“The 2nd IMSc Complex Systems School”, Chennai Jan 7-8, 2010

Mathematical Models in Biology

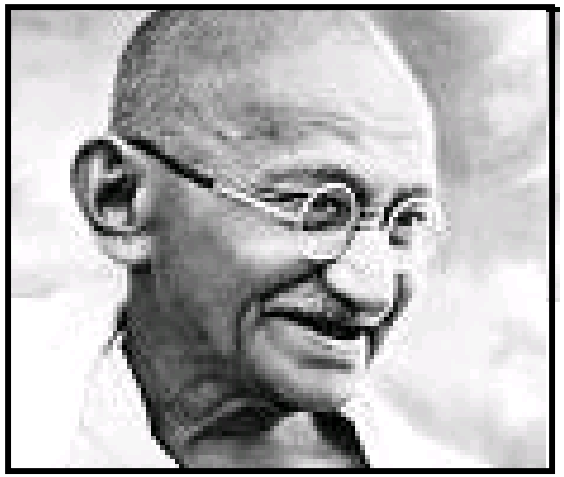
Somdatta Sinha

Centre for Cellular & Molecular Biology (CSIR), Hyderabad

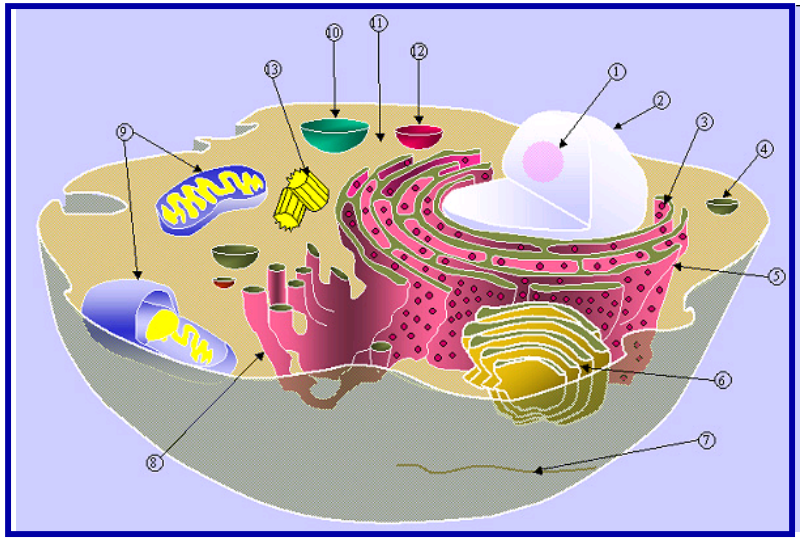
(Email: sinha@ccmb.res.in)

Model

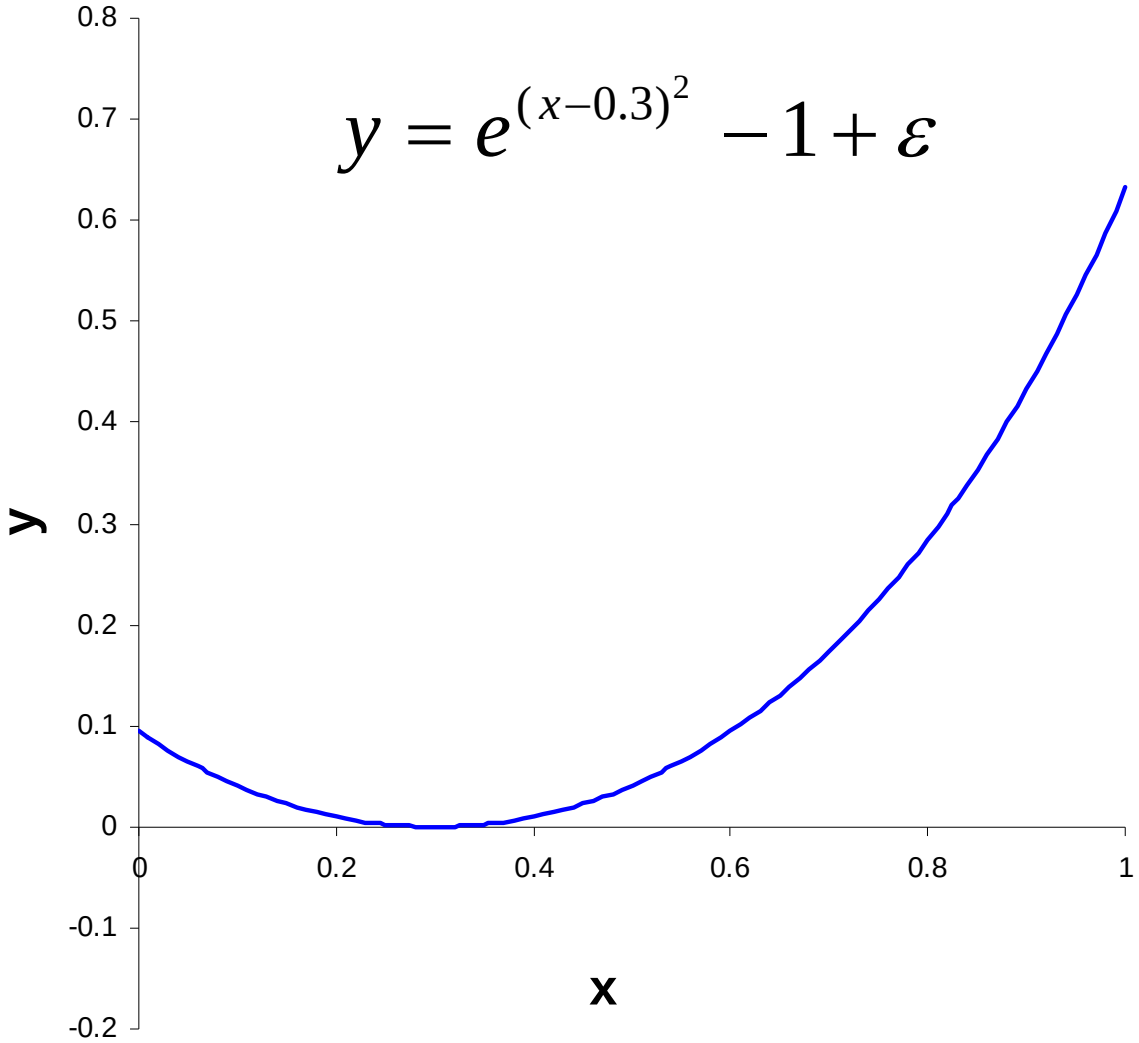
A simplified abstract view of the complex reality



Reality and Abstraction



Reality: Actual data



A set of candidate models

Approximate this model using a polynomial function. Fit three models: One is a simple linear model, one a quadratic model and the last one is a fifth order polynomial.

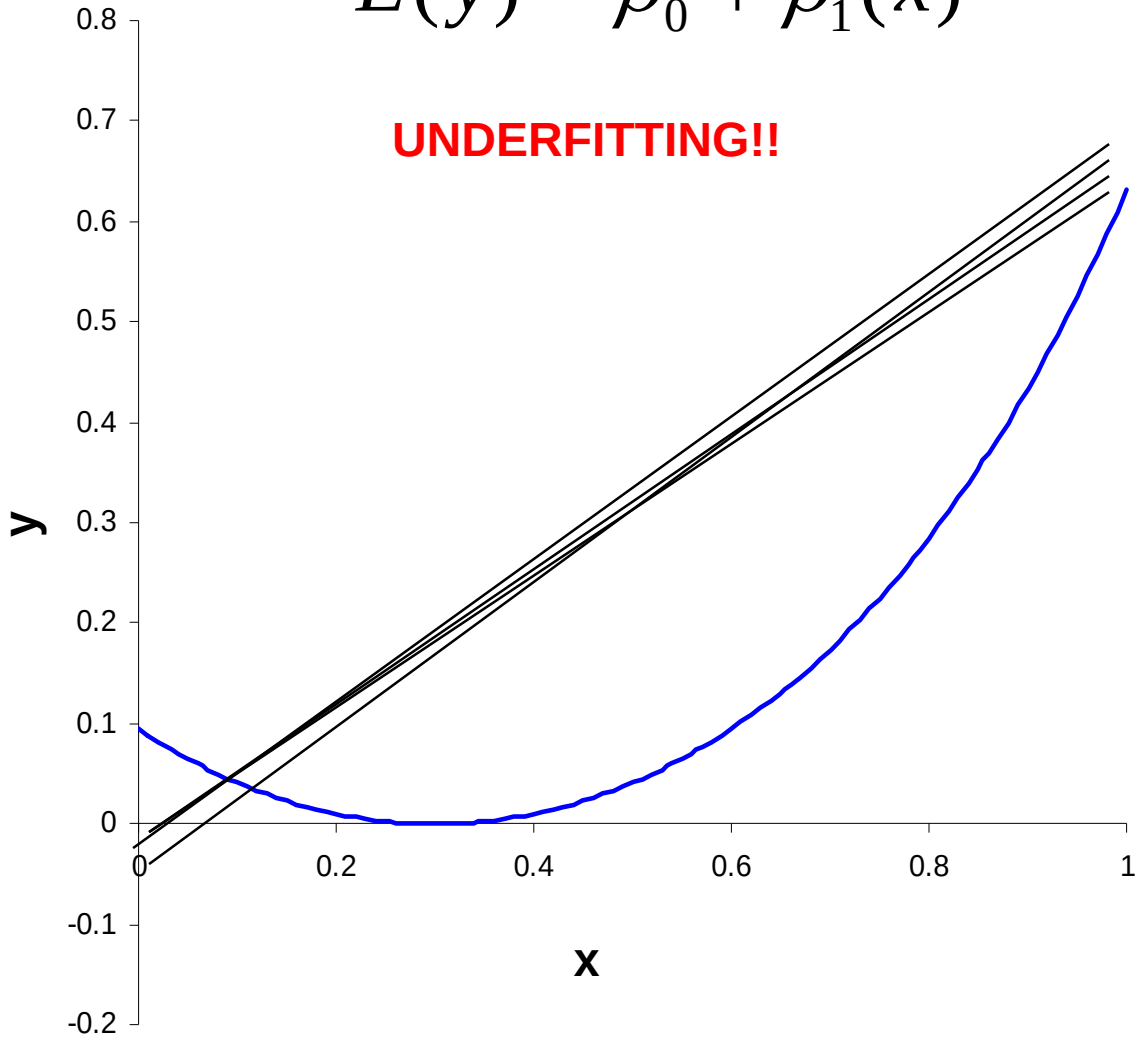
$$E(y) = \beta_0 + \beta_1(x)$$

$$E(y) = \beta_0 + \beta_1(x) + \beta_2(x^2)$$

$$E(y) = \beta_0 + \beta_1(x) + \beta_2(x^2) + \beta_3(x^3) + \beta_4(x^4) + \beta_5(x^5)$$

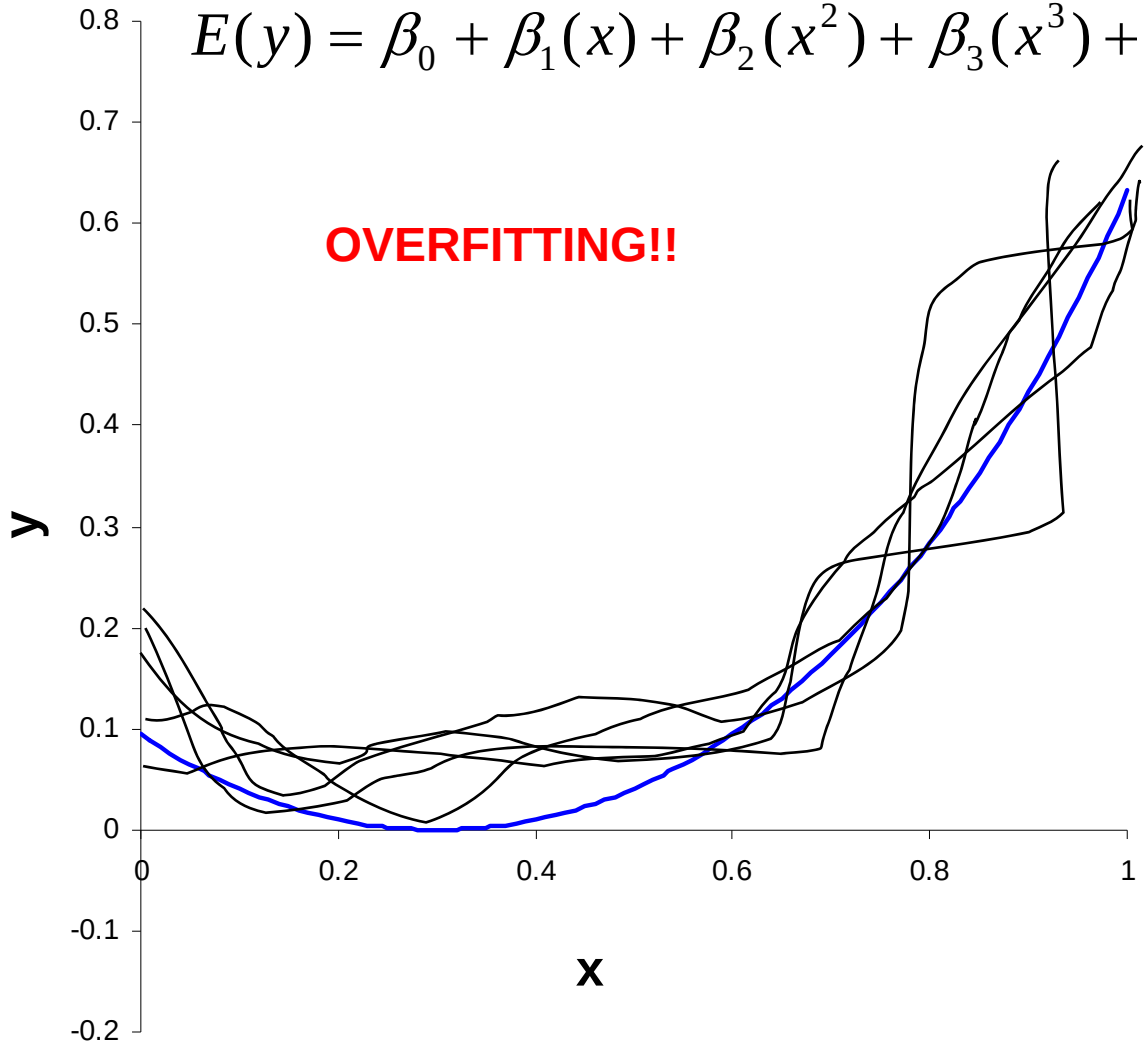
Too simple: High bias (low accuracy)

$$E(y) = \beta_0 + \beta_1(x)$$



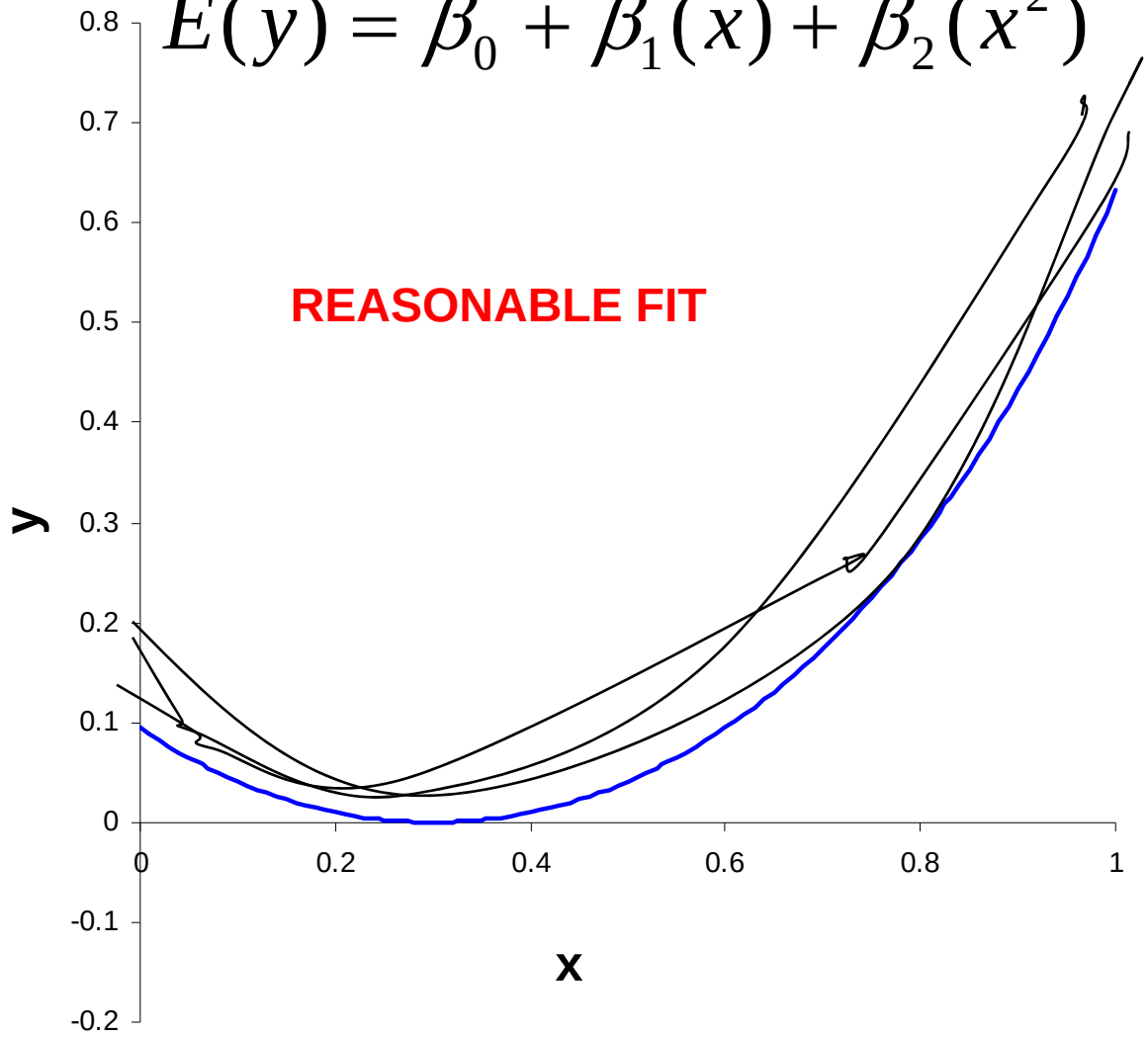
Too complicated: High variance (low precision)

$$E(y) = \beta_0 + \beta_1(x) + \beta_2(x^2) + \beta_3(x^3) + \beta_4(x^4) + \beta_5(x^5)$$



The compromise: a parsimonious model

$$E(y) = \beta_0 + \beta_1(x) + \beta_2(x^2)$$



How are interactions framed into functional forms

Functional form specification in a mathematical model

- *Relationship between the variables and the processes we are trying to understand need to be mathematically formalised*

- *The functional form should clarify the verbal description of the mechanisms driving the process under study.*

- *Choosing a functional form is a skill that needs to be developed over time.*

Principle of parsimony applied to model selection

- We typically **penalize added complexity**.
- A more complex model has to exceed a **certain threshold of improvement** over a simpler model.
- Added complexity usually makes a model more unstable.
- Complex models spread the data too thinly over data.
- Model selection is **not about whether something is true or not** but about whether we have **enough information to characterize it properly**.

