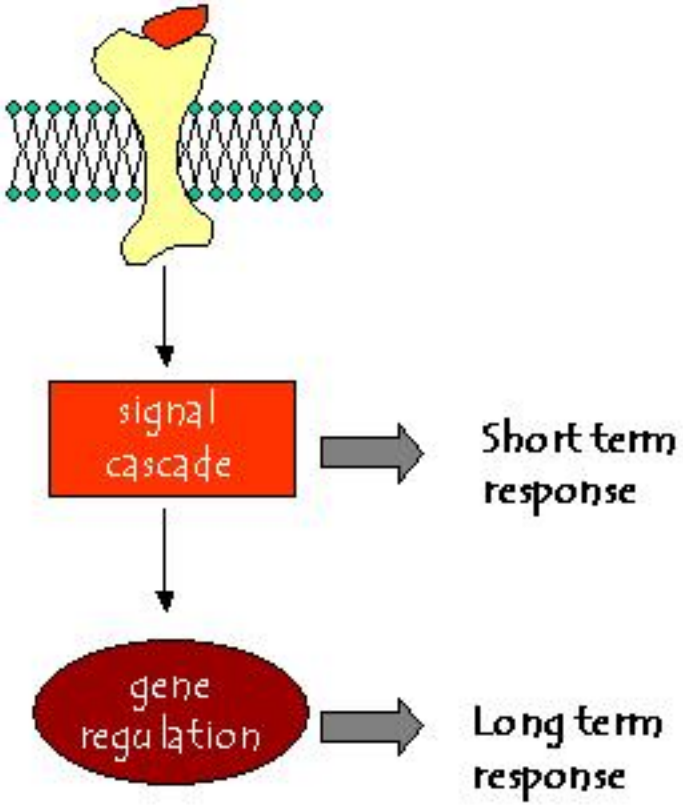
Cell Talk

Cell Talk



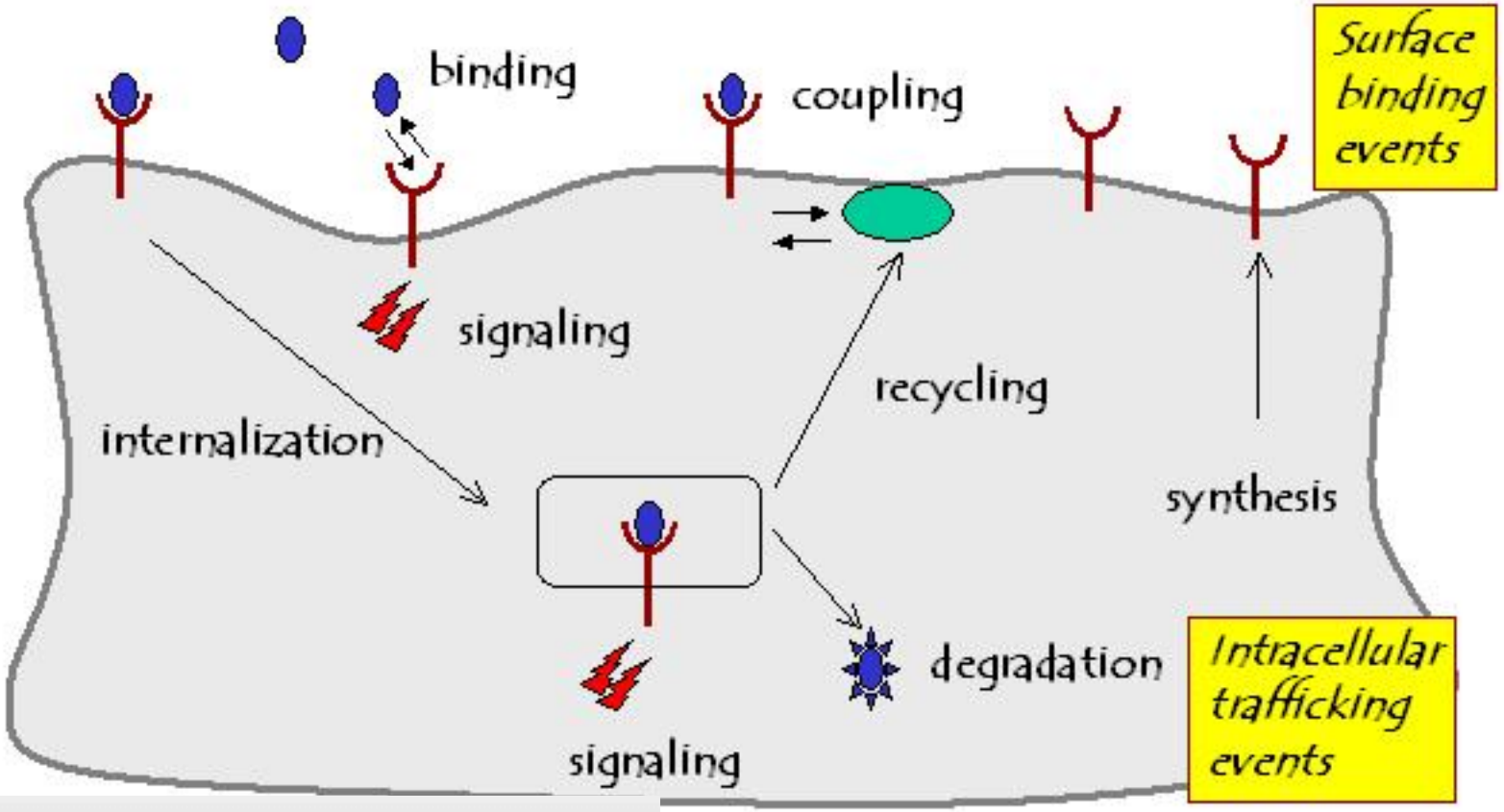
- Cell Surface Receptors
  - Extracellular domain for binding ligands (e.g., growth factors, adhesion molecules, etc.)
  - Transmembrane domain
  - Intracellular cytoplasmic domain
- Receptor driven cellular behavior are extremely important
  - E.g., Growth, Secretion, Contraction, Motility and Adhesion

Receptors and Gene Regulation



- Ligands bind to receptors at the cell surface.
- Bound receptors activate various intracellular enzymes and initiate entire cascades of intracellular reactions
  - Some of these regions trigger short term (of the order of milliseconds to minutes) responses.
  - Some eventually trigger long-term responses..e.g., requiring protein synthesis and additional molecular interactions

