Fundamentals of Computational Biology – U1

R G Brajesh
CSVTU, Bhilai

Pre-requisite of this course

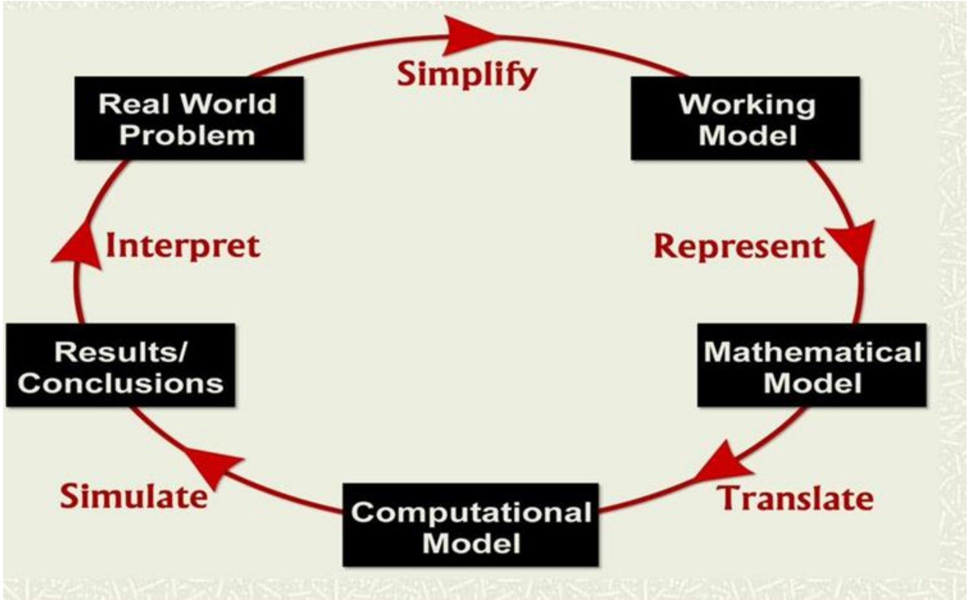
- Mathematical concepts related to building models:
  - Solving Matrix algebra
  - Solving ODE computationally
  - Solving systems of ODE computationally

- Basic Biology
  - Concepts of DNA, RNA, Proteins
  - Game theories in biology
  - Bacterial growth

- Computational tools
  - Matlab operations
  - Python operations

Mathematical Modeling

- Mathematical modeling is the process of **constructing**, **testing**, and **improving** mathematical models.
- Mathematical models should be general in the sense of containing parameters that can be adjusted to strengthen, weaken, or modify the behavior of each process.

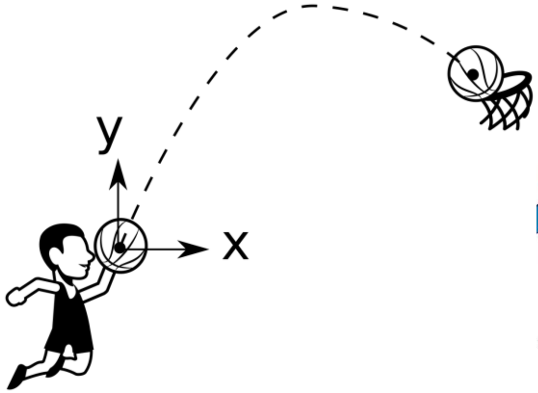


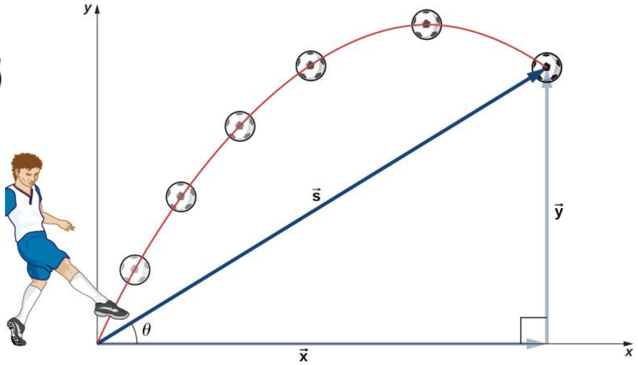
1st step in model building

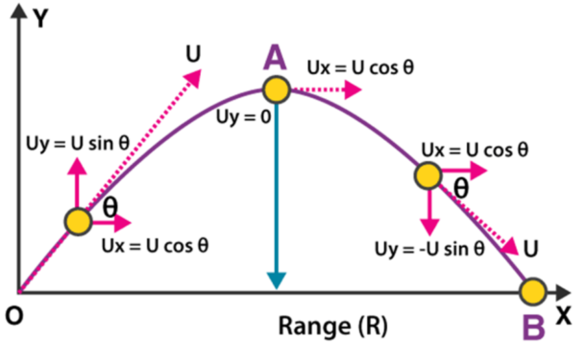
- Observational science: Observe any pattern and try to build a mathematical model.

Example 1: Newton’s law of gravitation

Example 2: Projectile motion

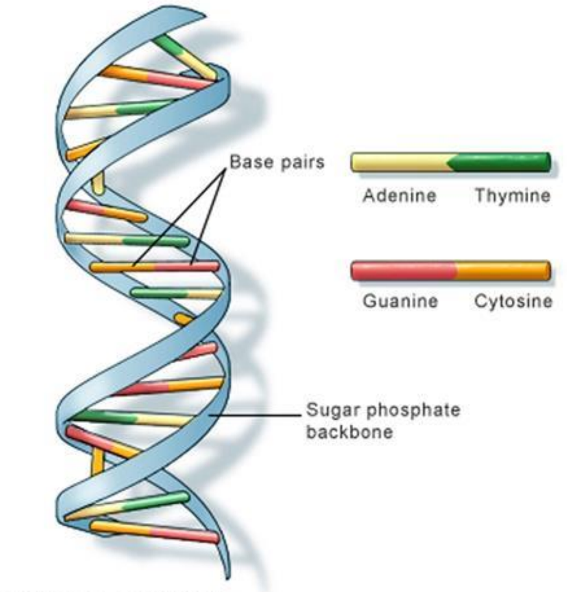






Computational Biology

- Computational Biology encompasses all computational methods and theories applicable to biology and areas of computer based techniques for solving biological problems.



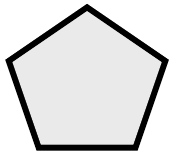
To find out coded information hidden in piece of DNA

Formulate models that can solve higher order problems
Such as mutation, evolution etc.

DNA

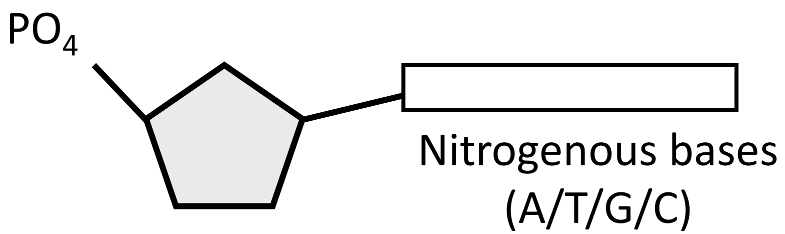
- **DNA** stands for **deoxyribose nucleic acid**
- This chemical substance is present in the nucleus of all cells in all living organisms
- DNA controls all the chemical changes which take place in cells.
- DNA controls the external and internal feature of any cell types (Brain cells, heart cells, liver cells etc).
- DNA controls the features which can be transferable from parents to next generations.

- **DNA** is a very large molecule made up of a long chain of sub-units called as Nucleotides.
- Each nucleotide is made up of a sugar called **deoxyribose** a phosphate group **-PO_4** and an **organic base.**
- **Deoxyribose** sugar is a five carbon based sugar molecule which form the backbone of the DNA molecule.