Basic Modelling Techniques

Mathematical models can take many forms

Dynamical systems –

Deterministic & Stochastic

Differential equations, Discrete equations

Game theoretic models,

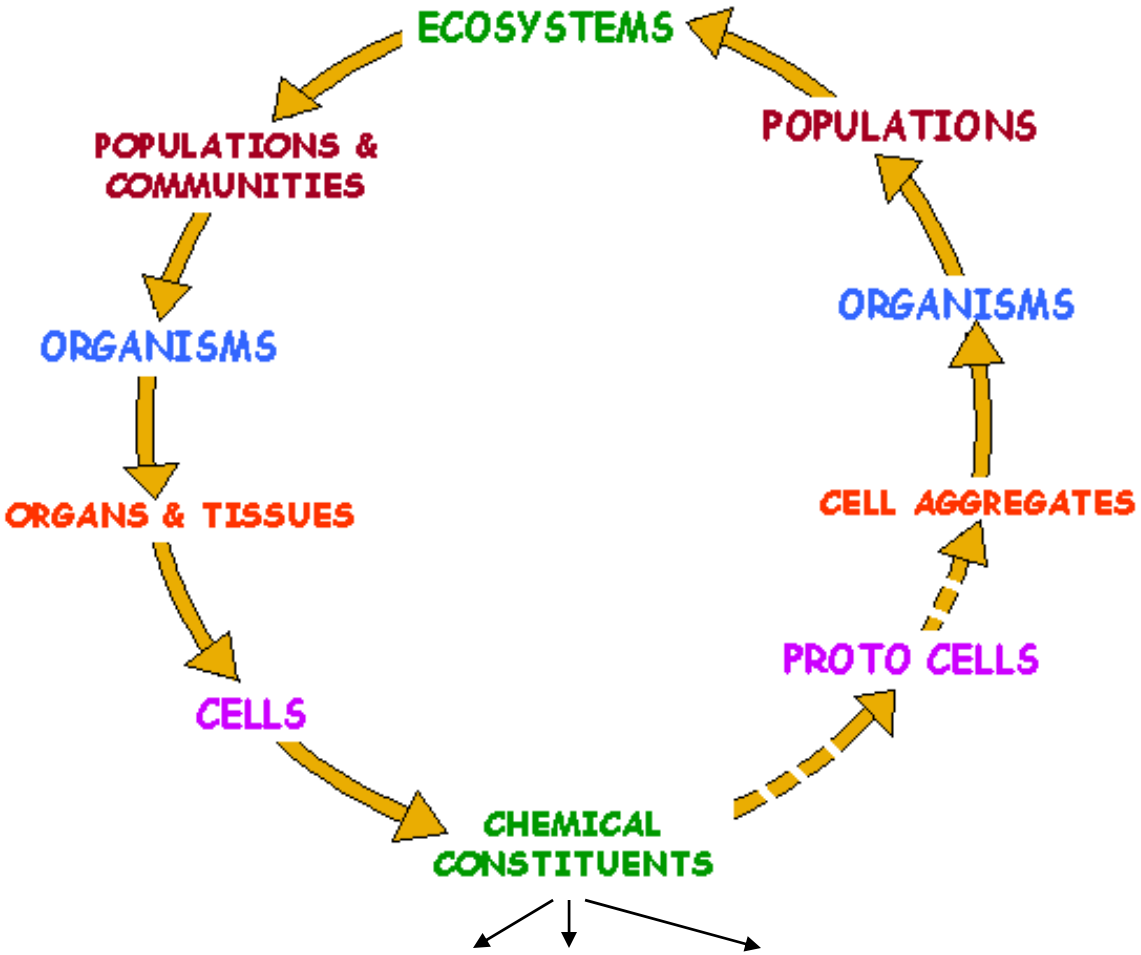
Cellular automata,

Genetic Algorithm,

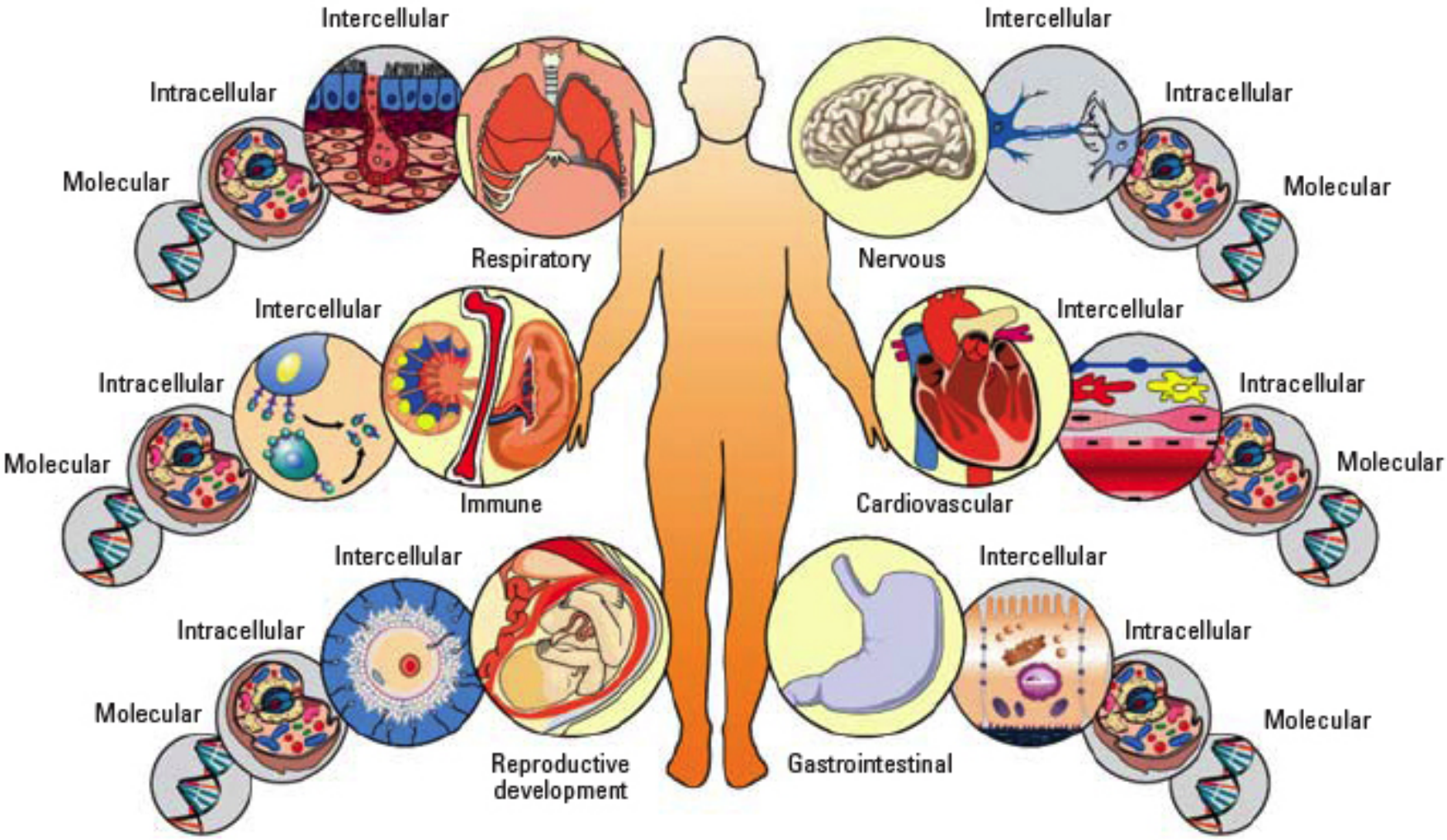
Graph Theory

Statistical models

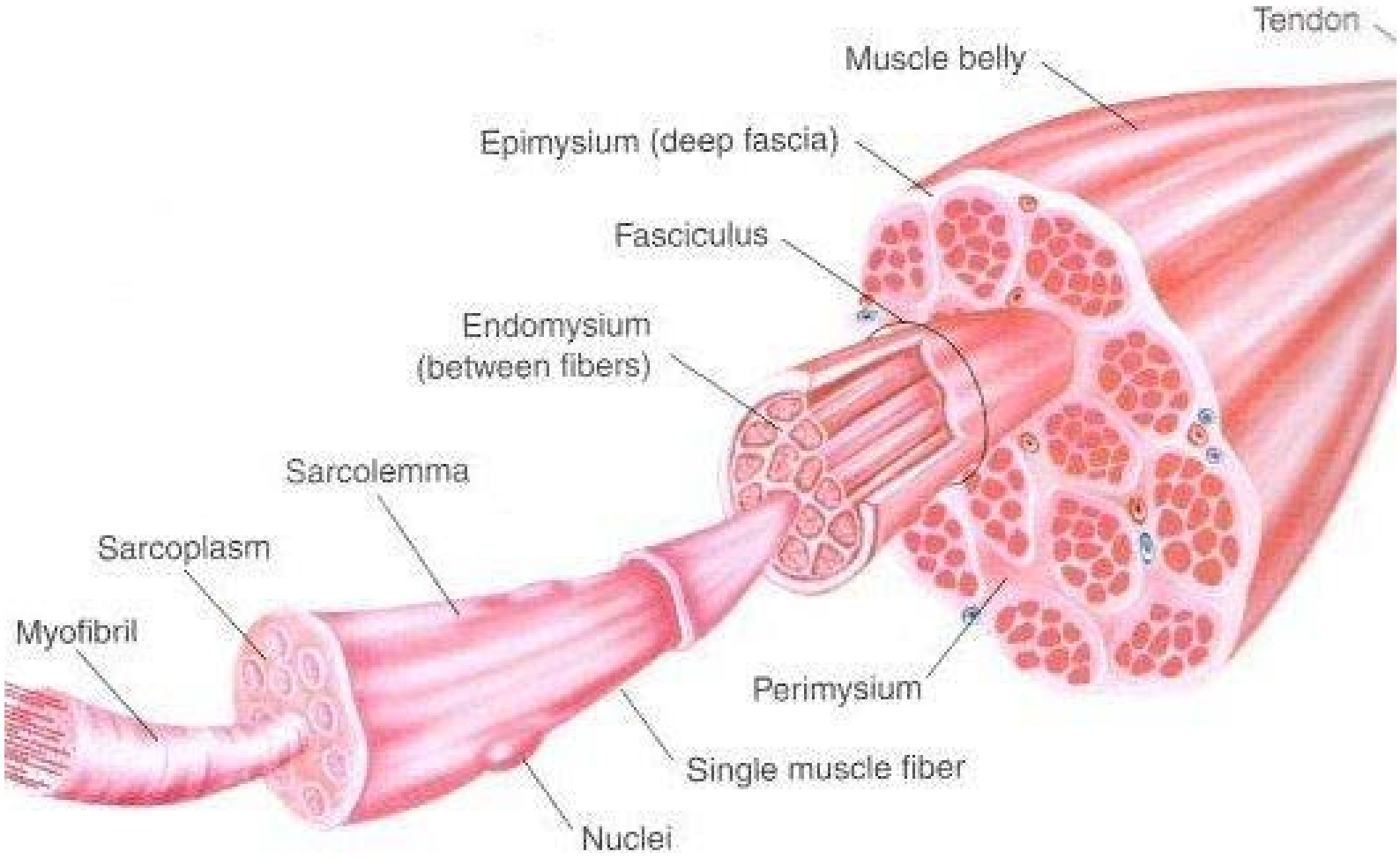
LEVELS OF ORGANISATION IN BIOLOGICAL SYSTEMS



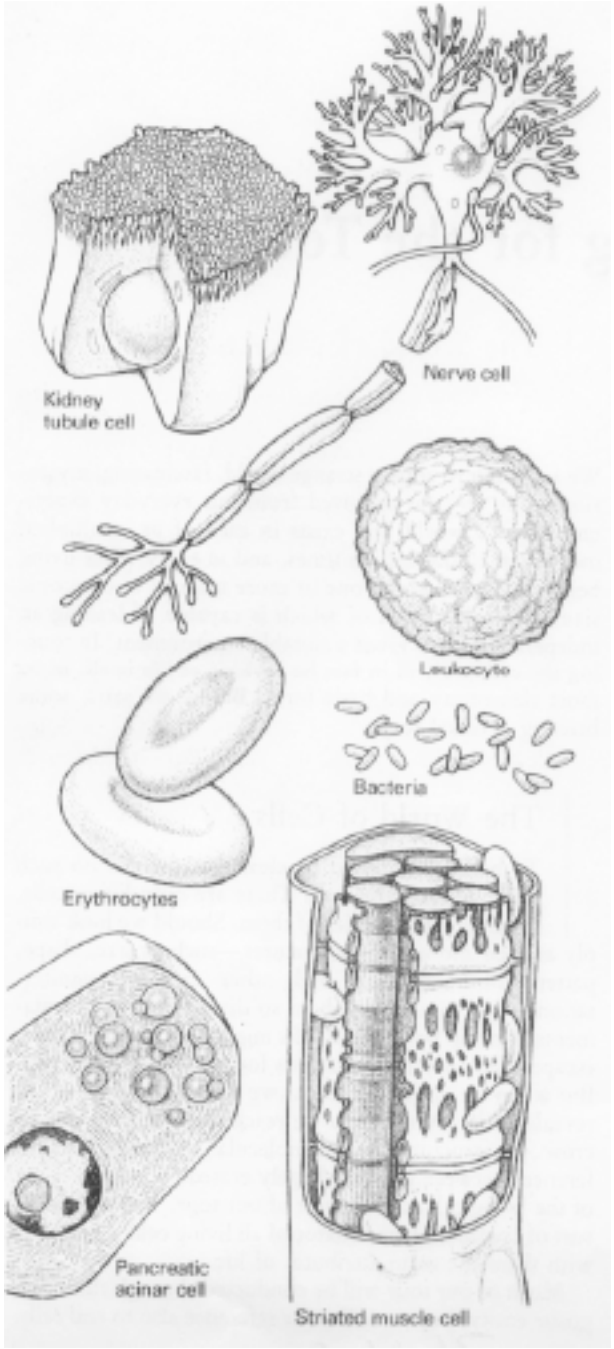
DNA, RNA, Proteins, Lipid bilayer, mitochondria, etc

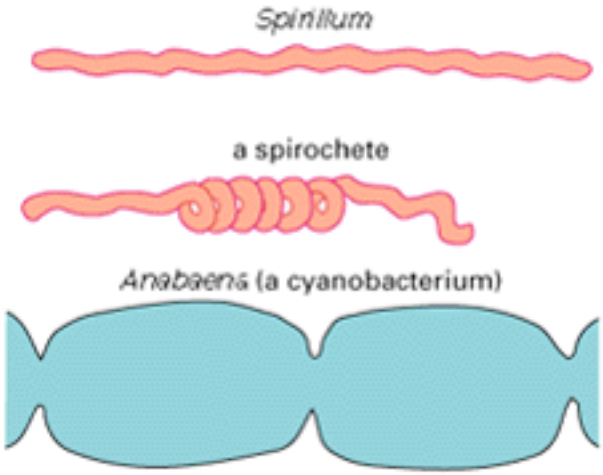

Intercellular
Intracellular
Molecular
Respiratory
Intercellular
Intracellular
Molecular
Nervous
Intercellular
Intracellular
Molecular
Immune
Intercellular
Intracellular
Molecular
Cardiovascular
Intercellular
Intracellular
Molecular
Reproductive development
Intercellular
Intracellular
Molecular
Gastrointestinal

