Course Name: Mathematical Biology
Programme Name: M.A./M.Sc. Mathematics
Course Code: MM18307DCE

Written by

Dr. Mukhtar Ahmad Khanday
Professor & Head
Department of Mathematics
University of Kashmir, Srinagar

Unit-I

Diffusion in Biology

Introduction:

Mathematical modeling is nowadays considered one of the important research areas to deal with the real life problems by using various tools and techniques of mathematics. The main objective of using mathematical methods in solving day today problems is to get better understanding of the system and to provide optimal solutions of the problem. There is a huge utility of mathematical concepts in handling problems arising from life sciences, technology, social sciences, environment, machine learning or artificial intellegence etc. Even mathematics is treated as a backbone of technology. Medical science has made advances by way of utilizing mathematical concepts since last many decades. To bridge the gap between biological systems and mathematical tools. The mathematical biology (or Biomathematics) is a branch of mathematical modeling that deals with the problems arising from biological systems.

Mathematical/Theoretical biology is a discipline which employs theoretical analysis, mathematical methods and abstractions of living organisms to investigate the principles that govern the structure, development and behavior of the systems. The field is also called as mathematical biology or biomathematics to focus on the mathematical side, or theoretical biology to stress towards biological side. Theoretical biology focuses more on the development of theoretical principles for biology while mathematical biology focuses on the use of mathematical tools to study biological systems, even though the two terms are sometimes interchanged.

Mathematical biology aims at the mathematical representation and modeling of biological processes, using techniques and tools of applied mathematics. It can be useful in both theoretical and practical research. Describing systems in a quantitative manner means their behavior can be better simulated, and hence properties can be predicted that might not be evident to the experimenter. This requires precise mathematical models. Because of the complexity of the living

systems, theoretical biology employs several fields of mathematics, and has contributed to the development of new techniques.

During 13th century the disclosure of Fibonacci sequence for the population development of rabits is cosidered as one of the imperative applications of mathematical sciences. In the 18th century, Daniel Bernoulli applied mathematical methods to describe the effect of smallpox on the human population. Thomas Malthus in 1789 essay on the growth of the human population was based on the concept of exponential growth. Pierre François Verhulst formulated the logistic growth model in 1836. The term theoretical biology was first used as a monograph title by Johannes Reinke in 1901, and soon after by Jakob von Uexküll in 1920. One founding text is considered to be On Growth and Form (1917) by D'Arcy Thompson, and other early pioneers include Ronald Fisher, Hans Leo Przibram, Vito Volterra, Nicolas Rashevsky and Conrad Hal Waddington.

Need of Mathematical Biology

The interest in the field has grown rapidly from the last seven decades due to the following reasons:

- The rapid growth of data-rich information sets, due to the genomics revolution, which are difficult to understand without the use of analytical tools.
- Recent development of mathematical tools such as chaos theory to help understand complex, non-linear mechanisms in biology.
- An increase in computing power, which facilitates calculations and simulations not previously possible
- An increasing interest in in silico experimentation due to ethical considerations, risk, unreliability and other complications involved in human and animal research.

Mathematical Methods

A model of a biological system is converted into a system of equations, although the word 'model' is often used synonymously with the system of corresponding equations. The solution of the equations, by either analytical or numerical

means, describes how the biological system behaves either over time or at equilibrium. There are many different types of equations and the type of behavior that can occur is dependent on both the model and the equations used. The model often makes assumptions about the system. The equations may also make assumptions about the nature of what may occur.

Diffusion in Biology

The movement of many substances into and out of cells is by simple diffusion. For example, if a cell is using oxygen, and oxygen molecules are plentiful around the cell, they continuously diffuse into the cell. The cell membrane here is entirely passive; it is not pumping or pulling on the oxygen molecules. They enter the cell by their own energy as a result of their constant random movement. When they bounce into the cell, they are gobbled up in chemical reactions and usually do not have a chance to bounce out again. As long as oxygen is used on the inside and is plentiful on the outside, a concentration (activity) gradient will be maintained, and will be a net movement inward.

Diffusion processes play an important role in almost any biological phenomena. Through diffusion many metabolites are exchanged between a cell and its environment or between the blood stream and tissues.

Fick's Diffusion Laws: Fick's first law relates the diffusive flux to the concentration field, by postulating that the flux goes from regions of high concentration to regions of low concentration, with a magnitude that is proportional to the concentration gradient (spatial derivative). In one (spatial) dimension, this is

$$J = -D\frac{\partial \phi}{\partial x}$$

where

- J measures the amount of substance that will flow through a small area during a small time interval.
- D is the diffusion coefficient or diffusivity
- ϕ (for ideal mixtures) is the concentration
- x is the position vector.
- D is proportional to the squared velocity of the diffusing particles, which depends on the temperature, viscosity of the fluid and the size of the

particles. For biological molecules the diffusion coefficients normally range from 10^{-11} to 10^{-10} m^2/s.

In two or more dimensions we must use ∇, the del or gradient operator, which generalizes the first derivative, obtaining

$$J = -D\nabla\phi$$

The driving force for the one-dimensional diffusion is the quantity

$$-\partial\phi/\partial x$$

Fick's Second Law: Fick's second law predicts how diffusion causes the concentration field to change with time:

$$\frac{\partial\phi}{\partial t} = D\frac{\partial^2\phi}{\partial x^2}$$

where

ϕ - is the concentration

t - is time

D - is the diffusion coefficient

x - is the position vector.

It can be derived from Fick's First law and the mass balance:

$$\frac{\partial\phi}{\partial t} = -\frac{\partial}{\partial x}J = \frac{\partial}{\partial x}\left(D\frac{\partial}{\partial x}\phi\right)$$

Assuming the diffusion coefficient D to be a constant we can exchange the orders of the differentiating and multiplying by the constant:

$$\frac{\partial}{\partial x}\left(D\frac{\partial}{\partial x}\phi\right) = D\frac{\partial}{\partial x}\frac{\partial}{\partial x}\phi = D\frac{\partial^2\phi}{\partial x^2}$$

and, thus, receive the form of the Fick's equations as was stated above.

For the case of diffusion in two or more dimensions the Second Fick's Law is:

$$\frac{\partial\phi}{\partial t} = D\nabla^2\phi$$

If the diffusion coefficient is not a constant, but depends upon the coordinate and/or concentration, the Second Fick's Law becomes:

$$\frac{\partial \phi}{\partial t} = \nabla.(D \nabla \phi)$$

An important example is the case where φ is at a steady state, i.e. the concentration does not change by time, so that the left part of the above equation is identically zero. In one dimension with constant D, the solution for the concentration will be a linear change of concentrations along x. In two or more dimensions we obtain

$$\nabla^2 \phi = 0$$

which is Laplace's equation, the solutions to which are called harmonic functions by mathematicians.

Heat transfer

Heat transfer is the transition of thermal energy or simply heat from a hotter object to a cooler object ("object" in this sense designating a complex collection of particles which is capable of storing energy in many different ways). When an object or fluid is at a different temperature than its surroundings or another object, transfer of thermal energy, also known as heat transfer, or heat exchange, occurs in such a way that the body and the surroundings reach thermal equilibrium. Heat transfer always occurs from a higher-temperature object to a cooler temperature one as described by the second law of thermodynamics. Where there is a temperature difference between objects in proximity, heat transfer between them can never be stopped; it can only be slowed.

Conduction is the transfer of heat by direct contact of particles of matter. The transfer of energy could be primarily by elastic impact as in fluids or by free electron diffusion as predominant in metals or phonon vibration as predominant in insulators. In other words, heat is transferred by conduction when adjacent atoms vibrate against one another, or as electrons move from atom to atom. Conduction is greater in solids, where atoms are in constant contact. In liquids (except liquid metals) and gases, the molecules are usually further apart, giving a lower chance of molecules colliding and passing on thermal energy.
Heat conduction is directly analogous to diffusion of particles into a fluid, in the situation where there are no fluid currents. This type of heat diffusion differs

from mass diffusion in behaviour, only in as much as it can occur in solids, whereas mass diffusion is mostly limited to fluids.

As density decreases so does conduction. Therefore, fluids (and especially gases) are less conductive. This is due to the large distance between atoms in a gas: fewer collisions between atoms means less conduction. Conductivity of gases increases with temperature. Conductivity increases with increasing pressure from vacuum up to a critical point that the density of the gas is such that that molecules of the gas may be expected to collide with each other before they transfer heat from one surface to another. After this point in density, conductivity increases only slightly with increasing pressure and density.

Formulation of the model

The temperature u(x,t) at any point of the bar depends on the distance x of the point from one end and the time t. Also the temperature of all points of any cross section is the same. Then by fundamental principle of heat conduction, the amount of heat crossing any section of the bar per second depends on the area A of the cross section, the conductivity k of the material of the bar and temperature gradient $\partial u / \partial x$, i.e., the rate of change of temperature with respect to distance x normal to the area.

i.e., Q_1, the quantity of heat flowing into the section at distance x.

$$Q_1 = -kA\left(\frac{\partial u}{\partial x}\right)_x \ per \ \sec$$ (-ve sign is taken because as x increases, u decreases)

Q_2 = the quantity of heat flowing out of the section at a distance x + δx

$$= -kA\left(\frac{\partial u}{\partial x}\right)_{x+\delta x} \ per \ \sec$$

Hence the amount of heat retained by the slab with thickness δx is

$$Q_1 - Q_2 = kA\left[\left(\frac{\partial u}{\partial x}\right)_{x+\delta x} - \left(\frac{\partial u}{\partial x}\right)_x\right] per \ \sec$$

But the rate of increase of heat in the slab

$$= \rho c A \delta x \frac{\partial u}{\partial t}$$

where c is the specific heat and ρ is the density of the material, thus we have

$$\rho cA\delta x\frac{\partial u}{\partial t}=kA\left[\left(\frac{\partial u}{\partial x}\right)_{x+\delta x}-\left(\frac{\partial u}{\partial x}\right)_{x}\right]$$

or $$\rho c\frac{\partial u}{\partial t}=k\frac{\left[\left(\frac{\partial u}{\partial x}\right)_{x+\delta x}-\left(\frac{\partial u}{\partial x}\right)_{x}\right]}{\delta x}$$

Now taking the limit as $\delta x\to 0$, we have

$$\rho c\frac{\partial u}{\partial t}=k\frac{\partial^2 u}{\partial x^2} \quad \text{or} \quad \frac{\partial u}{\partial t}=\frac{k}{\rho c}\frac{\partial^2 u}{\partial x^2}$$

or $$\frac{\partial u}{\partial t}=s^2\frac{\partial^2 u}{\partial x^2} \quad \text{where} \quad s^2=\frac{k}{\rho c}$$

Convection: It is the transfer of heat energy through a moving fluid at various temperatures. It is mainly a combination of diffusion and the bulk motion of molecules. When the mass of the fluid is in contact with a hot surface, its molecules expand and scatter, causing the mass of fluid to become less dense. When this happens, the fluid is displaced vertically or horizontally while the cooler fluid gets denser and sinks. Thus the hotter volume transfers heat towards the cooler volumes of that fluid.
There are two types of Convective Heat Transfer:

Natural Convection: When the fluid motion is caused by the density variations that result from the temperature distributions in heat transfer and gravitational forces. These variations in density can be described by buoyancy forces (body of force proportional to density gradient), thus buoyancy is the driving force for natural convection. The change in density of the boundary layer while heating will cause the fluid to rise and be replaced by a cooler fluid.

Example: Transfer of heat from hot water to a radiator in a room.

Forced Convection: Unlike natural convection, forced convection is not related to the use of heat transfer between fluids but rather an external source such as pumps and fans. It creates an artificially induced convection current.

Internal and external flow can also classify convection. Internal flow occurs when the fluid is enclosed by a solid boundary such as a flow through a pipe. An external flow occurs when the fluid extends indefinitely without encountering a solid surface. Both these convections, either natural or forced, can be internal or external as they are independent of each other.

The formula for convection is:

$$q = hA(T_S - T_\infty)$$

where A is the surface area of heat transfer and $T_S - T_\infty$ is the difference between the final and the initial temperatures. The h is the heat transfer coefficient which depends on the physical properties of the fluid (such as temperature) and the physical situation in which convection occurs. Therefore, the heat transfer coefficient must be derived or found experimentally for every system analyzed. Formulae and correlations are available in many references to calculate heat transfer coefficients for typical configurations and fluids. For laminar flows the heat transfer coefficient is rather low compared to the turbulent flows. This is due to the turbulent flows having a thinner stagnant fluid film layer on the heat transfer surface.

Radiation: It is the transfer of heat energy through empty space. All objects with a temperature above absolute zero radiate energy at a rate equal to their emissivity multiplied by the rate at which energy would radiate from them if they were a black body. No medium is necessary for radiation to occur; radiation works even in and through a perfect vacuum. The energy from the Sun travels through the vacuum of space before warming the earth.

Both reflectivity and emissivity of all bodies is wavelength dependent. The temperature determines the wavelength distribution of the electromagnetic radiation as limited in intensity by Planck's law of black-body radiation. For any body the reflectivity depends on the wavelength distribution of incoming electromagnetic radiation and therefore the temperature of the source of the radiation. The emissivity depends on the wave length distribution and therefore the temperature of the body itself. For example, fresh snow, which is highly

reflective to visible light, (reflectivity about 0.90) appears white due to reflecting sunlight with a peak energy wavelength of about 0.5 micrometers. Its emissivity, however, at a temperature of about -5C, peak energy wavelength of about 12 micrometers, is 0.99.

Gases absorb and emit energy in characteristic wavelength patterns that are different for each gas.

Visible light is simply another form of electromagnetic radiation with a shorter wavelength (and therefore a higher frequency) than infrared radiation. The difference between visible light and the radiation from objects at conventional temperatures is a factor of about 20 in frequency and wavelength; the two kinds of emission are simply different "colors" of electromagnetic radiation.

Fick's Perfusion Principle

If Q is the quantity of heat in a tissue element at time t, then rate of change $\partial Q / \partial t$ is sum of rates of change due to diffusion (D), perfusion (P) and metabolic heat generation (M). The heat supplied or heat removal inside the body through perfusion can be sought by using the Fick's principle, according to which rate of inflow of any substance is the product of concentration of substance in the arterial blood and the rate of inflow of arterial blood. It can be put in mathematical terms as:

$$\left.\frac{\partial Q}{\partial t}\right|_P = F_A C_A - F_V C_V$$

Here Q is the quantity of the substance perfused, F_A and F_V are rates of blood flow in arterial and venous blood vessels and C_A and C_V are concentrations of substances in it. For small volume δV, the above equation can be put in the following form.

$$\left.\frac{\partial(\delta Q)}{\partial t}\right|_P = \delta F_A C_A - \delta F_V C_V$$

Here δQ is the amount of substance in the volume element δV due to perfusion, δF_A is the sum total of rate of blood flow in all the arterioles intersecting the surface of δV and C_A is the arterial concentration of substance associated with δV, measurable as the sum of the rates of inflow of substance in the individual

arterioles intersecting the surface of δV divided by δF_A. Similar definitions apply for δF_V and C_V. Dividing both sides of the equation by δV, we get

$$\left.\frac{\partial C}{\partial t}\right|_P = \phi_A C_A - \phi_V C_V$$

where

$\frac{\delta Q}{\delta V} = C$, tissue concentration of substance, a function of position and time

$\frac{\Delta F_A}{\Delta V} = \phi_A$, arteriole blood perfusion rate

$\frac{\Delta F_V}{\Delta V} = \phi_V$, venous blood perfusion rate

In the case of heat flow the concentration in thermal energy per unit volume equals product of density, specific heat and temperature. Hence, for heat flow the equation assumes the following form

$$\rho c \left.\frac{\partial T}{\partial t}\right|_P = \rho_b c_b (\phi_a T_A - \phi_V T_V)$$

where T_A, T_V are arterial blood and venous blood temperatures respectively, ρ_b, c_b denotes the density and specific heat of the blood.

Transcapillary Diffusion of Oxygen

In the circulatory system, arteries and veins are responsible for the transport of blood to and from the heart. An artery carries oxygen-rich blood until it encounters a capillary bed, where the oxygen is transported to the surrounding tissue via diffusion, at which point the deoxygenated blood goes back to the heart through a vein.

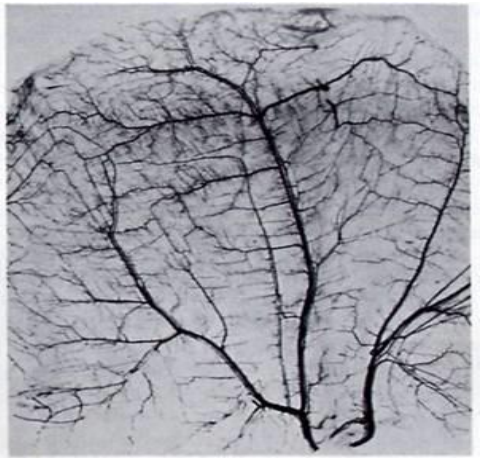


Figure: Pattern in Capillary tissue region

Here, we are interested to focuses on the diffusion that occurs in the capillary. Within the capillary, gaseous oxygen is released after being captured

by hemoglobin in the lungs, and the oxygen diffuses out through the capillary wall and into the surrounding tissue where it is used up in energy-producing reactions. This is by no means a comprehensive analysis of this subject and multiple assumptions and inferences shall be made to facilitate the development of a diffusion equation, but the general concepts are accurate and much of the work is based on information that has been verified over the years through experimentation.

The above figure shows the vascular pattern typical of skeletal muscle in mammals. While not all tissue regions exhibit the geometrical regularity of skeletal muscle, it is still tempting to introduce the concept of a repetitive unit structure as a representation of the capillary-tissue region.

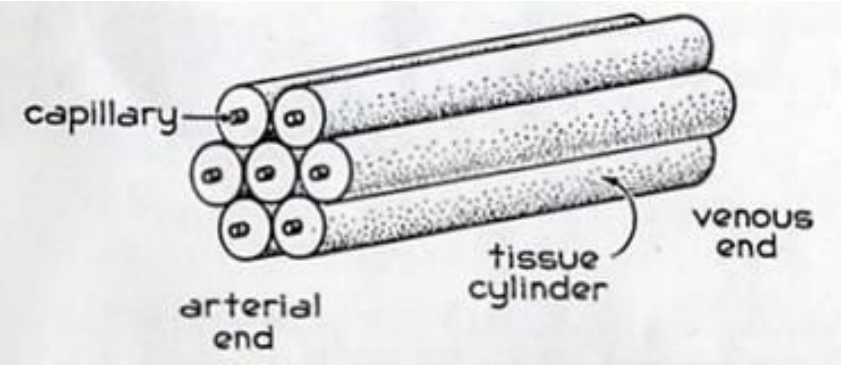
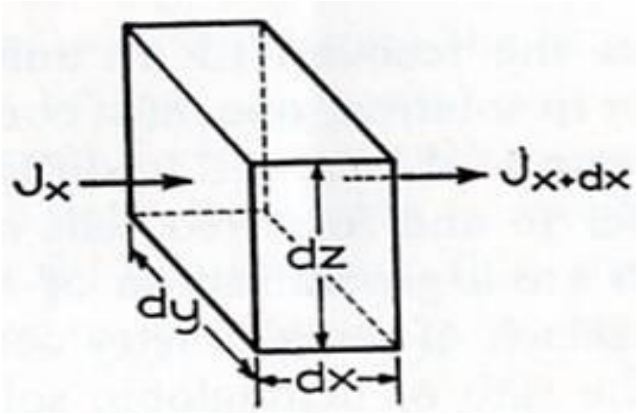


Figure: View of Krogh Cylinders

The Krogh Cylinder

The Krogh cylinder, a concept introduced by Nobel Laureate August Krogh is shown in the Figure. This unit structure implies that each section of capillary is responsible for the supply to a corresponding cylindrical section of surrounding tissue.



In order to yield a model amenable to solution, a few assumptions need to be made

- The Krogh cylinder is assumed to be an appropriate model for the geometry, even though some regions (such as in the neighborhood of branches) do not enjoy such geometrical simplicity,

- The tissue surrounding the capillary is actually a heterogeneous composite of materials with metabolic chemical reactions taking place. In order to idealize this situation, we will assume that the chemical reactions are continuously, evenly distributed,
- Symmetry about the axis of capillary-tissue cylinder will be assumed. This smoothing aspect of the model suggest that there is a cylindrical surface about a capillary across which transport from surrounding capillaries just balances transport toward these capillaries. Hence, net transport rates vanish along this surface, which can be called a no-flux cylinder.

There are two transport regions of concern here. Within the capillary, material is transported through the blood by both convection and diffusion. If $\mathbf{r_c}$ is the radius of the capillary, then a transport equation must be written for the material of interest over the cylindrical region $\mathbf{0<r<r_c}$. Outside of the capillary, in the tissue, another transport equation must be written over the region $\mathbf{rc <r<r_t}$, where $\mathbf{r_t}$ is the radius of the no-flux cylinder referred to previously.

Tissue Region Diffusion Equation

Suppose there is a flux J of some chemical species caused by diffusion in the x-direction. The rate at which this species crosses the boundary at x is just the product of flux times area perpendicular to the diffusion, $\mathbf{(Jdy/dx)_x}$ and it follows that the flux out of the region is simply $\mathbf{(Jdy/dx)_{x+dx}}$**.** Suppose that a chemical reaction is occurring within the volume at a rate per unit volume denoted by **r**. The amount of a particular species is simply the concentration times the volume, or **c*dV**.

Using the above information, the diffusion equation is obtained from the statement that any change in amount, with respect to time, must be due to the net diffusion rate across the boundaries plus the rate of reaction within the boundaries.

$$\frac{\partial}{\partial t}(c\ dV) = [(J\ dy\ /dx)_x - \left(J\frac{dy}{dx}\right)_{x+dx}] + r dV$$

Dividing out the volume from each side gives

$$\frac{\partial c}{\partial t} = \frac{J_x - J_{x+dx}}{dx} + r$$

In the limit of a vanishing dx, the first term on the right-hand side approaches

the derivative of J with respect to x. Thus the balance equation of the species considered takes the form

$$\frac{\partial c}{\partial t} = -\left(\frac{\partial J}{\partial x}\right) + r$$

and is valid at every point within the volume.