- **organic base or Nitrogenous bases:** four kinds of Nitrogenous bases present in DNA.
  (i) Adenine or A (ii) Guanine or G
  (iii) Cytosine or C (iv) Thiamine or T

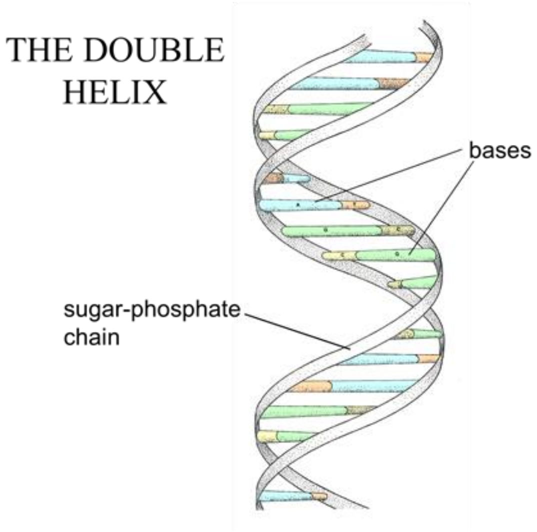
This is how a single unit of nucleotide look like:



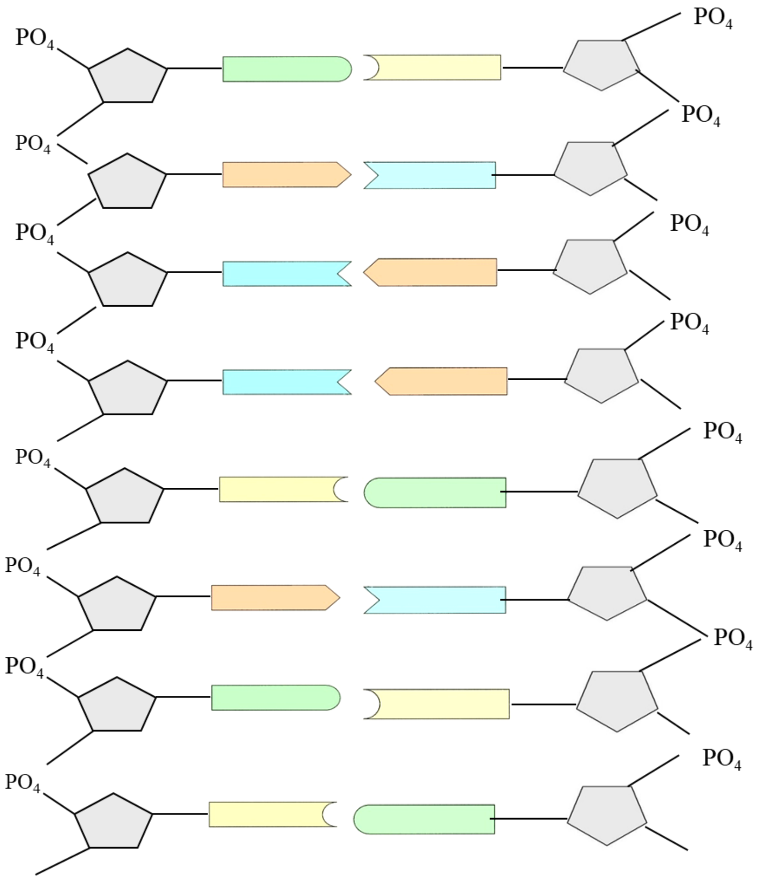
DNA

- Base pairing in among nitrogenous bases are always conserved.
- Adenine always pairs with Thymine with the help of two hydrogen bonds
- Cytosine always pairs with Guanine with the help of three hydrogen bonds

This is how we represent a DNA molecule



- Chemically this is how a double stranded DNA actually look like



RNA

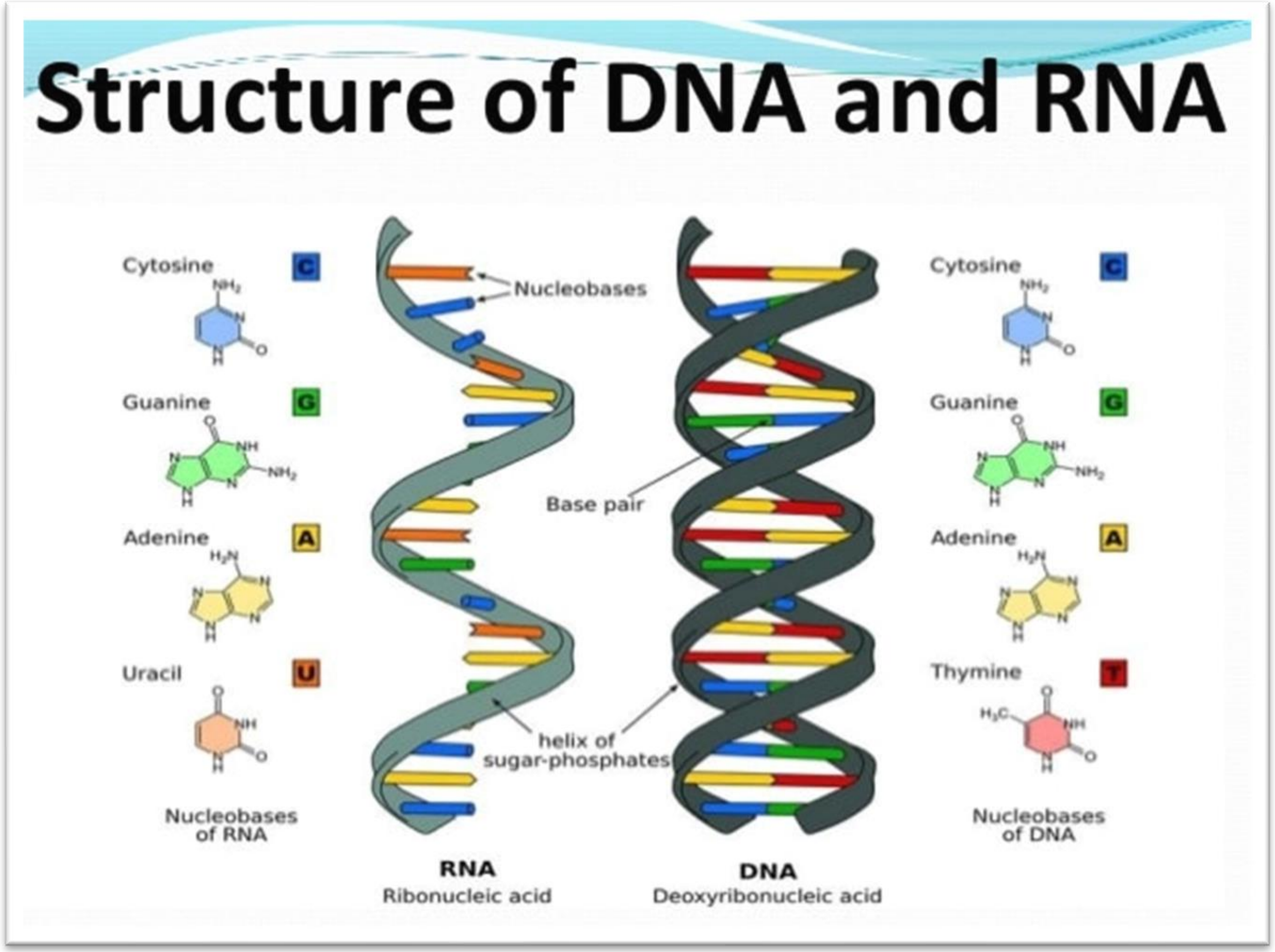
- RNA is chemically very similar to DNA but it exist as single stranded form.
- There are two important differences
- Four bases present in RNA are:

  Adenine(A)

  Guanine(G)