Muscle as a System

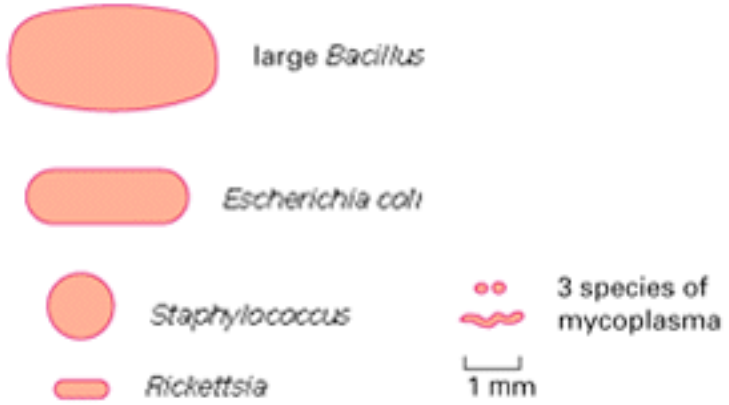


Living systems are made up of cells - *single or multi-cellular*





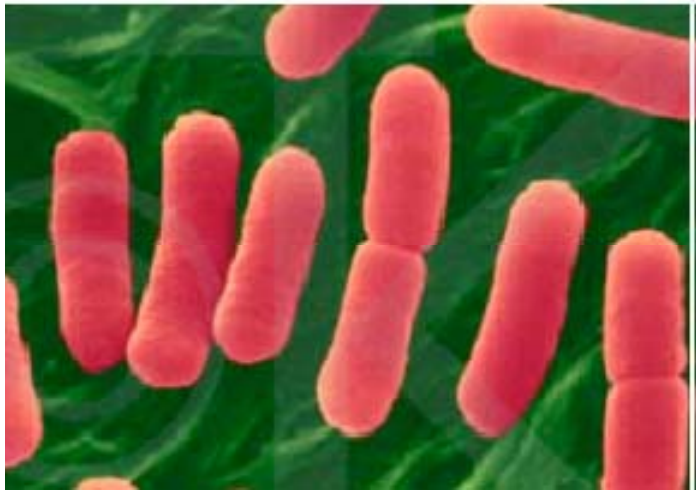
Some bacteria



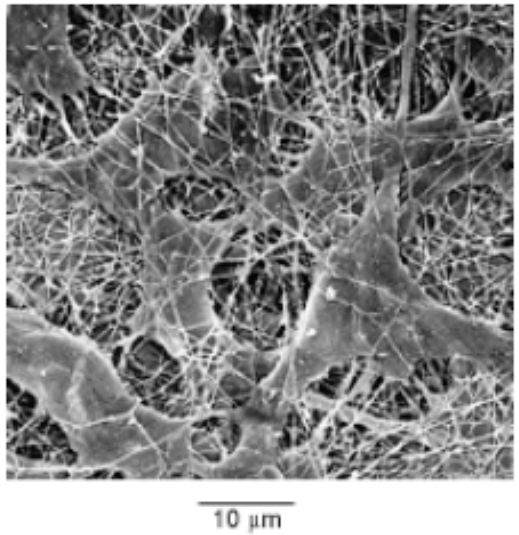
Smallest cell (mycoplasma) 0.0001 mm diameter

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

Electron micrograph of E.coli.



Dividing E.coli.



Collagen fibrils in ECM

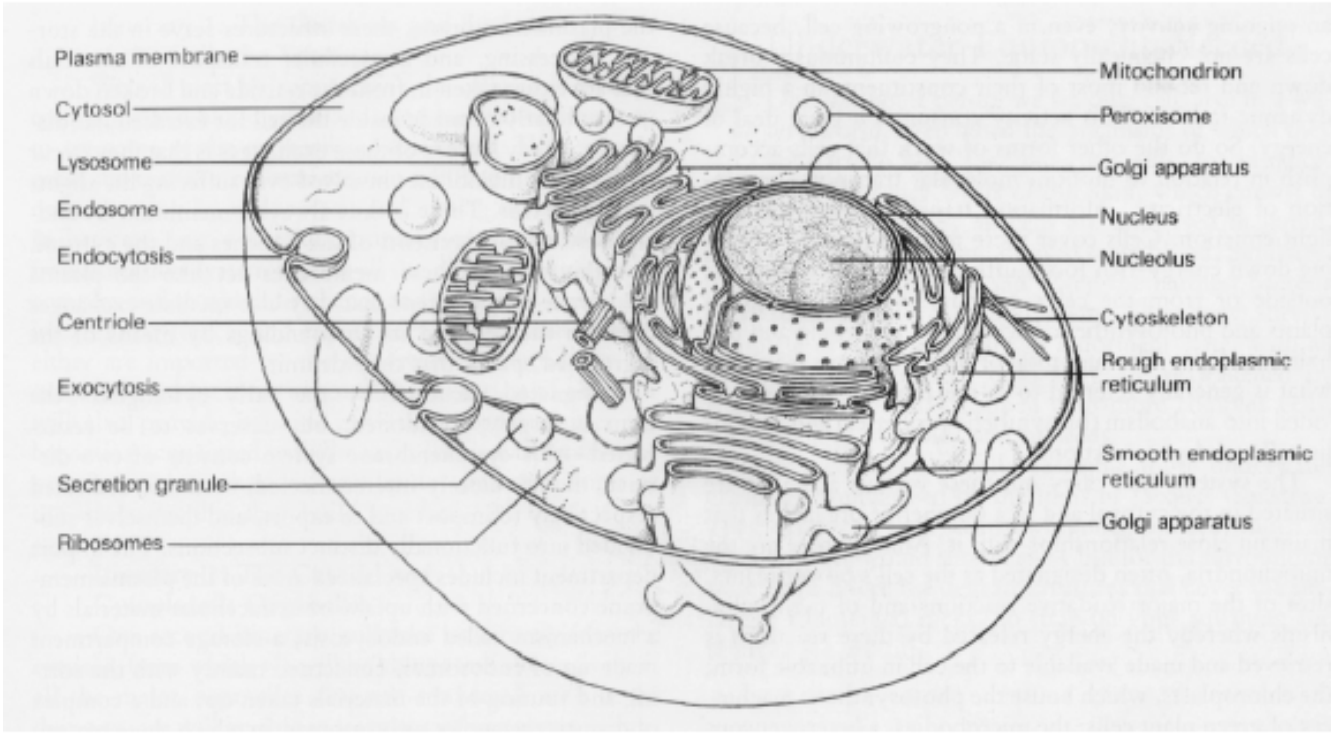
Fibroblast cells in cornea of rat
(*Scanning electron micrograph*)

Doubling time:

The bacterium *E. coli* can reproduce in 20 minutes

Cell cycle of single-celled yeast is 90 -120 minutes

A rapidly dividing mammalian cell cycle is about 24 hours.



The Approximate Chemical Composition of a Bacterial Cell	Percent of Total Weight
Water	70
Inorganic Ions	1
Sugars and precursors	1
Amino acids and precursors	0.4
Nucleotides and precursors	0.4
Fatty acids and precursors	1
Other small molecules	0.2
Macromolecules (proteins, nucleic acids, and polysaccharides)	26

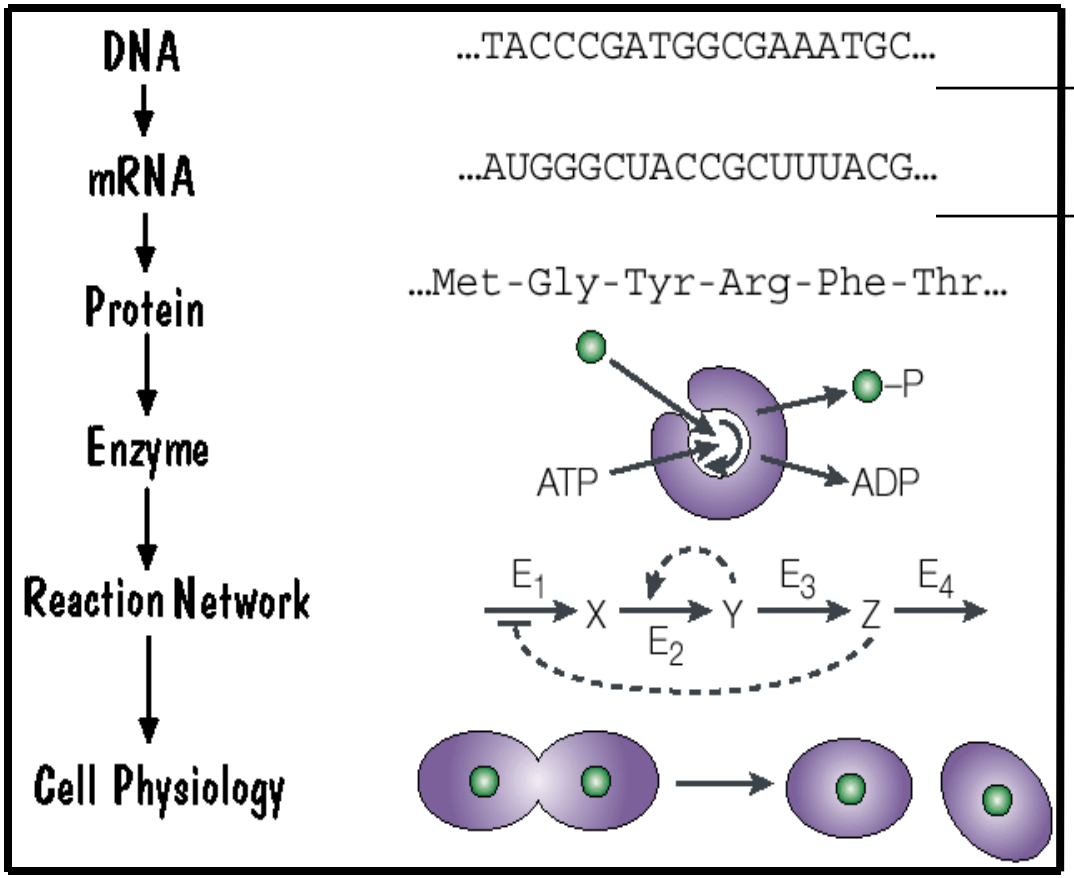
The Human genome length is about 2 metre (3,000 million base pairs).

E.coli genome is about 1.2 mm (4 million bp).

The smallest genes are ~10,000 bases long - Ovalbumin (7.7Kbp)

The largest gene is about 2 million bases (for a human muscle protein)

FLOW of INFORMATION



Transcription rate -
~ 1,000 nucleotides/minute

Translation rate -
~ 900 amino acids/minute

Production of the protein to the binding of dimer - about 3 min

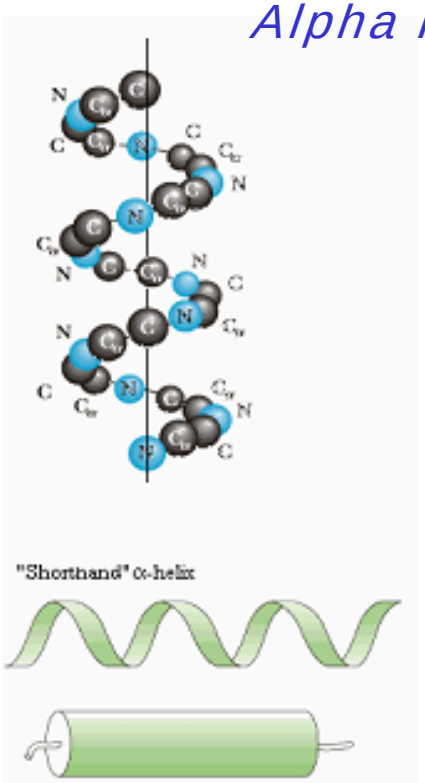
Biochemical pathways underlie cellular functions

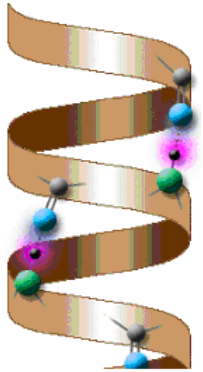
CELL → Collection of interacting biochemical pathways integrated into an overall reaction network through metabolic and genetic control elements

- *Co-ordination*
- *Regulation*
- *Decision-making*
- *Ability to evolve*

Amino acid sequences in proteins form patterns of specific secondary structures useful for their function

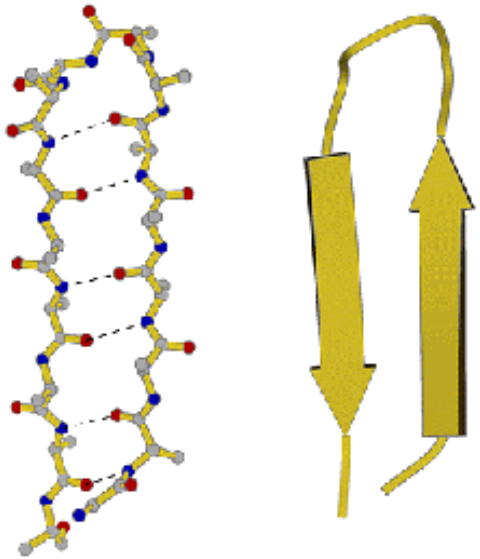
Alpha Helix





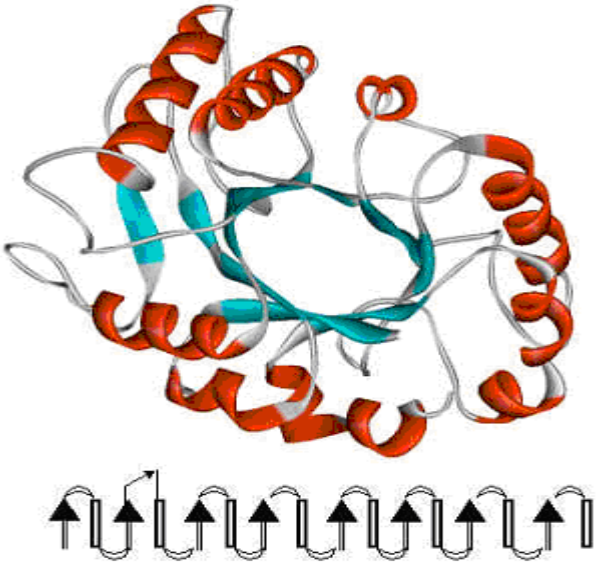
α - Helix
Only the N—C_α—C backbone is represented. The vertical line is the helix axis.

Beta Conformation



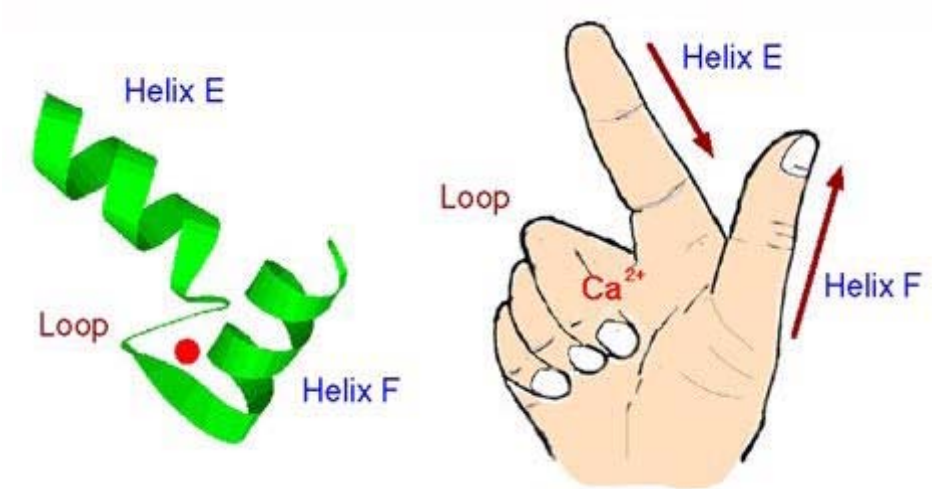
Parallel chains

Combinations of secondary structural elements form different patterns of *Super Secondary Structures* (folds) that perform specific functions



TIM Barrel

An eight-stranded α/β domain (first found in Triose phosphate isomerase). A central barrel formed by parallel β-strands surrounded by seven or eight α helices which shield the barrel from solvent.

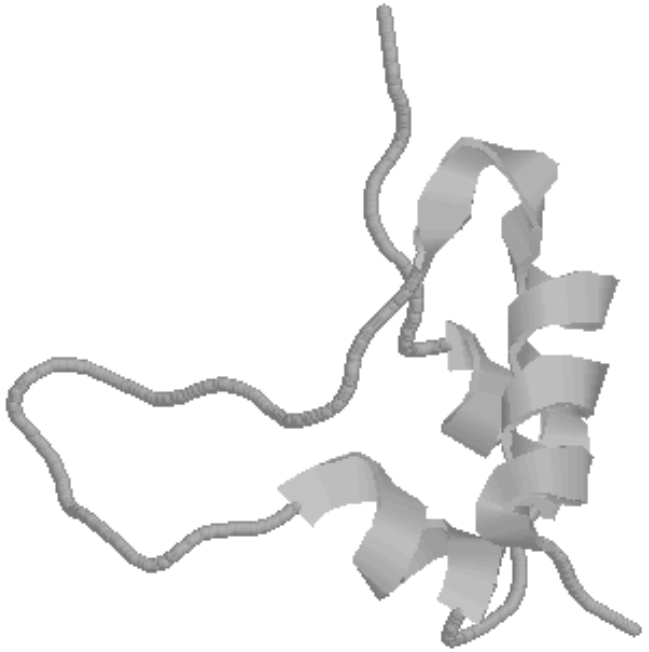


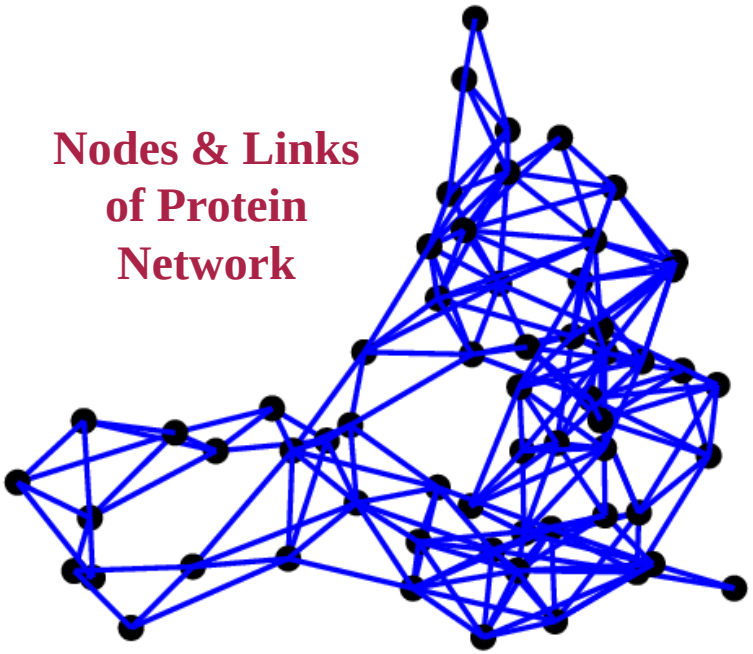
EF Hand

The loop region in Calcium binding proteins are enriched in Asp, Glu, Ser, and Thr.