A Complex Picture



A Complex Picture

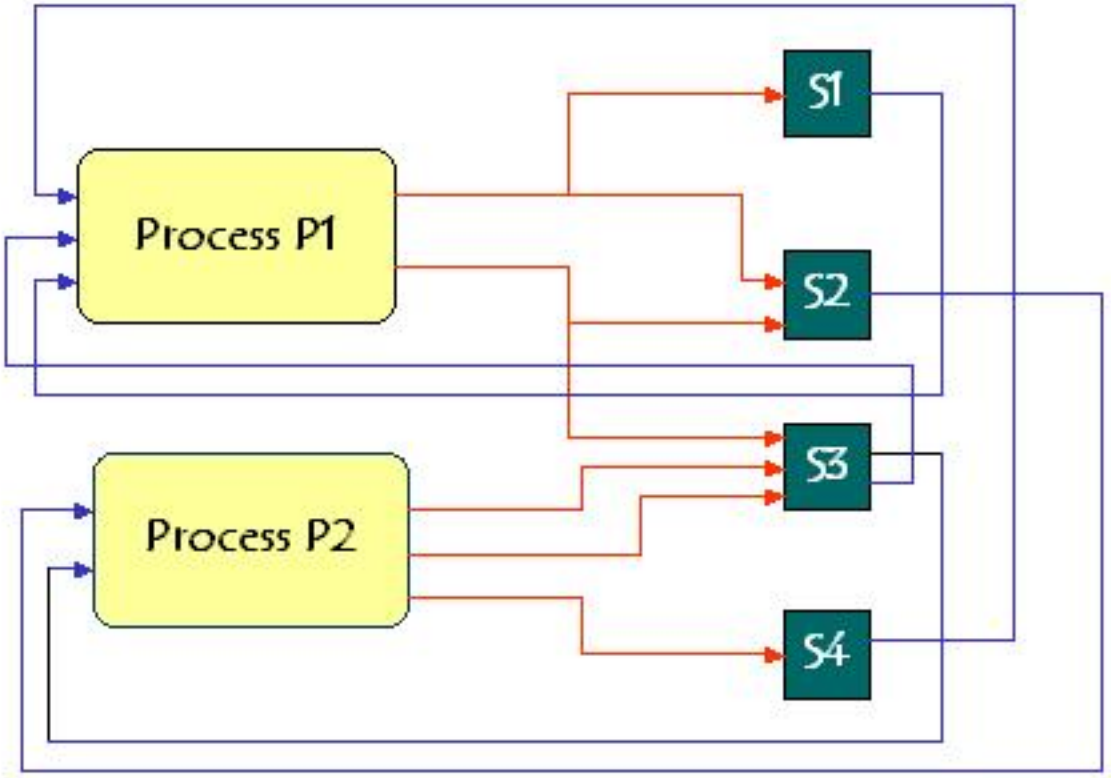
- Trafficking
  - Receptor population undergoes many complex events of coupling with other cell surface molecules
  - Internalization (RME: receptor-mediated endocytosis)
  - Recycling
  - Degradation
  - Synthesis

Modeling

Modeling Biomolecular Networks

- Agents and Modes:
  - **Species and Processes:** There are two kinds of agents:
    - S-agents (representing species such as proteins, cells and DNA): S-agents are described by concentration (i.e., their numbers) and its variation due to accumulation or degradation. S-agent's description involves differential equations or update equations.
    - P-agents (representing processes such as transcription, translation, protein binding, protein-protein interactions, and cell growth.) Inputs of P-agents are concentrations (or numbers) of species and outputs are rates.

P-agents and S-agents



Agents & Modes

- Each agent is characterized by a state $x \in \mathbb{R}^n$ and
- A collection of discrete modes denoted by Q
- Each mode is characterized by a set of differential equations ($q_i \in Q$ & $z \in \mathbb{R}^p$ is control)

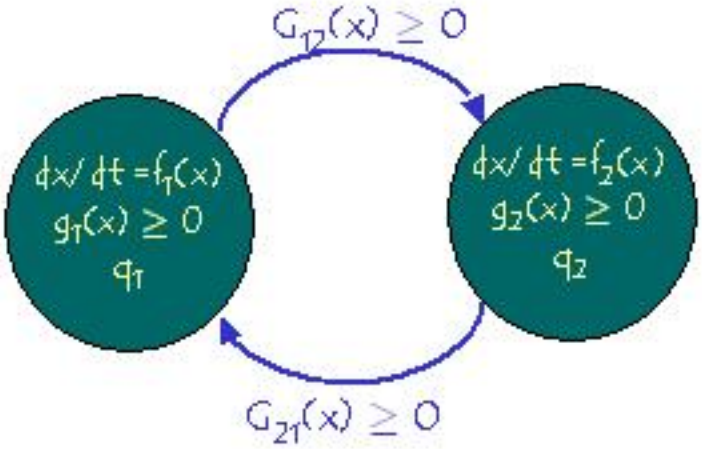
$$dx/dt = f_{q_i}(x,z),$$

  - and a set of invariants that describe the conditions under which the above ODE is valid…
  - these invariants describe algebraic constraints on the continuous state…

Mode Definition

- Modes are defined by the transitions among its submodes.
- A transition: specifies source and destination modes, the enabling condition, and the associated discrete update of variables.
- Modes and submodes are organized hierarchically.