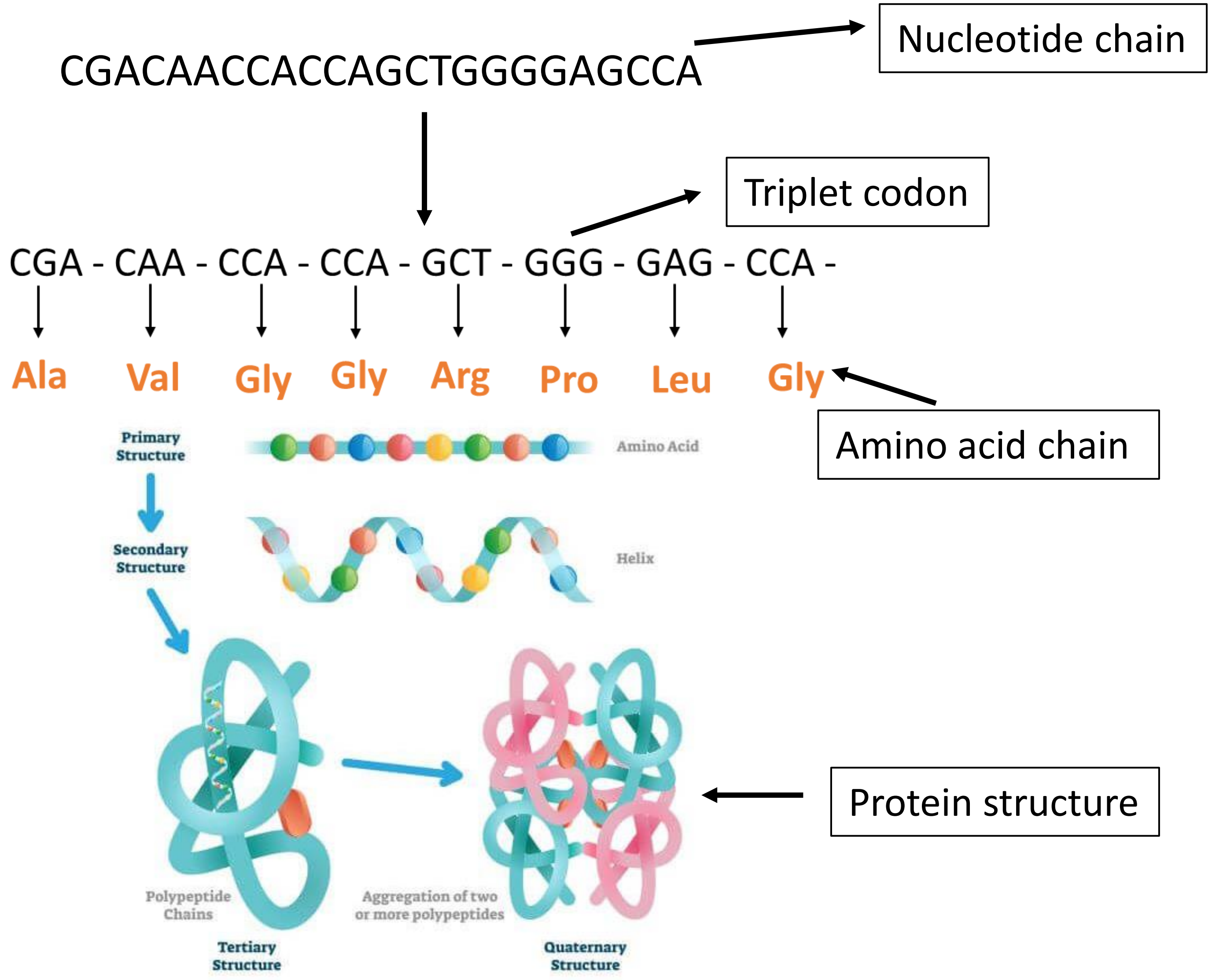
  Cytosine (C)

  Uracil (U)
- RNA nucleotides contain a different sugar molecule(ribose)



Proteins

- They are building blocks of living organism
- It is a large molecule that is composed of sequences of amino acids.
- There are 20 amino acids which are divided into classes
- Information related to amino acid sequence is stored in DNA in the form of triplet code also called as codons. Three of the nitrogenous bases forms a single unit codon and it gives information related to specific amino acid. For example:



Genetics and Evolution

- Mutation
  - The changing of the structure of a gene, resulting in a variant form that may be transmitted to subsequent generations, caused by the alteration of single base units in DNA.
- Natural selection
  - The process whereby organisms better adapted to their environment tend to survive and produce more offspring.
- Genetic Drift
  - Variation in the relative frequency of different genotypes in a small population.

Models in Biology and Physics

- Most physical processes are well described by “physical laws” valid in a wide variety of settings.
  - It is easy to get physical science models.

- Most biological processes are too complicated to be described by simple mathematical formulas.
  - It is hard to get good models for biology.
  - A model that works in one setting may fail in a different setting.

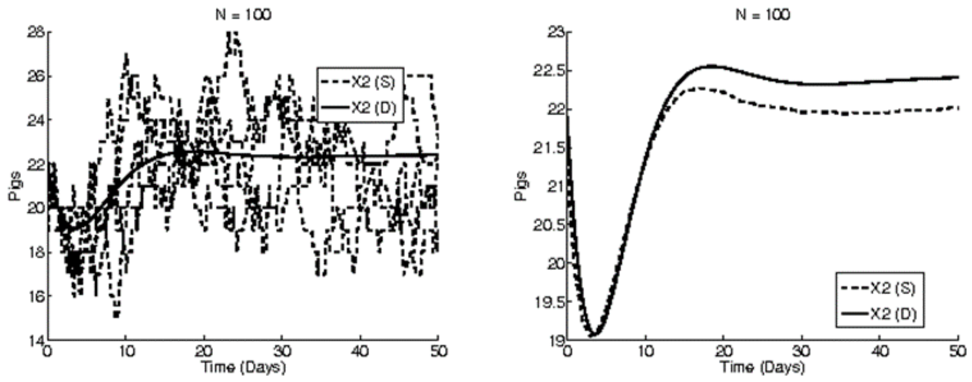
Models in Biology

Mathematical models

Live models Rats/pigeons

Modeling by Discovery

- Mathematical modeling requires good scientific intuition.
- Scientific intuition can be developed by observation.
- Detailed observation in biological scenarios can be very difficult or very time-consuming, so can seldom be done in a math course.
- Such system where it can be represented via simple mathematical expression often called as Deterministic models.
- More realistic models where randomness affects the system at greater extent is called as stochastic models.



(a) (b)

Evolutionary game theory models

- Prisoner's Dilemma
- Snow drift model
- Hawk and Dove decision models.

Prisoner's Dilemma: It is one of most popular game among decision making, two prisoner's were given choice to confess. Based on their choices there are four different outcomes, these outcomes are arranged in the form of matrix also called as payoff matrix. Most of the time both prisoner's end up confessing the crime and gets maximum sentences.

Snow drift model: This game talks about decision of cooperation or defection. In one hand cooperation is advantageous to both the parties at equal cost. But, if one party choose to defect/cheat he gets equal benefit at minimum cost. So it appears that being cheaters have more evolutionary advantage.