Pattern of contacts in protein structure indicate its ability to interact and fold

GLY PRO **GLU** **THR** LEU CYS GLY ALA GLU LEU VAL ASP ALA LEU
GLN PHE VAL CYS GLY ASP ARG GLY PHE TYR PHE ASN LYS PRO
THR GLY TYR GLY SER SER SER ARG ARG ALA PRO GLN THR GLY
ILE VAL ASP GLU CYS CYS PHE ARG SER CYS ASP LEU ARG ARG
LEU GLU MET TYR CYS ALA PRO LEU LYS PRO ALA LYS SER ALA

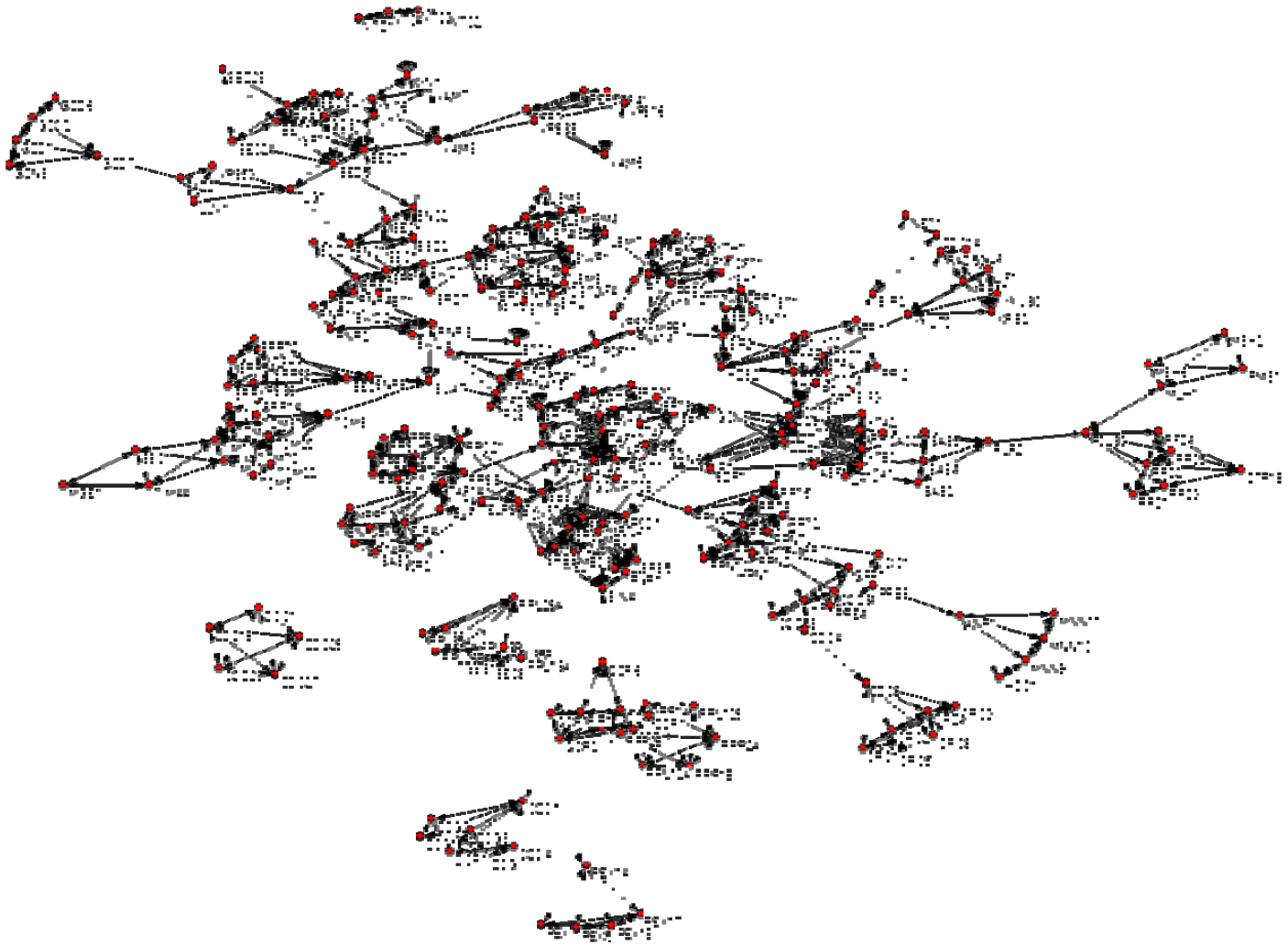




1GZR - *HUMAN INSULIN-LIKE GROWTH FACTOR*

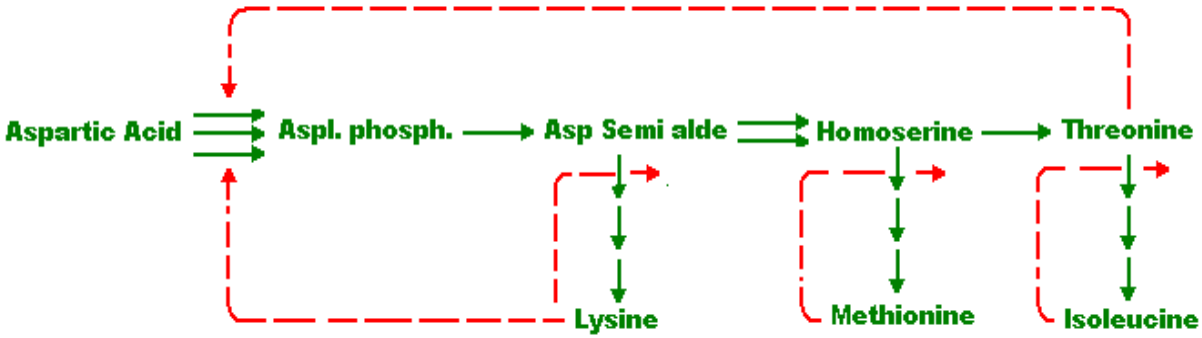
Pattern of contacts in protein interaction network

- scale free nature



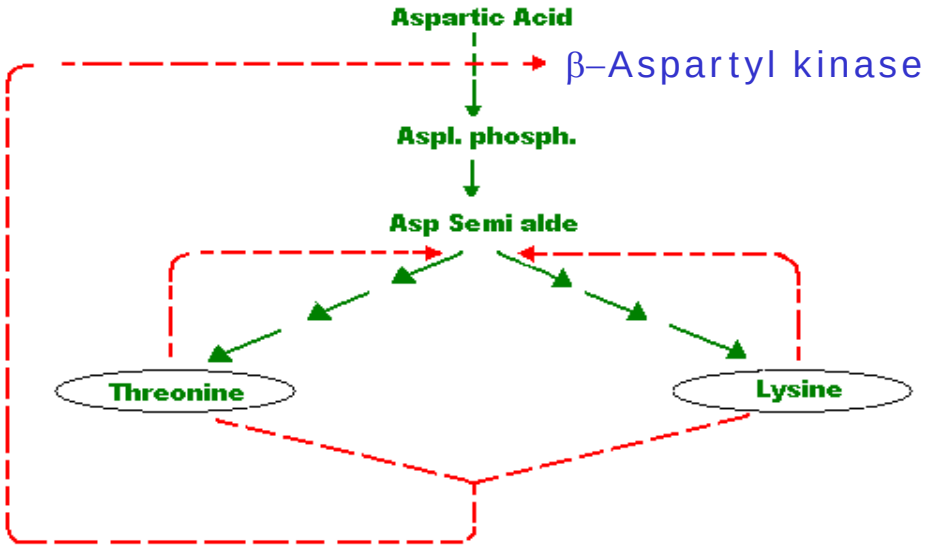
A small part of the budding yeast (*Saccharomyces cerevisiae*) protein-protein interaction network.

Biosynthesis of Aromatic Amino Acids



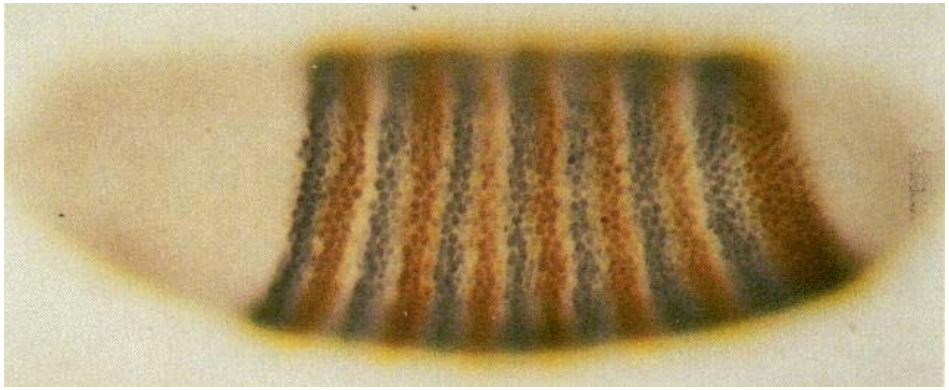
Divalent, Nested ***and*** Sequential control ***in bacteria***

Concerted control ***of b - Asp. kinase in Rhodopseudomonas capsulatus***

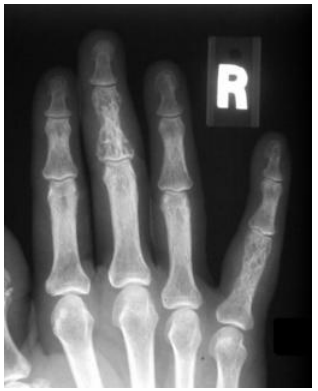


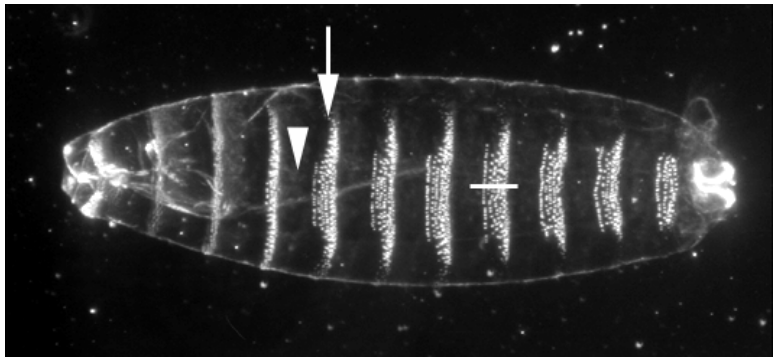
Patterns of Gene Expression give rise to structure

Pattern Formation (Segmentation) in Drosophila Embryo



Gene expression pattern

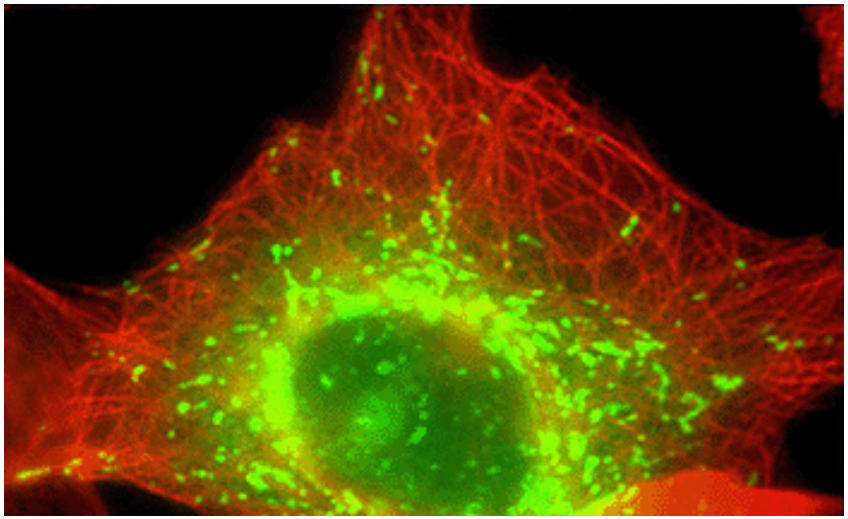




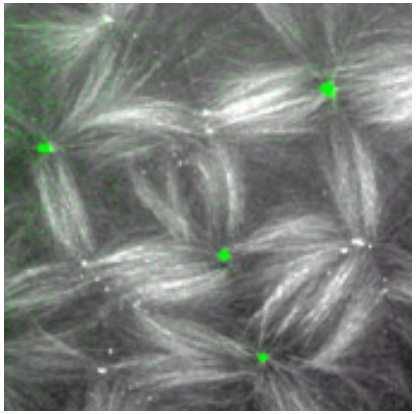
Epidermal pattern

Embryonic body pattern is organized into repeating, segmental units, visible in the cuticle of the first instar larva (ventral view). Smooth cuticle (arrrow head) alternates with bands of "denticles" (arrow) across each segment.

Differing Patterns of Microtubule Network
- aids in cellular functions

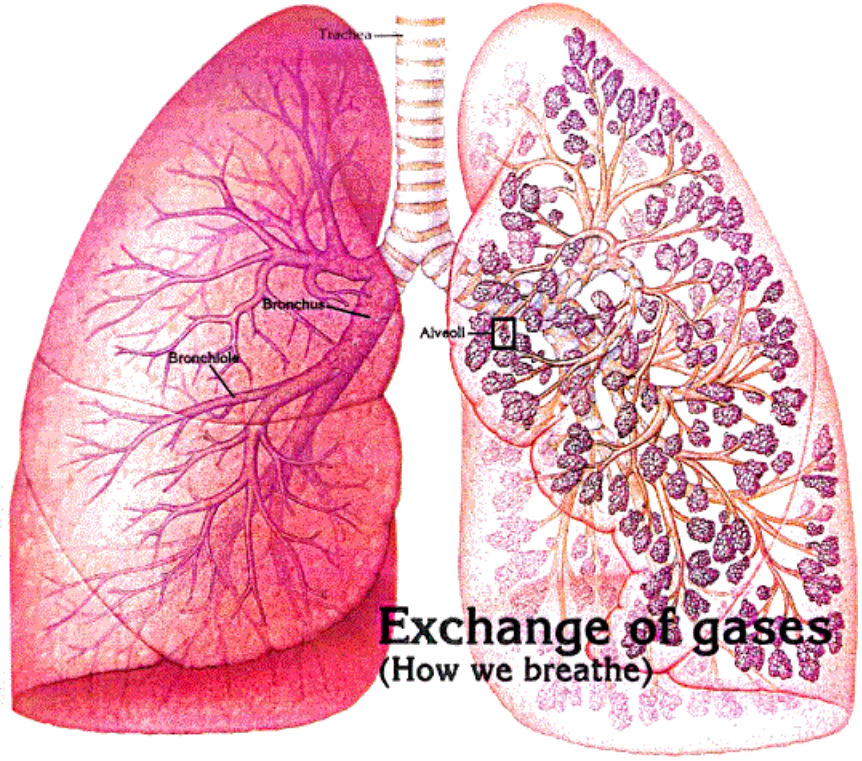


Network of microtubules forms the cytoskeleton (orange) in a cultured cell



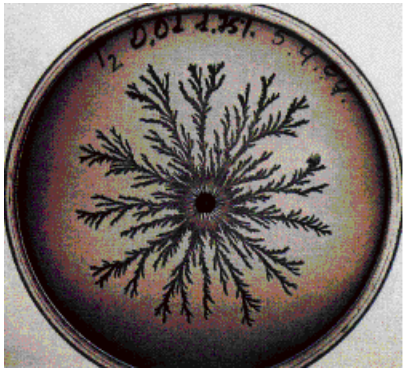
Network of microtubules (white) and two kinds of motor proteins (green) created by self-organization *in vitro*.

Differing Patterns of Cellular Processes
- aids in cellular functions

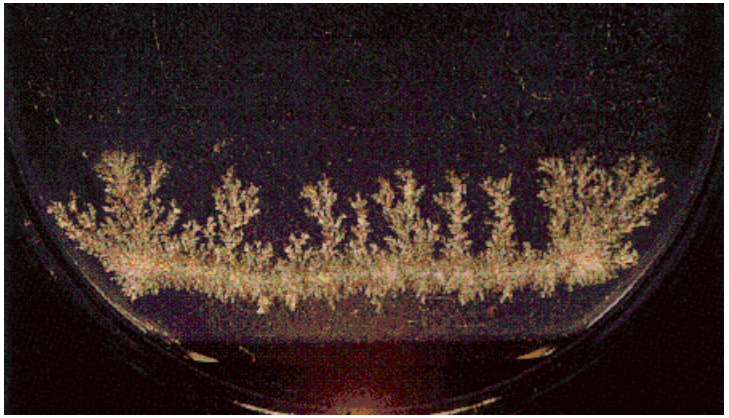


The lungs have an extensive network of blood vessels. This aids in excellent blood supply that is needed to transport oxygen away from the lungs efficiently.