The flux J must be related in some way to the concentration and the usual assumption for a diffusive flux is Fick's Law:

$$J = -D\left(\frac{\partial c}{\partial x}\right)$$

Thus, the diffusivity is defined, and if Fick's law is introduced to the balance equation and the diffusivity is assumed to be constant (since it usually depends on the concentration, which is constant for the present situation), the result is

$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial x^2}\right) + r$$

This is the diffusion equation developed for transport within a blood vessel. However, to deal with the Krogh cylinder by changing this equation, consider the cylindrical coordinates

$$\frac{\partial c}{\partial t} = \left[D_t\left(\frac{\partial^2 c}{\partial r^2}\right) + \frac{\frac{\partial c}{\partial r}}{r} + \left(\frac{\partial^2 c}{\partial z^2}\right)\right] - g(c)$$

To be specific, **c** is the oxygen concentration, $\mathbf{D_t}$ is the diffusivity of oxygen through tissue (independent of concentration), and **g(c)** is the rate of consumption of oxygen per unit volume of tissue. It's often observed that oxygen consumption follows Michaelis-Menten kinetics, which may be expressed as

$$g(c) = \frac{Ac}{B+c}$$

If the penetrant is passive and does not react in the tissue, then A = 0.

Capillary Region Diffusion Equation

In the capillary region, whole blood flows with a velocity profile **u(r)**. It's assumed here that the red cells travel with the same velocity as plasma at the same radial position. A "smoothing" assumption for diffusion through blood must be made, since blood is a heterogeneous fluid. The red cells occupy roughly half the volume of the blood. Within the red cells, hemoglobin reacts

with gases such as oxygen and carbon monoxide, and diffusion occurs. However, the diffusivity in the concentrated hemoglobin solution within the red blood cells is different than the diffusivity within the plasma surrounding those cells.

A reasonable way to approximate the overall diffusivity is by using an analogy to the calculation of overall transfer coefficients for heat and mass transfer through a series of resistances (similar to the calculation of total resistance within an electric circuit), with the result

$$\frac{1}{D_{blood}} = \frac{1}{2D_{Hemoglobin}} + \frac{1}{2D_{plasma}}$$

Using this, the equation found for diffusion in the tissue region becomes

$$\left(\frac{\partial \mathrm{c}}{\partial \mathrm{t}}\right) + u\left(\frac{\partial \mathrm{c}}{\partial \mathrm{z}}\right) = D_{blood}\left[\left(\frac{\partial^2 c}{\partial \mathrm{r}^2}\right) + \frac{\frac{\partial \mathrm{c}}{\partial \mathrm{r}}}{r} + \left(\frac{\partial^2 c}{\partial \mathrm{z}^2}\right)\right] + d(c)$$

The term d(c) represents the generation of oxygen within the blood due to the dissociation of oxyhemoglobin (oxygen bound to hemoglobin).

Note that the left-hand side is significantly different from the equation developed for the tissue region, namely that it has an extra term. This second term arises from the fact that, in a flowing system, the species flux along the axis would be

$$J_z = -D_{blood}\left(\frac{\partial \mathrm{c}}{\partial \mathrm{z}}\right) + uc$$

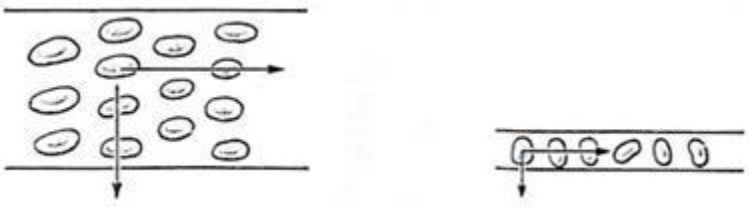
Assuming a constant cross-sectional area for the capillary, u is independent of z. The term

$$\frac{\partial \mathrm{J_z}}{\partial \mathrm{z}}$$

then gives, upon rearrangement, the convective term that appears in the diffusion equation.

However, this is only valid for a large capillary. In a small capillary, it is necessary to introduce two diffusivities; Dr = DHb for the radial diffusion process, and Dz = Dblood for the axial process. In this case, the diffusion equation would take the form

$$\left(\frac{\partial \mathrm{c}}{\partial \mathrm{t}}\right) + u\left(\frac{\partial \mathrm{c}}{\partial \mathrm{z}}\right) = D_{Hb}\left[\left(\frac{\partial^2 c}{\partial \mathrm{r}^2}\right) + \frac{\frac{\partial \mathrm{c}}{\partial \mathrm{r}}}{r}\right] + D_{blood}\left(\frac{\partial^2 c}{\partial \mathrm{z}^2}\right) + d(c)$$



Notice in the above figure how in the large capillary, the radial direction vector contains both red blood cells and plasma, while in the small capillary, only a red blood cell (and a very small amount of plasma) is included in the diffusion route. This is why the two equations are necessary, but for simplicity sake here we will only deal with the large capillary from here on out.

The equilibrium for the reaction depends upon the oxygen content of the plasma external to the red cell. Since the kinetics of this reaction are fast in comparison to diffusion rates, it is safe to assume that the fraction of oxygen bound to hemoglobin is always in equilibrium with respect to the oxygen content of the plasma. The fractional saturation, s, may be described as

$$s = \frac{Kc^n}{1 + Kc^n}$$

where K is a constant dependent on the ionic strength and hemoglobin concentration of the plasma, and n fluctuates between 2.5 and 2.6. If we take N to be the oxygen-binding capacity of blood (moles hemoglobin/volume blood), then the generation of oxygen by dissociation is given by

$$d(c) = -N\left(\frac{\partial \mathrm{s}}{\partial \mathrm{t}}\right) + u\left(\frac{\partial \mathrm{s}}{\partial \mathrm{z}}\right)$$

Hence, the diffusion equation (for a large capillary) becomes

$$\left(\frac{\partial}{\partial \mathrm{t}}(c + N_s)\right) + u\left(\frac{\partial}{\partial \mathrm{z}}(c + N_s)\right) = D_{blood}\left[\left(\frac{\partial^2 c}{\partial \mathrm{r}^2}\right) + \frac{\frac{\partial \mathrm{c}}{\partial \mathrm{r}}}{r} + \left(\frac{\partial^2 c}{\partial \mathrm{z}^2}\right)\right]$$

Boundary Conditions

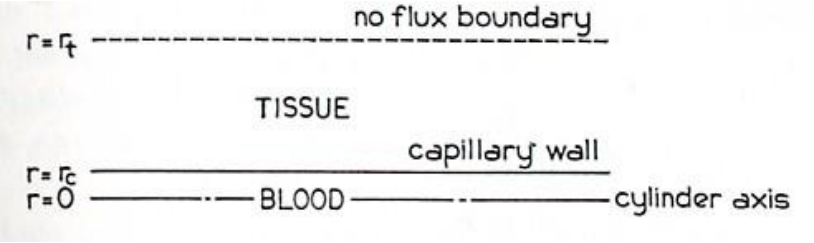
We now have our diffusion equations for both the tissue region and the capillary region

$$Tissue\_Region = \frac{\partial}{\partial \mathrm{t}}c = D_t\left[\left(\frac{\partial^2 c}{\partial \mathrm{r}^2}\right) + \frac{\frac{\partial \mathrm{c}}{\partial \mathrm{r}}}{r} + \left(\frac{\partial^2 c}{\partial \mathrm{z}^2}\right)\right] - g(c)$$

$$Capillary\_Region = \left(\frac{\partial}{\partial \mathrm{t}}(c + N_s)\right) + u\left(\frac{\partial}{\partial \mathrm{z}}(c + N_s)\right)$$

$$= D_{blood}\left[\left(\frac{\partial^2 c}{\partial \mathrm{r}^2}\right) + \frac{\frac{\partial \mathrm{c}}{\partial \mathrm{r}}}{r} + \left(\frac{\partial^2 c}{\partial \mathrm{z}^2}\right)\right]$$

The capillary-tissue region is shown schematically below.



Symmetry about the axis of the cylinder requires that the radial concentration gradient vanish along **r = 0**, and the no-flux condition we assumed at **r = rt** both have the same condition that

$$\frac{\partial c}{\partial r} = 0$$

Additionally, at the blood-tissue interface, the flux of the penetrant must be continuous, such that

$$D_{blood}\left(\frac{\partial c}{\partial r}\right) = D_t\left(\frac{\partial c}{\partial r}\right)$$

It is noteworthy that although the notation **c** is being used for both blood and tissue concentration, the gradients are not equal, one being calculated from **c** on the blood side and the other from **c** on the tissue side of **r = rc**. These conditions are based simply on notion of symmetry and continuity. In order to develop a steady-state model, additional conditions are required to yield a solvable set of equations, which are not easily derived and depend on additional assumptions, which serve to specify the type of model under consideration.

Overview of transcapillary exchange

To reiterate, what has been presented here is a very rough approximation of what actually happens in a capillary. There are many factors that affect the parameters used to develop these equations that are difficult to quantify. For example:

- Temperature has a strong affect on the size of a capillary (heat makes it wider, cold makes it thinner)
- Various hormones secreted by the brain can alter the size and function of the capillaries
- Diseases and toxins can have an effect on the diffusion of blood through the capillary walls

Human Blood Flow: Research in blood flow has a direct impact on our improved understanding and management of human health. Blood flow is a related measure of oxygen and nutrient concentrations in blood. Such concentration measures are an indication of general cell health.

This study begins by outlining the physiology involved in blood flow. This includes a discussion of the cardiovascular system, and of the regulation of cardiac output, blood flow, and blood pressure. It then moves on to present some of the available techniques for measuring blood pressure and flow. It will be shown that each technique provides slightly different, but still relevant, information for the targeted measurement. Lastly, a mathematical model for blood flow in arteries is highlighted, as well as a discussion on the assumptions and applications of such a model.

As will be seen through the course of this essay, the problem of mathematically modeling a human function is very involved. First, a solid foundation of the physiology to be modeled must be formed. Then, the technology required to detect the human function, in this case blood flow, must be thoroughly understood. It is only when this has been completed, that the researcher has enough insight to accurately derive a mathematical model. However, the process is not complete, for the model will still need to be verified. For this, the technology may again be used to generate data for validation of the model.

Once a satisfactory model has been generated, the benefits are unlimited. Modeling provides improved understanding of the human function it attempts to simulate. It also allows testing to be performed prior to human involvement, and leads to future developments for the management of human health.

Physiology: The first requirement to studying blood flow is to gain a general understanding of the physiology involved. A brief outline of the circulatory system and cardiac output is provided below to give the reader a better understanding of how blood is circulated and regulated in the human body.

The Circulatory System: The circulatory system is divided into two main components: the cardiovascular system, and the lymphatic system. The focus here will be on the cardiovascular system, which consists of blood, blood vessels, and the heart. Blood acts as a transport mechanism to the cells for nutrients and wastes. Blood vessels provide a tubular network to channel the

blood to every possible region of the body. And the heart creates the pressure required to push blood through the vessels.

The Composition of Blood: Blood is composed of fluid plasma, formed elements, and other elements either being carried to or away from cells. Blood plasma is a liquid that serves as the extracellular environment for all the cells in the body. Total volume and concentration of plasma is important in the regulation of blood pressure. The sodium ion is the major solute in plasma, and its concentration is what determines the amount of plasma water, and thus blood volume.

The formed elements of the blood are erythrocytes, leukocytes, and platelets. Erythrocytes, or red blood cells (RBCs), are flattened discs with a depressed centre, about 2.5 μm thick and 7.5 μm wide. The depressed centre provides increased surface area for the diffusion of gases. RBCs contain no nuclei or mitochondria, but they do contain hemoglobin. It is the hemoglobin, which binds to oxygen and allows the RBCs to transport oxygen to cells.

Leukocytes, or white blood cells (WBCs), are larger than RBCs. They have a nuclei and mitochondria which enable them to move around. WBCs are capable of squeezing through pores in capillary walls in order to reach sites of infection. This aids WBCs in their participation in the body's immune response.

Platelets are the smallest formed element and are actually fragments of large bone marrow cells. They contain no nuclei but are still capable of movement and function in blood clotting.

All formed elements of blood are produced in the myeloid tissue, or red bone marrow, in a process known as hemopoiesis.

Total blood volume in an average adult is about 5 litres, or 8% of the total body weight.

Formed elements constitute about 45% of the total blood volume and blood plasma makes up the remaining 55%. The percentage of formed elements closely approximates the percentage of RBCs per given volume of blood. This is an important measure in health care as it is an indicator of the oxygen-carrying capacity of blood.

Blood Vessels: The network of blood vessels allows the transport of blood to all living cells in the body. Their structure enables the exchange of blood plasma and dissolved molecules between the blood and surrounding tissues. Blood

travels away from the heart passing through a series of vessels progressively smaller in diameter: arteries to arterioles to capillaries. Blood returns to the heart through a series of vessels progressively larger in diameter: capillaries to venules to veins.

The walls of arteries and veins consist of three layers. The outermost layer is the tunica external; it is composed of loose connective tissue. The middle layer is the tunica media; it is composed of smooth muscle. The innermost layer is the tunica intima; it is composed of the endothelium and connective tissue. The endothelium lines all inner walls of vessels, and capillaries consist only of the endothelium.

Arteries contain more muscle than comparably sized veins. Large arteries stretch when the pressure of the blood rises during systole and recoil during diastole. The elastic recoil of the walls helps to produce a smoother flow of blood in the smaller arteries and arterioles.

However, the result is a cardiac cycle-dependent artery diameter. Smaller arteries and arterioles are less elastic than larger arteries, and contain a proportionally thicker layer of smooth muscle. Thus they maintain a relatively constant diameter.

Capillaries are the simplest structured vessel. They are composed of a single cell layer of endothelium and are about 8 mm in length. They permeate the entire body in a fine mesh\ to provide the surface area for blood and interstitial fluid transfer. Capillaries only contain about 250 ml of blood at any given time. The amount of blood in a capillary bed is regulated by the pre-capillary sphincter muscles and by the resistance to blood flow provided by the small arteries and arterioles.

Blood is transported back to the heart by venules, which empty into progressively larger and larger veins. The pressure in the veins is around 2 mmHg which is insufficient to return blood to the heart. However, veins pass between skeletal muscles, which contract during motion and naturally provide a massaging action on the veins. The squeezing effect increases venous pressure and helps to push the blood back up to the heart. One directional flow is maintained by venous valves, which close in response to increased pressure.

The Structure and Function of the Heart:

The heart acts as a double pump to keep blood circulating through blood vessels. The heart consists of upper right and left atria and lower right and left ventricles. It is surrounded by the pericardium whose inner lining produces fluid to lubricate the hearts motion.

The heart is composed of three layers: the epicardium forms an outer protective sheath; the myocardium forms the thick middle layer composed mainly of cardiac muscle tissue; and the endocardium forms the inner layer which is continuous with the endothelium of the blood vessels.

Blood enters the heart via the right atrium. The right atrium fills the right ventricle which pumps blood to the lungs and back to the left atrium. Blood is then passed to the left ventricle which pumps to the aorta and systemic arteries. Atrioventricular and semilunar valves open based on pressure gradients and allow the heart to pump blood in a unidirectional flow.

The cardiac cycle refers to the pattern of systole and diastole, and is produced by a wave of depolarization through the nodal tissues of the heart. The sinoatrial (SA) node initiates the cardiac cycle by depolarizing from about - 60 mV to - 40 mV. This depolarization moves across the atria, causing them to contract, to the atrioventricular (AV) node and into the Bundle of His. The wave then travels to the Purkinje fibres, which initiate the contraction of the ventricles.

The heart is a two-step pump first the atria, then the ventricles contract. During diastole, while the ventricles are relaxed, they are filled with blood by the atria. The ventricles are about 80% full before the atria actually contract. The amount of blood ejected by the ventricles is about two-thirds that of the volume prior to contraction, and is referred to as the stroke volume.

Cardiac Output and Cardiac Rate:

The amount of blood ejected from the ventricles, or cardiac output, is a measure of the pumping ability of the heart.

It is determined by the following cardiac rate and stroke volume

Cardiac output	=	cardiac rate	x	stroke volume
(ml/min)		(beats/min)		(ml/beat)

Cardiac output is related to total blood volume. An increase in cardiac output requires an increase in the rate blood flows through the vessels. This increased flow is caused by the regulation of cardiac rate and stroke volume.
Cardiac rate is based on the natural rhythm of the SA node. Sympathetic and parasympathetic stimulation continuously modify the cardiac rate by increasing or decreasing the rate of depolarization. The cardiac control centres of the medulla oblongata coordinate this autonomic activity. Other effects of this autonomic nerve activity include changes in the AV node conduction rate, and changes in the strength of contraction for both the atrial and ventricular muscles.

Regulation of Stroke Volume: Recall that stroke volume is the amount of blood pumped per beat by each ventricle. It is regulated by three parameters:

- end-diastolic volume (EDV),
- contractility, and
- total peripheral resistance.

Stroke volume is directly proportional to EDV since an increase in the volume of blood in the ventricles prior to contraction, results in an increase in the ejection volume. Contractility is also directly proportional to stroke volume: a more forceful ejection results in a more complete emptying of the ventricles.
Total peripheral resistance, however, is inversely proportional to stroke volume. A higher resistance, or increase in arterial pressure, results in less blood entering the aorta. This is due to the fact that blood is only ejected from the ventricles while a pressure gradient exists.
Higher arterial pressure means that the time it takes for the pressures to equalize is shorter, and thus less blood will transfer from ventricle to aorta.

Regulation of Blood Flow:

Blood flow to an organ is controlled by the constriction and dilation of vessel walls.

The changes in vessel diameter are regulated by sympathetic interaction and by local conditions within the blood vessel and organ.

Sympathetic nerve stimulation causes vasoconstriction in the viscera and skin, and vasodilation in the skeletal muscles. Myogenic regulation, in the cerebral

arteries, causes the dilation of vessels in response to a decrease in blood pressure, and vice versa. Metabolic regulation, in the skeletal and cardiac muscle vessels, promotes vasodilation based on local conditions such as oxygen concentrations, carbon dioxide concentrations, tissue pH, and the release of adenosine. The increase in blood flow in skeletal muscles and the heart as a result of increased metabolism is called active hyperemia.

Regulation of Blood Pressure:

Blood pressure is regulated by changes in the cardiac rate, blood volume, and total peripheral resistance - all of which are controlled by the sympatho-adrenal system. An increase in any of these parameters results in an increase in blood pressure. This effect, however, can be offset by a decrease in the other parameters. Baroreceptors in the aortic arch and carotid sinuses detect an increase in pressure when the arterial walls of these regions are stretched. The sensory nerve activity of the baroreceptors is a negative feedback control for the medulla oblongata. Vasomotor control centres regulate the vasoconstriction, vasodilation, and total peripheral resistance in the vessels. Cardiac control centres regulate the cardiac rate.

Heart Rate Measurement:

One of the best indicators of a patient's health is the concentration of oxygen and other nutrients in cells. This measurement is fairly difficult to achieve. A secondary measurement that correlates to the concentration of nutrients is that of blood flow and changes in blood volume. In the doctor's office, however, such measurements are not available, and a tertiary measurement of blood pressure must suffice. Blood pressure, however, is only a satisfactory indicator of blood flow. Finally, in the case where blood pressure is immeasurable, the ECG may be used, which correlates with blood pressure.

ECG and blood pressure are the easiest measurements to attain. However, measurement of blood flow, most closely relates to the concentration of oxygen in the patient. First, a discussion of ECG and blood pressure measurement is provided for completeness. Then, various techniques for measuring blood flow are presented.

The simplest measurements of heart activity are cardiac rate and heart sounds. The pulse can be measured using a finger and watch. With a stethoscope, a physician can hear the sounds associated with the opening and closing of the semilunar and atrioventricular valves. Such measurements provide information on the health of the heart, but provide little information on general body cell health, i.e. the concentration of nutrients in blood. Electrocardiograms and blood pressure measurements can be used to indirectly gain such knowledge.

Electrocardiogram: ECG is the measurement of electrical signals associated with the motion of the heart. As the heart contracts, an electrical signal propagates from the SA node to the Purkinje fibres. Surface electrodes can measure the potentials as seen at various positions on the skin. Voltages measured between electrodes create a signal indicative of the heart's electrical activity. By changing the position of the electrodes, different signals can be acquired.

A typical electrocardiograph signal consists of a P wave, a QRS wave, and a T wave.

The P wave is produced by the depolarization of the atria. The QRS wave is produced primarily be the depolarization of the ventricles. The time interval from P-R is caused by the conduction delay at the AV node. The re-polarization of the ventricles produces the T wave. The time interval S-T represents the period between ventricular depolarization and initiation of re-polarization. Artefacts in the normal ECG are indicators of possible arrhythmias, or disturbances in the hearts normal rhythm.

The heart can be viewed as an electrical generator, and the electrical activity of the heart can be represented by a net current dipole located at the electrical centre of the heart. The thoracic tissues can be considered the resistive load for this electrical generator model. The dipole field can be represented by a vector whose magnitude and orientation change as a function of time.

If two electrodes are placed at different locations on the body, a potential difference will be recorded. This potential difference depends on where the electrodes are placed. Therefore, standard positions for electrodes have been designed for clinical ECG.

To predict the voltage across a set of electrodes at an instant of time, lead vectors are defined for every pair of electrodes. The lead vector represents the

direction that a constant magnitude cardiac vector must have in order to generate a maximal voltage across the pair of electrodes. In clinical electrocardiography, multiple leads must be recorded to adequately describe the electrical activity of the heart. Several leads are taken in the frontal plane and transverse plane of the body.

The shape of an ECG signal can be used to determine the forcefulness of ventricular ejection. During the QRS wave, the ventricle contracts to eject blood into the aorta. A malformed QRS wave indicates abnormalities in the ventricles. For instance, unusually large ventricles will produce a stronger ejection force, and a resulting enlarged R spike. This information can be used to infer blood pressure, and thus provides some information on the flow of blood.

Blood Pressure: Blood pressure levels in various chambers of the heart and arteries indicate the health of the cardiovascular system. There are a number of direct (invasive) and indirect (non-invasive) techniques for measuring blood pressure.

Direct measurements involve some sort of invasive element. Blood pressure sensor systems are divided into two categories -dependent on where the sensor is placed. Extra-vascular sensors directly measure pressure by coupling the vascular pressure to an external sensor via a liquid-filled catheter. Intravascular sensors incorporate the sensor into the tip of a catheter that is inserted into the vascular system.

The standard method used in doctor's offices is an auscultatory method, which is based on arterial sounds produced by turbulent flow. When the pressure applied by the cuff is in between the diastolic and systolic pressures, sounds are produced by the turbulent flow through the partially constricted, vibrating artery. These sounds can be heard using a stethoscope. Other automatic techniques have been developed using a sphygmomanometer and a sensitive detector to measure the cuff pressure.

Blood pressure waveforms can be analyzed using Fourier analysis to quantify the dynamics of blood pressure and flow. These analyses have lead to improved understanding of the cardiovascular system. However, blood pressure remains a tertiary measure of a patient's general cell health.