Prisoner's Dilemma payoff matrix

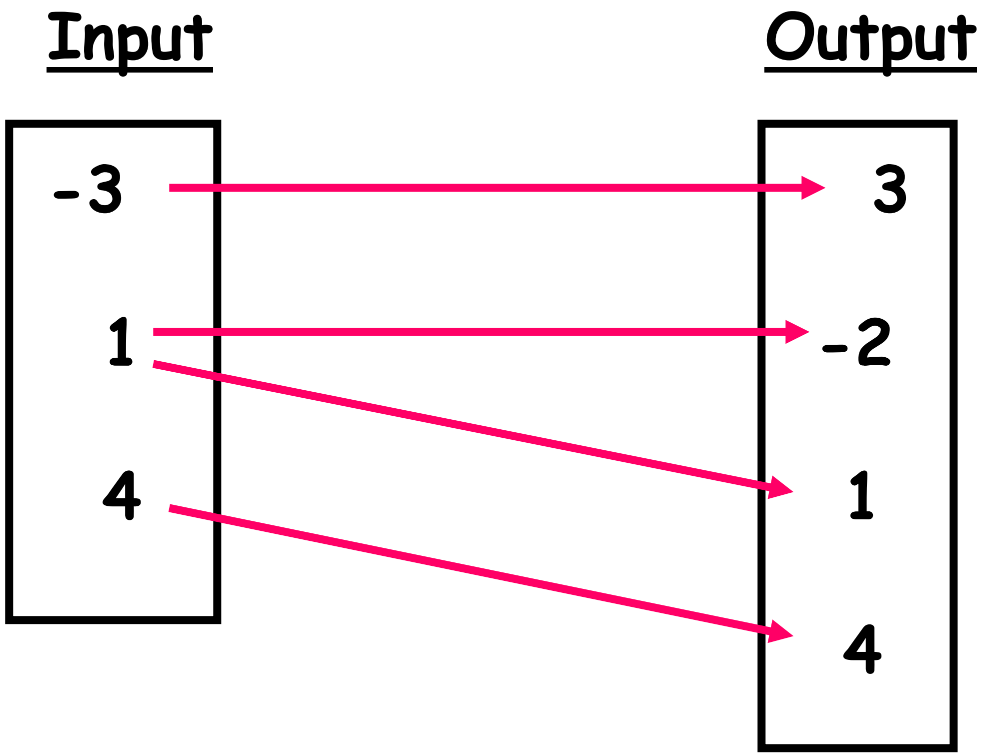
	Don not confess	Confess
Don not confess	1 Year / 1 Year	Parole / Life
Confess	Life / Parole	20 Year / 20 Year

Snow drift payoff matrix

	Cooperate	**Defect**
Cooperate	$\left(b - \frac{c}{2}, b - \frac{c}{2}\right)$	$(b - c, b)$
Defect	$(b, b - c)$	$(0, 0)$

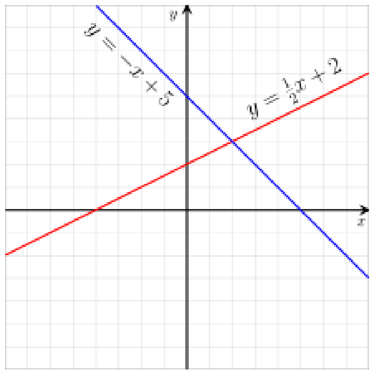
Functions

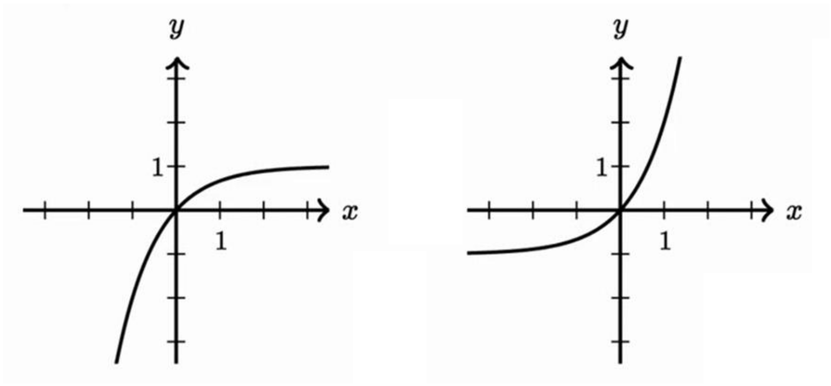
- A **function** is a *relation* with the property that for each input there is one output.

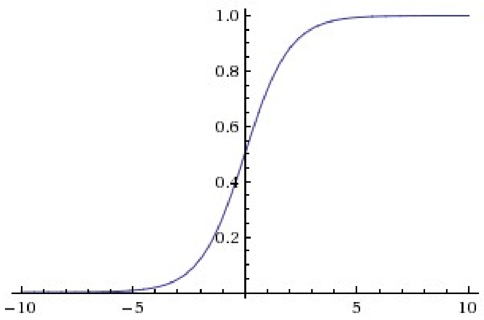


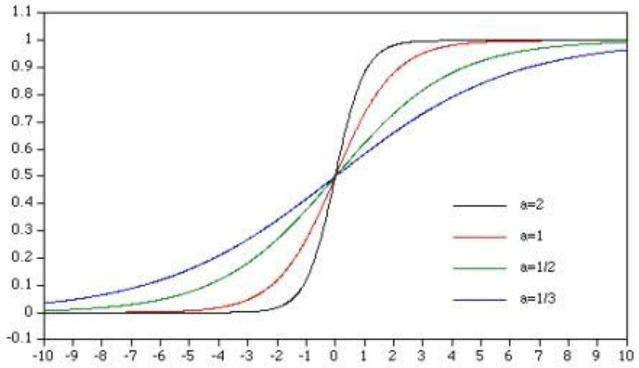
Functions

- Representation of a function
- $y = f(x)$
- $y = f(x_1, x_2 \dots \dots .. x_n)$
- It is difficult to draw a direct relation with each independent variable in biology, so we deduce the relation of each variable with its output and represent graphically.









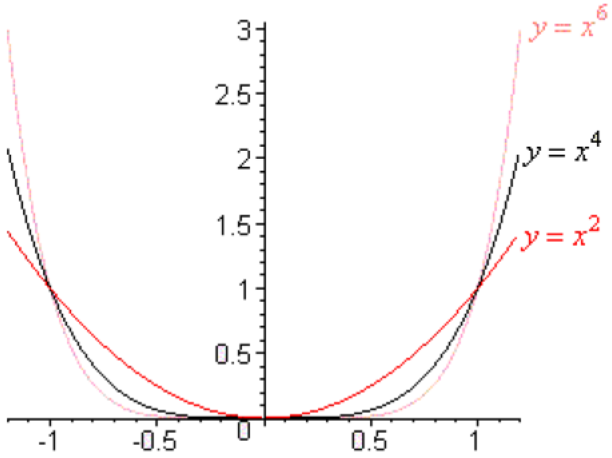
Functions

$y = f(x) = mx + c$

Range for this is {-5,5}

$y = f(x) = ax^2 + bx + c$

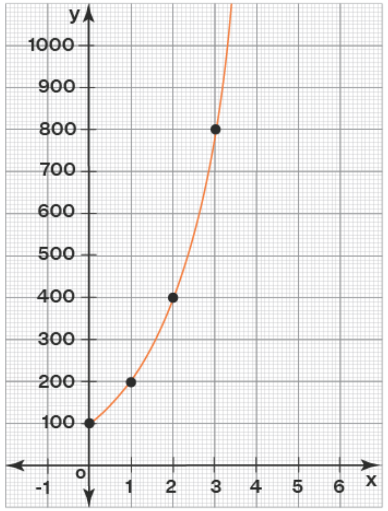
Range for this is {-5,5}



$y = f(x) = x^n$

or

x	y
0	100
1	200
2	400
3	800



ODE

In mathematics, an **ordinary differential equation** (**ODE**) is a differential equation containing one or more functions of one independent variable and the derivatives of those functions.