Differing Patterns of Growth under Stressed Environment - aids in survival



Colonies of the bacteria *Bacillus subtilis* (*B* 168), common in food, under stressed conditions causing the colonies to adopt a fractal form.



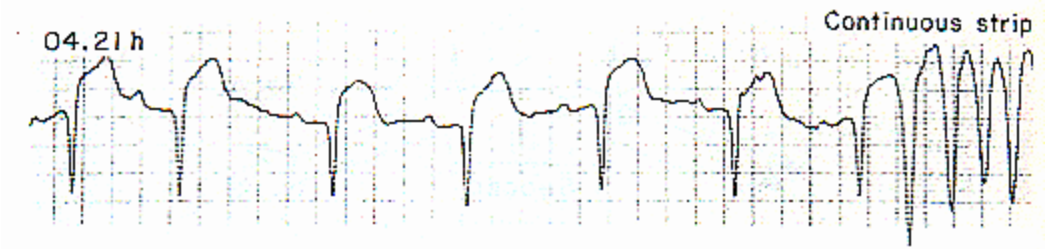
Aspergillus oryzae grown under decreasing nutrient concentrations

Diffusion-Limited Aggregation, or DLA, is a simple computer simulation of the formation of clusters by particles diffusing through a medium that jostles the particles as they move.

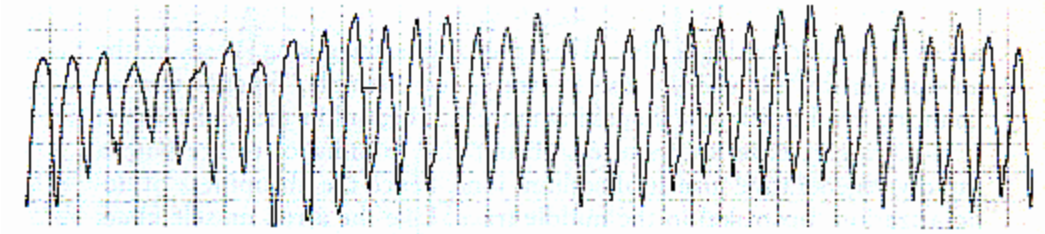
Eshel Ben-Jacob

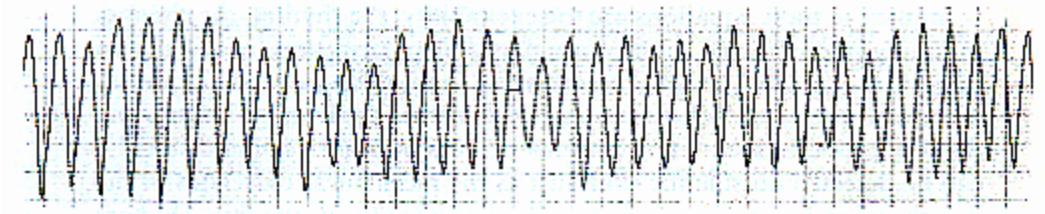
Temporal Patterns - alterations lead to disease

Normal sinus rhythm ~1 sec

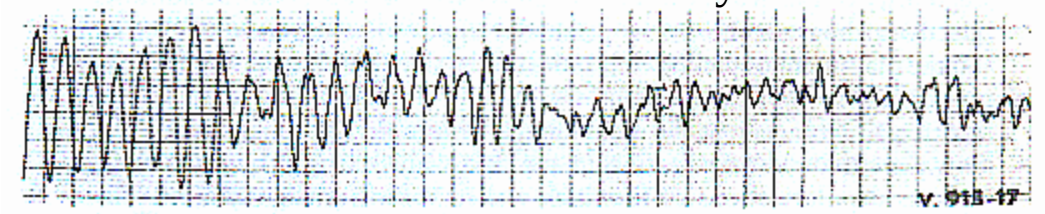


Tachycardia at 5Hz





Terminal electrical activity



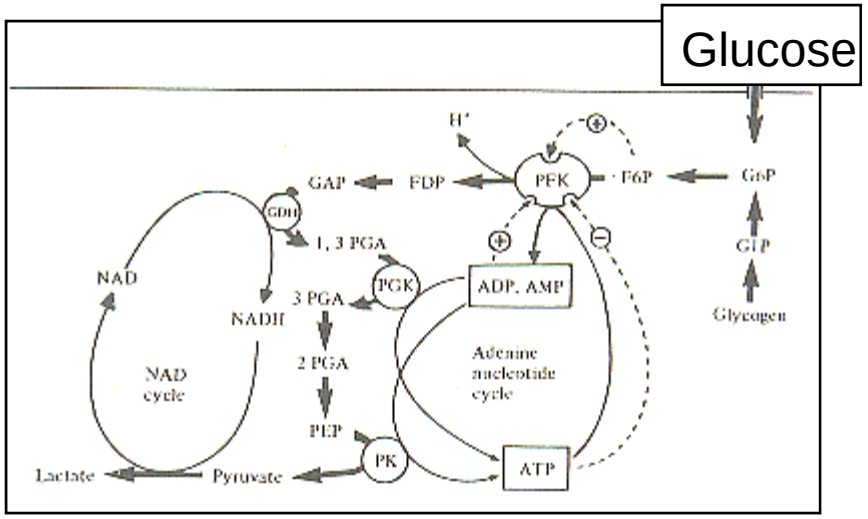
Change in the temporal pattern of rhythm

disease and death

Sudden cardiac death

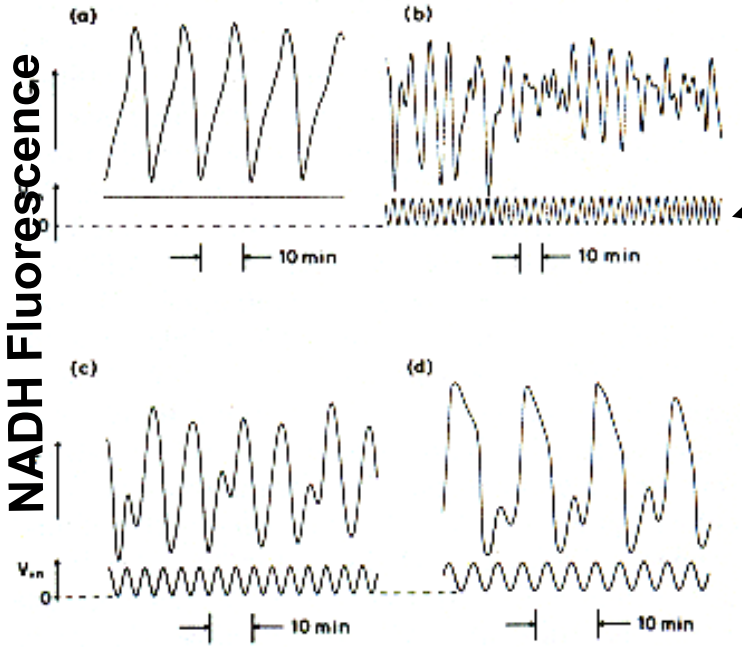
Temporal Patterns - Driven by external agents

Control structure of
glycolytic pathway
Material transport ____
Control loops -----



Experiments on cell-free yeast extracts:

measurements of NaDH fluorescence (Y axis) with time for different glucose input

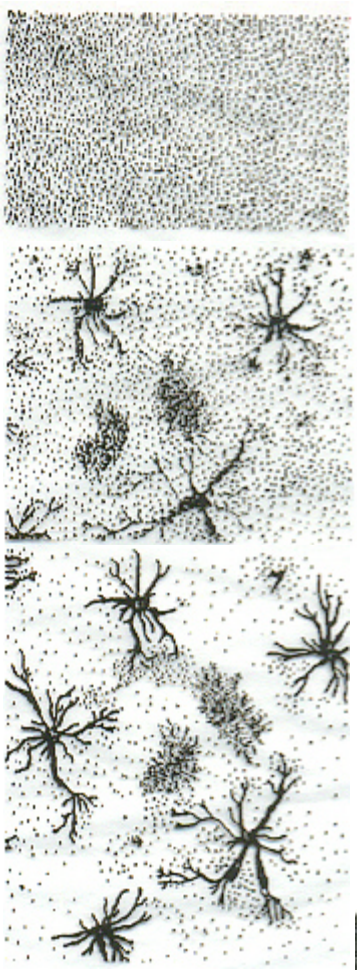


Glucose input

Constant input → oscillations
Periodic input → chaos
1:5 entrainment
1:3 entrainment

Spatio-Temporal Patterns
change lead to developmental irregularities

social amoebae



Free-living cells - eat and reproduce

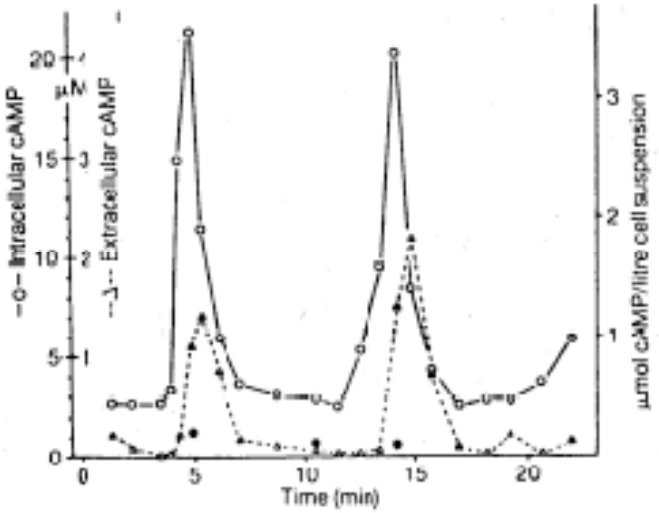
Cells aggregate (signal (cAMP) relay)

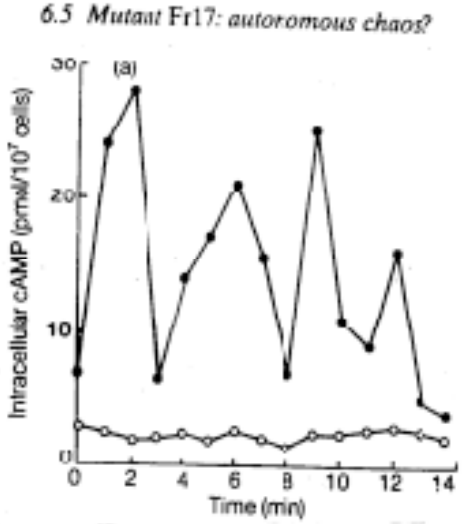
Large aggregation centres formed

Mound of cells form slug

Fruiting body with stalk and spore cell

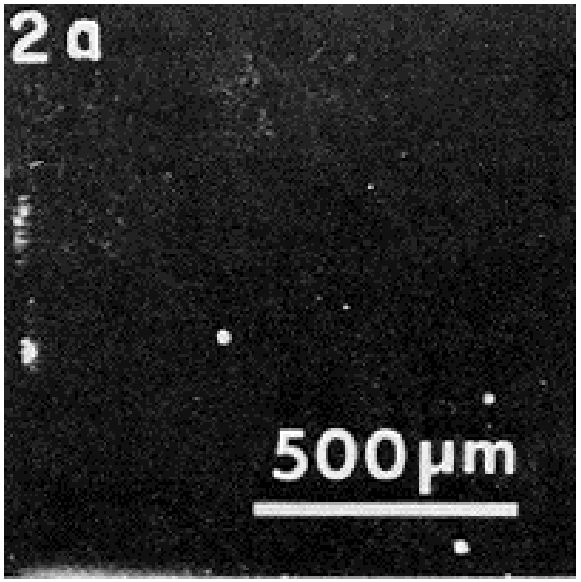
Active communication and interaction among the cells





This mutant strain does not form slug and fruiting body

Spatio-Temporal Patterns - waves



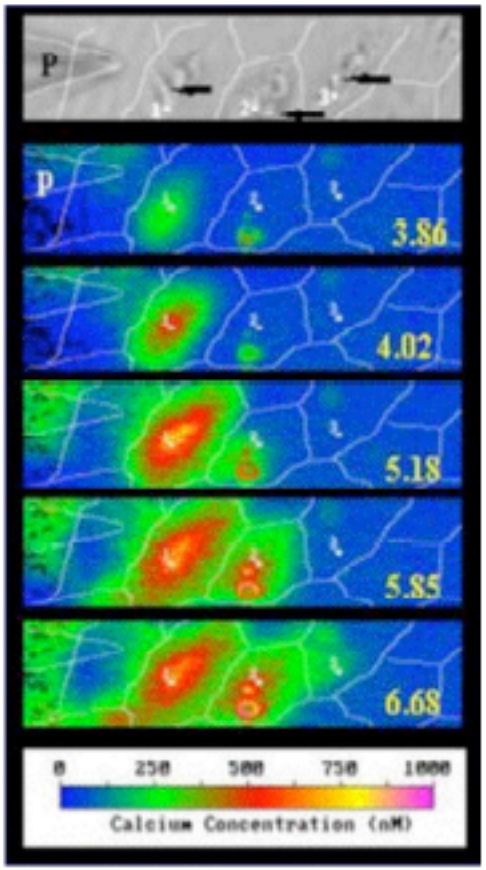
A free calcium wave propagating across a sperm-activated medaka egg

Successive photographs are 10 s apart.
Egg axis horizontal with sperm entry point to the left.

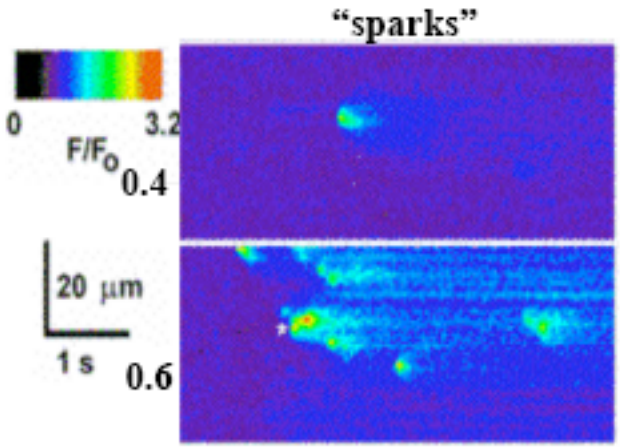
(Gilkey et al, J. Cell Biol. 1978)

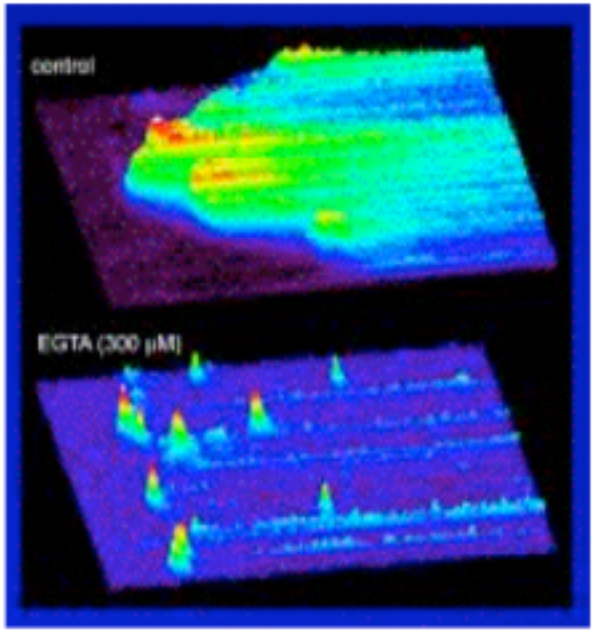
Transient Temporal Patterns

Intra- and Intercellular Calcium waves



Imaging Calcium micro domains in single cells in a tissue



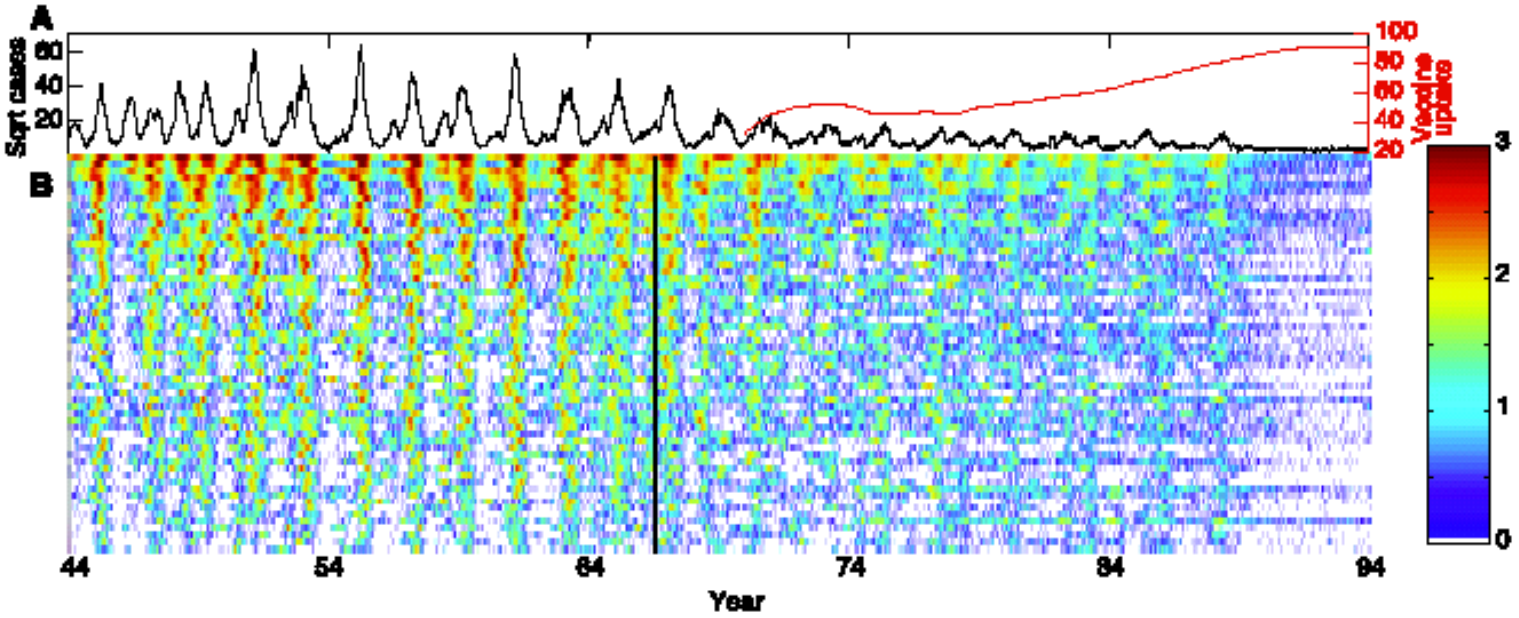


Ant Foraging Trails

As the colony size (number of ants) increases these trail networks increase in size.