Measurement of Blood Flow and Volume of Blood:

The best measurements available to indicate oxygen and nutrient concentrations in blood, i.e. cell health, are the flow of blood and the changes in blood volume. Commonly used flowmeters are impractical in medical situations, as they would require cutting the vessel to insert a turbine. Because of this, specialized techniques have been developed to measure blood flow.

Indicator-dilution Methods:

Measuring Average Flow: Indicator-dilution methods measure the flow averaged over a number of heartbeats. The general idea is to insert an indicator into the blood flow and measure the rate at which it is carried downstream. Common indicators used are colored dyes and cold saline. Two types of indicator-dilution methods will be discussed: continuous infusion and rapid injection.

Continuous infusion techniques involve the continuous addition of an indicator into a flowing stream. For a fixed change in indicator concentration, ΔC, and a known rate of injection, $\frac{dm}{dt}$, the average flow, F, can be calculated:

$$F = \frac{dV}{dt} = \frac{dm/dt}{\Delta C}$$

In clinical use, continuous infusion has largely been replaced by rapid injection. Rapid injection is more convenient and involves a bolus of indicator being rapidly injected into the vessel. The variation in downstream concentration of indicator, C(t), is measured until the bolus has passed. The average flow is then obtained by:

$$F = \frac{m}{\int_0^{t_1} C(t)dt}$$

where m, is the quantity of indicator injected, and t_1 is the time after which the indicator has completely passed.

The most common method for measuring cardiac output is thermo-dilution, where a bolus of cold saline is rapidly injected into the right atrium. The blood and saline mix in the right ventricle and the temperature difference is recorded in the pulmonary artery. The four-lumen catheter used for this procedure can stay

inside the patient for up to 24 hours; this allows a series of cardiac output measurements to be recorded.

Electromagnetic flow meters:

Measuring Instantaneous Pulsatile Flow: An improvement over the indicator-dilution method is the electromagnetic flowmeter. Electromagnetic flowmeters measure the instantaneous pulsatile flow of blood, rather than the average flow.
The flow meter depends on the fact that blood flowing with velocity, u, through a magnetic field, B, induces an electromagnetic flux, e, given by:

$$e = \int_0^{L_1} u \times B.dL$$

where L is the length between the recording electrodes. For a uniform magnetic field and uniform velocity profile, the induced emf is $e = BLu$.
In the case of parabolic flow profiles, resulting from laminar flow in tubes, the flow velocity u is replaced by the cross-sectional average flow velocity $\bar{u}$. The volumetric flow can thus be attained by $A\bar{u}$, where A is the cross-sectional area of the vessel.
There are both DC and AC flow meters, with AC flow meters overcoming the unsatisfactory performance of their DC counterparts. AC flow meters are typically driven by sine or square waves, and use a toroidal-shaped cuff probe to create the magnetic field and properly position the electrodes. The probe must fit snugly around the vessel during diastole in order to record a valid signal. This means that the vessel is slightly constricted during systole, and that this method is not applicable to veins.

Ultrasonic flow meters:

Measuring Flow Profiles: Another method for measuring the instantaneous pulsatile flow in vessels is the ultrasonic flow meter. The ultrasonic flowmeter can measure flows non-invasively. Advanced types of these devices can also measure flow profiles. These advantages make ultrasonic flow meters a hot topic in current research.
Continuous-wave flow meters provide information on blood flow. Pulsed Doppler flow meters, however, can detect flow profiles. The transmitter exerts a pulse, which travels in a single packet to the source. The wave reflects off of

RBCs in the blood stream and is received by the transmitter with a time delay proportional to the distance travelled. Analyzing the Doppler shift at various delays creates the velocity profile across the vessel.
Pulses are usually transmitted at 8 HMz and last about 1μs. The intensity of the packet is convolved with the velocity profile to create the reflected signal. Thus, the received signal must be mathematically deconvolved before it adequately represents the velocity profile across the vessel.
The rate at which pulses are sent must satisfy the sampling theorem as well as a restriction imposed by the returning signal. The received signal should be analyzed prior to a new pulse being transmitted. The pulse train is multiplied by the carrier, which produces sidebands in the frequency domain. This spectrum can excite the transducer and cause spectral spreading or smear in the received signal.

Plethysmography:

Measuring changes in Blood Volume: Plethysmographs measure changes in blood volume. In an extremity, chamber plethysmographs are used to measure changes in blood volume, and thus to compute arterial flow, $F = dV/dt$ F. Electric impedance plethysmography also measures changes in volume, but uses impedance characteristics of the tissues, rather than conservation of volume.
In chamber plethysmography, the chamber is first calibrated by injecting a known volume of fluid. A venous-occlusion cuff is then applied at 50 mm Hg to inhibit the venous blood from leaving the limb. The cuff does not interfere with arterial flow, so the rate of change of volume in the limb is equal to the arterial inflow. If the chamber does not enclose the limb completely, an arterial-occlusion cuff must be applied, at 120 mm Hg, downstream from the chamber. This ensures volumetric changes are due only to arterial inflow. Once pressure in the venous-occlusion cuff exceeds 50 mm Hg, blood will begin to exit the limb and total limb volume will plateau.
Electric-impedance plethysmography uses the simple principle that electrodes placed on a tissue will measure the resulting impedance of that tissue. As the volume of the tissue changes, or the resistivity changes, the impedance changes. A change in blood volume, ΔV, causes an increase in cross-sectional area, ΔA,

and thus an increase in impedance, ΔZ. The relation for finding the change in blood volume is given by:

$$\Delta V = \frac{-\rho_b L^2 \Delta Z}{Z^2}$$

Here, ρ_b is the resistivity of blood and L is the distance between electrodes. Electric-impedance plethysmography is non-invasive and a relatively simple technique for measuring changes in blood volume.

Electric-impedance plethysmography is more versatile than chamber plethysmography. The recording electrodes can be placed virtually anywhere a measurement is desired. For example, electrodes placed on limbs can detect vessel obstructions, and when placed around the neck and waist can detect beat-by-beat changes in cardiac output. However, the accuracy of plethysmography is undetermined.

The above discussion mentions only a few of the techniques available to measure blood flow and changes in blood volume. The techniques differ in that they measure different aspects of the flow: the average flow, the instantaneous pulsatile flow, the flow profile, and changes in blood volume. These measurements are all-unique and provide valuable information to physicians and medical researchers.

Modeling: Studies of blood flow in vessels provide a greater understanding of human health and management. The area of blood flow in arteries is a mature subject and the highlights of which are presented below. For smaller vessels, such as arterioles and capillaries, many of the known principles break down as the scale of the problem is of the same order of magnitude as an RBC.

Dynamics of Blood Flow: In order to attempt to model blood flow, the problem must be restricted to a specific area. As the order of magnitude of the problem increases or decreases, many assumptions made initially fail to hold. Therefore, only blood flow in arteries will be considered henceforth.

Flow in arteries can be considered a continuum. That is, the elements of blood seem to be continuous with each other, with no empty spaces in between. In fact, even though blood is composed of many different elements, the continuum hypothesis implies that every point in the fluid represents a fluid element, and that the properties at that point, represent the properties of that fluid element.

Is Blood a Newtonian Fluid? A Newtonian fluid is one, which exhibits the following property. The shear stress, or resisting force τ , and velocity gradient, or rate of deformation du/dy , are linearly related by

$$\tau = \mu \frac{du}{dy}$$

where μ is the coefficient of viscosity.

The question of whether blood is a Newtonian fluid is still standing. The composition of blood would seem to indicate that it is indeed not a Newtonian fluid. In fact, studies seem to indicate that plasma by itself is a non-Newtonian fluid. However, under certain restrictions, such as those imposed on this problem, it is sufficient to assume blood acts like a Newtonian fluid. Indeed, all modelling of pulsatile blood flow accepts this assumption. It is only when the scale of the problem is so reduced, as in capillary modelling, that the Newtonian fluid model breaks down.

Dynamics of Pulsatile Flow:

Pulsatile flow in a rigid tube causes the fluid to oscillate in bulk. That is, with each pulse applied to the flow, there is a uniform increase and decrease in the flow profile for different positions, x, along the vessel. Arteries, however, are not rigid tubes, and it would be naive to presume so. As discussed earlier, the diameter of an artery is controlled to regulate cardiac output. Arteries also gradually taper and branch as they progress further down the network of vessels. Furthermore, arterial walls are elastic and stretch and recoil with each pulse of blood.

In order to proceed, more simplifying assumptions must be made. Thus it is assumed that the problem of modeling arterial blood flow is restricted to an arterial region where the gradual narrowing is negligible, and that no branching or vasoconstriction/vasodilation occurs.

Pulsatile Flow in Elastic Tubes: In elastic tubes, the movement of the walls implies that the velocity profiles will not only be dependent on the distance x, along the vessel, but also on the radial distance r.

The Navier-Stokes equations describe the axial symmetric motion.

$$\rho\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial r}\right)+\frac{\partial p}{\partial x}=\mu\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}\right)$$

or $$\rho\left(\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial r}\right)+\frac{\partial p}{\partial r}=\mu\left(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial r^2}+\frac{1}{r}\frac{\partial v}{\partial r}-\frac{v}{r^2}\right)$$

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial r}+\frac{v}{r}=0$$

Here, u is axial velocity and v is radial velocity.
Since the pressure gradient and velocities inside the tube are a function of both x and t, wave motion exists within the tube. If it is assumed that the length of the propagating wave, L, is much larger than the radius, r, and that the wave speed is much higher than the average flow velocity, the above Navier-Stokes equations simplify to give the following equations of motion.

$$\rho\frac{\partial u}{\partial t}+\frac{\partial p}{\partial x}=\mu\left(\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}\right)$$

$$\rho\frac{\partial v}{\partial t}+\frac{\partial p}{\partial r}=\mu\left(\frac{\partial^2 v}{\partial r^2}+\frac{1}{r}\frac{\partial v}{\partial r}-\frac{v}{r^2}\right)$$

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial r}+\frac{v}{r}=0$$

Looking for solution of the form

$$p(x,r,t)=P(r)e^{iw(t-x/c)}$$
$$u(x,r,t)=P(r)e^{iw(t-x/c)}$$
$$v(x,r,t)=V(r)e^{iw(t-x/c)}$$

where w is the frequency of oscillation of the input pressure, and c is the propagation speed, result in ordinary differential equations for P, U, and V . Solutions to these ODEs can be found using Bessel functions. To equate the motion of the arterial wall to that of the tubular flow, fluid pressure and shear

stress can be used. After much algebra and non-dimensionalizing, the axial and radial velocities can be determined as

$$\frac{u(x,r,t)}{\hat{u}_S} = \frac{-4}{\Lambda^2}\left[1 - G\frac{J_0(\zeta)}{J_0(\Lambda)}\right]e^{iw(t-x/c)}$$

$$\frac{v(x,r,t)}{\hat{u}_S} = \frac{2aw}{i\Lambda^2 c}\left[\frac{r}{a} - G\frac{2J_1(\zeta)}{\Lambda J_0(\Lambda)}\right]e^{iw(t-x/c)}$$

where $\hat{u}_S = -\frac{k_S a^2}{4\mu}$ is the axial velocity in steady flow, $\Lambda = \left(\frac{i-1}{\sqrt{2}}\right)$, $\zeta = \Lambda\frac{r}{a}$, a is the neutral position of the arterial wall, and G is the elasticity factor.

The rate of blood flow can be modeled from the above result. Flow rate is given by:

$$q(x,t) = \int_0^a 2\pi r\, u dr$$

Since the radius of the artery, r, is not constant, the neutral radius is used to gain an approximation. This assumption holds only if the radial movements are sufficiently small.

Non-dimensionalizing in terms of the flow rate for steady flow, qs, gives the expression below.

$$\frac{q(x,t)}{q_S} = \frac{-8}{\Lambda^2}\left(1 - G_g\right)e^{iw(t-x/c)}$$

This model for flow rate in elastic tubes provides a simplified prediction model of the dynamics of pulsatile flow in arteries. It is interesting to note that the wall movements of the arteries make it \easier" for blood to flow.

Wave Reflections, Branching and Tethering:

The above model does not adequately represent the true physical flow in arteries due to the fact that too many simplifying assumptions were made.

An important factor in pulsatile blood flow in arteries is reflection. Because the pulsatile flow propagates down the elastic tube as a wave, any obstacle will create reflections in the tube. When an artery branches into two smaller arteries, for example, an approaching wave is reflected. Reflected waves interfere with

forward moving waves to create a very complex flow field. Incorporating these wave dynamics into the above discussion creates a mathematical model that sufficiently resembles pulsatile blood flow in arteries.

Medical imaging can be used to determine specific branching structures in arteries. From these images, models can be derived to predict blood flow in particular branching structures.

These models provide an improved understanding of blood flow dynamics and of flow diseases, such as atherosclerosis.

As it turns out, the concept of tethering does not complicate the above derived model. Tethering refers to the way in which the arteries are \tied down" to nearby tissues. The assumption of no tethering implied that arteries were free to move around under the forces of the flow field. However, the actual effect of tethering in the cardiovascular system is added mass and stiffness to artery walls, rather than restriction of artery movement. Thus, tethering does not significantly alter the mathematical model.

This lecture notes has reviewed the basic physiology of the cardiovascular system. The main emphasis was placed on blood circulation through the network of vessels, and regulation of cardiac output. Focus was placed on the transport of nutrients to living cells as a measure of general cell health. Blood flow is controlled by the constriction or dilation of vessel walls, whose action is regulated by the sympathetic nervous system and by local conditions within blood vessels and surrounding tissues.

Measurement of oxygen and nutrient concentration in blood is very hard to obtain. The tertiary measurement, blood pressure, was discussed, both via ECG, and direct and indirect pressure measurement methods. Multiple flowmeters were then presented for the secondary measure of blood flow. Each technique measured a different aspect of the flow including instantaneous flow and flow profile.

New techniques are currently being developed for indicator-dilution methods. One such method is the trans-cerebral double-indicator-dilution technique, which determines cerebral blood flow.

Electromagnetic flowmeters are continually being improved upon but no major changes in the basic technique have been suggested. New designs for probes and sensors allow for improved blood flow monitoring during surgery.

Ultrasonic flowmeters include direct, indirect and laser Doppler techniques. Direct ultrasonic flowmeters are useful in surgery whereas indirect ultrasonic flowmeters can detect blood flow non-invasively. Pulsed ultrasonic flowmeter methods are very state-of-the-art as they detect flow profiles across vessels. Laser Doppler techniques are also new and can detect capillary flows.

Plethysmography can be used to measure volumetric changes in many different fashions. Body plethysmography, as well as chamber, and impedance plethysmography are staples in clinical diagnosis.

The problem of modeling arterial blood flow, as well as the many simplifying assumptions, was discussed. Highlights of a well-known mathematical model were presented. It was determined that the addition of wave reflections to the model, would enable it to accurately predict the dynamics of real pulsatile blood flow, even with the simplifying assumptions.

The problem of modeling arterial blood flow is a mature subject. Current investigations are now focusing on complex branching structures and on the smaller-scaled problems of arteriole and capillary flow. In particular, the assumption that blood acts like a Newtonian fluid fails for this microscopic-scaled problem. Mathematicians are thus trying to model blood flow in capillaries as a viscoelastic substance, instead of a Newtonian fluid.

Applications: Obtaining a valid mathematical model for branching structures of the arterial network allows virtual investigations to be conducted. Improved understanding of localized flow diseases can be gained through simulations. Also, flow shear stresses and other flow dynamics can be approximated to determine specifications for prosthetic vascular grafts, and other biomedical materials.

Pulsed ultrasonic flowmeters can be used to verify mathematical models of flow profiles across specified arterial cross-sections. Perhaps even more exciting, laser Doppler flowmeters can aid in the understanding of capillary blood dynamics. Modeling of capillary flow would provide great insight into the transport of oxygen and nutrients to tissue cells.

Modeling Oxygen transport in Human Body:

The mechanism by which oxygen enters and carbon dioxide exits living organism is an example of diffusion. Both oxygen and carbon dioxide as freely

through moist cell membranes, and all movement of these gases into and out of cells is solely by means of simple diffusion. Examples of diffusion include diffusion of oxygen from the lumen of the respiratory passages across two epithelial linings and into the blood, the diffusion of oxygen from the tissue fluid into a nerve cell of the brain, and the diffusion of carbon dioxide from a skeletal muscle cell into the tissue fluid. The other examples are secretion of waste materials by kidney (purification of blood), supply of nutrients to various tissues of the livings system. The process in all the examples as mentioned above is governed by one or other forms of diffusion phenomena.

We consider the problem of transport of molecular oxygen from the lungs to different cells of the body. We know that blood is the primary transport vehicle. This study is very important since it answer the questions, namely 1) what factors affect the supply of oxygen for tissue cell respiration? What happens when we inhale oxygen at low concentration, say at high altitudes? What happens when we inhale pure oxygen at high pressure?

We first try to understand how oxygen is transported through blood. We know that blood is a suspension of cells in an aqueous solution called 'plasma' as discussed earlier. The functionally important cell among these cells is red cells, which makes about 45% of the blood volume. These red cells are responsible for the transport of oxygen. The normal red cell has a disc shape of dimensions shown in the Figure below.

The cellular contents are separated from the surrounding plasma by the cellular membrane, which functions to maintain the integrity of the cell and to control the transport of material between the cell interior and plasma.

Now we shall see briefly how transport of oxygen (oxygenation) takes place. This involves several steps. Oxygen diffuses through a layer of blood plasma to the surface of the red cell. The oxygen then passes through the cell membrane. Once within the cell the oxygen diffuses through the hemoglobin (Hb) as dissolved oxygen (O_2) and as oxy-hemoglobin (HbO_2). The extent as well as the rate of consumption of oxygen by hemoglobin depends upon such factors as the equilibrium for the reactions of oxygen with hemoglobin, the reaction rates of oxygen with hemi groups, and the rates of diffusion oxygen through plasma, membrane and hemoglobin. O_2 is also transported as dissolved oxygen in plasma.

Exercises

1. State the problem in Biology for which modeling is sought.
2. Write the one dimensional diffusion equation, which describes the diffusion of a substance in a solvent (or medium).

Oxygen transport through a single red cell:

Let us now formulate a simple model for the diffusion transport of O_2 in a single red cell. We consider only pure diffusion and not go into other complexities like reaction within the cell.

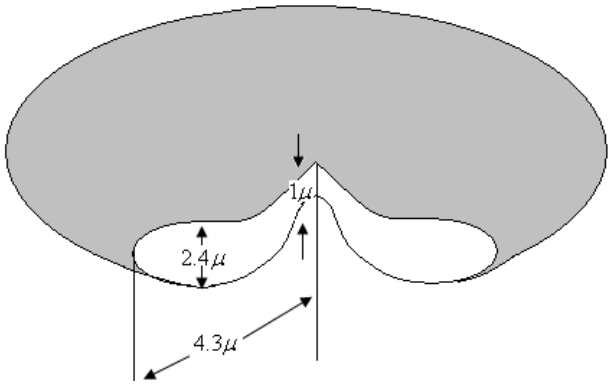


Figure: Schematic diagram of the normal red cell

Formulation of the model:

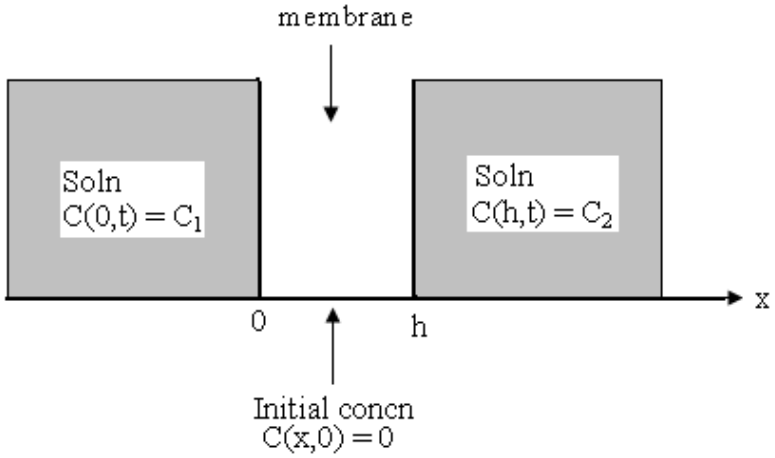
As in case of blood flows, here again, we consider a very simple model. Let us consider a planar section of a red cell. This will look like a plane infinite slab of hemoglobin solution enclosed within a membrane of thickness h as shown in the Figure below.

The Figure shows the schematic diagram of the model. The model considers the diffusion of oxygen through a membrane of thickness h, the two ends of which are maintained at concentration C_1 and C_2 respectively.

Since the cells are of very small dimensions, the slab can be considered as very thin and hence can be modeled as one dimensional diffusion problem.

Thus, we look for the concentration of O_2 within the membrane, i.e., C(x,t) in the region $0 < x < h$ which satisfies the diffusion equation.

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \;;\; 0 < x < h,\, t > 0$$



and the boundary conditions

$$C(0,t)=C_1\ ,\ t>0\ ,\ \ C(x,0)=0$$

$$C(h,t)=C_2\ ,\ t>0$$

Hence the equation together with the boundary conditions gives a formulation of the problem under consideration.
The next step is to find a solution of the formulated problem. The solution of the above diffusion equation together with the boundary condition is given by

$$C(x,t)=C_1+(C_2-C_1)\frac{x}{h}+\frac{2}{\pi}\sum_{n=1}^{\infty}\left(\frac{C_2\cos n\pi-C_1}{n}\right)\sin\frac{n\pi x}{h}e^{-\frac{n^2\pi^2Dt}{h^2}}$$

We can write this in the following form also

$$C(x,t)=C_1+(C_2-C_1)\frac{x}{h}+\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\left[C_2(-1)^n-C_1\right]\sin\frac{n\pi x}{h}e^{-\frac{n^2\pi^2Dt}{h^2}}$$

Exercise: Verify that the equation established above is a solution of the formulated equation with given initial and boundary conditions.
Therefore, gives an analytical solution of the formulated problem.