Example of a Hybrid System



- q_1 and q_2 = two discrete modes
- x = continuous variable evolving as
  - $dx/dt = f_1(x)$ in mode q_1
  - $dx/dt = f_2(x)$ in mode q_2
- Invariants: Associated with locations q_1 and q_2 are
  - $g_1(x) \geq 0$ and $g_2(x) \geq 0$, resp.
- The hybrid system evolves continuously in disc. mode q_1 according to $dx/dt = f_1(x)$ as long as $g_1(x) \geq 0$ holds.
- If ever x enters the "guard set" $G_{12}(x) \geq 0$, then mode transition from q_1 to q_2 occurs.

Generic Equation

- Generic formula for any molecular species (mRNA, protein, protein complex, or small molecule):

$$dX/dt = \text{synthesis} - \text{decay} \pm \text{transformation} \pm \text{transport}$$

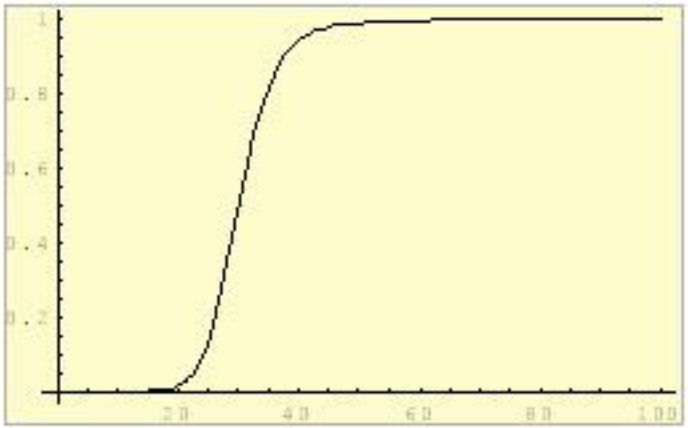
- Synthesis:
  - replication for DNA,
  - transcription of mRNA,
  - translation for protein
- Decay: A first order degradation process
- Transformation:
  - cleavage reaction
  - ligand binding reaction
- Transport: Diffusion through a membrane..

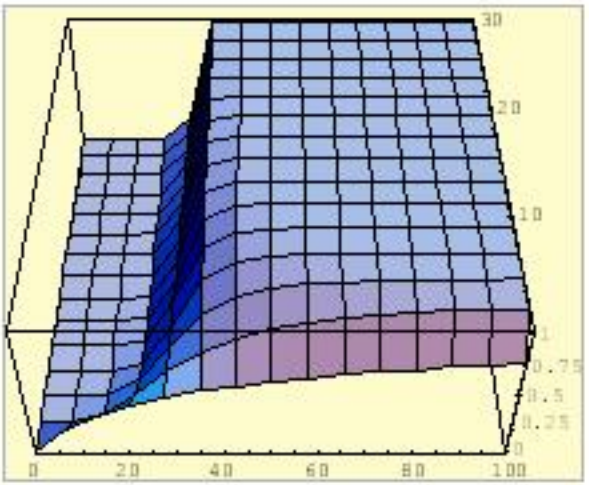
Model of transcription

- ν_{Xm} = Cooperativity coefficient
- κ_{Xm} = Concentration of X at which transcription of m is "half-maximally" activated.