Holldobler and Wilson *The Ants* (1990)

Patterns in Processes *effect of intervention*

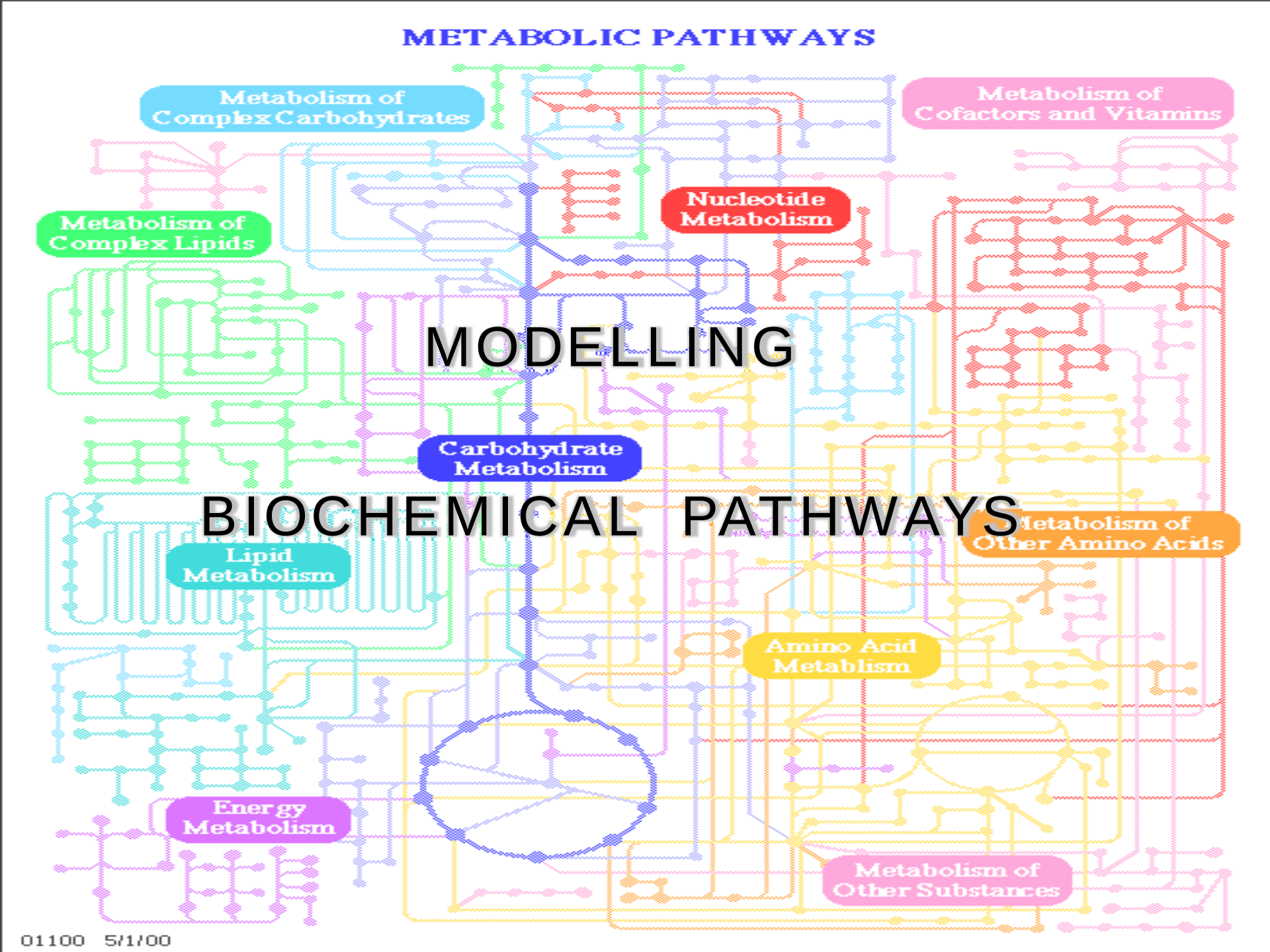


Measles and whooping cough notifications in England and Wales from 1944 to 1994, obtained from the Registrar General's Weekly Returns.

(A) Time series for measles in London (black line) together with the published vaccine uptake levels (percentage of infants vaccinated) for England and Wales, starting in 1968 (red line).

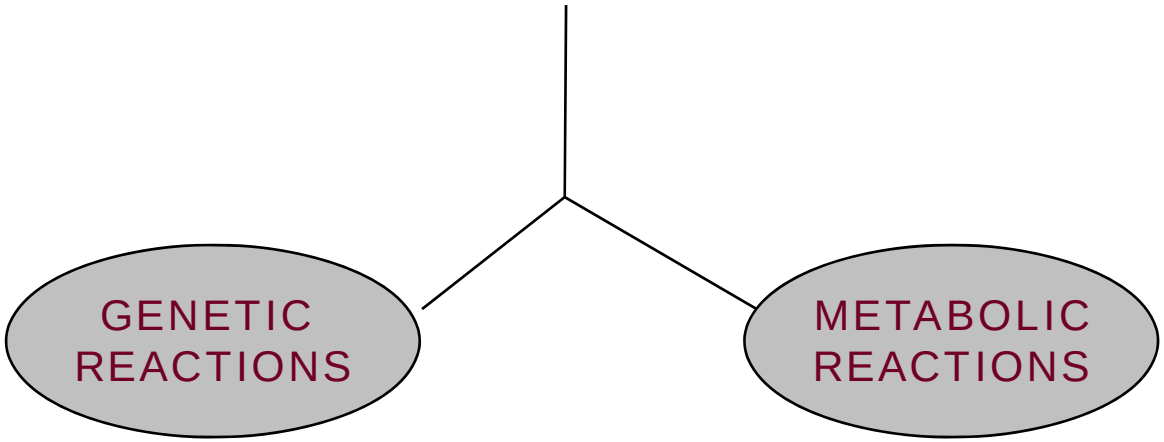
(B) The spatial distribution of log10 (1 + measles cases) with cities arranged in descending order of population size (from top to bottom) and colors denoting epidemic intensity (white regions highlight periods with no reported cases).

The vertical black line represent the onset of vaccination.


METABOLIC PATHWAYS
Metabolism of Complex Carbohydrates
Metabolism of Cofactors and Vitamins
Metabolism of Complex Lipids
Nucleotide Metabolism
MODELLING
Carbohydrate Metabolism
BIOCHEMICAL PATHWAYS
Metabolism of Other Amino Acids
Lipid Metabolism
Amino Acid Metablism
Energy Metabolism
Metabolism of Other Substances
01100 5/1/00

Complex network of biochemical reactions in cells coordinate and control cellular functions

two interacting sets



Slow
(minutes, hours)

TIME SCALE

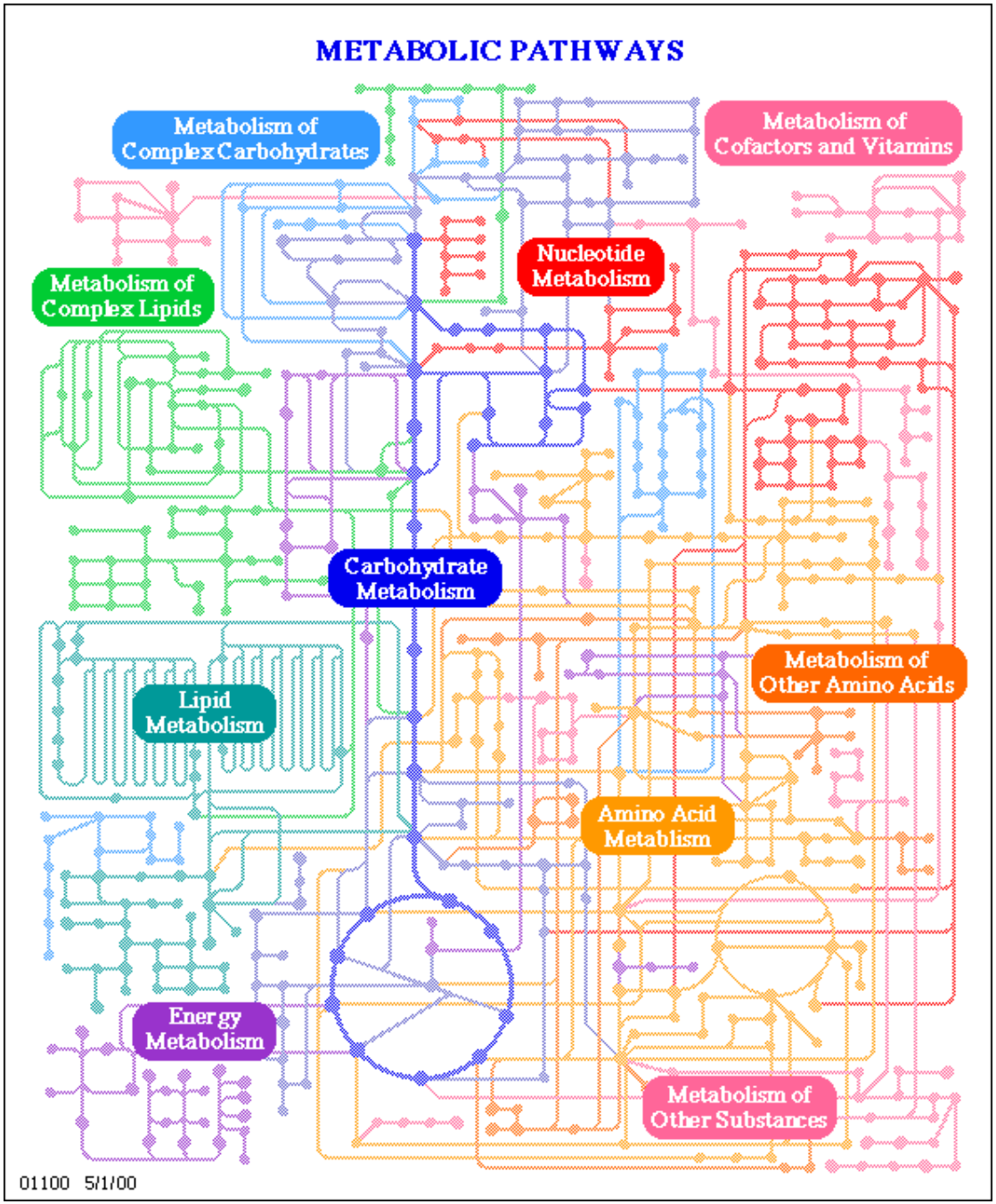
Fast
(seconds)

Gene induction, repression, replication, transcription

Conversion of substrate molecules by enzymes,

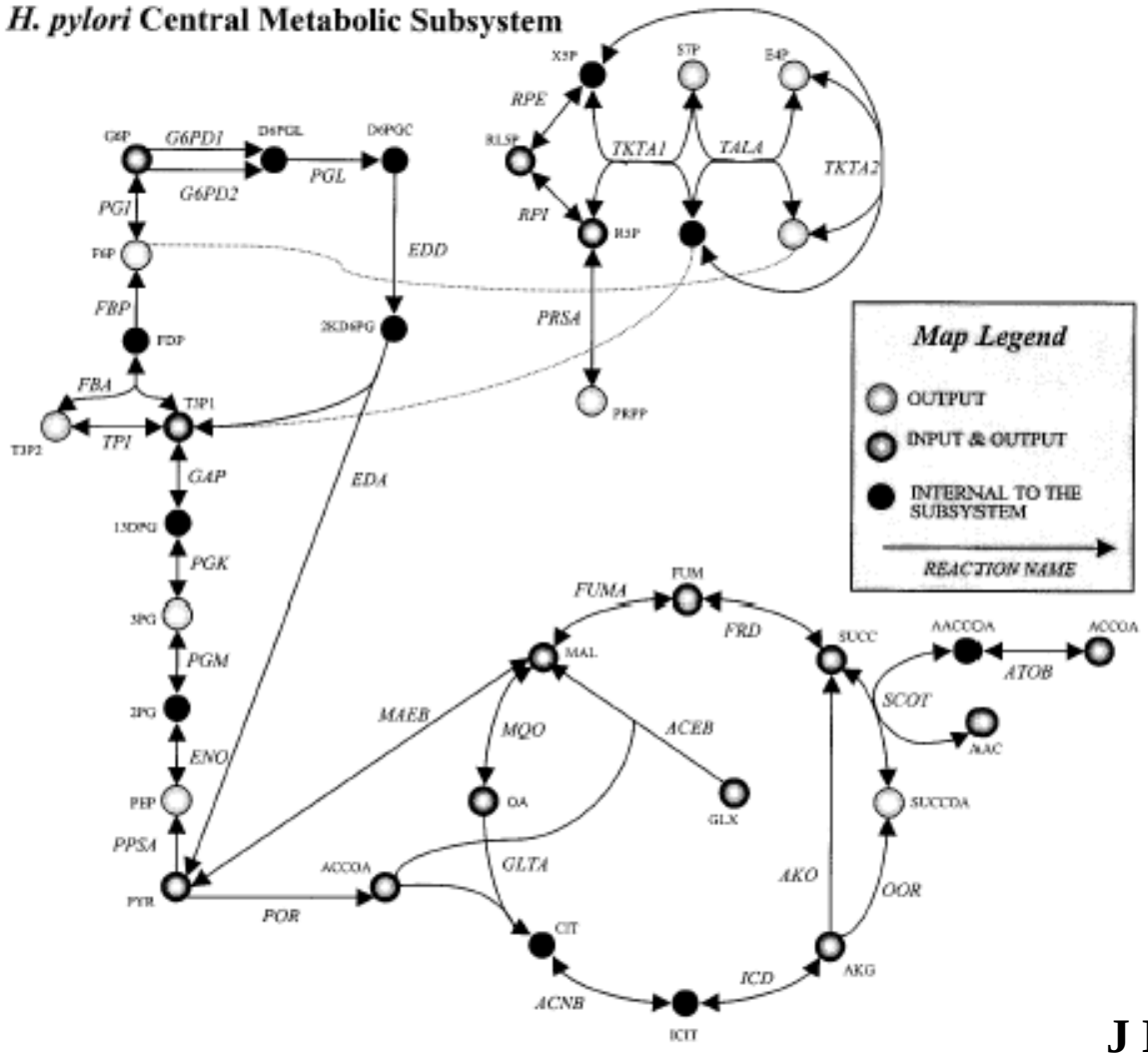
enzyme inhibition or activation

Wire diagram of the metabolic pathways in *E.coli*



www.genome.ad.jp/kegg/

Genome-Scale Metabolic Model of *Helicobacter pylori*



J Bacteriology 2002
Science 2001

BIOCHEMICAL PATHWAYS ARE SEQUENTIAL REACTIONS

Cellular behaviour is the emergent property of many complex biochemical reactions networked through feedback/feed-forward processes with overlapping and wide-ranging time scales

Biochemical details of each pathway may be different, but they possess certain general features which can be described using the rules of chemical kinetics.

The resultant dynamics underlie different functional behaviour

Stability – ***Homeostasis***

Multistability – ***Ability to operate on alternative conditions***

Threshold Sensitivity – ***Switching behaviour***

Oscillatory – ***Rhythmic and cyclic processes***

Chaotic – ***Bursting activity & irregular behaviour***

Biochemical reactions are highly networked reactions

Primary mode of regulation to co-ordinate and control is through

Single, Multiple & Coupled
Negative and Positive Feedback Processes

Negative feedback

desensitize the system to perturbations

It ensures stability and conservation of energy and are, therefore, naturally selected to be the most common form of regulation in pathways

Positive feedback

are potentially destabilizing

Employed for excitable dynamics & amplification in switching & rapid response processes.

Functional dynamics in cells

is a consequence of the

non-linearity inherent in multiple modes of biochemical regulation

Types of dynamics

Homeostasis, Multi-stable, Multi-rhythmic, Oscillatory, Chaotic and Transient processes

Examples –

Sustained oscillations, Multiple stable states, Birhythmicity, Bursting & Complex oscillations

Glycolysis, Peroxidase-Oxidase reactions, Hormonal systems, cAMP oscillations in cellular slime mould, Neuronal systems & Insulin secretion in pancreatic beta cells

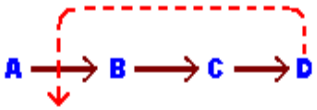
Transient processes → ***"quarks", "sparks" and "puffs" in calcium oscillations, spatial waves in many cell types.***

Both negative and positive feedback processes are useful for optimal performance requiring stability, sensitivity and multiplicity of dynamics.