Solution: We shall use the method of separation of variables here.
Let $C(x,t)=X(x)T(t)$, substituting this in equation and dividing by C, we get

$$\frac{D}{X}\frac{d^2X}{dx^2}=\frac{1}{T}\frac{dT}{dt}$$

Since the left hand side and right hand sides are purely functions of x and t respectively, so equating them to a constant, say k. The constant has to be negative, otherwise the model will predict and exponential growth which will not be realistic. We take $k=-\lambda^2$. We know that in this case the equation has a non-trivial solution. We now get the following equations which determines $X(x)$ and $T(t)$.

$$\frac{1}{X}\frac{d^2X}{dx^2}=-\lambda^2$$

$$\frac{1}{T}\frac{dT}{dt}=-\lambda^2 D$$

which are ordinary differential equations for x and t. For $\lambda=0$, the solutions are

$$X(x)=Ax+B$$

and $T(t)=const.$

For $\lambda\neq 0$, $X(x)=C\sin\lambda x+E\cos\lambda x$

and $T(t)=Fe^{-\lambda^2 Dt}$

where A, B, C, E and F are constants to be determined. Now we apply the boundary conditions followed by equation. Then we get that

$B=C_1$, and $A=\frac{C_2-C_1}{h}$,

Note that h is the thickness of the slab.

$$C(x,t)=C_1+(C_2-C_1)\frac{x}{h} \quad \text{for } \lambda=0$$

As equation is linear, the principle of superposition can be used and the required solution in general is

$$C(x,t)=C_1+(C_2-C_1)\frac{x}{h}+(C\sin\lambda x+E\cos\lambda x)e^{-\lambda^2 Dt}$$

where $\lambda\neq 0$. Again using boundary conditions

$$C(0,t)=C_1=C_1+Ee^{-\lambda^2 Dt}$$

which gives E = 0.

$$C(h,t) = C_2 = C_2 + Ce^{-\lambda^2 Dt} \sin \lambda h$$

Therefore,

$$Ce^{-\lambda^2 Dt} \sin \lambda h = 0$$

If C = 0 then as already E = 0, equation does not satisfy initial condition. So it is satisfied for all t if

$$\sin \lambda h = 0$$

This condition in turn requires that λ has certain characteristic values, also called eigenvalues and are given by

$$\lambda = \lambda_n = \frac{n\pi}{h}, \qquad n = 1,2,3,\cdots$$

Corresponding to each λ_n, there is a separable solution having specific constant C, denote it by S_n. Again, by superposition principle

$$C(x,t) = C_1 + (C_2 - C_1)\frac{x}{h} + \sum_{n=1}^{\infty} S_n \sin \lambda_n x e^{-\lambda_n{}^2 Dt}$$

Again, the equation has to satisfy initial condition, we get

$$C_1 + (C_2 - C_1)\frac{x}{h} + \sum_{n=1}^{\infty} S_n \sin \lambda_n x = 0$$

or $$-C_1 + (C_1 - C_2)\frac{x}{h} = \sum_{n=1}^{\infty} S_n \sin \lambda_n x$$

The right hand side of the above equation is the Fourier representation of the function on left side. We know that the following orthogonality property of trigonometric functions

$$\frac{2}{\pi}\int_0^{\pi} \sin mx \sin nx dx = h_{mn} = 0 \quad \text{for } m \neq n$$

$$= 1 \text{ for } m = n$$

where h_{mn} is the Kronecker delta.

For $\lambda = \lambda_n = \frac{n\pi}{h}$, we get

$$\frac{2}{h}\int_0^h \sin\lambda_n x.\sin\lambda_m x dx = \frac{2}{h}\int_0^h \sin\frac{n\pi x}{h}.\sin\frac{m\pi x}{h}dx = h_{mn}$$

For finding the constants S_n in the above equation, we multiply both sides of the equation by $\sin\lambda_m x$ and integrate it over x from 0 to h.

Only one term for which n = m is non-zero and rest are zero and this enables us in finding S_n and are given by

$$S_n = \frac{2}{h}\int_0^h \left[-C_1 - (C_2 - C_1)\frac{x}{h}\right]\sin\frac{n\pi x}{h}dx$$

$$= \frac{2}{n\pi}(C_2\cos n\pi - C_1); \qquad n = 1,2,3,\cdots$$

Therefore, the solution of one-dimensional diffusion equationsatisfies the boundary and initial condition is

$$C(x,t) = C_1 + (C_2 - C_1)\frac{x}{h} + \frac{2}{\pi}\sum_{n=1}^{\infty}\left(\frac{C_2\cos n\pi - C_1}{n}\right)\sin\frac{n\pi x}{h}e^{\frac{-n^2\pi^2 Dt}{h^2}}$$

Example 3: Let the slab represent a biological cell in a large bathing space solution of solute with fixed (given) concentration C_0. Then find the concentration distribution inside the cell at any given position $0 < x < h$ and time $t > 0$.

The thin biological cell is given by $0 \le x \le h$ and the region $x < 0$ and $x > h$ represent the cell exterior.

Since the solute has a fixed concentration C_0, with $C_1 = C_2 = C_0$ we find that the even numbered terms vanish.

Therefore,

$C_2\cos n\pi - C_1 = 0$ for even n and $C_2\cos n\pi - C_1 = -2C_0$ for odd n,

with these values the equation reduces to

$$C(x,t) = C_0\left[1 - \frac{4}{\pi}\sum_{n=odd}^{\infty}\frac{1}{n}\sin\frac{n\pi x}{h}e^{-\frac{n^2\pi^2 Dt}{h^2}}\right]$$

Next we shall consider an application.

The current density which is defined as the rate at which the diffusing substance emerges at the interface x = 0 per unit area per unit time, is one of the important factors in the analysis of diffusion. It is denoted by J_0 and calculated by

$$J_0 = D\frac{\partial C}{\partial x}\bigg|_{x=0}$$

In the following examples we shall calculate the current density for the concentration C obtained in equation.

Example 4: Calculate J_0 and discuss the nature of terms in the series expansion of J_0, which influence J_0.

We have

$$J_0 = D\frac{\partial C}{\partial x}\bigg|_{x=0}$$

At x = 0,

$$D\frac{\partial C}{\partial x} = \frac{D}{h}(C_2 - C_1) + \frac{2D}{h}\sum_{n=1}^{\infty}(C_2\cos n\pi - C_1)e^{-\frac{n^2\pi^2 Dt}{h^2}}$$

First term in the RHS of the above series corresponds to steady state (time independent) flux and the remaining terms are negligible for large values of time i.e. these terms are significant only for short times, because of the exponential term in it. Since the exponential term also contains a factor of $-n^2$, the terms in equation, vanishes for large n. Therefore; we retain only the first and the second term only.

Hence, we have

$$J_0 \approx \frac{D}{h}(C_2 - C_1) - \frac{2D}{h}\sum_{n=1}^{\infty}(C_1 + C_2)e^{-\frac{\pi^2 Dt}{h^2}}$$

Interpretation and limitations of the model

From equation above, the RHS is an infinite series and can be solved numerically with the help of computers. The graph of the calculations of different values of Dt/h^2 is given in the Figure below.

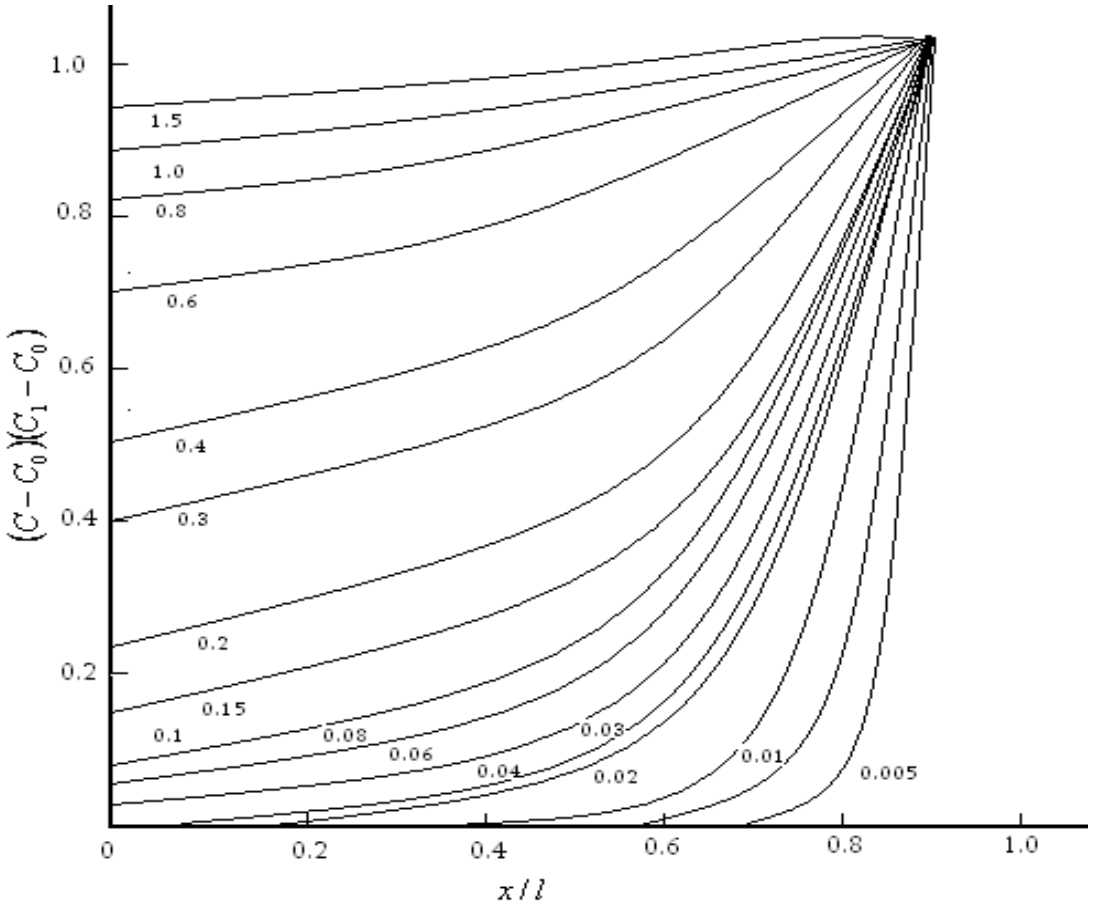


Figure: Pattern of Oxygen concentration

Now let us see the limitations of the model.

- The diffusion constant D may not be a constant – what means it may vary from position to position.
- This model does not consider the chemical reaction of oxygen with hemoglobin: we have discussed earlier that during the process of the transport of oxygen to the red cell in the reaction of oxygen with hemoglobin takes place.
- The model deals with only one cell. Modeling diffusion of O_2 in whole blood flow is more complicated since O_2 is carried by blood as (a) dissolved O_2 in plasma (b) oxyhemoglobin and model should give allowance to the two phases- plasma and red cells. The transfer of O_2 will not only be by diffusion but also by the bulk motion/ movement of blood.

Blood Circulation in Skin and Subcutaneous Tissues:

The vascular network in a tissue is idealized into a system of circulatory loops. Each loop consists of functional elements, arteriole, capillary and venule. The capillary is that portion of a loop where the substance of interest is exchanged between the flowing blood and the surrounding medium. The arteriole is that portion of a loop in which blood flows towards a capillary without exchanging substance of interest with the surrounding medium and the venule is that portion of a loop in which blood flows away from the capillary without exchanging substance of interest with the surrounding medium.

Circulation through skin serves following two major functions (i) Nutrition to the skin tissues and (ii) carrying heat from the internal body to the skin so that heat can be removed from the body. To perform these two functions, the circulatory system of the skin is divided into two major types of vessels. (a) The nutritive arteries, capillaries and veins (b) Vascular structures concerned with the heating of the skin. It consist principally extensive subcutaneous veins plexus which holds large quantities of blood that can heat the surface of the skin and in some skin areas, arteriovenous anastomosis, which are large vascular communications directly between the arteries and venous plexuses.

The blood flow in the epidermis is negligible due to absence of blood vessels in epidermis. The number of blood vessels in dermis is very thin near the interface of epidermis and dermis but it increases gradually and becomes large and almost uniformly distributed in the subcutaneous tissue. The rate of blood flow in SST region is most variable in comparison to any other part of the body.

When the skin is heated until maximum vasodilatation, the blood flow can be as much as seven times of its normal value. When the skin is exposed to cold, the blood vessels constrict and at a temperature of about 15^0C, they reach their maximum degree of constriction. The blood may conserve (due to vasoconstriction) or dissipate (due to vasodilatation) heat, which constitutes an important factor in temperature regulation in skin and subcutaneous tissues.

The Ideal Gas Law:

In fluid flow is playing a pivotal role in human physiology and anatomy. To understand the dynamics of fluid flow in human body, it is essential to develop and relationship between various parameters responsible for the fluid transprt in human body.

The Ideal gas law is appropriate to deal with the situation of fluid flow modeling in human body. The law describes the relationship between Pressure, Temperature, Volume, and Amount of a gas; and how also these relationships can be combined to give a general expression that describes the behavior of a gas. The ideal gas law, also called the general gas equation, is the equation of state of a hypothetical ideal gas. It is a good approximation of the behavior of many gases under many conditions, although it has several limitations. It was first stated by Benoît Paul Émile Clapeyron in 1834 as a combination of the empirical Boyle's law, Charles's law, Avogadro's law, and Gay-Lussac's law. The ideal gas law is often written in an empirical form:

$$PV = nRT$$

where P, V and T are the pressure, volume and temperature respectively; n is the amount of substance; and R is the ideal gas constant. It can also be derived from the microscopic kinetic theory, as was achieved (apparently independently) by August Krönig in 1856 and Rudolf Clausius in 1857.

Derivation of the Ideal Gas Law:

Any set of relationships between a single quantity such as V and several other variables (P, T and n) can be combined into a single expression that describes all the relationships simultaneously. The three individual expressions are as follows:

Boyle's law: $V \propto \frac{1}{P}$; with constant n and T.

Charles' law: $V \propto T$; with constant n and P.

Avogadro's law: $V \propto n$; with constant P and T.

Combining these three expressions gives

$$V \propto \frac{nT}{P}$$

which shows that the volume of a gas is proportional to the number of moles and the temperature and inversely proportional to the pressure. This expression can also be written as

$$V = R\frac{nT}{P}$$

where R is a constant called gas constant.

The resulting equation is

$$PV = nRT$$

and is the required ideal gas law.

An ideal gas is defined as a hypothetical gaseous substance whose behavior is independent of attractive and repulsive forces and can be completely described by the ideal gas law. In reality, there is no such thing as an ideal gas, but an ideal gas is a useful conceptual model that allows us to understand how gases respond to changing conditions. As we shall see, under many conditions, most real gases exhibit behavior that closely approximates that of an ideal gas. The ideal gas law can therefore be used to predict the behavior of real gases under most conditions. The ideal gas law does not work well at very low temperatures or very high pressures, where deviations from ideal behavior are most commonly observed.

The most frequently introduced forms are:

$$PV = nRT = nk_BN_AT = Nk_BT$$

where

- P is the absolute pressure of the gas,
- V is the volume of the gas,
- n is the amount of substance of gas (also known as number of moles),
- R is the ideal, or universal, gas constant, equal to the product of the Boltzmann constant and the Avogadro constant,
- k_B is the Boltzmann constant,
- N_A is the Avogadro constant,
- T is the absolute temperature of the gas,
- N is the number of particles (usually atoms or molecules) of the gas.

Alveolar Gas Equation:

The Alveolar gas equation (AGE) is a mathematical equation describing the partial pressure of oxygen (PAO_2) in the alveoli of the lungs. It takes into account the inspired partial pressure of oxygen (PIO_2), the partial pressure of carbon dioxide ($PACO_2$) and the respiratory quotient (R).

The following formula is used to approximate the partial pressure of oxygen in the alveolus (PAO_2):

$$PAO_2 = (PB - PH_2O)F_iO_2 - (P_aCO_2 \div R)$$

where PB is the barometric pressure, PH_2O is the water vapor pressure (usually 47 mmHg), F_iO_2 is the fractional concentration of inspired oxygen and R is the gas exchange ratio. (The rate of CO_2 production to O_2 used is usually around 0.8 at rest). At sea level

$$PAO_2 = (760 - 47)0.21 - (40 \div 0.8) = 100\text{mmHg}.$$

Henry's Law:

Henry's law is a gas law which states that at the amount of gas that is dissolved in a liquid is directly proportional to the partial pressure of that gas above the liquid when the temperature is kept constant. The constant of proportionality for this relationship is called Henry's law constant (usually denoted by 'k_H'). The mathematical formula of Henry's law is given by:

$$P \propto C \text{ (or) } P = k_H.C$$

where,

- P denotes the partial pressure of the gas in the atmosphere above the liquid.
- C denotes the concentration of the dissolved gas.
- k_H is the Henry's law constant of the gas.

The solubility of gases in liquids:

Henry's Law gives a quantitative relation between pressure and gas solubility in a liquid. It states that:

The solubility of a gas in a liquid is directly proportional to the partial pressure of the gas present above the surface of liquid or solution.

The most general way of using Henry's Law is that the partial pressure of a gas above a solution is proportional to the mole fraction of the gas in the solution.

$$P = K_Hx$$

where,

p = partial pressure of the gas

x = mole fraction of the gas in solution

K_H = Henry's law constant

Unit-II
Biofluid Mechanics

Introduction:

This unit is intended to be of an introductory nature to the vast field of biofluid mechanics. Here, we shall consider the ideas and principles of the preceding unit in the context of fluid motion in biological systems.

The human body is a complex system that requires materials such as air, water, minerals, and nutrients for survival and function. Upon intake, these materials have to be transported and distributed around the body as required. The associated bio-transport and distribution processes involve interactions with membranes, cells, tissues, and organs comprising the body. Subsequent to cellular metabolism in the tissues, waste byproducts have to be transported to the excretory organs for synthesis and removal. In addition to these functions, biotransport systems and processes are required for homeostasis (physiological regulation e for example, maintenance of pH and of body temperature), and for enabling the movement of immune substances to aid in the body's defense and recovery from infection and injury. Furthermore, in certain other specialized systems such as the cochlea in the ear, fluid transport enables hearing and motion sensing. Evidently, in the human body, there are multiple types of fluid dynamic systems that operate at macro-, micro-, nano-, and pico-scales. Systems at the micro and macro levels, for example, include cells (micro), tissue (microemacro), and organs (macro). Transport at the micro, nano, and pico levels include ion channeling, binding, signaling, endocytosis, and so on. Tissues constitute organs, and organs as systems perform various functions. For example, the cardiovascular system consists of the heart, blood vessels (arteries, arterioles, venules, veins, capillaries), lymphatic vessels, and the lungs. Its function is to provide adequate blood flow and to regulate that flow as required by the various organs of the body. In this chapter, as related to the human body, we shall restrict attention to some aspects of the cardiovascular system for blood circulation.

The Circulatory System in the Human Body:

The primary functions of the cardiovascular system are:

1) to pick up oxygen and nutrients from the lungs and the intestine, respectively, and deliver them to tissues (cells) of the body,
2) to remove waste and carbon dioxide from the body for excretion through the kidneys and the lungs, respectively, and
3) to regulate body temperature by advecting the heat generated and transferring to the environment outside the skin. The circulatory system in a normal human body (as in all vertebrates and some other select group of species) can be considered as a closed system, meaning that the blood never leaves the system of blood vessels. The motive mechanism for blood flow is the prevailing pressure gradient.

The circulations associated with the cardiovascular system may be considered under three sub-systems. These are the 1) systemic circulation, 2) pulmonary circulation, and 3) coronary circulation. In the systemic circulation, blood flows to all of the tissues in the body except the lungs. Contraction of the left ventricle of the heart pumps oxygen-rich blood to a relatively high pressure and ejects it through the aortic valve into the aorta. Branches from the aorta supply blood to the various organs via systemic arteries and arterioles. These, in turn, carry blood to the capillaries in the tissues of various organs. Oxygen and nutrients are transported by diffusion across the walls of the capillaries to the tissues. Cellular metabolism in the tissues generates carbon dioxide and byproducts (waste). Carbon dioxide dissolves in the blood and waste is carried by the bloodstream. Blood drains into venules and veins. These vessels ultimately empty into two large veins called the superior vena cava (SVC) and inferior vena cava (IVC) that return carbon dioxideerich blood to the right atrium. The mean blood pressure of the systemic circulation ranges from a high of 93 mm Hg in the arteries to a low of few mm Hg in the venae cavae. From the systems it is evident that pressure falls continuously as blood moves farther from the heart. The highest pressure in the vessels of the circulatory system is in the aorta and in the systemic arteries while the lowest pressure is in the venae cavae. In pulmonary circulation, contraction of the right atrium ejects carbon dioxideerich blood through the tricuspid valve into the right ventricle. Contraction of the right ventricle pumps the blood through the pulmonic valve (also called semilunar valve) into the pulmonary arteries. These arteries bifurcate and transport blood into the complex network of pulmonary capillaries in the lungs. These capillaries lie between and around the alveoli walls. During respiratory inhalation, the

concentration of oxygen in the air is greater in the air sacs of the alveolar region than in the capillary blood. Oxygen diffuses across capillary walls into the blood. Simultaneously, the concentration of carbon dioxide in the blood is higher than in the air and carbon dioxide diffuses from the blood into the alveoli. Carbon dioxide exits through the mouth and nostrils. Oxygenated blood leaves the lungs through the pulmonary veins and enters the left atrium. When the left atrium contracts, it pumps blood through the bicuspid (mitral) valve into the left ventricle.

Literature Survey of Biofluid Mechanics:

The motion of fluid in the human body has been a subject of scientific exploration since the very earliest stages of development of the scientific study of natural phenomena. Attempts to understand the mechanism were made by Leonardo da Vinci and Descartes. Harvey explained the nature of the circulation of blood in the cardiovascular system as early as 1628. On the basis of experiments with flow of blood serum in glass tubes, Poiseuille derived the relationship for laminar flow in 1842, which became well known as Poiseuille's law. In 1899, Frank established the first mathematical model of propagation of arterial pulse wave.

Later in the second half of this century a new and vigorous development of biomechanics of fluids took place, which led to an interdisciplinary approach to the study i.e., a co-operative study in which Physicians, physiologists, mathematicians, physicists, technicians and chemists were joined together. These studies later helped in the development of artificial organs like artificial kidney, or artificial heart or of its components, for example heart valves etc. The main topics covered in this unit are based on basic cell biology, nutrient uptake by a cell; Growth of microbial colony; growth of a chemostat; enzyme kinetics (Michaelis-Menten theory), allosteric enzymes, cooperative properties of enzymes etc.

Some Elements of Fluid Mechanics:

Fluid mechanics is the study of behaviour of fluids-both at rest and in motion. The term fluid refers to substances that are capable of flowing and that has no definite shape but rather assume the shape of the container in which they are placed. Fluids, therefore, refers to both liquids and gases. Gases are

differentiated from liquids in that gases expand to fill the container in which they are placed while liquids do not. If a volume of liquid less than the volume of a container is poured into the container, the liquid will rest toward the bottom of the container, will occupy a volume of the container equal to the volume of the liquid, and will have a free (level) surface.
Gases are compressible, their volume changes as their pressures and/or temperature changes. Liquids on the other hand, are practically incompressible; their volumes do not change appreciably to pressures and/or temperature changes. Accordingly, the analysis of gases will obviously differ from the analysis of liquids in some regard.