- $\Phi(X, \kappa_{Xm}, \nu_{Xm}) = X^{\nu}/[\kappa^{\nu} + X^{\nu}]$
- $\Psi(X, \kappa_{Xm}, \nu_{Xm}) = \kappa^{\nu}/[\kappa^{\nu} + X^{\nu}] = 1 - \Phi(X, \kappa_{Xm}, \nu_{Xm})$
- A graph of function Φ = Sigmoid Function

Transcription Activation Function

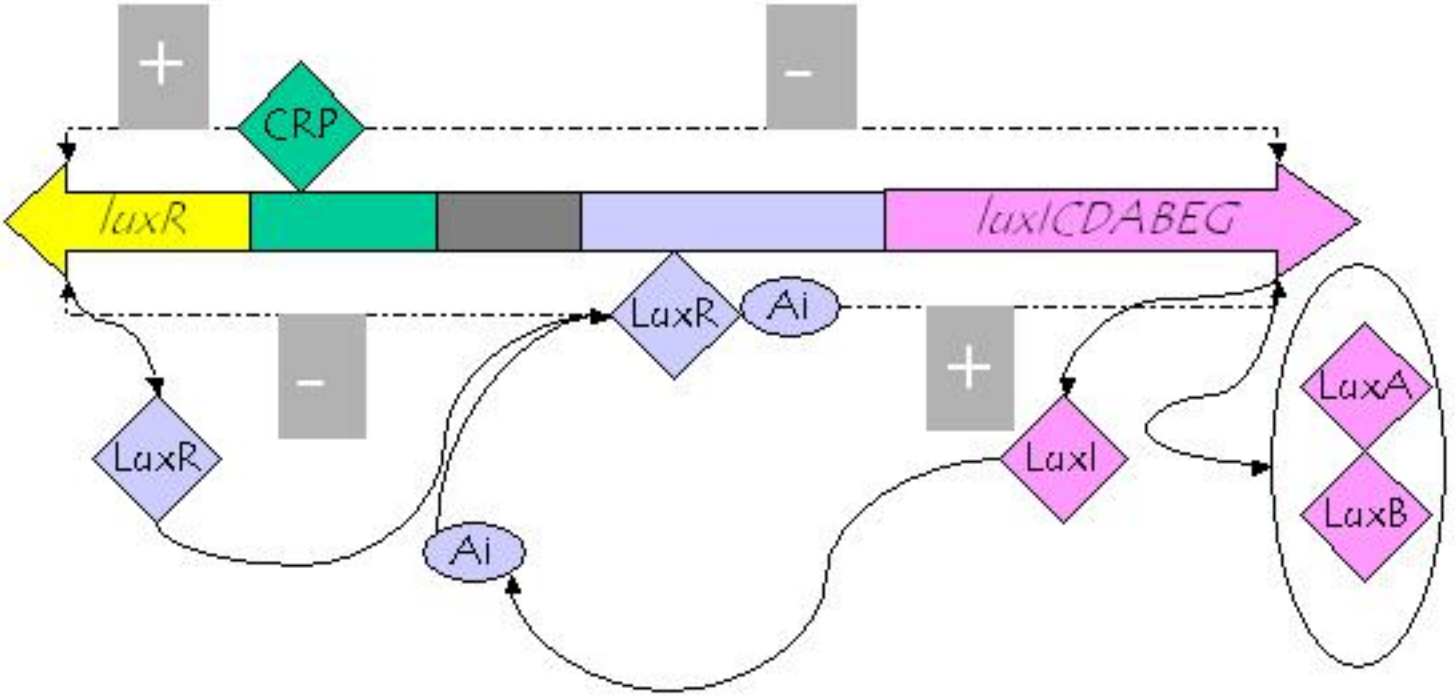




Quorum Sensing in V. fischeri

- Cell-density dependent gene expression in prokaryotes
  - Quorum = A minimum population unit
- A single cell of V. fischeri can sense when a quorum of bacteria is achieved—leading to bioluminescence…
- Vibrio fiscehri is a marine bacterium found as
  - a free-living organism, and
  - a symbiont of some marine fish and squid.
    - As a free-living organism, it exists in low density is non-luminescent..
    - As a symbiont, it lives in high density and is luminescent..
    - The transcription of the lux genes in this organism controls this luminescence