The **order** of an ordinary differential equations is the order of the highest order derivative

$Examples:$

$$\frac{dy}{dx} - y = e^{x}$$

First order ODE

$$\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 2y = \cos(x)$$

Second order ODE

$$\left(\frac{d^2y}{dx^2}\right)^3 - \frac{dy}{dx} + 2y^4 = 1$$

Second order ODE

ODE model building

- Here is an example of chemical kinetics:

$$a \underset{k2}{\overset{k1}{\leftrightarrow}} b \underset{k4}{\overset{k3}{\leftrightarrow}} c \overset{k5}{\rightarrow} d$$

- Rate of reaction can be written as difference in rate of formation and rate of reduction.
- For example: $a \underset{k2}{\overset{k1}{\leftrightarrow}} b$

$$\frac{d[a]}{dx} = -k1[a] + k2[b]$$

$$\frac{d[b]}{dx} = k1[a] - k2[b]$$

- Rate of change of each different constituents in above equation can be written as:

$$a \underset{k2}{\overset{k1}{\leftrightarrow}} b \underset{k4}{\overset{k3}{\leftrightarrow}} c \overset{k5}{\rightarrow} d$$

$$\frac{d[a]}{dx} = -k1[a] + k2[b]$$

$$\frac{d[b]}{dx} = k1[a] - k2[b] - k3[b] + k4[c]$$

$$\frac{d[c]}{dx} = k3[b] - k4[c] - k5[c]$$

$$\frac{d[d]}{dx} = k5[c]$$

Solving ODE

- By Integration factors

- Write ODE in general formula

$$\frac{dy}{dx} + P(x)y = Q(x)$$

$$IF = I(x) = e^{\int P(x)dx}$$

General solution can be written as:

$$y = \frac{1}{I(x)}[\int I(x)Q(x)dx + C]$$

Solve for:

$$\frac{dy}{dx} + 2y = 2e^x$$

$$xy' + 4y = 2x^3$$

$$y' + \frac{2}{t}y = t + 1 + \frac{1}{t}$$

Solving ODE

$$\frac{dy}{dx} + 2y = 2e^x \longrightarrow \frac{dy}{dx} + P(x)y = Q(x) \longrightarrow P(x) = 2; = Q(x) = 2e^x$$

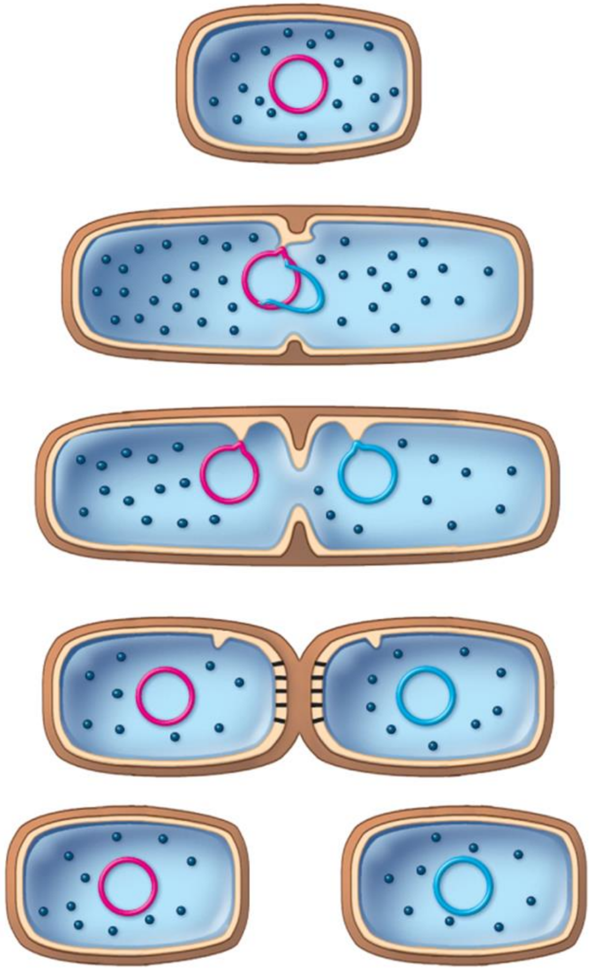
$$IF = I(x) = e^{\int P(x)dx} \longrightarrow I(x) = e^{\int 2\,dx} \longrightarrow I(x) = e^{2x}$$

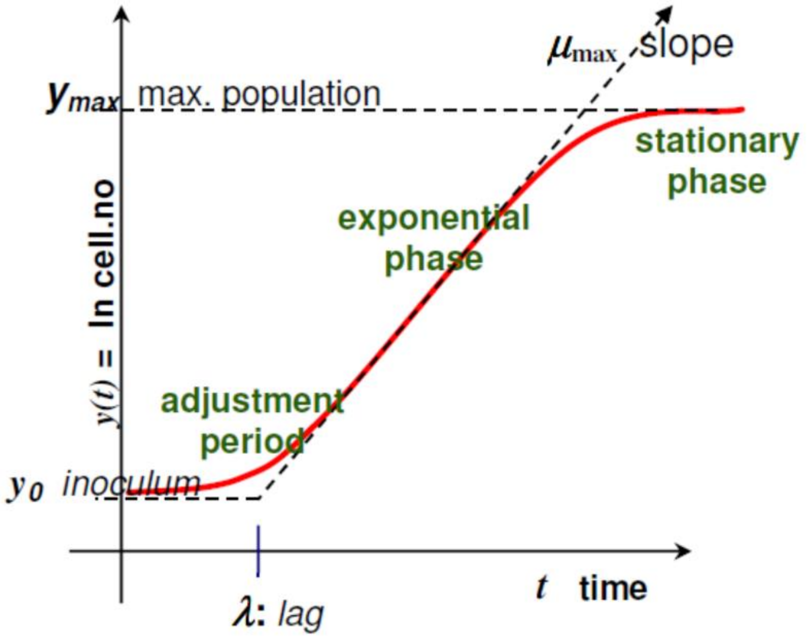
General solution :

$$y = \frac{1}{I(x)}[\int I(x)Q(x)dx + C] \qquad y = \frac{1}{e^{2x}}[\int e^{2x}2e^x dx + C] \longrightarrow y = \frac{1}{e^{2x}}[\int 2e^{3x}dx + C]$$

$$y = \frac{1}{e^{2x}}[2\frac{e^{3x}}{3} + C] \longrightarrow y = 2\frac{e^x}{3} + Ce^{-2x}]$$

Bacterial growth model





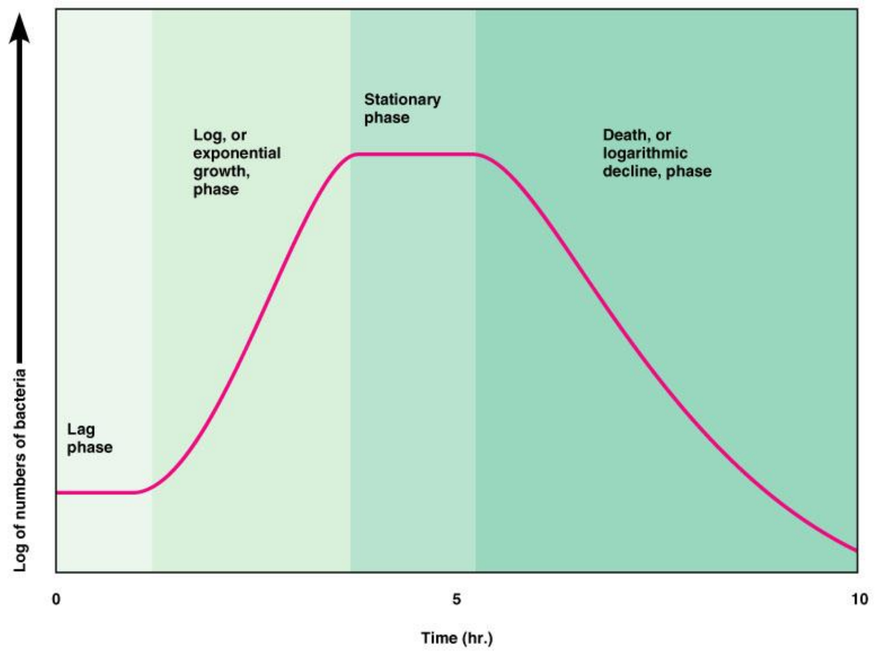
Stages in the Normal Growth Curve

Data from an entire growth period typically produce a curve with a series of phases

- Lag Phase
- Exponential Growth Phase
- Stationary Growth Phase
- Rapidly Declining Phase
- Death Phase