Viscosity:

Gases and liquids are collectively known as fluids since they are both made to flow if a force is applied. While all fluids will flow, we know that pouring water out of a bottle is a faster process than pouring, for example cream or honey. One obvious distinction between these three fluids is their thickness, which is an intuitive measure of how close the fluid is to a rigid object. Here we are using the word 'thicknesses as in the phrase 'thickened cream' or in the saying 'blood is thicker than water'.

Pressure:

The term 'pressure' refers to the effects of a force acting against and distributed over a surface. A solid, liquid or a gas may exert the force. Often the phrase causing a pressure is simply the weight of a material. Pressure is a very important factor in many fluid mechanics problems. Pressure exerted by a liquid varies directly with depth. For practical purposes, the pressure is quantified in terms of unit pressure, which is the amount of force per unit area. The unit pressure is dynes/cm^2 or N/m^2. Fluid pressure is the pressure exerted by a fluid. Fluid pressure along any plane at right angles to this plane satisfies the three basic laws, which are given below:

(i) Fluid pressure is equal in all directions. If we consider a plane lying in any direction, in a fluid at rest, the fluid pushes it sideways, upward or downward with the same pressure. The kinetic theory of liquid and

gases where the molecules of fluid move in all directions randomly confirms this first law.

(ii) Pressures at points lying in the same horizontal plane in a fluid are equal. The free surface of the fluid is a special case where the pressures are everywhere atmospheric. If the pressures are not equal (at points lying in the same horizontal plane) then by the definition of fluid there would be a shear force which would cause movement of fluid until pressures become equal.

(iii) The pressure increases with depth under the free surface. In a fluid at test under gravity, pressure increases uniformly with depth. If ρ is the density of fluid (gram per ml), g is the acceleration of gravity 980 cm per sec^2, and h is the depth (say of given point) in cm then the uniformly increasing pressure with depth is equal to ρgh dynes per cm^2.

The pressure gradient is the pressure drop per unit length.

Velocity and flow rate:

Velocity and flow rate are two fundamental parameters that are used in dealing with fluids in motion. The parameter velocity is a common one that tells how fast and in what direction something, such as an automobile, is moving. Velocity is such as ft/sec (or fps) or meters per second (m/s).

In dealing with moving fluids, velocities of individual moving particles that make up the fluid are ordinarily not equal; therefore, the velocity is not necessarily constant throughout a cross section. Because of resistance to flow by the walls of a conduit or by a channel bed, fluid velocity is generally lesser near such walls or channel beds and greater away from them. Accordingly, the velocity at a cross section of a closed conduit will generally vary from zero at the wall to a maximum near the centre of the conduit; while the velocity at a channel cross section will generally vary from zero at the bottom to some value at the liquid surface with a maximum value at some distance nearer the surface than the bottom of the channel. It is usually sufficient, and introduces no serious errors in flowing problems, to use a mean velocity at a cross section. Hence, the

term velocity as used in fluid flow generally indicates mean velocity unless otherwise specified.

It is important, in many contexts; to know "how much" fluids is flowing in addition to (or instead of) how fast it is moving. The parameter that tells "how much" is known as flow rate, or discharge. Flow rate is the quantity of fluid flowing per unit time.

Classification of Fluid Flows:

Fluid flow can be classified in a number of ways: (1) steady or unsteady flow, (2) one- two- or three- dimensional flow, (3) uniform or non-uniform flow and (4) laminar or turbulent flow. Certain classifications are explained briefly here.

Steady flow occurs when various flow parameters (flow rate, velocity, pressure, etc.) at a point are constant with time. If flow conditions vary with time, the flow is unsteady. Thus, water flowing through pipe to an open faucet may be steady flow; but if the faucet is turned on and off repeatedly, the flow would be unsteady. An unsteady flow is said to be pulsatile if it is periodic.

Flow is said to be laminar when adjacent fluid layers move at the same (or nearly the same) velocity and the paths of individual fluid particles do not cross or intersect. Laminar flow tends to occur with low fluid velocities and with fluid having high viscosities. Flow is said to be turbulent when adjacent fluid layers move at different velocities and paths of individual fluid particles are erratic and cross each other. Turbulent flow tends to occur with higher fluid velocities and with fluids having lower viscosities.

Continuity Equation:

The continuity equation is an equation that describes the transport of some quantities like fluid or gas. It is also known as the transport equation. The continuity equation is very simple and powerful when it is applied to a conserved quantity. When it is applied to an extensive quantity it can be generalized. Physical phenomena are conserved using continuity equations like energy, mass, momentum, natural quantities, and electric charge.

According to the continuity equation:

$$A_1 V_1 = A_2 V_2$$

where,

A_1 = cross-sectional area of region 1
V_1 = flow velocity in region 1
A_2 = cross-sectional area of region 2
V_2 = flow velocity in region 2.

Remark:
A continuity equation becomes useful if a flux can be defined. To explain flux, first, there must be a quantity q that can flow or move, such as energy, mass, electric charge, momentum, number of molecules, etc. Let us assume ρ is the volume density of this quantity (q), that is, the amount of q per unit volume.
The way by which this quantity q is flowing is described by its flux.

In Continuity equation, flux is of two types:

- **Volumetric flux -** Across a unit area, the rate of volume flow is known as Volumetric flux. It is calculated by the formula Volumetric flux = liters/ (second*area). Its SI unit is ($m^3 s^{-1} m^{-2}$)
- **Mass flux -** It is the rate of mass flow. Its SI unit is ($kg\ m^{-2} s^{-1}$. It is represented by the symbols j, J, Q, q.

Continuity Equation in Diffusion Systems:
Continuity equations are a local and stronger form of conservation laws. For example, a weak version of the law of conservation of energy states that energy can neither be created nor destroyed which means that the total amount of energy in the universe is fixed. It means energy can neither be created nor destroyed nor can it teleport from one place to another—it can only move by continuous flow.
A continuity equation is nothing but a mathematical way to explain this kind of statement. The continuity equation consists of many other transport equations like the convection-diffusion equation, Navier–Stokes equations, and the Boltzmann transport equation.

- **Convection–Diffusion Equation -** It is a combination of convection and diffusion equations. It describes the physical phenomena where particles,

energy, and other physical quantities are transferred with the help of 'diffusion and convection' inside a physical system.

- **Boltzmann Transport Equation -** Boltzmann transport equation describes the behavior (statistical in nature) of the thermodynamic system, which is not in the state of rest or equilibrium.

Continuity Principle:

Continuity principle refers to the principle of fluid mechanics. The principle of continuity equation is a consequence of the law of conservation of mass. Through the continuity equation, the behavior of fluid is described when it is in motion. Whereas, the second equation is based on Newton's law of motion (which describes the motion of an object and the force acting on its flow) and the third equation is based on 'the law of conservation of energy (which states that mass can be neither created nor destroyed.)

Integral Form:

The integral form of the continuity equation says that:

- When additional q flows inward through the surface of the region, the amount of q in a region increases and decreases when it flows outward;
- When new q is created inside a region the number of q increases and decreases
- When q is utilized;
- Apart from these two methods, there is no other way for the amount of q in a region to change.

In terms of mathematics, the integral form of the continuity equation expressing the rate of increase of q within a volume V is:

$$\frac{dq}{dt} + \oiint s_j . ds = \sum$$

- Here, S denotes an imaginary closed surface, that encloses a volume V,

- $\oiint_S dS$ is a surface integral over that closed surface,
- q denotes the total amount of the quantity in volume V,
- J is the flux of q,
- t denotes time.
- And Σ is the net rate that q is being produced inside the volume V.

Flow Rate Formula:

This equation gives very useful information about the flow of liquids and their behavior when it flows in a pipe or hose. The hose, a flexible tube, whose diameter decreases along its length has a direct consequence. The volume of water flowing through the hose must be equal to the flow rate on the other end. The flow rate of a liquid means how much a liquid passes through an area in a given time.

The formula for the flow rate using equation of continuity can be written as:

$$m = \rho_{i1} v_{i1} A_{i1} + \rho_{i2} v_{i2} A_{i2} + \cdots + \rho_{in} v_{in} A_{in}$$

$$m = \rho_{O1} v_{O1} A_{O1} + \rho_{O2} v_{O2} A_{O2} + \cdots + \rho_{On} v_{On} A_{On}$$

where,

m = Mass flow rate
ρ = Density
v = Speed
A = Area

With uniform density equation of the above equation is

$$q = v_{i1} A_{i1} + v_{i2} A_{i2} + \cdots + v_{im} A_{im}$$

$$q = v_{O1} A_{O1} + v_{O2} A_{O2} + \cdots + v_{Om} A_{Om}$$

where

q = flow rate

$$\rho_{i1} = \rho_{i2.= \ldots =} \rho_{in} = \rho_{O1} = \rho_{O2} = \ldots = \rho_{Om}$$

Continuity Equation in Fluid Dynamics

The continuity equation in fluid dynamics says that in any steady-state process, the rate at which mass leaves the system is equal to the rate at which mass enters a system including the accumulation of mass within the system.

The differential form of the continuity equation is:

$$\frac{\partial \rho}{\partial t} + \nabla.(\rho u) = 0$$

where

t =Time

ρ =Fluid density

u =Flow velocity vector field

The derivative time can be understood as the loss of mass in accumulation inside the system, while the divergence term means the difference in flow in and flow out. The above-mentioned equation is also one of the (fluid dynamics) Euler equations. The equations of Navier–Stokes form a vector continuity equation expressing the conservation of linear momentum.

Derivation of Continuity Equation:

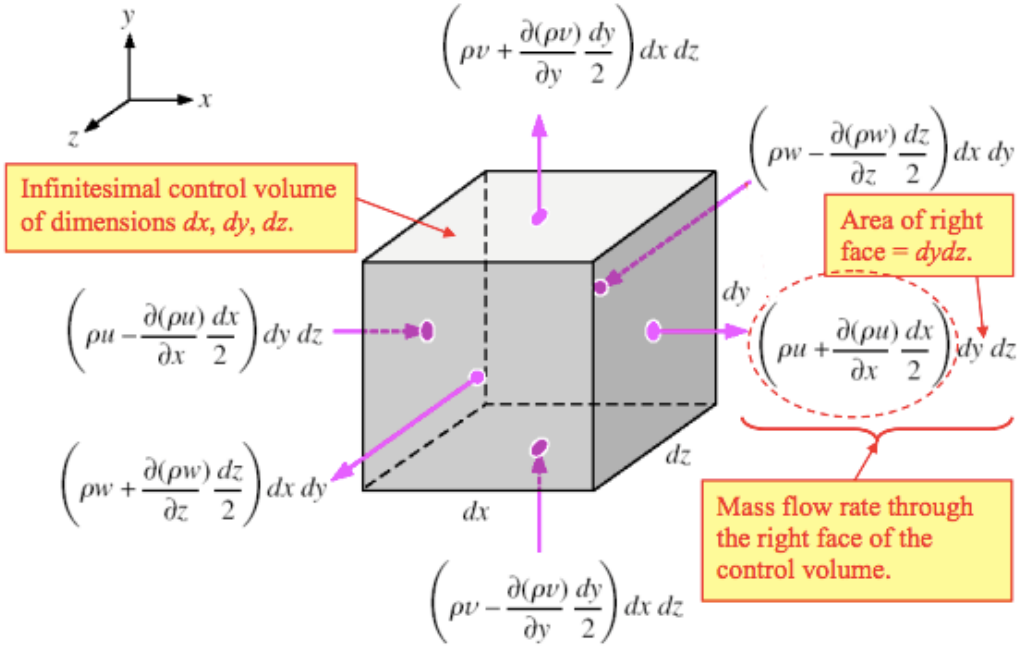
In thermodynamics, a *control volume* is defined as a fixed region in space where one studies the masses and energies crossing the boundaries of the region. This concept of a control volume is also very useful in analyzing fluid flow problems. The boundary of a control volume for fluid flow is usually taken as the physical boundary of the part through which the flow is occurring. The control volume concept is used in fluid dynamics applications

To approximate the mass flow rate into or out of each of six surface of the control volume, we shall use Taylor series expansion around the centre point, where the velocity components and density are u, v, w and ρ.

$$(\rho u)_{\text{center of right face}} = \rho u + \frac{\partial(\rho u)}{\partial x}\frac{dx}{2} + \frac{1}{2!}\frac{\partial^2(\rho u)}{\partial x^2}\left(\frac{dx}{2}\right)^2 + \cdots$$

Ignore terms higher than order *dx*.

Consider the right surface of the given control voume.
The mass flow rate through each face is equal to ρ times the normal component of velocity throught the face times the area of the face. The mass flow rate equations through all six faces in the diagram are given below.



Now adding all the mass flow rates through all six faces of the control volume (CV) to establish general equation of continuity (unsteady, incompressible).

Net mass flow rate into CV: ← all the *positive* mass flow rates (into CV)

$$\sum_{\text{in}} \dot{m} \cong \underbrace{\left(\rho u - \frac{\partial(\rho u)}{\partial x}\frac{dx}{2}\right) dy\, dz}_{\text{left face}} + \underbrace{\left(\rho v - \frac{\partial(\rho v)}{\partial y}\frac{dy}{2}\right) dx\, dz}_{\text{bottom face}} + \underbrace{\left(\rho w - \frac{\partial(\rho w)}{\partial z}\frac{dz}{2}\right) dx\, dy}_{\text{rear face}}$$

Net mass flow rate out of CV: ← all the *negative* mass flow rates (out of CV)

$$\sum_{\text{out}} \dot{m} \cong \underbrace{\left(\rho u + \frac{\partial(\rho u)}{\partial x}\frac{dx}{2}\right) dy\, dz}_{\text{right face}} + \underbrace{\left(\rho v + \frac{\partial(\rho v)}{\partial y}\frac{dy}{2}\right) dx\, dz}_{\text{top face}} + \underbrace{\left(\rho w + \frac{\partial(\rho w)}{\partial z}\frac{dz}{2}\right) dx\, dy}_{\text{front face}}$$

We plug these into the integral conservation of mass equation for our control volume:

$$\int_{\text{CV}} \frac{\partial \rho}{\partial t}\, dV = \sum_{\text{in}} \dot{m} - \sum_{\text{out}} \dot{m}$$

This term is approximated at the center of the tiny control volume, i.e.,

$$\int_{\text{CV}} \frac{\partial \rho}{\partial t}\, dV \cong \frac{\partial \rho}{\partial t}\, dx\, dy\, dz$$

The conseravation of mass equation thus becomes

$$\frac{\partial \rho}{\partial t} dx\, dy\, dz = -\frac{\partial(\rho u)}{\partial x}\, dx\, dy\, dz - \frac{\partial(\rho v)}{\partial y}\, dx\, dy\, dz - \frac{\partial(\rho w)}{\partial z}\, dx\, dy\, dz$$

Dividing both sides by the volume $dx\, dy\, dz$ of the CV, we have

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0$$

Applying the divergence of a vector, we have

$\vec{\nabla}.\vec{G} = \frac{\partial G_x}{\partial x} + \frac{\partial G_y}{\partial y} + \frac{\partial G_z}{\partial z}$; where $\vec{\nabla} = \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}\right)$ and $\vec{G} = (G_x, G_y, G_z)$

Using $\vec{G} = \rho\vec{V}$ in the above equation, where $\vec{V} = (u, v, w)$

The resulting equation is given by

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}.(\rho\, \vec{V}) = 0$$

which is the required continuity equation.

Uses of the Continuity Equation

The continuity equation is commonly used in pipes, tubes, and ducts. These structures have flowing fluid or gasses etc. which need a specific flow to be moved. Continuity equation can also be applied to huge water sources such as rivers, lakes, etc. This equation can also be applied in diaries, power plants, road logistics, etc.
Along with this, the modern application of continuity equations includes computer networking and semiconductor technologies, cell and tissue interaction etc. which uses a specific path to move data from one location to another. It is also used in gas pipelines and underground connections to transport gas.

Volumetric Flow Rate:

The volumetric flow rate ($\dot{\boldsymbol{V}}$) of a system is a measure of the volume of fluid passing a point in the system per unit time. The volumetric flow rate can be calculated as the product of the cross-sectional area (A) for flow and the average flow velocity (v).

$$\dot{\boldsymbol{V}} = \boldsymbol{Av}$$

It is the area measured in sqft and velocity in feet/sec. The equation above results in volumetric flow rate measured in cf/sec. Other common units for volumetric flow rate include gallons per minute, cubic centrimeters per second, liters per minute and gallons per hour.

Example:

A vessel with an inner diameter of 4 mm contains blood that flow at an average velocity of 3 meters per second. Calculate the volumetric flow rate of blood in the vessel.

Solution: From the equation below, we have

$$\dot{\boldsymbol{V}} = \boldsymbol{Av}$$

$$\dot{\boldsymbol{V}} = (\boldsymbol{\pi r^2})\boldsymbol{v}$$

or $\dot{V} = (3.14)\left(\frac{2}{1000}m\right)^2\left(3\frac{m}{sec}\right) \approx 0.02\, m^3/sec$ is the volumetric flow rate of blood in the given vessel.

Examples of Continuity equation (Piping Expansion):

Steady- state flow exists in a that blood vessel that undergoes a gradual expansion from a diameter of 6 mm to a diameter of 8mm. The desnity of the fluid in the vessel is constant at 6.08 lbm/m^3. If the blood flow velocity is 2.24 m/sec in the 6mm section, what is the flow velocity in the 8mm section.

Solution: From the continuity equation, we have the mass flow in the 6mm section must equal the mass flow rate in the 8 mm section. Let the subscript 1 represents the 6 mm section and 2 represent 8mm section, we have

$$\dot{m}_1 = \dot{m}_2$$

$$\rho_1 A_1 v_1 = \rho_2 A_2 v_2$$

or $$v_2 = \frac{\rho_1 A_1 v_1}{\rho_2 A_2} = v_1 \frac{A_1}{A_2}\frac{\rho_1}{\rho_2} = v_1 \frac{\pi r_1^2}{\pi r_2^2}$$

$$= \left(2.24\frac{m}{sec}\right)\frac{(3\ mm)^2}{(4\ mm)^2}$$

$$= 1.26\frac{m}{sec}.$$

So by using the continuity equation, we find that the increase in vessel diameter from 6 to 8 mm caused a decrease in flow velocity from 2.24 to 1.26 m/sec.

The continuity equation can also be

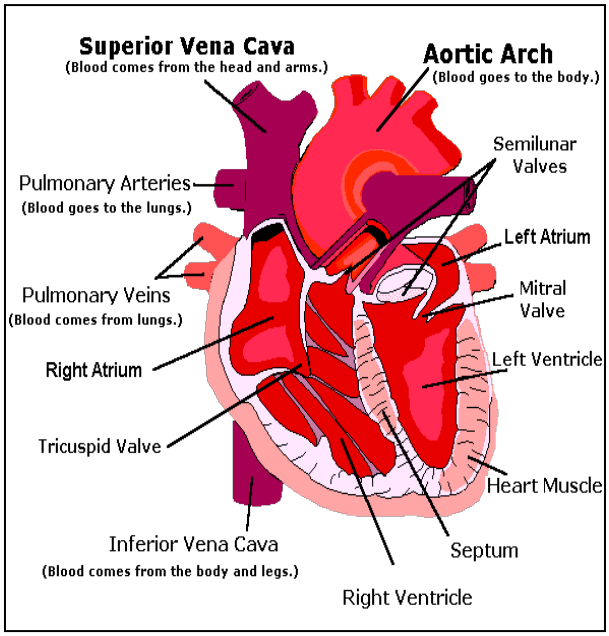


Figure: The structure of human Heart

used to show that the decrease in pipe diameter can increase in flow velocity.

Blood Flow (Circulation of Blood):

Human circulatory system is a transport mechanism by which the circulating fluid, called blood is moved through the body in a regular fashion by the pumping action of the muscular heart. Seventeenth century mathematician William Harvey showed that the blood circulates from the heart in vessels, through the body and taken back to the heart. He also showed that the blood circulates through the body and it passes through the vessels in only one direction. That means the blood does not flow back and forth through the same vessels.

Blood vessels are of three types: the arteries (and arterioles) carry blood away from the heart: the capillaries exchange material with the tissues; and the veins (and venules) return blood to the heart.

The heart is actually two pump housed within a single structure. The right pump receives blood from all of the body except the lungs and pumps it to the lungs as shown in the Figure.

The left pump receives blood from the lungs and pumps it to the remainder of the body. The chambers of the right side of the heart consist of one atrium and one ventricle. The atrium is a relatively small chamber, and its walls are thin. Emptying into the right atrium are two large veins; the superior vena cava, which brings blood from the upper body, and the inferior vena cava, which drains the lower body. Blood flows from the right atrium into the right ventricle through an opening guarded by a one- way valve, the tricuspid valve, which prevents the backflow of blood from the ventricle. The blood leaves the right ventricle by way of the large pulmonary artery, which branches and services the lungs. A one-way valve, the semi-lunar valve, at the junction of this artery with the heart prevents blood from flowing backward into the ventricle, which the ventricle is relaxing.

Blood returns from the lungs through the pulmonary veins, which drain into the left atrium. From the thin walled left (atrium), blood flows into the larger and very muscular left ventricle. The bicuspid (mitral) valve on this side of the heart prevents the flow of blood back into the atrium once the ventricle begins to contract. Blood flows from the left ventricle into the aorta, the largest artery of the body. A semilunar valve also guards the entrance to the aorta. The aorta

gives rise to the coronary arteries and the major arteries that service the head, arms and upper chest areas. The aorta then turns downward, paralleling the backbone and gives rise to arteries that serve the abdominal organs and body wall. In the pelvis the aorta branches and sends arteries into each leg. Blood returns from the various parts of the body, other than the lungs, by way of veins that ultimately empty into the right atrium as shown in Figure.

Constituents of Blood:

Blood is a complex mixture of cells and cell fragments, proteins and small molecules of many types all dissolved or suspended in water. The cellular components of blood are considered to represent connective tissue, but certainly connective tissue of a very special kind. The matrix of blood cells is obviously a liquid; however, the cells do not produce the liquid. The cells and cell fragments, the microscopically visible objects, are referred to as the formed elements as distinguished from the liquid material called the plasma. Formed elements constitute about 40-45 % of the blood and the plasma makes up the remaining 55-60 %.