lux gene



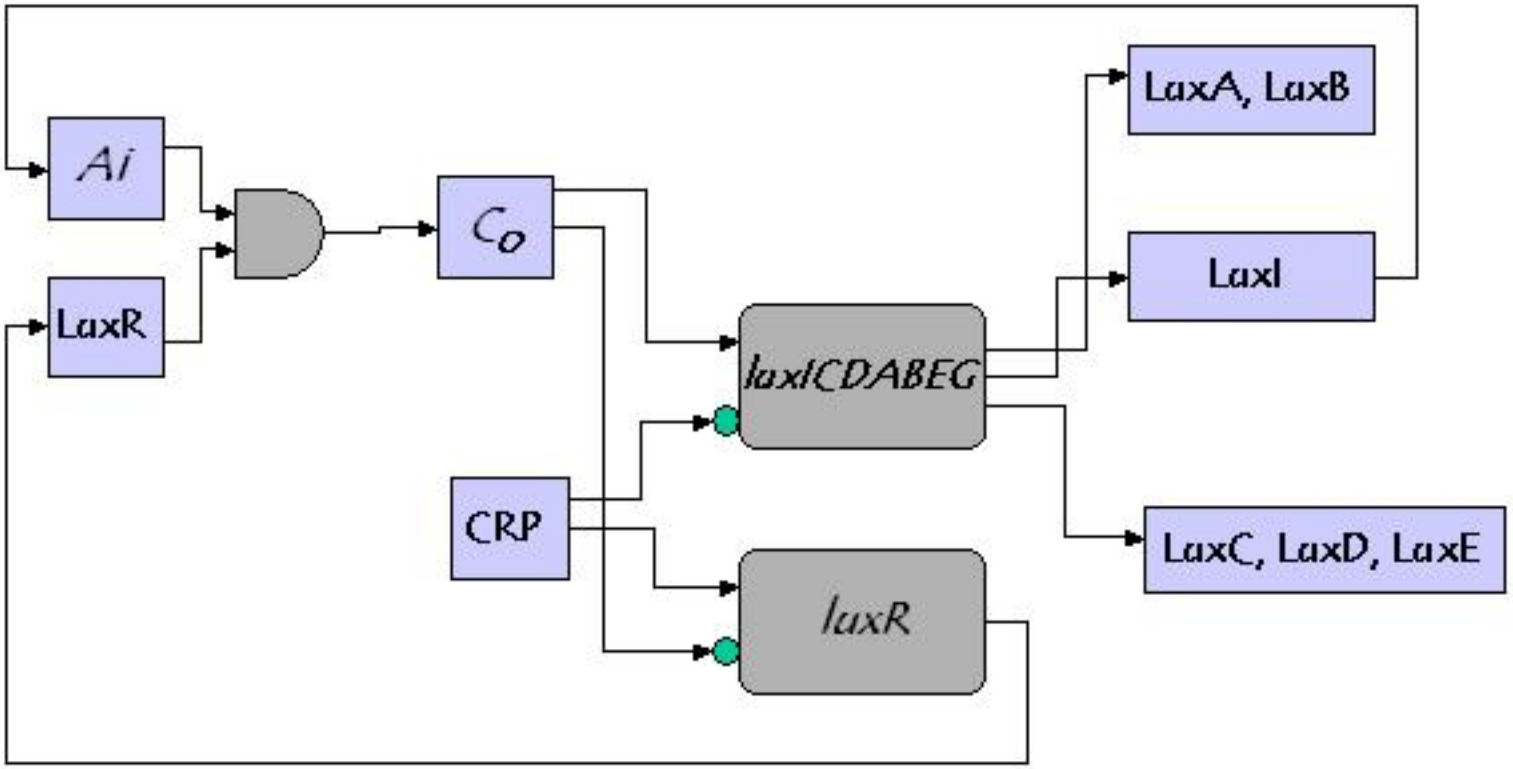
Quorum Sensing

- The *lux* region is organized in two transcriptional units:
  - O_L: containing *luxR* gene (encodes protein LuxR = a transcriptional regulator)
  - O_R: containing 7 genes *luxICDABEG*.
    - Transcription of *luxI* produces the protein LuxI, required for endogenous production of the autoinducer *AI* (a small membrane permeable signal molecule (acyl-homoserine lactone).
    - The genes *luxA* & *luxB* code for the luciferase subunits
    - The genes *luxC*, *luxD* & *luxE* code for proteins of the fatty acid reductase, needed for aldehyde substrate for luciferase.
    - The gene *luxG* encodes a flavin reductase.
    - Along with LuxR and LuxI, cAMP receptor protein (CRP) controls luminescence.

Biochemical Network

- The autoimmune inducer *Ai* binds to protein LuxR to form a complex C_0 which binds to the *lux box*.
- The *lux box* region (between the transcriptional units) contains a binding site for CRP.
- The transcription from the *luxR* promoter is activated by the binding of CRP.
- The transcription from the *luxICDABEG* is activated by the binding of C_0 complex to the *lux box*.
- Growth in the levels of C_0 and cAMP/CRP inhibit *luxR* and *luxICDABEG* transcription,

Biochemical Network



Notation

- x_0 = scaled population
- x_1 = mRNA transcribed from O_L
- x_2 = mRNA transcribed from O_R
- x_3 = protein LuxR
- x_4 = protein LuxI