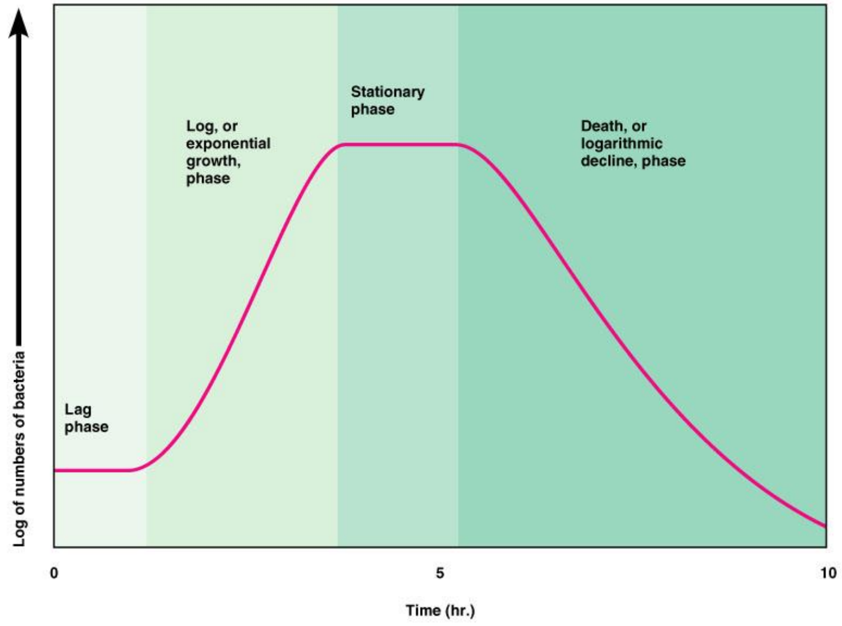
Lag Phase

- Relatively “flat” period
- Newly inoculated cells require a period of adjustment, enlargement, and synthesis
- The cells are not yet multiplying at their maximum rate
- The population of cells is so sparse that the sampling misses them
- Length of lag period varies from one population to another



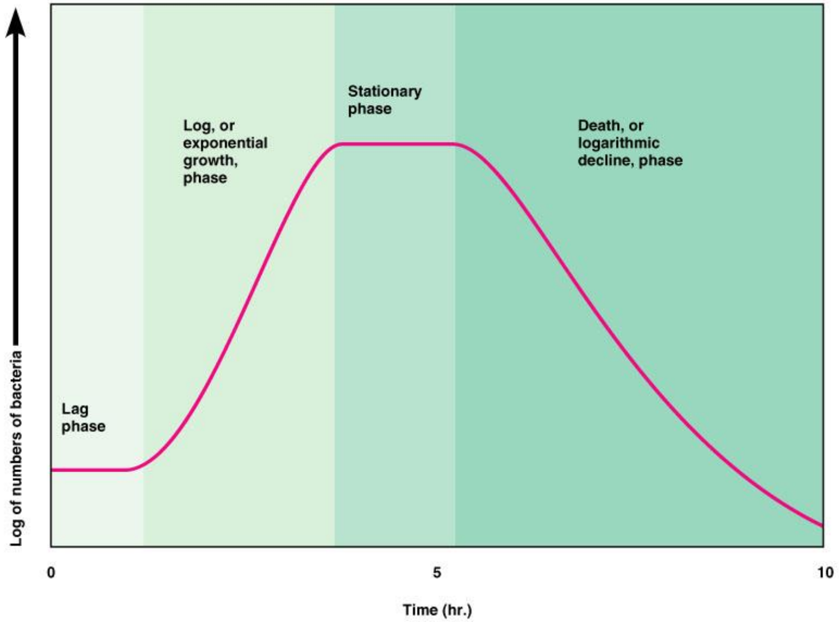
Exponential Growth (Logarithmic or log) Phase

- When the growth curve increases geometrically
- Cells reach the maximum rate of cell division
- Will continue as long as cells have adequate nutrients and the environment is favorable
- The number of cells growing greatly out number the number of cells dying.



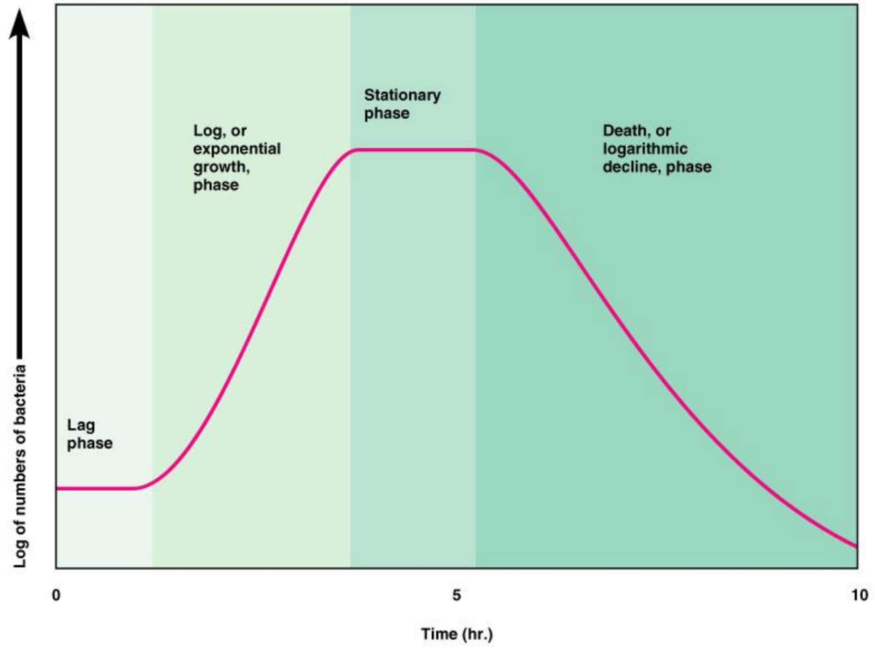
Stationary Growth Phase

- The population enters a survival mode in which cells stop growing or grow slowly
  - The rate of cell inhibition or death balances out the rate of multiplication
  - Depleted nutrients and oxygen
  - Excretion of organic acids and other biochemical pollutants into the growth medium
  - The number of cells growing will equal the amount of cells dying.
  - Endospores begin to form in this phase.



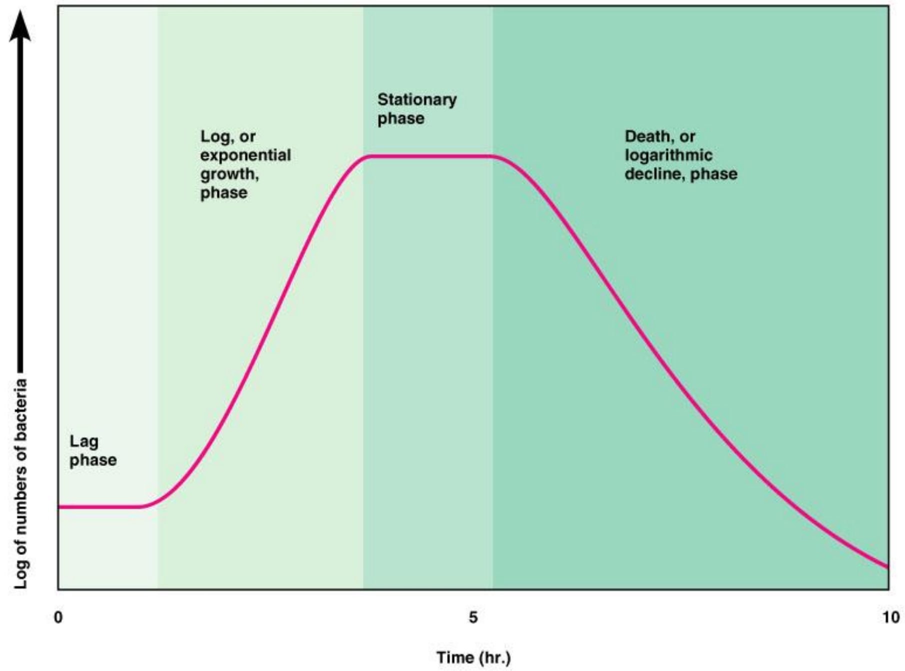
Rapidly Declining Phase

- The curve dips downward
- Cells begin to die at an exponential rate
- The amount of cells dying out numbers the amount of cells growing.
- The dead cells become nutrients for the growing cells.