The blood plasma is a complex liquid consisting 90-92 % water and 7-8 % solutes. In addition to water and proteins, amino acids, carbohydrates, fats, waste materials, salts, hormones and dissolved gases are present in specific quantities. Even though the plasma is continually delivering a variety of materials to different target sites, there are mechanisms that constantly control its composition.

For example, the liver keeps blood sugar level constant by

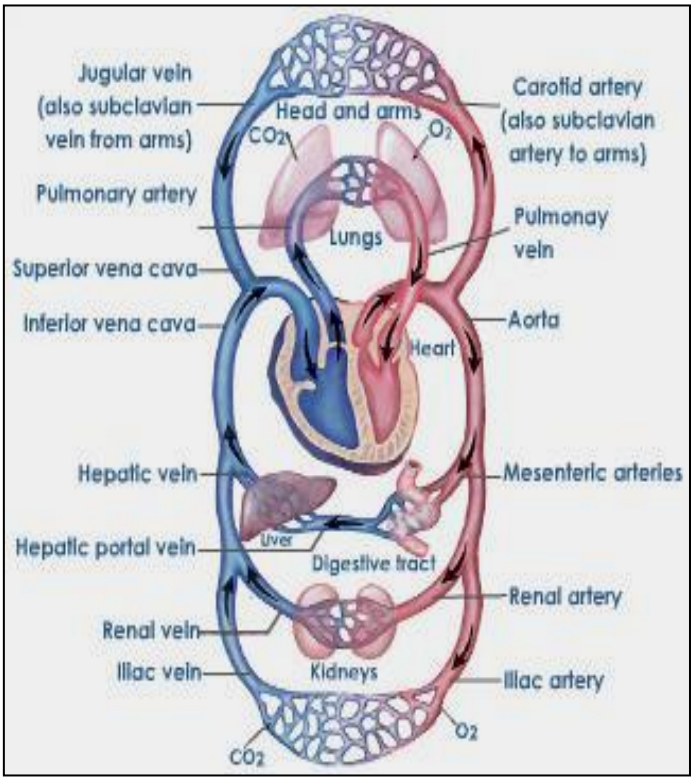
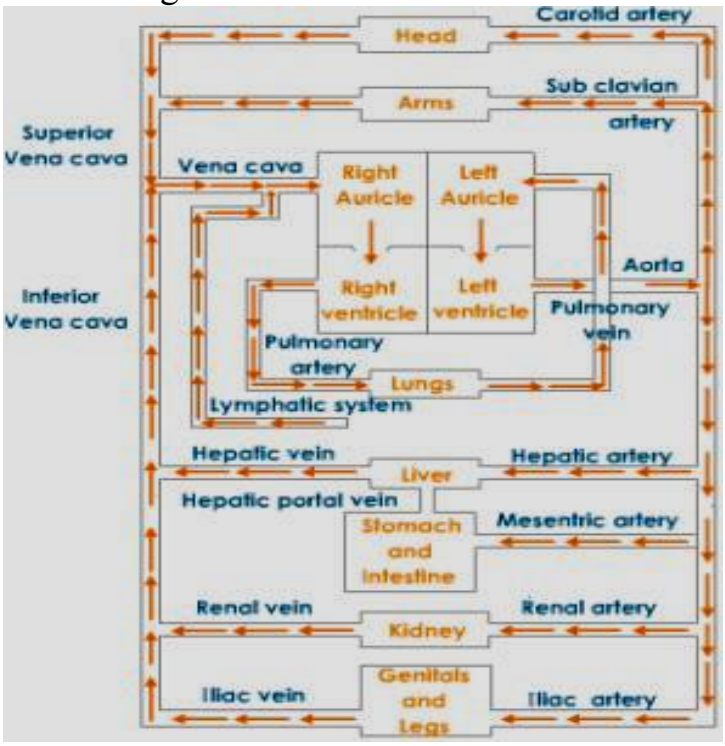


Figure: Blood circulation in in Human Body

releasing glucose into the blood as it is needed, the lungs remove carbon dioxide and add oxygen and thus maintain the concentration of these gases in the blood and the kidneys remove nitrogenous wastes, water and many other substances as they accumulate. All of these activities and many others produce a blood fluid of remarkable stability upon which the fragile leaving cells of the body are absolutely dependent.

***Blood is indeed thicker than water*:**

The proteins of the plasma especially albumin, give it a viscosity about three times that of water. The plasma proteins play several vital roles. Some are involved in clotting. Others regulate the water content of the blood and tissue cells by affecting osmosis. Proteins called immunoglobulin are sources of antibodies. One of these proteins known as gamma globulin is especially well known for its ability to form antibodies that will complex (join) with and inactivate the measles and hepatitis viruses and the tetanus bacterium. Some play a part in regulating the acid base balance of the blood. All these activities are essential to a healthy existence. Blood flows from regions of higher pressure to regions of lower pressure. The pressure is greater in arteries near the heart than in other parts of the body. Close to the heart the pressure also fluctuates greatly as the heart alternatively contracts and relaxes.



The vessels of the human circulatory system consist of arteries, capillaries and veins. These tubular passageways can easily be distinguished from one another on the basis of their structure. Capillaries are microscopic tubes that connect the arteries to the veins. Their

Figure: Network of cardiovascular system in human body

walls are extremely thin, consisting of simple (once cell thick) squamous epithelium. This epithelial layer is usually referred to as endothelium. The diameter of capillaries is so small that red blood cells are often bent in squeezing through the vessel. There are two ways in which substances can pass through the capillary walls and reach the surrounding tissue cells: Diffusion through the endothelial cells and a bulk flow through the cracks between the endothelial cells. Materials are returned from the tissue spaces back into the capillary by the same two routes. White blood cells escape from the capillary by squeezing between the endothelial cells. However, non-motile cells such as red blood cells, platelets and most proteins in the blood cannot normally leave the capillary.

The arteries are relatively thick walled vessels. The tough but flexible outer wall consists of a combination of connective tissues in which many elastic and collagenous fibers are present. The middle layer of the wall is made up of smooth muscle and some elastic connective tissue fibers. The inner wall consists of the same thick layer of endothelium that constitutes the capillary wall. The inner surface of the endothelium is especially slick and smooth, some times said to be glassy and it plays an important role in preventing the spontaneous formation of clots that tend to form as blood flows over a rough surface.

The smooth muscle in the artery wall is supplied with nerve fibers that regulate its contraction. Stimulation of the muscle results in a narrowing of the arterial lumen, the bore of the tube, which raises the blood pressure within the vessel. Such constriction in small arteries greatly reduces the volume of blood that can pass through the vessels and permits blood flow to be shunted away from certain body parts when it is needed elsewhere.

A reduction in nervous stimulation results in a relaxation of arterial muscle and a lowering of blood pressure. A greater volume of blood is transported through relaxed arteries. Certain hormones produce the same effects as nerve impulses in the smooth muscle of arteries indeed; normal contraction in these vessels is usually controlled by a combination of nervous and hormonal mechanisms.

The structure of veins is similar to that of arteries. However, the outer layer of connective tissue contains fewer elastic fibers and the smooth muscle layer is considerably thinner. Because they have less muscle, veins do not undergo nearly as much change in size due to muscle contraction and relaxation, they can stretch considerably. However, veins are also different from arteries in having

valves that prevent the backflow of blood. Such a backflow is never a problem in arteries since the blood there is under considerable pressure from the heart. But in veins the pressure is much reduced and under certain circumstances, the blood would flow in a reverse direction if it were not for the valves.

Another important distinction between arteries and veins involves the direction of blood flow. Blood in arteries always flows away from the heart and towards a capillary bed in some part of the body. Blood in veins is always flowing away from the capillary bed in the tissue it has served and back towards the heart. The arteries, veins and capillaries in an average adult consist of about 60,000 miles of vessels. The volume of blood in the different types of vessels at any given time varies considerably. Excluding the lungs, the veins contain about 59 % of the blood, arteries 13%, and capillaries 7%. The vessels involved in lung circulate on contain about 12% of blood and the heart about 9%.
The following table gives the diameter and blood velocity of the different blood vessels.

Structure	Diameter (cm)	Blood Velocity (cm/Sec)
Ascending aorta	2.0-3.2	63
Descending aorta	1.6-2.0	27
Large arteries	0.2-0.6	20-50
Capillaries	0.0005-0.001	0.05-0.1
Large veins	0.5-1.0	15-20
Vena cava	2.0	11-16

Modeling Blood Flow:

We know that blood is carried from heart to various parts of the body and eventually returned to heart. In fact, blood is carried through system of elastic tubes-the arteries, capillaries and veins. The blood returns to the heart without actually leaving the system. This process is known as circulation of blood or flow of blood as discussed above.
We also know that proper flow of blood is essential to transmit oxygen and other nutrients to various parts of the body in human beings as well as in all other animals. Any constriction in the blood vessel or any change in the characteristics

of blood vessels can change the flow and cause damages ranging from minor discomfort to death, in worst case. Therefore a better understanding of the physiology of the system is essential. Mathematical modeling of the system is aimed at this:

As a first step in modeling, we shall first identify the essential characteristics of blood flow. We list them below:

i) Blood is a non-homogeneous fluid
ii) Blood vessels are elastic, they branch repeatedly
iii) Blood flow is unsteady or pulsatile
iv) Blood flow is generally laminar except for flow near heart

Viscosity:

Suppose a force is applied to a portion of a mass of a fluid it will begin to flow but if the force is removed the movement will be brought to rest. On the other hand, if a similar portion of a fluid is kept in moving, the movement will be transferred to the rest of the fluid. This property is analogous to that of friction between solid bodies.

Now we shall explain the concept of Viscosity of a fluid based on the following simple experiment. Consider the motion of a fluid between two long parallel plates one of which is at rest and the other one is moving with a constant velocity U parallel to itself as shown in Figure.

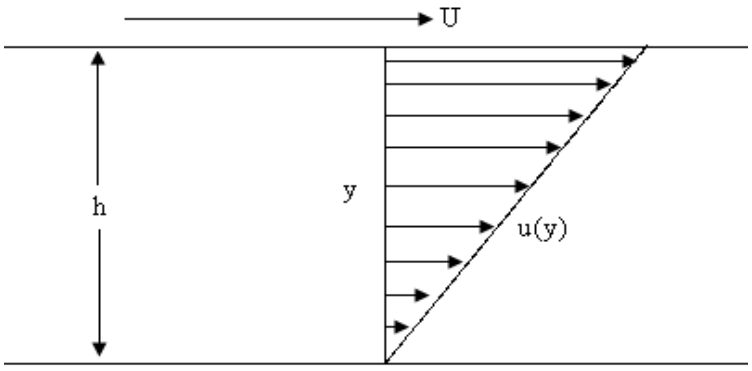


Figure: Motion of fluid between parallel plates

Let the distance between the plates be h and the fluid velocity be u. Assume that the fluid pressure is constant throughout the fluid. Due to cohesive nature of fluid it adheres to the plates. The fluid velocity at the lower plate is zero and that

at upper plate is U. This is because the upper plate is moving and the lower plate is at rest. So we get

$u = 0$, when $y = 0$

$u = U$, when $y = h$

Experimentally, it is observed that the fluid velocity distribution is linear and as such it is given by

$$u(y) = \frac{U}{h} y$$

where y is the direction at right angles to the flow. In order to support the motion it is necessary to apply a tangential force to the upper plate. Experimentally it is observed that this force, taken per unit area, is proportional to the velocity U of the upper plate and inversely to the distance h. If τ denote the force, then τ is directly proportional to U/h. This is denoted by

$$\tau \propto U/h$$

Many researchers have studied this property; the first theoretical consideration was made by Newton in which he considered the motion imparted to a large volume of fluid by the rotation of a long cylinder suspended in it. The hypothesis on which he based his derivation was that the resistance, which arises from the defect of slipperiness of the parts of the liquid, other things being equal is proportional to the velocity with which the parts of the liquid are separated from one another. 'Defect of slipperiness' was the term used to describe what we now call viscosity. This hypothesis emphasizes immediately that in a fluid moving relative to a surface there are laminar slipping on one another and so moving at different velocities. There is thus a velocity gradient i.e., du/dy in this case in a direction perpendicular to the surface. This gradient is usually called the rate of shear. In modern terms, the velocity gradient is written as du/dy, where y is the distance from the axis. The resistance or force is denoted as τ.

Then by Newton's hypothesis

$$\tau = \mu(du/dy)$$

where μ is a constant. Note that when we differentiate the expression given in equation above and substitute for du/dy in the subsequent equation, we get the expression given in another equation. μ is called the proportionality constant which gives the measure of the viscosity of the fluid, μ is also called the coefficient of viscosity.

Exercise: Find the dimension of viscosity

Solution: We know that

$$\text{Viscosity} = \frac{Force}{Velocity\ \ gradient}$$

Also, Force = mass x acceleration
Here we are considering force per unit area
Therefore, the dimension of the force = $ML^{-1}T^{-2}$
Also dimension of velocity = LT^{-1}
Therefore, dimension of velocity gradient = $\frac{LT^{-1}}{L} = T^{-1}$
Hence dimension of viscosity = $ML^{-1}T^{-1}$

Remark: It is common experience that in hot weather honey flows rapidly in comparison with its behavior in cold weather. This shows that the coefficient of viscosity is dependent on temperature.

The unit of viscosity is called Poise. The viscosity of water at $20.2\,^0C$ is 0.01 Poise and at $37\,^0C$ it is 0.0069 Poise. At $20\,^0C$, water it is approximately 60 times more viscous than air. Over the range $0\,^0C$ - $30\,^0C$ the viscosity of air increases by about 9% while that of water decreases by 45%. Using this we can say that the viscous components of resistance to motion (the frictional drag) is about 9% higher for birds fling in tropics than for the same bird flying in arctic. Also fish (and other marine creatures) have considerably easier way for moving about in tropics than in arctic water.

For convenience the viscosity of any fluid is expressed relative to the viscosity of water. This viscosity of fluid is known as relative viscosity.

The equation is known as Newton's law of friction. Fluids obeying this law are called Newtonian otherwise non-Newtonian. Most of the common fluids obey this law.

Poiseuille's Law:

Poiseuille's law is the relation between flow rate and pressure gradient –for fluid flow in a rigid cylindrical tube under a pressure gradients.

(Note that the pressure gradient is the pressure drop per unit length $\frac{dp}{dz} = \lim_{\Delta z \to 0} \frac{\Delta p}{\Delta z}$).

In order that we can understand the flow properties of biological fluids such as blood, which may exhibit non-Newtonian properties it is first necessary to discuss the behavior of simple or Newtonian fluids.

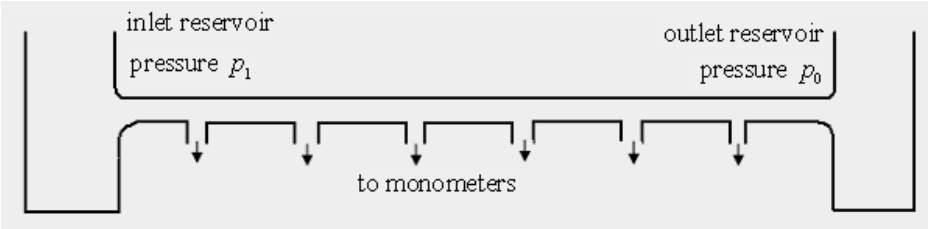


Figure: Flow properties of simple fluid

Let us look at the flow properties of a simple liquid like water in a very long horizontal pipe. Imagine that the pipe is circular in cross-section and d units in diameter as shown in Figure above.

Its entrance and exit are connected to large reservoirs so that the pressure drops between the ends of the tube may be maintained constant and a steady flow of water through the pipe is achieved. Small side hole, or lateral, pressure tappings are made in the pipe at frequent intervals along its length and these tappings are connected to a series of manometers. It is thus possible to measure the pressure drop per unit length or pressure gradient along the pipe.

If the pressure at the inlet to the pipe is p_1 and that at the outlet p_0, then we shall observe that, as $p_1 - p_0 (or\ \Delta p)$ is increased by raising the level in the upstream reservoir, so is the flow rate V through the pipe.

It was Poiseuille in 1840, who as a first step towards understanding the mechanics of the circulation, published a quantitative study of the flow properties in a pipe very remote from the entrance, and flow conditions in this region are now named after him. In addition to varying the flow rate and tube size, Poiseuille also studied the effect of viscosity on the flow conditions. Here we found that as viscosity was increased so was the pressure gradient necessary to maintain a given flow-rate.

Now to derive Poiseuille's formula we make use of Newton's second law of motion, which says that

Mass x acceleration

= body force + pressure gradient force + Viscous force

Let us consider the fluid flow through a circular tube with length L and diameter D = 2R, which is small compared with the length. We assume that the rate of flow is constant i.e. flow is steady. We also assume that the fluid velocity everywhere inside the tube is laminar/ stream lined. As we know for a laminar flow, the velocity is purely in the direction parallel to the axis of the cylinder. The fluid velocity at the inner surface is zero and it reaches maximum value on the axis (here axis means axis of the cylinder.)

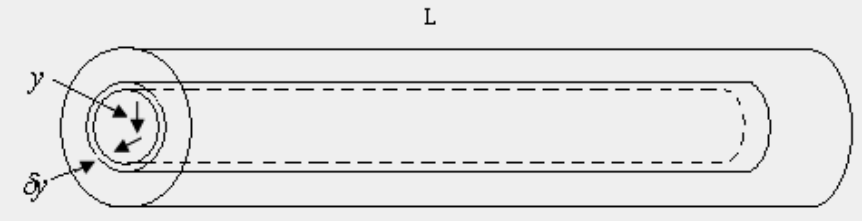


Figure: Fluid flow through cylindrical tube

We can consider the flow of fluid as the simultaneous movement of several layers, which are in the form of hollow cylinders one inside the other. If we assume that y is the radius of any one of these cylinders, then y varies from 0 to R, i.e., $0< y < R$ as shown in Figure.

If we consider the fluid flow to be due to pressure differences at the ends of the tube from the higher to the lower one then the only force opposing this flow is viscous resistance. We know that this force is $\mu(du/dy)$, and we find that the fluid particles are accelerated by the pressure difference and retarded by viscous resistance. If we look at the equation, then we will find that the only forces present are pressure gradient force and viscous force. This is because, since the flow is a steady flow in a straight tube, the fluid is not subjected to any acceleration (i.e. when the flow is steady, things do not change with respect to time). Therefore, LHS of equation is zero also, since we are considering the flow in horizontal pipes, gravitational forces are not relevant and therefore the body force term also vanishes. Thus, equation reduces to Pressure gradient force = - Viscous force. Now if F_{visc} denote the viscous force and F(P) denote the pressure difference, then we have

$$F(P) = - F_{visc}$$

(The negative sign indicates one force accelerates the motion, the other retards.)

Now we will calculate the LHS and RHS of the above equation, we first consider the RHS of equation. Here note that each flow is in the form of cylindrical layer of length L and radius y, y varying from 0 to R. The viscous force acts on the surface and it is given by the following formula:

F_{visc} = Surface area of the cylinder x viscosity x the velocity gradient

We have denoted the viscosity as μ and we know that the velocity gradient is given by du/dy. Therefore, we can write equation above as

$$F_{visc} = 2\pi\, yL\left(\mu \frac{du}{dy}\right)$$

Next we shall find the pressure difference.

Note that the force exerted by the pressure at an end of the cylinder is pressure at that end multiplied by the cross sectional area. Now if P_1 and P_2 respectively denote the pressures at either end of length L of the cylinder considered, then the required pressure difference is

$$F(P) = \pi y^2 (P_1 - P_2)$$

Substituting for F(P) and F_{visc} in equation mentioned above, we get

$$\pi y^2 (P_1 - P_2) = -2\pi yL\left(\mu \frac{du}{dy}\right)$$

or

$$y(P_1 - P_2) = -2L\mu \frac{du}{dy}$$

This gives the velocity gradient du/dy as

$$\frac{du}{dy} = \frac{-y(P_1 - P_2)}{2L\mu}$$

(the negative sign here implies u decreases when y increases. Also, note $P_1 > P_2$).

Now substituting this value of the velocity gradient in equation shown above, we get the shear stress as

$$\tau = \mu(du/dy)$$
$$= \frac{\mu \times (-y) \times (P_1 - P_2)}{2L\mu}$$
$$= \frac{-y(P_1 - P_2)}{2L}$$

Now if we consider the wall of the tube, then the radius y of the wall is R, therefore from equation above, we get

Shear stress at the wall of the tube $= \dfrac{-R(P_1 - P_2)}{2L}$

Thus, we have derived the equation describing the flow of fluid in a thin tube of length L with pressures P_1 and P_2 at the ends.

Now we have to solve equation for the velocity u.

Since the equation is a first order linear ordinary differential equation. To find the solution we integrate on both sides of the equation and we get the velocity as

$$u(y) = -\frac{(P_1 - P_2)}{4\mu L} y^2 + C$$

where C is the constant of integration which is to be evaluated. To evaluate C, it is necessary to prescribe the boundary conditions. Here, we make use of the assumption made by Newton that the fluid in contact with the wall is at rest,
i.e,

u = 0, when y = R

Substituting the condition in equation, we get

$$C = -\frac{(P_1 - P_2)}{4\mu L} R^2$$

So that equation reduces to

$$u(y) = -y^2 \frac{(P_1 - P_2)}{4\mu L} + R^2 \frac{(P_1 - P_2)}{4\mu L}$$

$$= (R^2 - y^2)\frac{(P_1 - P_2)}{4\mu L}$$

where u is the velocity component parallel to the axis, R is the radius of the cylinder, L is the length of the tube, μ is the viscosity of the fluid and $P_1 - P_2$, the pressure drop.

Therefore, the above equation describes the velocity of the fluid in a steady laminar flow.

Now, let us see what this equation represents geometrically? Since equation it is an equation of a parabola where u = 0 when y = R and u is maximum when y = 0 i.e. at the axis of the tube as shown in Figure below.

Our boundary conditions say that the velocity is zero at the wall. If the principle of conservation of mass is to hold well, the same amount of fluid should come out of every cross-section. The loss of velocity at the wall has to be compensated by maximum velocity at the centre.

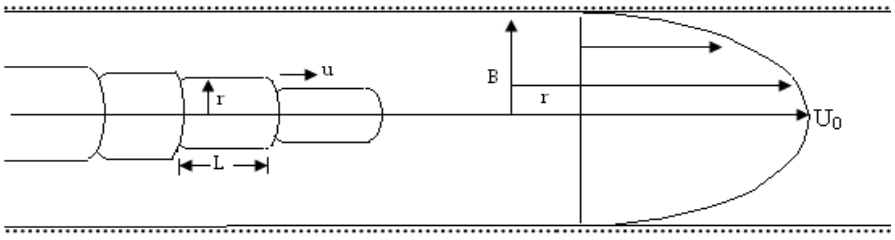


Figure: The parabola shows the velocity profile in a steady laminar flow

Thus, we find that the velocity distribution in a tube, with given pressure gradient is parabolic.
So, we have got an equation, which gives velocity distribution in a tube.
Let us now find the volume of fluid, flowing through a section of the tube per unit time. Here we shall see how we will use equation along its axis. That is, we have to determine the volume of the solid of revolution of parabola.