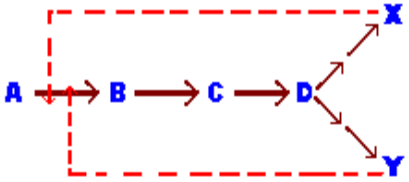
Biochemical reactions are controlled and co-ordinated mainly through feedback processes.

Most biosynthetic pathways have multiple levels of feedback control

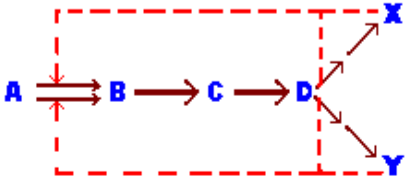
Monovalent Control



Divalent Control

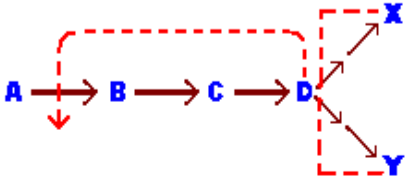


Nested Control

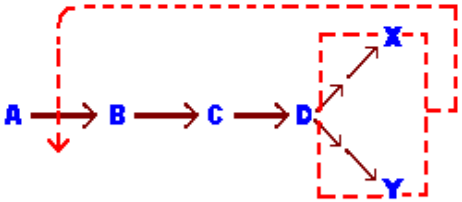


isoenzyme

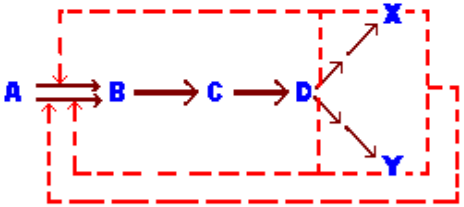
Sequential Control

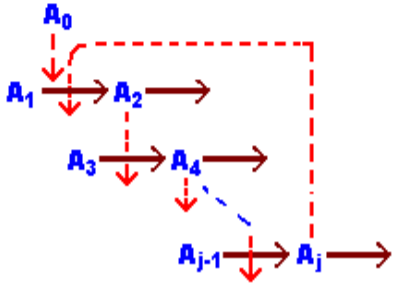


Concerted Control



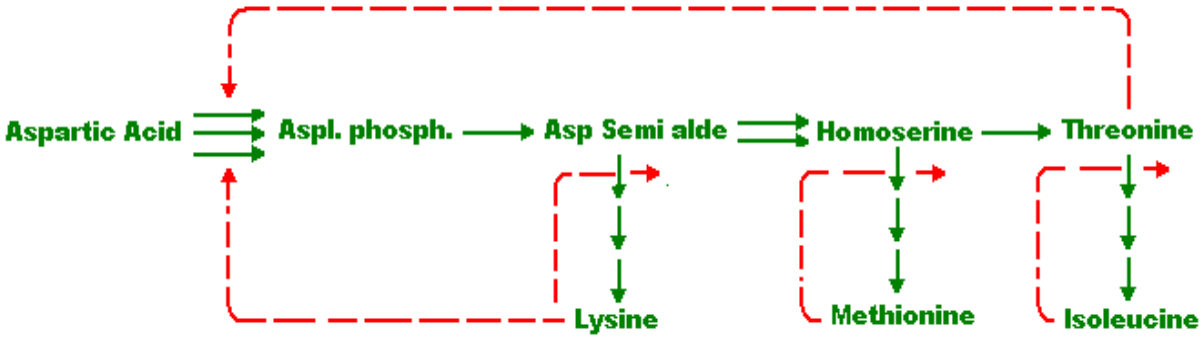
Cumulative Control





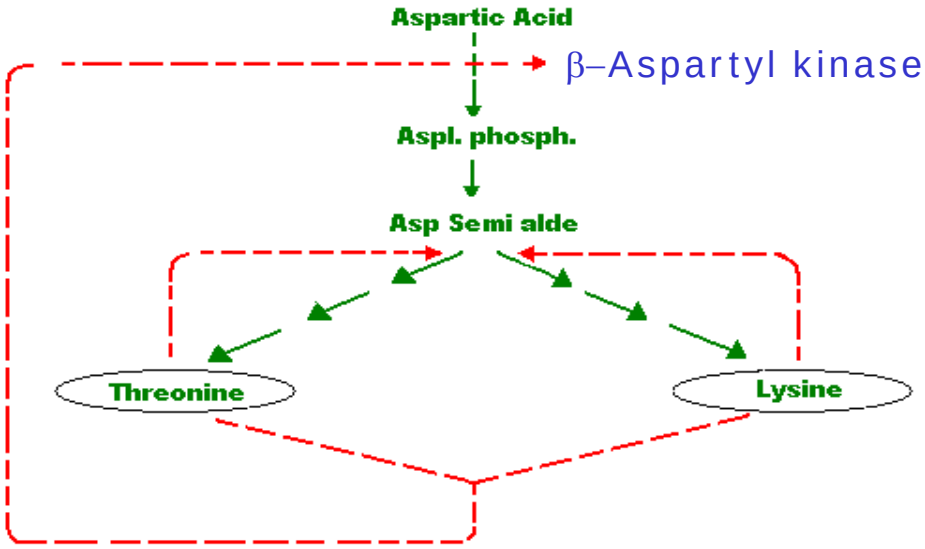
Cascaded Control

Biosynthesis of Aromatic Amino Acids



Divalent, Nested ***and*** Sequential control ***in bacteria***

Concerted control ***of β - Asp. kinase in Rhodopseudomonas capsulatus***



Modelling regulatory modules in biochemical networks

The complex circuitry of biochemical reactions function in a reliable manner under a fair amount of unpredictability in the internal and external environment

Are there sets of regulatory designs that confer robustness in the "functional effectiveness" ?

But biological systems also evolve new structures and functions.

If function is decided by the dynamical state of the output of a pathway, then how much does an increase in noise level lead to changes in the observable of the dynamical system ?

UNDER NOISE

For highly robust systems – attractor is unchanged but fuzzy

For less robust system - attractor shape changes or approaches a different attractor

EXPECTATIONS FROM THEORETICAL STUDIES

Identification of common patterns of regulation in different pathways

Identification of differences in similar pathways in different organisms *(functional implications)*

Study the role of regulation on the temporal pattern of pathway behaviour

Predict new behaviour

Design new pathways

Correct pathological states

MODELLING BIOCHEMICAL PATHWAYS

Three complementary approaches

REVERSE ENGINEERING

Model existing pathways based on information derived from –

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

LARGE NETWORKS

Construction of functionally related pathways from large scale gene expression and protein interaction data

FORWARD ENGINEERING

→ *'Rational Network Design'*

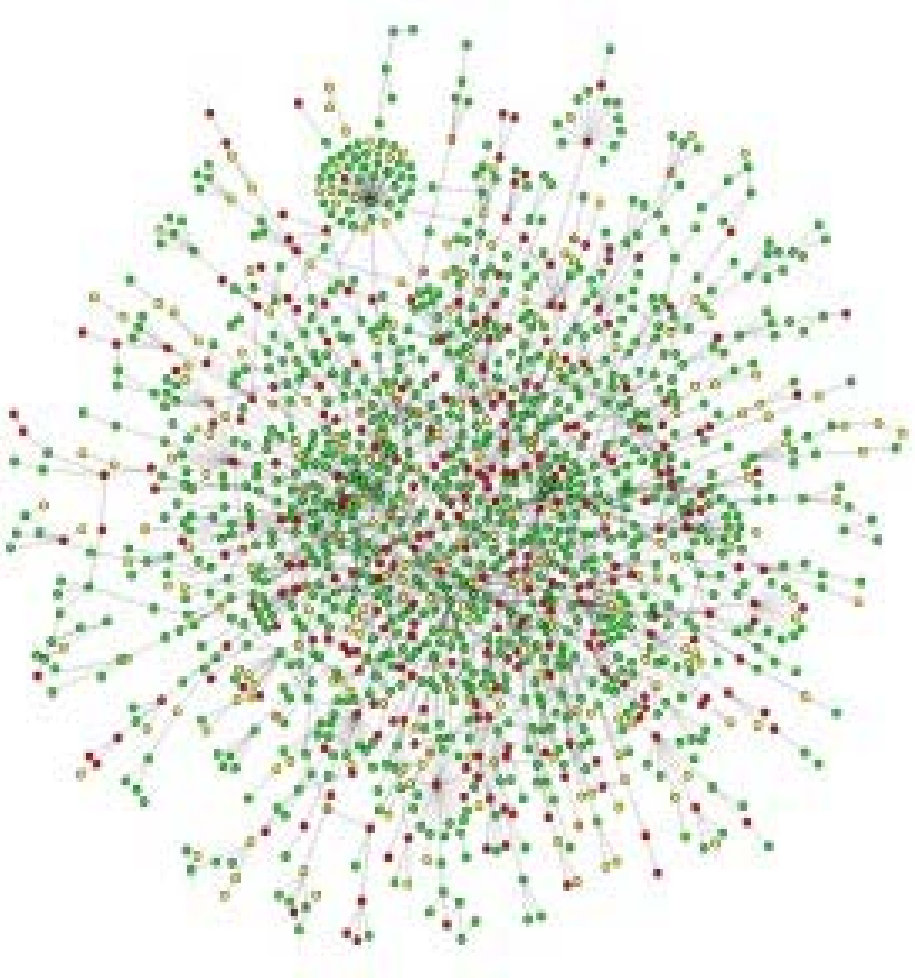
All designs that are not physically forbidden are realizable, but not all realizable designs are functionally effective (in relation to context and constraints of the system and environment).

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

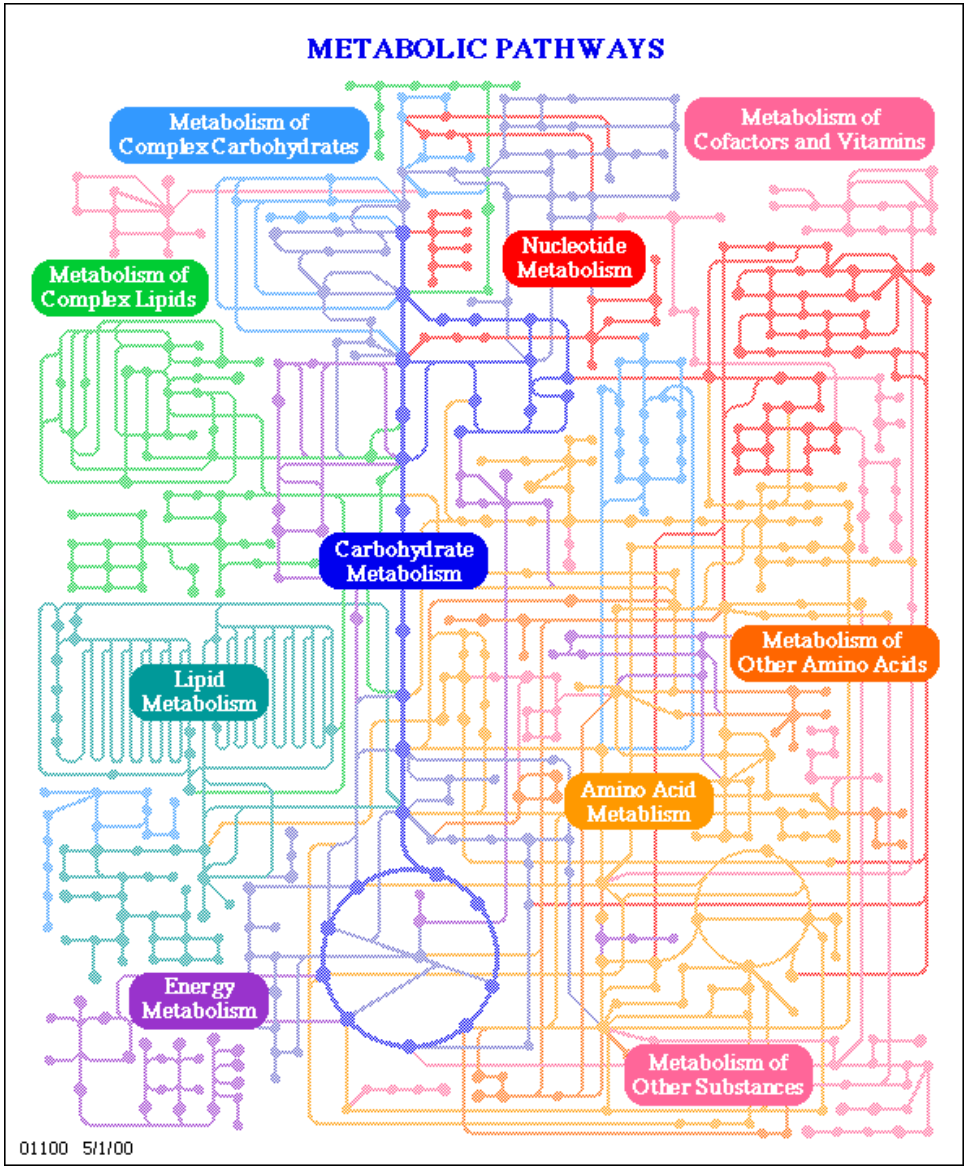
Synthetic oscillatory circuit; Toggle switch in bacteria; Amplifiers of gene expression.

Large networks

***S. cerevisiae* protein–protein interaction network**



Wire diagram of the metabolic pathways in (*E.coli*)



Do these large networks that look similar have similar properties ?

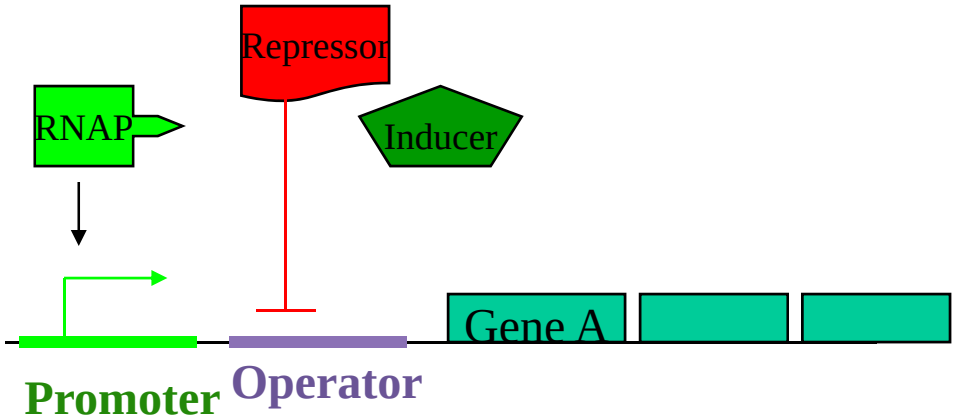
Forward Engineering of Networks

(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. Some aspects of biological systems can be considered as Boolean.



Jacob & Monod Model of the prokaryotic operon (1961)

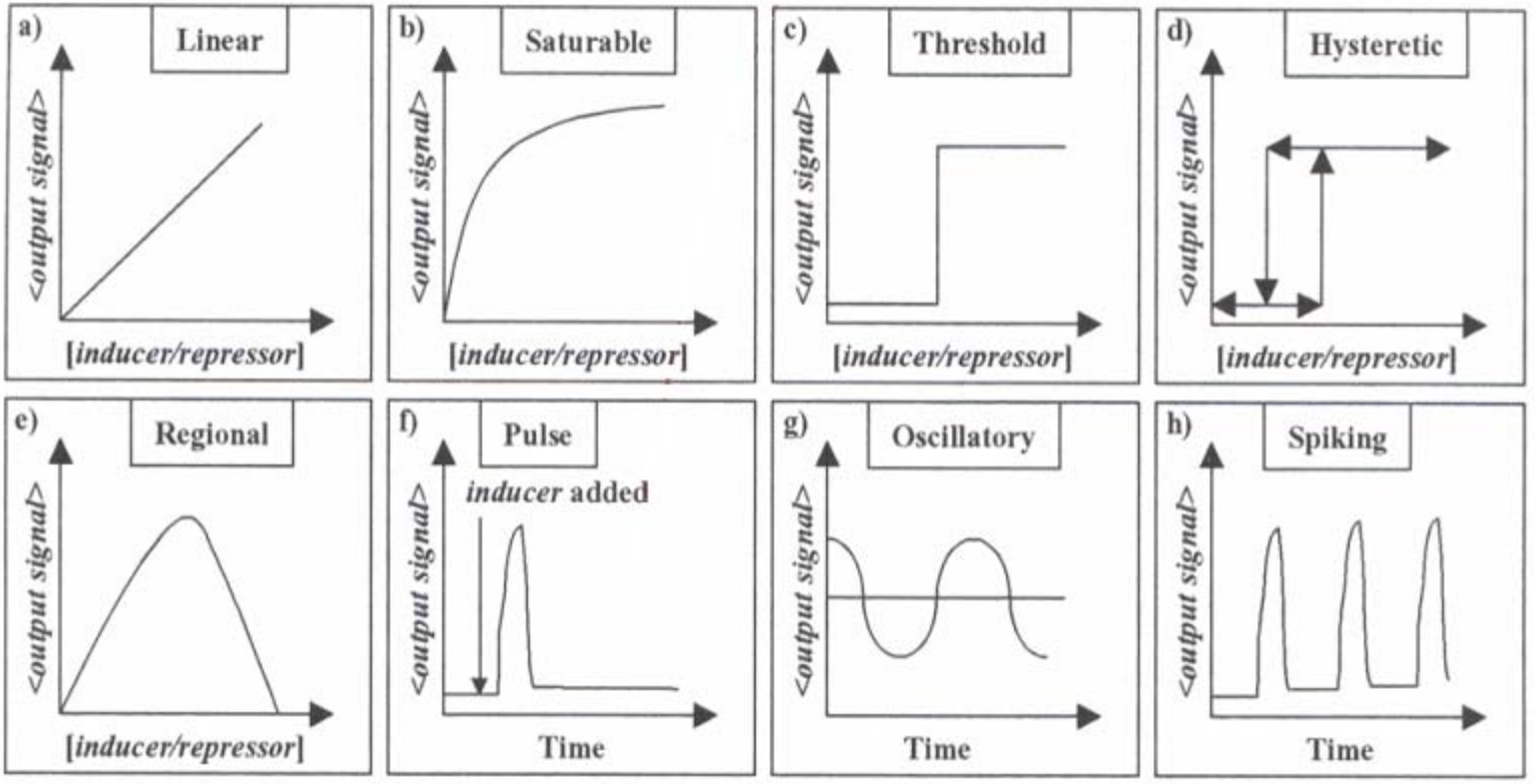
Such " Rational Network Design " can -

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

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)*

may help in controlled gene expression in gene therapy



Develop integrated computational infrastructure for

Computer Aided Design (CAD) of genetic circuits

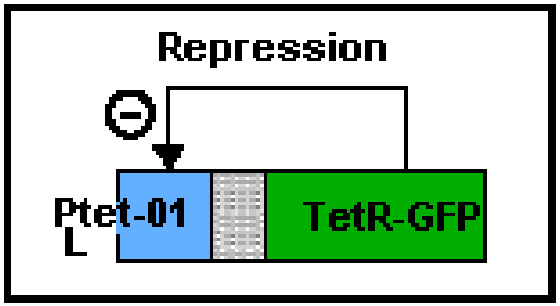
Simulation and dynamic analysis

Build increasingly complex genetic circuits using well-characterized parts

Synthetic transcriptional regulatory networks

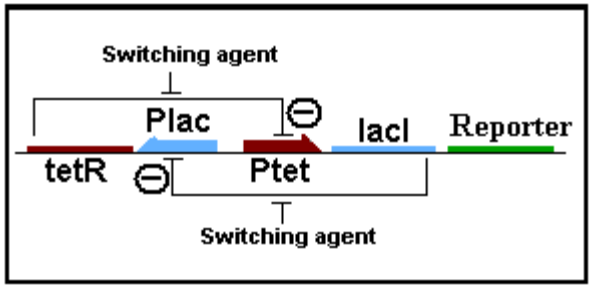
1.Single negative feedback:

Induced by ATC - an analog of tetracycline. Nature(2000).