Death Phase

- The curve continues to dips downward
- Most cellular activity stops
- Endospores are formed and released from the parent cells.



Growth model

- Exponential phase growth can be represented as:
- $\frac{dN}{dt} = \mu N$ *(where N is population size and μ is specific growth rate)*

Solving above equation give rise to:

- $N_t = N_0 e^{\mu t}$ *(where N_t is population size at time t, N_0 and is initial population)*

- For doubling time expression
- $2N_0 = N_0 e^{\mu t}$
- $t_d = \frac{ln2}{\mu}$ *(td is the doubling time)*

- If the growth of a population is being limited by limiting agent for example a substrate then specific growth rate of the organism can be written as:

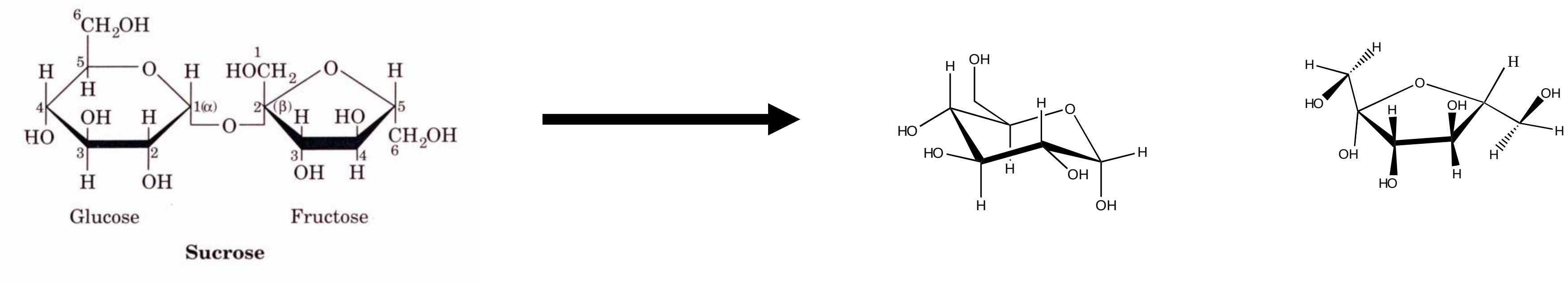
$$\mu = \mu_{max} \frac{S}{k_s + S}$$

Enzyme kinetics

Enzymes follow zero order kinetics when substrate concentrations are high. Zero order means there is no increase in the rate of the reaction when more substrate is added.

Given the following breakdown of sucrose to glucose and fructose

Sucrose + H_20 $\longrightarrow$ Glucose + Fructose



Enzyme kinetics

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

E = Enzyme   S = Substrate P = Product

ES = Enzyme-Substrate complex

k_1 rate constant for the forward reaction

k_{-1} = rate constant for the breakdown of the ES to substrate

k_2 = rate constant for the formation of the products

Enzyme kinetics

$$\frac{d[\mathrm{ES}]}{dt} = k_1[\mathrm{E}][\mathrm{S}] - k_{-1}[\mathrm{ES}] - k_2[\mathrm{ES}]$$ 1

$$v = \frac{d[\mathrm{P}]}{dt} = k_2[\mathrm{ES}]$$

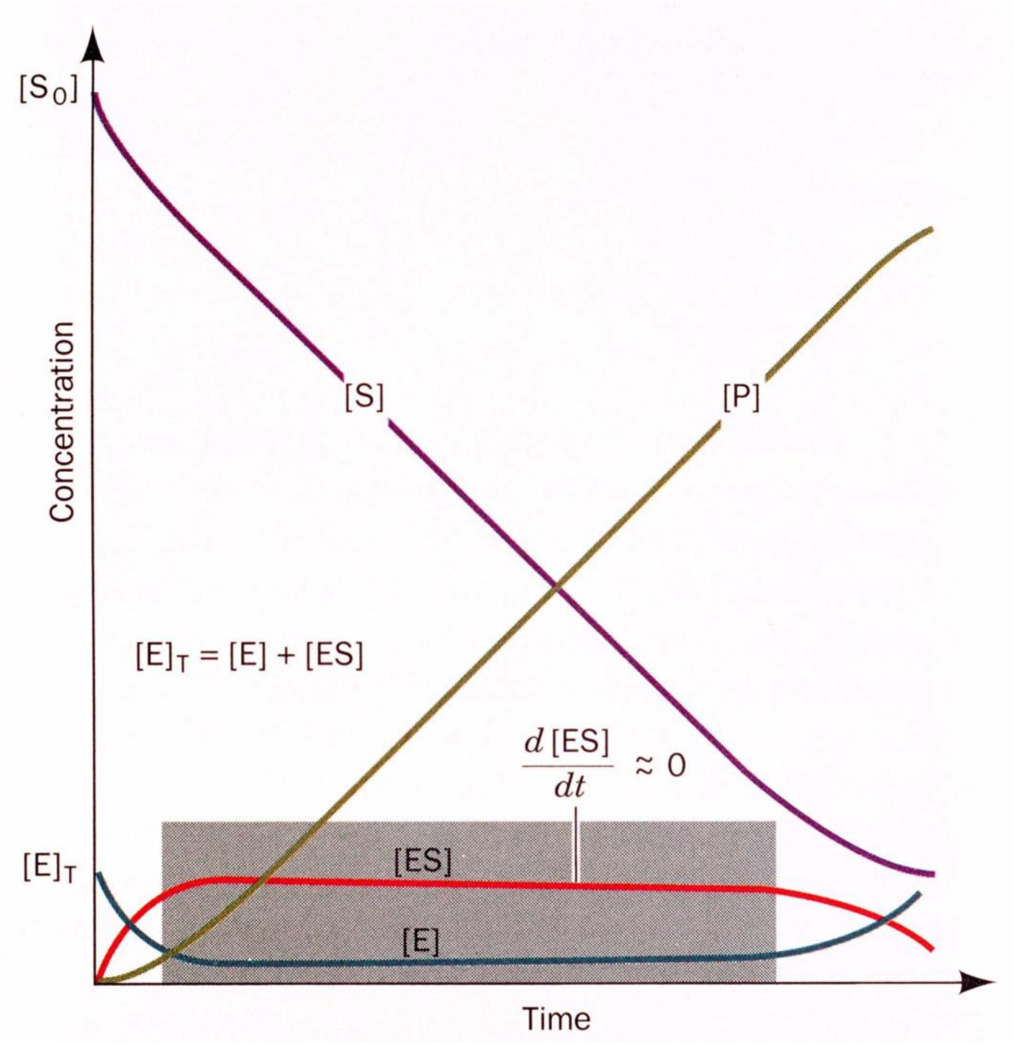
Assumption of equilibrium

$k_{-1} \gg k_2$ the formation of product is so much slower than the formation of the ES complex. That we can assume:

Enzyme kinetics

$$\frac{d[\mathrm{ES}]}{dt} = 0$$

Eq. 2



$$\frac{d[\mathrm{ES}]}{dt} = k_1[\mathrm{E}][\mathrm{S}] - k_{-1}[\mathrm{ES}] - k_2[\mathrm{ES}] \longleftrightarrow k_1[\mathrm{E}][\mathrm{S}] = [ES](k_{-1} + k_2)$$ Eq. 1

$$[\mathrm{E}]_\mathrm{T} = [\mathrm{E}] + [\mathrm{ES}]$$ Eq. 3

Combining 1 + 2 + 3

$$\mathrm{k}_1([\mathrm{E}]_\mathrm{T} - [\mathrm{ES}])[\mathrm{S}] = (\mathrm{k}_{-1} + \mathrm{k}_2)[\mathrm{ES}]$$

rearranging

$$[\mathrm{ES}](\mathrm{k}_{-1} + \mathrm{k}_2 + \mathrm{k}_1[\mathrm{S}]) = \mathrm{k}_1[\mathrm{E}]_\mathrm{T}[\mathrm{S}]$$