- x_5 = protein LuxA/B
- x_6 = protein LuxC/D/E
- x_7 = autoinducer *Ai*
- x_8 = complex C_0

Evolution Equations...

- $dx_0/dt = k_G x_0$
- $dx_1/dt = T_c[\Psi(x_8, \kappa_{C0}, \nu_{C0})\ \Phi(c_{CRP}, \kappa_{CRP}, \nu_{CRP})+b] - x_1/H_{RNA} - k_G x_1$
- $dx_2/dt = T_c[\Phi(x_8, \kappa_{C0}, \nu_{C0})\ \Psi(c_{CRP}, \kappa_{CRP}, \nu_{CRP})+b] - x_2/H_{RNA} - k_G x_2$
- $dx_3/dt = T_l x_1 - x_3/H_{sp} - r_{AiR} x_7 x_3 - r_{C0} x_8 - k_G x_3$
- $dx_4/dt = T_l x_2 - x_4/H_{sp} - k_G x_4$
- $dx_5/dt = T_l x_2 - x_5/H_{sp} - k_G x_5$
- $dx_6/dt = T_l x_2 - x_6/H_{sp} - k_G x_6$
- $dx_7/dt = x_0(r_{All} x_4 - r_{AiR} x_7 x_3 + r_{C0} x_8) - x_7/H_{Ai}$
- $dx_8/dt = r_{AiR} x_7 x_3 - x_8/H_{sp} - r_{C0} x_8 - k_G x_8$

Parameters

T_c	Max. transcription rate	ν_{CRP}	Cooperativity coef for CRP
T_l	Max. translation rate	κ_{CRP}	Half-max conc for CRP
H_{RNA}	RNA half-life	ν_{C0}	Cooperativity coef for C_o
H_{sp}	Stable protein half-life	κ_{C0}	Half-max conc for C_o
H_{up}	Unstable protein half-life	b	Basal transcription rate
H_{Ai}	$A/$ half-life	v_b	Volume of a bacterium
r_{AII}	Rate constant: LuxI $\rightarrow A/$	V	Volume of solution
r_{AiR}	Rate constant: $A/$ binds to LuxI	k_g	Growth rate
r_{C0}	Rate constant: C_o dissociates	x_{0max}	Maximum Population

Remaining Questions

- Simulation:
  - Nonlinearity
  - Hybrid Model (Piece-wise linear)
- Stability Analysis
- Reachability Analysis
- Robustness

To be continued…

. . .