The required volume V = Volume of parabola of revolution

Then

$$V = \int_0^{2\pi} \int_0^R u(y) y\,dy\,d\theta$$

Now substituting for u in the above integral, we get

$$V = \frac{2\pi(P_1 - P_2)}{4L\mu} \int_0^R \left(R^2 - y^2\right) y\,dy$$

$$= \frac{(P_1 - P_2)\pi R^4}{8L\mu}$$

i.e., $$V = \frac{(P_1 - P_2)\pi R^4}{8L\mu}$$

This is called Poiseuille's law and it says that the volume is proportional to the first power of the pressure drop per unit length, $(P_1 - P_2)/L$, and to the fourth power of the radius of the pipe R^4 and it is inversely proportional to the length of the tube as well as the viscosity of the fluid. This equation is a general solution for any problem of fluid flow through cylindrical pipes, provided that the fluid flow satisfies all the assumptions, which are assumed in obtaining equation above. The assumptions made are as follows:

1) The fluid is homogeneous and its viscosity is the same at all rates of shear.
2) The fluid does not slip at the wall of the tube. This was the assumption that u = 0 when y = R which was made in evaluating the constant of integration.
3) The flow is laminar, i.e. the fluid is flowing parallel to the axis of inner surface wall of the tube.
4) The rate of flow is steady.
5) The tube is along with length much greater than the diameter of the tube.

Note: What are the units of quantities given in Poiseuille's equation? R and L are in cm, P = dynes/cm^2 and is a Poise.

Exercise: Consider the flow of fluid due to pressure gradient in a tube of radius R and length L. Find the bounds for velocity distribution.

Solution: If μ is the viscosity of the fluid and P_1 and P_2 fluid pressures at the ends of the tube, then we have, from the equation

$$u(y) = \frac{(P_1 - P_2)}{4\mu L}\left(R^2 - y^2\right),\ 0 \le y \le R$$

$$= \frac{P}{4\mu L}\left(R^2 - y^2\right),\ \text{where } P = P_1 - P_2$$

At $y = 0,\ u(0) = \frac{PR^2}{4\mu L}$ and at $y = R,\ u(R) = 0$

Therefore, $0 \le u(y) \le \frac{PR^2}{4\mu L}$

Formulation: In the previous section we have used Poiseuille's law, which gives a relation between the rate of flow and the pressure difference existing while a fluid flows in a rigid tube of circular cross section.

Let us now formulate a simple mathematical model for blood flow in arteries. Since the real situation is quite complex and including all the essential characteristics will make the model very complicated, let us make certain assumptions:

i) Blood is a homogeneous fluid
ii) The flow is steady and laminar
iii) The tube is rigid, long and straight

With these assumptions, the formulation leads to the steady flow of blood in a long rigid blood vessel for which Poiseuille's law is applicable. Thus, the velocity of blood in such a configuration corresponding to this simple model is

$$u(y) = \frac{(P_1 - P_2)}{4\mu L}\left(R^2 - y^2\right),\ 0 \le y \le R$$

and the rate of flow is

$$V = \frac{\pi R^4 (P_1 - P_2)}{8\mu L}$$

We can calculate the shear stress here by using the formula

$$\tau = \mu \frac{du}{dy} = -\frac{y(P_1 - P_2)}{2L}$$

Correspondingly the shear stress on the wall i.e. $y = R$ is

$$\tau = -\frac{R(P_1 - P_2)}{2L}$$

$P = P_1 - P_2$ **Exercise:** For a blood vessel of constant radius R, length l and driving force show that the average velocity of flow is equal to half the maximum velocity and the resistance (ratio of driving force to the flow.

i.e. $\frac{P_1 - P_2}{V}$) is proportional to $\frac{l}{R^4}$.

We have from equation,

$$u(y) = \frac{(P_1 - P_2)}{4\mu L}(R^2 - y^2)$$

$$= \frac{P}{4\mu l}(R^2 - y^2); \quad 0 \le y \le R \quad \text{where } P = P_1 - P_2$$

Also, from equation above, we have

$$V = \frac{\pi R^4}{8\mu l}(P_1 - P_2)$$

Average velocity of blood in a vessel =

$$\frac{V}{\pi R^2} = \frac{R^2}{8\mu l} P = K \frac{R^2}{2}$$

where $K = \frac{P}{4\mu l}$

Also, maximum velocity

$$u_m = u|_{y=0} = \frac{P_1 - P_2}{4\mu L} R^2 = KR^2$$

From the above equations, Average velocity of blood in a vessel

$$= \frac{1}{2} \text{ maximum velocity}$$

Resistance to the flow = $\dfrac{P_1 - P_2}{V}$

$$\frac{8\mu L}{\pi R^4} = C\frac{L}{R^4}, \quad \text{where } C = \frac{8\mu}{\pi} \text{ is constant.}$$

This shows that the resistance to the flow is directly proportional to $\frac{L}{R^4}$.

Exercises:

a) List some objectives of modelling blood flow.
b) Find the dimension of viscosity.
c) Consider arterial blood viscosity $\mu = 0.027$ poise. If the length of the artery is 2cm
d) (wide arterial capillary), and radius 8×10^{-3} cm and $P = P_1 - P_2 = 4\times10^3$ dynes/cm^2, then
e) find u(y) and the maximum peak velocity of blood
f) find the shear stress at the wall

Interpretation and Limitations of the model:

Let us consider the equation described above for the flow. The terms appearing as L and $P_1 - P_2$ are suitable, but the only unexpected term is the geometric term R^4 (fourth power of the radius of the tube), which is important in physiological flows. This would enable the effective control of blood flow by arteries. The blood vessels and the radius of the lumen go on decreasing towards the peripheral regions. For a given pressure and a fluid of constant viscosity a decrease to half the radius in the decrease of flow to sixteen of the original value. In turn it enables regulation of normal metabolic activities of the body.

If $u = \left(V/\pi R^2\right)$ is the mean velocity over the cross section, then from equation, we have

$$P_1 - P_2 = \frac{8\mu L}{R^2}u$$

Poiseulle's equation is one of the useful formulas commonly used in Laboratories for finding the viscosity of a given fluid. The method just involves in the measurement of the rate of flow and of the pressure drop across a fixed portion of a tube of given radius.

Next we shall discuss some limitations of the model. The assumptions made here are not satisfied completely. This will be clear from the following list of limitations:

i) The assumption that "flow is steady" is not satisfied always. Since the flow in all large arteries and the intra-thoracic veins is shown to be pulsative, and hence time dependent the flow is not steady in these situations. Therefore, we cannot directly apply the method in such situations. This is a major assumption and should be got rid of in any realistic model.

ii) Blood is treated as homogeneous and obeys Newton's law of friction. But it has been observed that this law can be used only if the flow occurs in tubes in which the internal diameter is large as compared with the size of the red cells. In smaller vessels, the cell structure should be included.

iii) The next assumption is flow is laminar. This is not always true. At rate of flow above a critical value, which has been calculated, the flow is turbulent. The turbulent flow may occur in very large blood vessels, especially near the heart and aorta. Flow here is also pulsative, the flow will not be steady so that assumption 4 will not be satisfied. This is not a serious limitation since the studies on the turbulent behavior of flow of blood in circulation shows that laminar flow occurs in almost all vessels where flow is sufficiently steady.

iv) The tubes are assumed to be long. But in the circulating system, where there are frequent branching's, no tube is sufficiently long enough for the model to be valid.

Though the assumptions and limitations are serious, still, this model is the basis of all other models for the study of blood flow. Therefore, understanding this model is very important for the study of blood flow.

This study of modeling flow is quite advanced now. The recent models have tried to remove the limitations by including elasticity of the vessel, pulsatility of the flow near the branching and the suspensions blood.

Pulse Wave:

A pulse wave is the progressive expansion of the arteries occuring with each contraction of the left ventricle of the heart.

The blood flow in the human body is mostly caused by the deformation of the vessels and other flow conduits. The interaction between the forces acting on fluid and vessel walls often leads to a non-linear pressure drop, wave-flow phenomena and the complex instabilities in the circulation.

Consider an unsteady, laminar blood flow through the thoracic aorta segment in human. Figure 1 shows the considered flow diagram. The model is formatted by taking into account the following assumptions:

- blood vessel has the shape of a circular cylinder with changing cross sectional area S = S(s, t), constant wall thickness b and fixed length L,
- the vessel wall undergoes substantial elastic deformations in the radial direction only,
- in a vessel filled with blood pressure p(s, t) and velocity u = u(s, t) wave propagation occur,
- the blood filling the vessel is a Newtonian fluid demonstrating a constant viscosity and a slight compressibility.

Given the above assumptions, the considered blood flow in the thoracic aorta is described by the following equations:

$$\frac{du}{dt} + \frac{1}{\rho}\frac{\partial p}{\partial s} + g\frac{\partial z}{\partial s} + \frac{\lambda}{2D}u|u| = 0$$

$$\alpha^2 \frac{du}{ds} + \frac{1}{\rho}\frac{\partial p}{\partial t} = 0$$

The equations result from the analysis of the overall mass and momentum balance equations for blood and from the equilibrium equations for an elastic blood vessel wall material. The symbols in above equations mean:

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial s}, \frac{dp}{dt} = \frac{\partial p}{\partial t} + p\frac{\partial p}{\partial s} \text{ and } a = \frac{\sqrt{\frac{K}{\rho}}}{\sqrt{1+\frac{K}{E}\frac{D}{b}C}}$$

For the thick blood vessel, constrained at both ends, the parameter C takes the form

$$C = \frac{1}{1+\frac{b}{D}}\left[(1-\mu^2) + 2\frac{b}{D}(1+\mu)(1+\frac{b}{D})\right]$$

where μ is the Poisson's ratio for the material of the arterial wall.

To solve this system of equations the method of characteristics is most commonly used procedure. The main idea behind the method of characteristics is an effective reduction of a system of partial differential equations to an equivalent system of ordinary differential equations.

$$\frac{du}{dt} + \frac{g}{a}\frac{dH}{dt} - \frac{g}{a}u\frac{dz}{ds} + \frac{\lambda}{2D}u|u| = 0 \qquad \ldots(a)$$

$$\frac{du}{dt} - \frac{g}{a}\frac{dH}{dt} + \frac{g}{a}u\frac{dz}{ds} + \frac{\lambda}{2D}u|u| = 0 \qquad \ldots(b)$$

The first equation of the above is valid for a characteristic C^+, described by the relationship $\frac{ds}{dt} = u + a$ and the second equation is valid along the characteristic C^-: $\frac{ds}{dt} = u - a$.

From the equations above, the value H(s, t) is the piezometric height of the liquid column expressing the pressure changes:

p □ □g□H □ z□

From the above characteristic curves, it is apparent that the slope of the lines of characteristics depends on the fluid velocity $u(s,t)$.

The solution of the equation (a) and (b) can be established using numerical methods.

***

Unit-III

Tracers in Physiological Systems

Introduction:

A living organism is not isolated from its environment but is rather in continual interactions with it. First and foremost, it requires energy to maintain itself. Thus, it is continually obtaining energy from its environment and excreting waste products to the environment. The energy is in the form of radiation for plants, and oxidative nutrients for animals. This process requires the passage of matter across the boundary between the organism and its environment, as well as the passage of matter across boundaries contained within the organism. Because of this exchange, a living organism is said to be an open system. The interaction of a living organism with its environment can be characterized approximately as a steady state, in which there are constant interchanges of energy and matter within the environment.

At the same time, organisms are unstable in the sense that they produce large response to relatively insignificant alterations in their environment, as, for example, the response of an owl to a field mouse observed in its peripheral vision. However, the higher multi-cellular animals, when viewed internally are very stable. They attempt to maintain their process and characteristics such as temperature and blood sugar level remain constant within narrow limits, despite small disturbances. The maintenance of the many internal steady state processes that characterize the higher animals is called homeostasis.

Compartment models:

Compartment models are often used to describe transport of material in biological systems. A compartment model contains a number of compartments, each containing well-mixed material. Compartments exchange material with each other following certain rules. Figure below shows a sketch of such a system. In this figure, boxes represent compartments and arrows represent the connections between the compartments. Every compartment (that is every box) has a number of connections leading to the box (inflows) and a number of arrows leading from the box (outflows). Material can either flow from one compartment to another, it can be added from the outside through a source, or it can be removed through a drain or a sink. Think of a bathtub, where water (the

well-mixed material) is added through the faucet and leaves through the drain. In the example above, the material was water, but it can be used in a more abstract way. Generally, the material represents the amount of something that we wish to account for. To account for the material the model must fulfill some conservation law. In the bathtub example, we could develop a model based on conservation of mass. Most compartment models (as the one shown in Figure below) have more than one compartment and equations for such a model are obtained by describing a conservation law for each compartment. Conservation laws state that the difference between what flows in and what flows out amounts to how much will be stored in. the compartment.

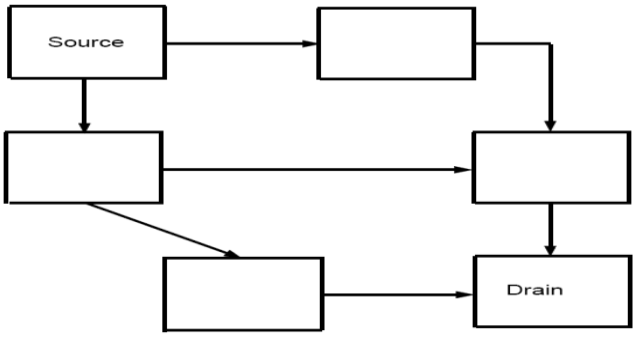


Figure: Sketch of a compartment model. Material can be stored in the boxes and transported between boxes following arrows.

A compartment model could also represent an ecological system where the material could be energy, the compartments could represent different species of animals and plants, and the flow between compartments could account for uptake and loss of food (or energy). In this case we would base the equations on laws describing conservation of energy. Compartment models also arise in physiology, where the material could be oxygen that is transported with the blood between different organs (compartments) in the body.

It should be emphasized, that one couldn't think of compartments and the flows in and out of compartments as individual components where each part can be described independently of each other. Both the in- and the outflow from any compartment may depend on the volume inside the compartment. Similarly, the inflow into a compartment may depend on the outflow from another compartment. In other words, it is important to think of the system as a whole,

where the parameter representing the material in the compartment (the state-variable) can dependent on what flows in and what flows out. In addition, since what flows into one compartment typically flows out of another compartment, the state variables depend on each other and on the state of the system as a whole.

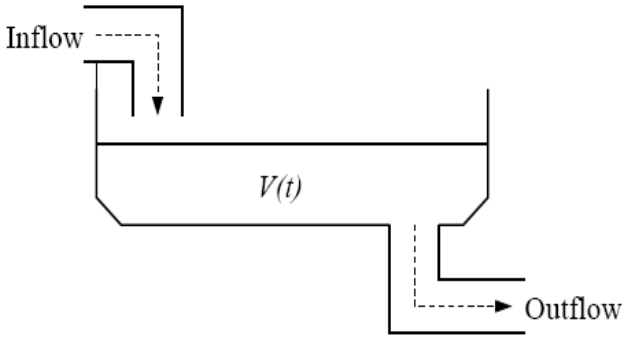
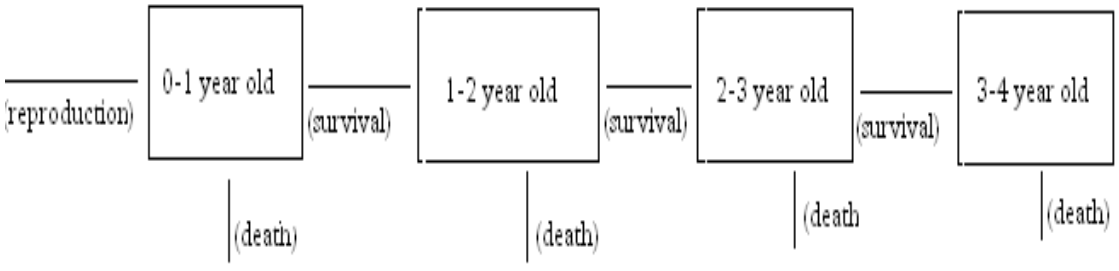


Figure: A bathtub with an external source (the faucet) and sink (the drain).

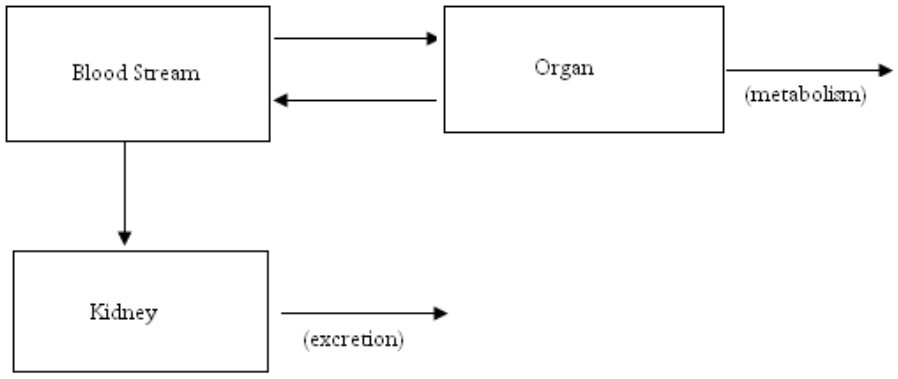
The important point to remember is that it is the person modeling the system who chooses how the model parameters and variables, in a consistent way, depend on each other.

In constructing a compartment model of a physical system, we first conceptually separate the system into a number of distinct compartments or states between which 'material' is transported. It is not necessary that the compartments be spatially distinct, but a clear set of rules must be given from which we can determine whether a given part of the system is contained in a compartment.

Example 1: The members of a population are divided into age classes and as time goes on, they either move from age class or pass out of the system entirely. A model that might be appropriate for a fish population is illustrated in figure below.



Example 2: A radioactive tracer (such as iodine 131) is injected into the blood stream. Some of the tracer is metabolized by a particular organ, (e.g. the thyroid gland), while the rest eventually makes its way to the kidneys, where it is excreted.



The model shown below in figure might be suitable. We can artificially create a fourth compartment labeled losses to keep track of the amount of tracer lost to the system.

Example 3: Strontium-90 is deposited into pastureland by rainfall. To study how this material is cycled through the ecosystem, we might divide the system into the five compartments atmosphere grasses soil, dead organic matter and stream (Figure). The model of passing of material from one compartment to another is also known.

Example 4: It is not necessary that the material being transported be a physical substance. As an example, suppose that we classify the weather in a certain locale as 'rainy', 'overcast' or 'clear'. These are the three states of the system. we next keep track of the probability that the system is in a particular state at time t. Thus at a given moment the total probability of 1 will be distributed among the three compartments (Figure).

On the arrow from compartment 1 to 2, for example, we place the conditional probability that it will rain tomorrow given that today is clear. Such probability modes are called Markov chains.

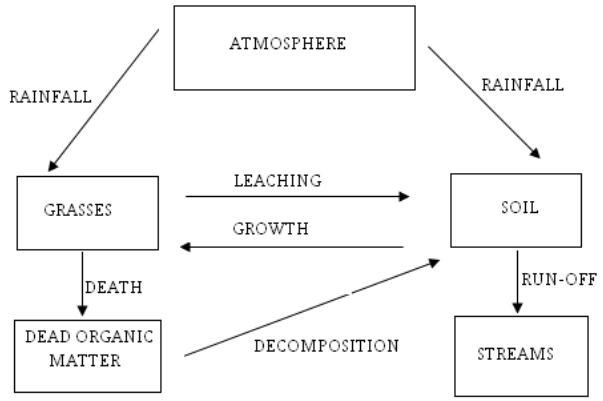


Figure: Strontium 90 cycling in pastureland

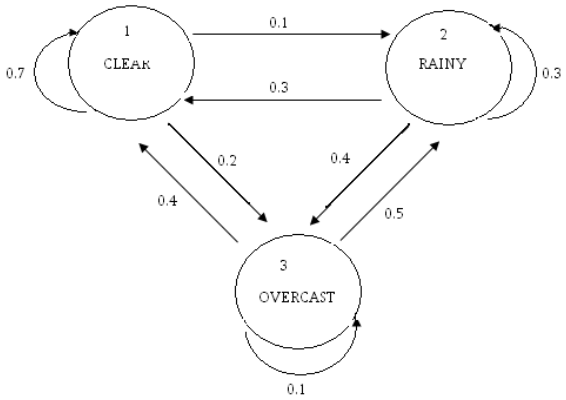


Figure : Markov process for weather forecasting

Transfer coefficients

Although the above examples show how varied the interpretations of compartment and material can be as an aid in deriving the equation, let us imagine that a physical substance is being moved from box to box as shown in the Figure.

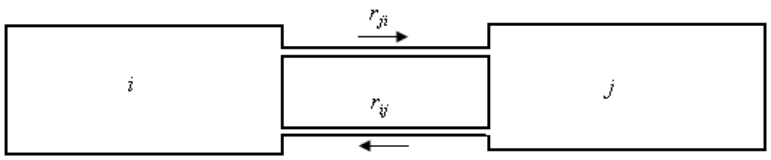


Figure: Fluxes between two compartments

Let $x_i(t) =$ amount of material in compartment i at time t and form the vector $X(t) = [x_1(t), x_2(t), \cdots, x_n(t)]$. We will regard $X(t)$ as either a row vector or a column

vector. $X(t)=[x_1(t),x_2(t),\cdots,x_n(t)]^T$, whichever is more convenient. $X(t)$ then specifies the state of the system at time t.

Let $r_{ij}(t)$ be the rate at which material passes to i from j at a particular time t. Typical units for r_{ij} might be grams/minute. The rate function r_{ij} is also called the flux to compartment i from j . r_{ij} may be a function not only of t but also of the states $x_1,x_2,\cdots,x_n$, so it is convenient to write $r_{ij}(t)$ to be the function $r_{ij}(x_1,x_2,\cdots,x_n,t)$.

Finally, $a_{ij}(t)=r_{ij}(t)/x_j(t)$, a_{ij} is called the transfer coefficient to i from j and is a percentage rate of the change. Typical units for a_{ij} might be (minutes)$^{-1}$. Thus, if $a_{12}=0.4$ per minute, then at a given moment, material passes from 2 to 1 at a rate $0.4x_2(t)$ gram per minute.