2.Two nested negative feedback:

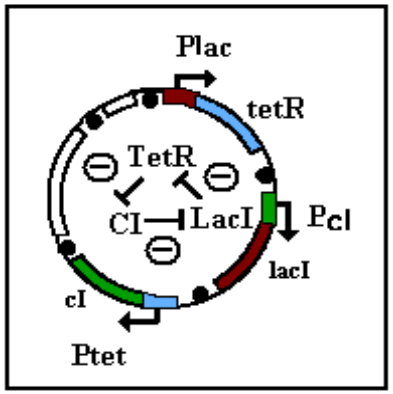
The Toggle:- switched by IPTG and ATC which induce the plac and ptet. Nature (2000)



3.Three negative feedback:

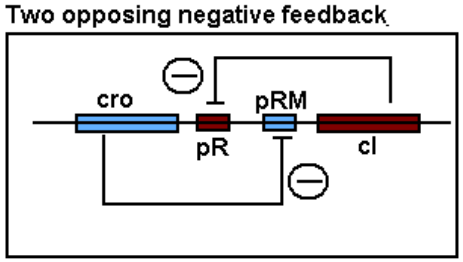
The repressilator:-

Oscillates due to mutual repression of the three repressors Nature(2000)

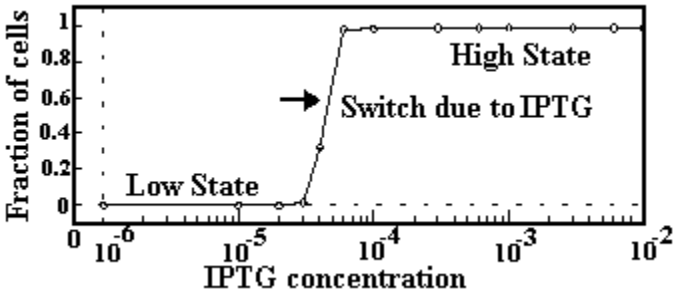


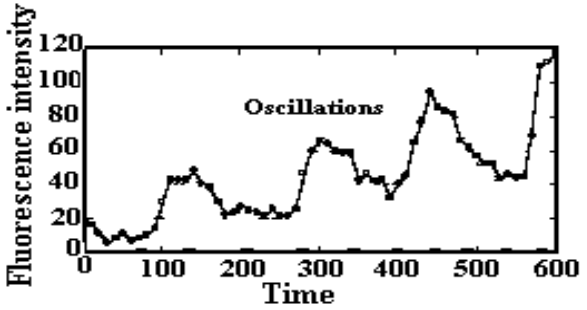
The bacteriophage λ paradigm

Two opposing negative feedback processes lead to switch-like behaviour



Switch like behaviour....





REVERSE ENGINEERING

Model existing pathways based on
biochemical & genetic information

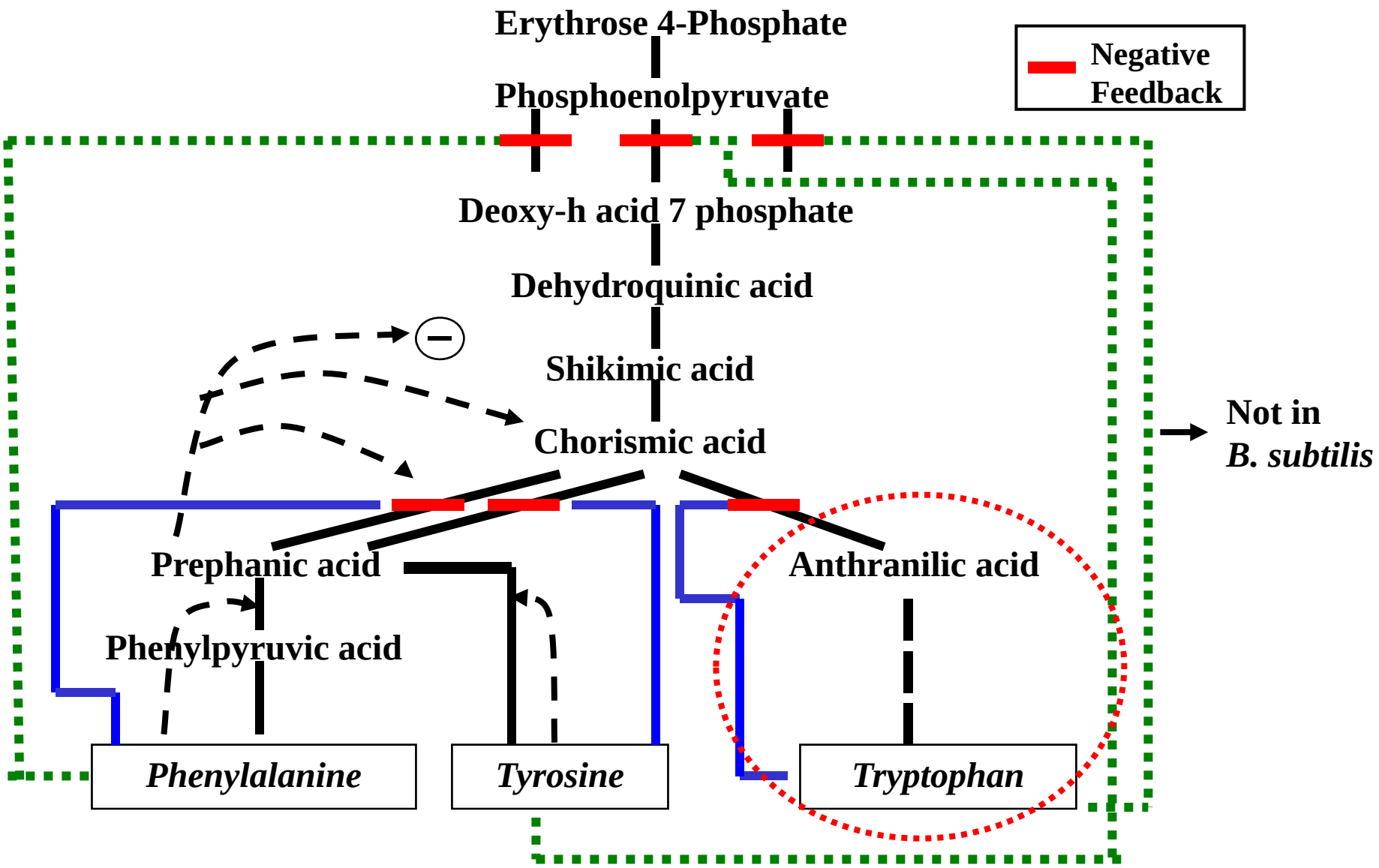
Construct simple mathematical models based on real biological pathways .

Study the dynamics of simple pathways having different structural designs/arrangements of feedback regulation

Study the dynamic behaviour of these pathways under realistic changes (mutations) and stochastic variation in reaction rates or concentration of substrates.

HOW DO PATTERNS OF REGULATION AFFECT FUNCTIONAL DYNAMICS ?

Biosynthesis of Aromatic Amino Acids



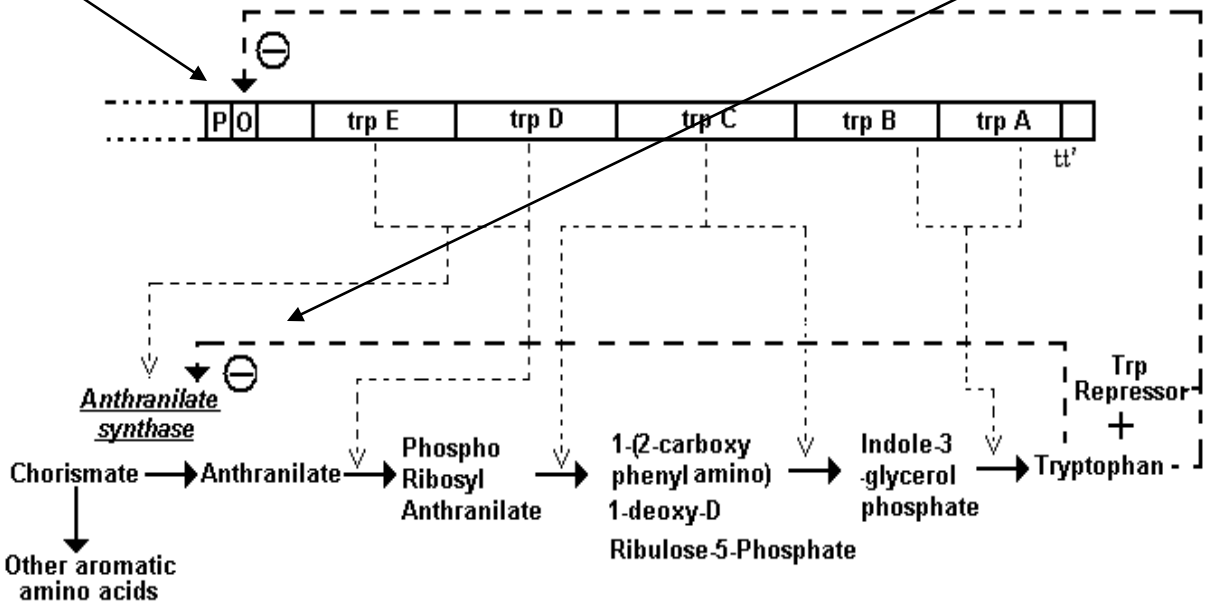
Variation in regulation of the same pathway in different microorganisms

Dual nested feedback in *E.coli* Tryptophan biosynthetic pathway

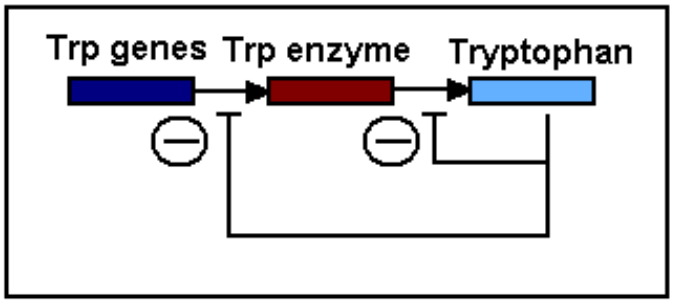
GENETIC — **Two levels of control** — **METABOLIC**

Gene repression through Trp repressor

Enzyme inhibition by end product



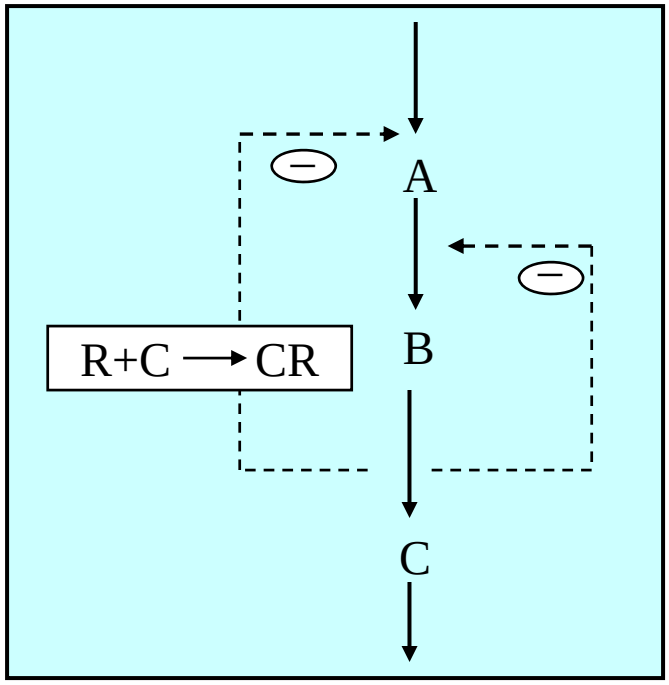
Two nested negative feedback



Modelling Tryptophan Biosynthetic Pathway

Facts & Assumptions

- *5 contiguous structural genes (trpEDCBA) code for the enzymes*
- *Single polycistronic mRNA (7000 nucl.) under normal transcription*
- *All enzymes act in a concerted manner and regulation is on the first enzyme*



mRNA (A),

Enzyme (B),

Tryptophan (C)

Repressor (R) – *inactive*

Bound/Active Repressor (CR)

Modelling Tryptophan Biosynthetic Pathway

Rate of change of

mRNA (A), Enzyme (B), and Tryptophan (C)

concentrations with time

$$\frac{dA}{dt} = F(C) - K_1A$$

$$\frac{dB}{dt} = K_eA - K_2A$$

$$\frac{dC}{dt} = G(B,C) - K_DC - F'(C)$$

Each function represents a biochemical process

F(C) - synthesis of mRNA depends on the repressor-mediated process – a function of *C*

G(B,C) - endproduct synthesis is a function of enzyme inhibition which depends on enzyme (B) and endproduct (*C*) concentrations

F'(C) - utilisation of *C* in cellular processes (protein synthesis)

K_eA - enzyme synthesis $\propto$ to conc. of *A*

K_1A, K_2A, K_DC - degradation kinetics of *A*, B, *C* are first order processes

F(C) represents the genetic repression process - 2 step process

- Tryptophan - Repressor binding (*active repressor*)
- active repressor-operator binding

Fraction of repressor bound to tryptophan

If binding is co-operative

n = Hill co-efficient, K_R = pseudo-Michaelis constant

$$\frac{C^n}{K_R^n + C^n}$$

$$\frac{nC}{K_d + C}$$

If the binding sites are identical and non-interacting

n = binding sites, K_d = dissociation constant

F(C) for the two cases would be

Cooperative binding

$$F(C) = DK_m\left(\frac{r}{1+r}\right)\frac{K_R^n}{K_R^n + (1+r)C^n} + \frac{DK_m}{1+r}$$

Non-cooperative binding

$$F(C) = DK_m\left(\frac{r}{1+r}\right)\frac{K_d}{K_d + (1+r)C} + \frac{DK_m}{1+r}$$

G(B,C) represents the metabolic inhibition process

Tryptophan - Anthranilate Synthase binding follows Michaelis-Menten kinetics with two binding sites

(K_I = *pseudo michaelis constant*)

$$G(B, C) = \frac{K_I^2}{K_I^2 + C^2}$$

F'(C) represents the rate of utilisation of tryptophan in cellular processes *(e.g., protein synthesis)*

Hyperbolic saturation function

V_{max} = *maximum rate of utilisation*

K_G= *pseudo michaelis constant*

$$F'(C) = \frac{V_{max} C}{K_G + C}$$

Approximation: $F'(C) = V_{max}$

(Bliss et al 1983, Tyson,1983, Painter & Tyson 1984)

Tryptophan Biosynthetic Pathway Model

The time variation of concentrations of A,B, and C are -

$$\frac{dA}{dt} = \frac{r}{1+r}\frac{K_m D K_d}{K_d + (1+r)C} + \frac{K_m D}{1+r} - K_1 A$$

$$\frac{K_R^{\,n}}{K_R^{\,n} + (1+r)C^n}$$

$$\frac{dB}{dt} = K_e A - K_2 B$$

$$\frac{dC}{dt} = K_P \frac{K_I^{\,2} B}{K_I^{\,2} + C^2} - K_D C - \frac{V_{max} C}{K_G + C}$$

V_{max}

A ***trp*** **mRNA**
B **Enzyme (*Asase*)**
C **Tryptophan**