Divide by k_1 and solve for [ES]

$$[\mathrm{ES}] = \frac{[\mathrm{E}]_\mathrm{T}[\mathrm{S}]}{\mathrm{K}_M + [\mathrm{S}]}$$

Where

$$\mathrm{K}_M = \frac{\mathrm{k}_{-1} + \mathrm{k}_2}{\mathrm{k}_1}$$

$$v_o = \left(\frac{d[\mathrm{P}]}{dt}\right)_{t=0} = k_2[\mathrm{ES}] = \frac{k_2[\mathrm{E}]_\mathrm{T}[\mathrm{S}]}{\mathrm{K}_M + [\mathrm{S}]}$$

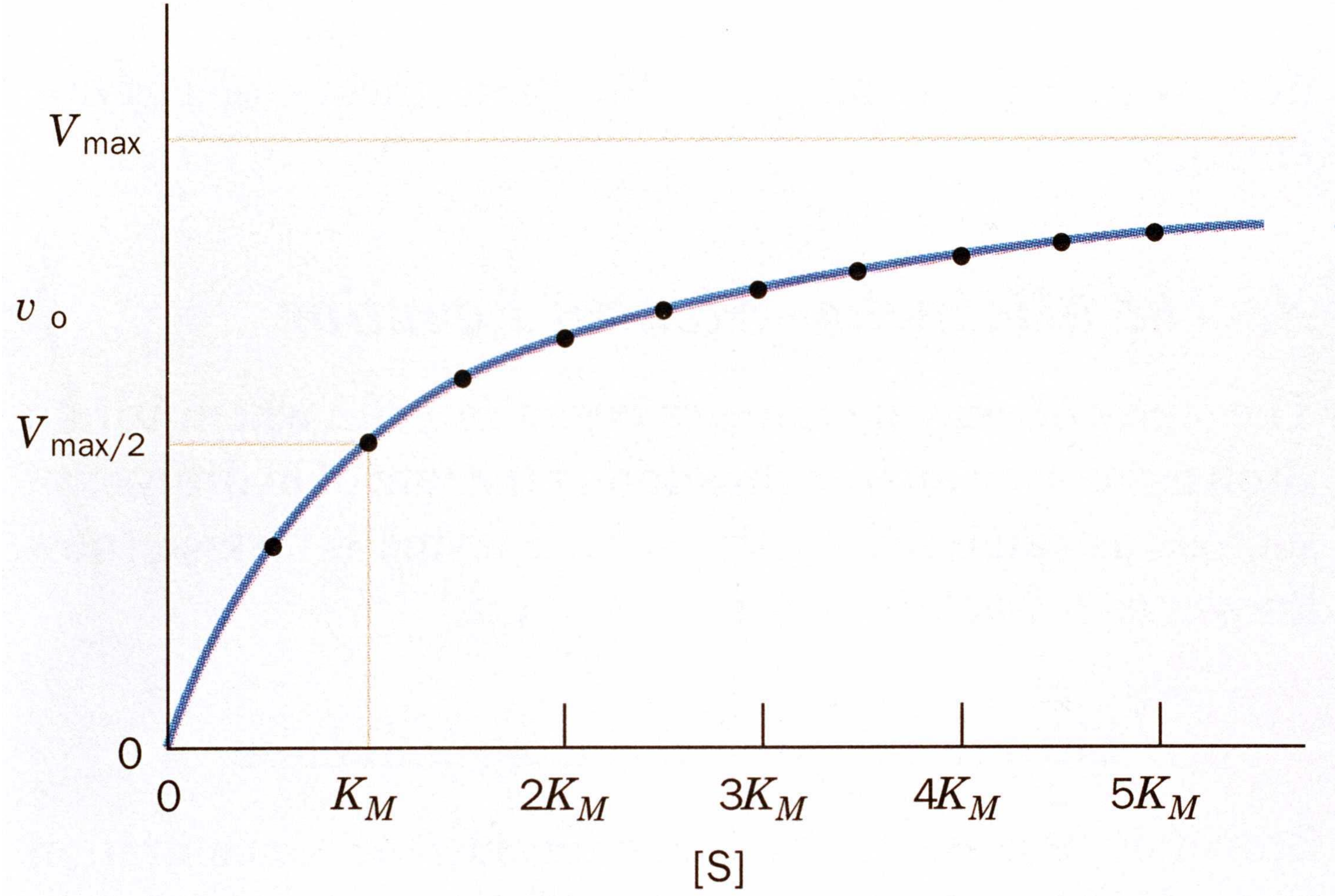
v_o is the initial velocity when the reaction is just starting out.

and Vmax is the maximum velocity

$$\mathrm{V}_{\mathrm{max}} = \mathrm{k}_2[\mathrm{E}]_\mathrm{T}$$

$$v_o = \frac{\mathrm{V}_{\mathrm{max}}[\mathrm{S}]}{\mathrm{K}_M + [\mathrm{S}]}$$

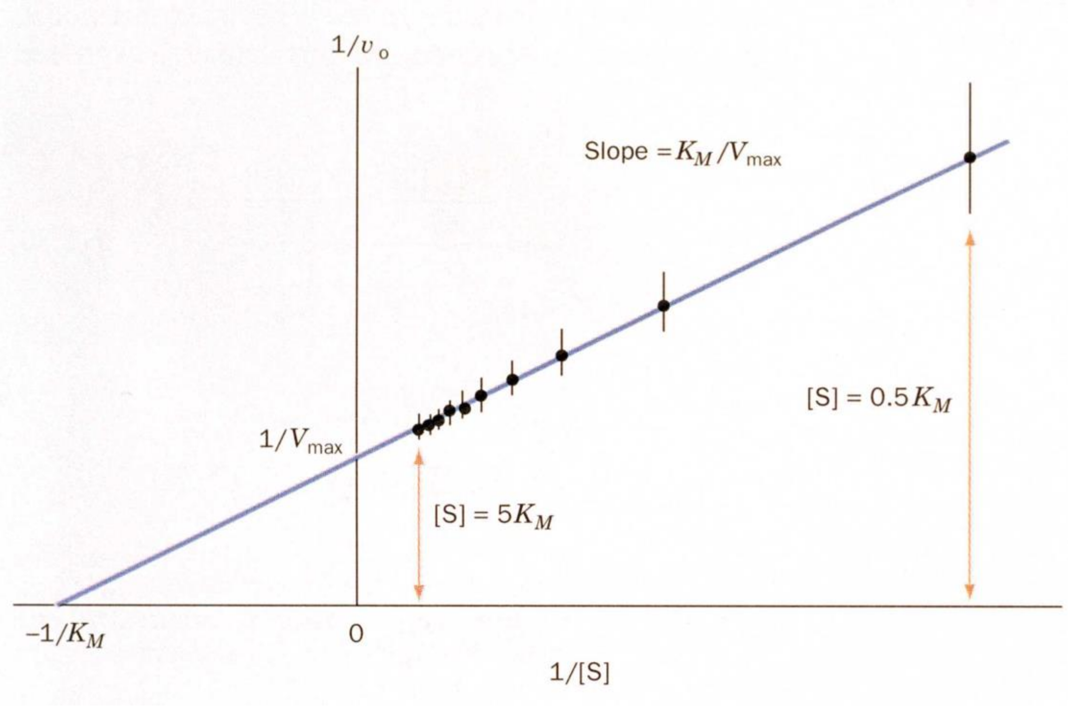
The Michaelis - Menten equation



The K_m is the substrate concentration where v_o equals one-half V_{max}

The double reciprocal plot

$$\frac{1}{v_o} = \left(\frac{K_M}{V_{max}}\right)\frac{1}{[S]} + \frac{1}{V_{max}}$$



Competitive Inhibition

$$
\begin{array}{l}
\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{P} + \mathrm{E} \\
+ \\
\mathrm{I} \\
K_{\mathrm{I}} \Big\updownarrow \\
\mathrm{EI} + \mathrm{S} \longrightarrow \text{NO REACTION}
\end{array}
$$