The assumptions that $a_{ij}(t)=a_{ij}$ is constant for all i and j , is known as the linear donor-controlled hypothesis. Thus $r_{ij}(t)=a_{ij}.x_j(t)$ and the rate at which material passes to i from j is determined only by the donor $x_j(t)$. Although rather simplistic in nature, this assumption is quite common. In many cases there is simply not enough known data about the system to make alternate and more realistic assumptions. In any case it is sound from pedagogical point of view to start with these basic models and gradually increase the complexity.

Definition: If $a_{ik}=0$ for all $i\neq k$, then compartment k is called a sink for the system. If $a_{kj}=0$ for all $j\neq k$, we call compartment k a source for the system.

Thus, in above example-3, the atmosphere is a source while as the compartment 'streams' is a sink. One can always create a sink for a compartment model by adding an (n + 1)st compartment called losses which collects all materials leaving the system.

Discrete Transfers

There are two different approaches to answer how can $X(t)$ be found given the transfer coefficient a_{ij} and the initial state $X(0)$. The solution depends upon whether the material is continuously transported between the compartments or is

interchanged only at a discrete set of times $t_1 = \Delta(t)$, $t_2 = 2\Delta(t)$, . . . , $t_n = n\Delta(t)$, ...

If a_{ij} is the transfer coefficient and interchanges are made every $\Delta(t)$ units, then at time $t_n + \Delta(t)$, $f_{ij} = a_{ij}\Delta(t)$ is the fraction of compartment that is passed to compartment i.

Example: A trust fund contains $\$5000$ and transfers are to be made to a second account at the rate of 12% per year (Figure). Thus, $a_{21} = 0.12$

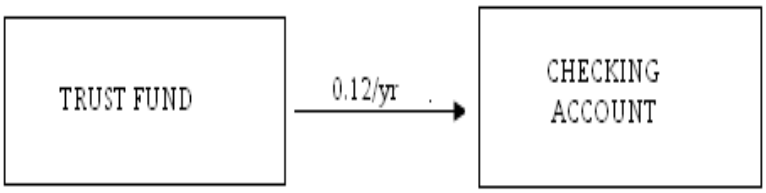


Figure: Transfers into a checking account

If transfers are made monthly ($\Delta(t) = 1/12$), then $a_{21}\Delta(t) = 0.01$ or 1% is transferred each month. Thus, $x_1(t + \Delta(t)) = x_1(t) - 0.01x_1(t) = 0.99x_1(t)$. If transfers are made daily (i.e. $\Delta(t) = 1/365$), then $a_{21}\Delta(t) = 0.12/365$ or 0.033% is transferred each day.

To develop the matrix equation relating $X(t + \Delta(t))$ to $X(t)$,

Note that

$$x_i(t + \Delta(t)) = x_i(t) + \text{(amount entering i) –(amount leaving i)}$$

$$= x_i(t) + \sum_{j \neq i} f_{ij} x_j(t) - \sum_{j \neq i} f_{ij} x_i(t)$$

$$= x_i(t) + \sum_{j \neq i} \left(a_{ij}\Delta(t)\right) x_j(t) - \left(\sum_{j \neq i} a_{ij} \right) \Delta(t) x_i(t)$$

If we define $a_{ii} = -\sum_{j \neq i} a_{ji}$, we have

$$x_i(t + \Delta(t)) = x_i(t) + \Delta(t) \sum_{k=1}^{n} a_{ik} x_k(t).$$

Letting $A = \lfloor a_{ij} \rfloor$, we have the matrix form of the above as

$$X(t + \Delta(t)) = X(t) + \Delta(t)AX(t)$$

or $$X(t + \Delta(t)) = [I + \Delta(t)A]X(t)$$

where I is the n x n identity matrix. Note that the sum of the entries in any column of the matrix A is 0 since

$$a_{ii} + \sum_{j \neq i} a_{ji} = 0$$

Such a matrix is called ecomatrix.

Exercise: In the compartment diagram given below, the transfers are made on daily basis

(a) Find the transfer matrix A

(b) If X(0) = [100, 0, 0], find the state of the system over the next five days.

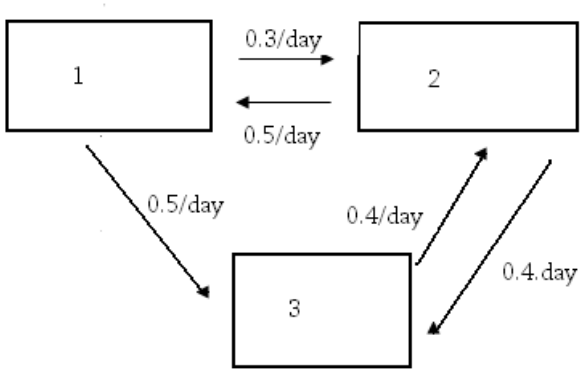


Figure: Compartment diagram

Solution: we are given that $a_{21} = 0.3$, $a_{12} = 0.5$, $a_{32} = 0.4$, $a_{23} = 0.4$, $a_{13} = 0$ and $a_{31} = 0.5$. Since the sum of the entries in any column must be zero, therefore

$$A = \begin{bmatrix} -0.8 & 0.5 & 0 \\ 0.3 & -0.9 & 0.4 \\ 0.5 & 0.4 & -0.4 \end{bmatrix}$$

and for $\Delta(t) = 1$,

$$I + \Delta(t)A = I + A = \begin{bmatrix} 0.2 & 0.5 & 0 \\ 0.3 & 0.1 & 0.4 \\ 0.5 & 0.4 & 0.6 \end{bmatrix}$$

Hence $X(1) = (I + A)X(0) = [20,\ 30,\ 50]$

$X(2) = (I + A)X(1) = [19,\ 29,\ 52]$.

Continuing with the recursion $X(t+1) = (I + A)X(t)$, we can obtain

$X(3) = [18.3,\ 29.4,\ 52.3]$,

$X(4) = [18.36,\ 29.35,\ 52.29]$

and $X(5) = [18.35,\ 29.36,\ 52.29]$

If we assume $B = I + \Delta(t)A$, then $X(t + \Delta(t)) = BX(t)$ and $X(\Delta(t)) = BX(0)$,

$$X(2\Delta(t)) = BX(\Delta t) = B(BX(0)) = B^2 X(0)$$

and in general

$$X(n\Delta(t)) = B^n X(0)$$

Hence $X(n\Delta t) = (I + \Delta tA)^n X(0)$ is the general solution for the discrete case.

Continuous transfers:

By letting $\Delta(t) \to 0$ in the discrete case, we can derive the matrix differential equation for $X(t)$, when the material flows continuously between the compartments.

For Δt very small $X(t + \Delta t) \approx X(t) + \Delta tAX(t)$ or $\frac{X(t + \Delta t) - X(t)}{\Delta t} \approx AX(t)$. Now letting $\Delta(t) \to 0$, we obtain the system of differential equations

$$\dot{X}(t) = AX(t)$$

where $\dot{X}(t) = [\dot{x}_1(t),\ \dot{x}_2(t), \cdots, \dot{x}_n(t)]$

Example: Water flows continuously between two 100 gallon tanks at the rate of 10 gallons/minute as shown in figure.

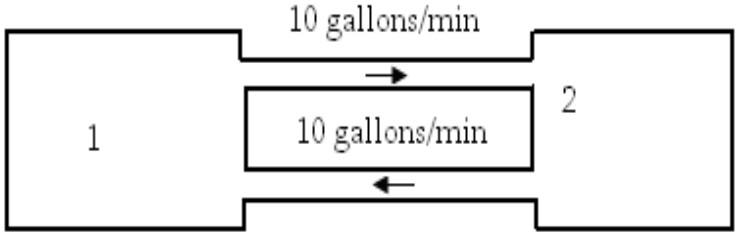


Figure: Water flow between two tanks

Twenty pounds of salt are mixed in tank 1 at time t = 0. Since the concentration of salt in tank 1 at time t $x_1(t)/100 lb/gallon$, Salt moves to tank 2 at the rate

$$r_{21} = (x_1(t)/100 \text{ lb/ gallon})(10 \text{ gallons/min})$$

$$= 0.1\, x_1(t) \text{ lb/min}$$

Take $a_{21} = 0.1$. Likewise $a_{12} = 0.1$ and so

$$A = \begin{bmatrix} -0.1 & 0.1 \\ 0.1 & -0.1 \end{bmatrix}$$

Hence $$\dot{x}_1 = -0.1x_1 + 0.1x_2$$

$$\dot{x}_2 = 0.1x_1 - 0.1x_2$$

with $x_1(0) = 20$ and $x_2(0) = 0$.

The system can be solved for $x_1(t)$.

$$\ddot{x}_1 = -0.1\dot{x}_1 + 0.1\dot{x}_2$$

$$= -0.1\dot{x}_1 + 0.1(0.1x_1 - 0.1x_2) \quad \text{using the second equation.}$$

$$= -0.1\dot{x}_1 + 0.01x_1 - 0.1(0.1x_2)$$

$$= -0.1\dot{x}_1 + 0.01x_1 - 0.1(\dot{x}_1 + 0.1x_1) \quad \text{using the first equation}$$

$$= -0.2\dot{x}_1$$

Thus $x_1(t) = c_1 + c_2 e^{-0.2t}$ and $x_2(t) = 10[\dot{x}_1(t) + 0.1x_1(t)] = c_1 - c_2 e^{-0.2t}$.

Since $x_1(0)=20$ and $x_2(0)=0$, it follows that $c_1=10$ and $c_2=10$.

Thus $\lim_{t\to\infty} x_1(t)=\lim_{t\to\infty} x_2(t)=10$, as expected

Remark: To motivate the general form for the solution of $\dot{X}(t)=AX(t)$, fix t > 0 and let $\Delta(t)=t/n$. Then for n very large, $X(t)=X(n\Delta t)$ should be approximated by $(I+\Delta tA)^n X(0)$. Thus,

$$X(t)=\left(1-\frac{1}{n}[tA]\right)^n X(0)+\varepsilon \text{ , where } \varepsilon\to 0 \text{ as } n\to+\infty$$

Recall that for x a real number, $\lim_{x\to\infty}\left(1+\frac{x}{n}\right)^n=e^x$. Our matrix expression resembles $\left(1+\frac{ta}{n}\right)^n$ which approaches e^{ta}. If we define a matrix e^B to be $\lim_{n\to\infty}\left(I+\frac{B}{n}\right)^n=e^B$, then assuming that the limit actually exists, we should have

$$X(t)=e^{tA}X(0)$$

Actually, it is easier to define e^{tA} by the matrix series

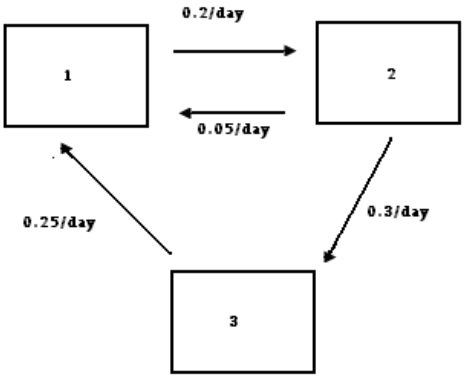
$$e^{tA}=I+tA+\frac{t^2}{2!}A^2+\cdots$$

and verify that $X(t)=e^{tA}X(0)$ is a solution by direct differentiation. The explicit solution $X(n\Delta t)=(I+\Delta tA)^n X(0)$ and $X(t)=e^{tA}X(0)$ arising from the discrete and continuous cases provide little insight into the long-range behaviour of the system.

Exercises:

(i) For the compartment diagram shown below

a) Find the transfer matrix A
b) If X(0) =[100, 250, 80], find X(1) if material is transferred once a day
c) Find X(1) if material is transferred hourly
d) Estimate X(1) if material is transferred continuously.



Introduction to Tracer Method in Physiology:

One of the earliest successful applications of mathematics to biology is the dye, dilution method for measuring cardiac output. This technique originated with the English physiologist George Stewart in 1890s and was refined by the American physiologist William Hamilton in the late 1920s. In such physiological applications, a dye or radioactive tracer is injected into the bloodstream. By monitoring the concentration of tracer in various parts of the organism, we can obtain enough information to compute the rates at which the tracer moves through the system. Closely related to the subject of pharmacokinetics, which dates back to the late 1930s and studies the rates at which drugs are distributed, metabolized, absorbed, and/or eliminated from the body.

Bath-Tub Models:

Compartment models are often called 'bath-tub' models. Suppose that we are given n bath-tubs with pipes connecting some of the tubs. As shown in figure below, tub (i) contains a constant volume of V_i liters and liquid flows to tub (j) from tub (i) at the rate of F_{ij} liters/minute. Also suppose that tracer (e.g. a dye or drug) is introduced into the system either through instantaneous injection or infusion via a driving function $I(t)=[I_1(t), I_2(t),\cdots, I_n(t)]$

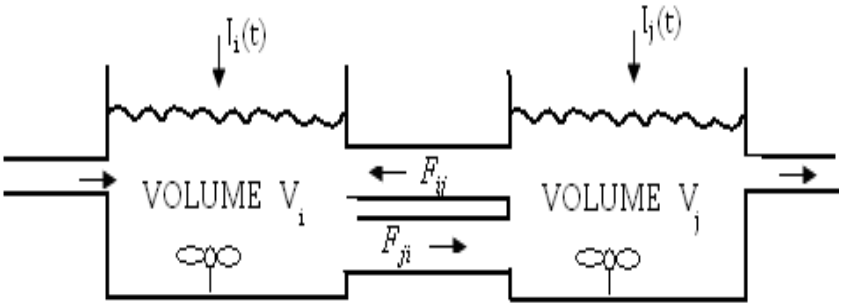


Figure: Bath-tub model

Let $x_i(t)$ = amount of tracer in tub (i) at time t. To insure that the concentration of tracer is uniform throughout the tub, we might imagine that blenders are placed in each tub to thoroughly mix its contents. Let $c_i(t)=x_i(t)/V_i$, the concentration in tub (i) at time t. The flux of tracer from tub (i) to tub (j) is then given by

$$r_{ji}(t) = F_{ji}(lit/\min)c_i(t)(mg/lit)$$
$$= (F_{ji}/V_i)x_i(t)(mg/\min)$$

Hence the transfer coefficient a_{ji} is given by F_{ji}/V_i. The dynamics of the systems are given by

$$\dot{X} = AX + I(t)$$

where $I(t) = [I_1(t), I_2(t), \cdots, I_n(t)]$ specifies the rate at which the tracer enters the system from the outside and $a_{ij} = F_{ij}/V_j$. We can create an additional compartment (usually labeled compartment (0)) that keeps track of the tracer that leaves the system.

Figure below is a three-tub system. In order that the volume of each tub remains constant, the flow rates into a tub must balance the flow rates out.

Thus $F_{13} + F_{12} = F_{21}$, $F_{21} = F_{12} + F_{32}$ and $F_{32} = F_{13}$. In general, we have the situation

$$\sum_{j \neq i} F_{ij} = \sum_{j \neq i} F_{ji}$$

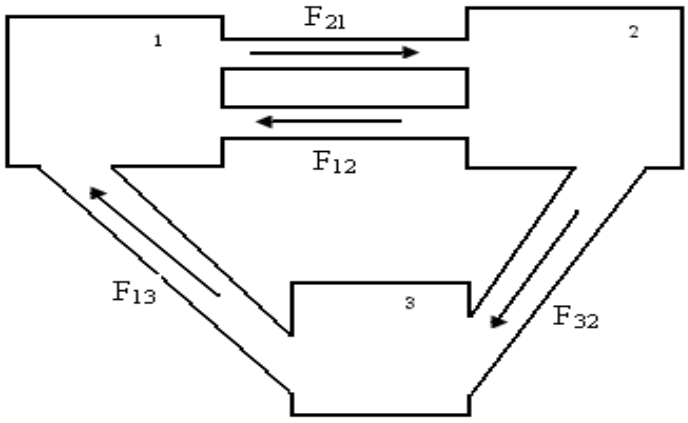


Figure: A three-tub system

Example: Shown in figure below is a single compartment model with driving function I(t). Hence $\dot{x} = -(F/V)x + I(t)$, it follows that $x(t) = ce^{-(F/V)t} + x_p(t)$, where x_p is the particular solution. For the case $I(t) \equiv 0$, then $x(t) = x_0 e^{-(F/V)t}$

and so the amount of tracer in the tank decays exponentially. The transfer coefficient F/V is often called the turnover rate.

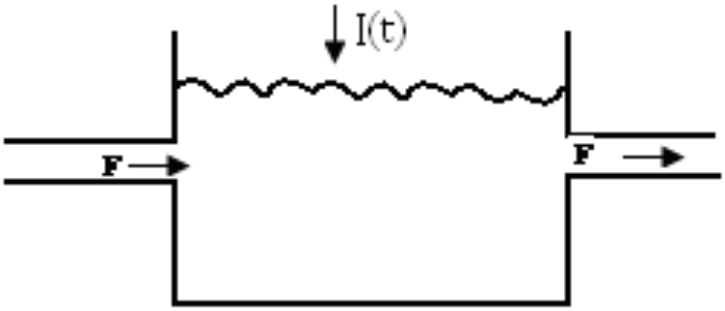


Figure: Single compartment model with infusion

This model has been used to study pollution in Great Lakes, by assuming the precipitation on a given lake equals evaporation and so the flow rate in and out of the lake is constant. In addition, we must take the rather unrealistic assumption that the pollutants are uniformly distributed throughout the lake. For example $V \approx 458\ km^3$ and $F \approx 175\ km^3 / year$, hence the turnover rate is F/V = 38.21% /year. If pollution were stopped completely, I(t) = 0 and so

$$x(t) = x_0 e^{-(0.3821)t}$$

To estimate the amount of time it would take to clear 90% of the waste, let x(t) = 0.1 x_0. This in turn gives $t \approx years$. On the other hand for another lake with $V \approx 12.221\ km^3$ and $F \approx 65.2\ km^3 / year$, has a turnover rate is 0.53% per year. It would take about 430 years to eliminate 90% of the waste.

Remark: The above example is typical of the majority of applications of compartment models in ecology. Flow rates and volumes are measured directly to compute transfer coefficients and make long range predictions about the state X(t) of the system. The situation in physiological applications is quite different. Here rates and volumes are in general unknown. In fact, the whole point of the model may be to estimate a particular turnover rate, flow rate, or volume in order to judge the health of the system. Most physiological processes, however, take place over a short enough time span that can collect sample concentration

$$c_i^*(t_1), c_i^*(t_2), \cdots, c_i^*(t_n)$$

from one or more compartments. From these partial observations of the system, we wish to estimate a_{ij} and then F_{ij} and V_j . This is the estimation of parameters problem. We might also want to measure the rate R at which a tracer is removed from the body by an organ. Thus some of the components $I_{ij}(t)$ in the driving function $I(t)$ may be unknown.

Stewart-Hamilton Method for measuring cardiac output:

Suppose that a known quantity x_0 of dye is rapidly injected into the tank shown in figure below

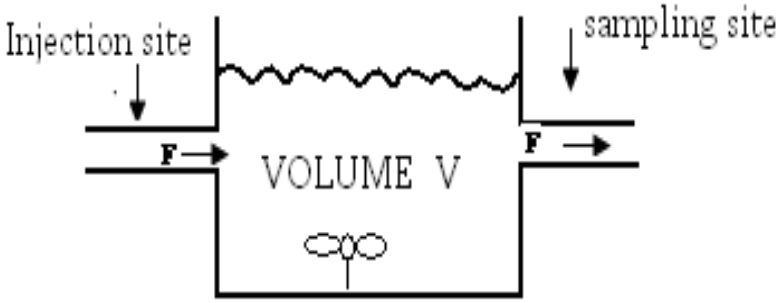


Figure: Injection and sampling sites in a one compartment model

We then estimate the concentration $c(t)$ of dye in the tank at subsequent times. The rate at which dye leaves the tank is given by $r(t) = Fc(t)$ and so the total amount of dye leaving the tank, namely x_0 is given by

$$x_0 = \int_0^\infty Fc(t)dt$$

(Actually there will be a time T when, practically speaking, all dye is cleared from the tank.) Solving for the unknown flow rate F, we have

$$F = \frac{x_0}{\int_0^\infty c(t)dt}$$

From sample concentrations $c^*(t_1), c^*(t_2), \cdots, c^*(t_n)$, we estimate the integral using a numerical integration procedure such as the trapezoid rule or Simpson's rule.

In applying this model to measure cardiac output, the left side of the heart can serve as the 'mixing chamber' and the bolus of dye can be injected in the pulmonary artery. Alternatively, the right atrium can serve as the mixing chamber with the dye placed in a nearby vein. The withdrawal or sampling site is a peripheral artery preferably near the aorta. A densitometer can make the withdrawals and dye concentration estimates as frequently as five times per second. Figure below is a typical concentration curve obtained with the aid of a densitometer.

Figure: A typical dye concentration curve

Unlike the actual cardiovascular system, our basic model assumes that once the dye has left the chamber it does not return. To minimize the effects of recirculation, it is important to select a dye that is rapidly removed from the circulatory system by an organ. In addition, because of the injection site, the dye cannot be removed or altered by the lungs. One such dye is indocyanine green (ICG), is metabolized by the liver.

Exercise: A 5mg bolus of ICG is injected into the right atrium. Concentration samples are then measured each second as reported in table below. Estimate the cardiac output in liters/minute.

t (sec)	0	1	2	3	4	5	6	7	8	9	10	11	12
c(t) (mg/l)	0	0	1.7	5.6	9.2	8.4	5.2	3.8	2.1	1.0	0.5	0.2	0

Solution: We are given that x_0 = 5mg and $\Delta t = 1$ sec. Applying Simpson's rule, we have

$$\int_0^{\infty} c(t)dt \approx \int_0^{12} c(t)dt \approx \frac{1}{3}[0 + 4(0) + 2(1.7) + 4(5.6) + 2(9.2) + 4(8.4) + 2(5.2) + 4(3.8) + 2(2.1) + 4(1.0) + 2(0.5) + 4(0.2) + 0]$$
$$= 37.8$$

Applying the formula of F, $F = 5/37.8$ liters per second = 7. 94 liters/minute. Such a cardiac output is considerably healthy.

The amount of dye in the mixing chamber is predicted to be $x(t) = x_0 e^{-(F/V)t}$ and so

$c(t) = (x_0 / V)e^{-(F/V)t}$. If we choose $C = \ln c(t)$, then $C = mt + b$, where $m = -F/V$ and $b = \ln(x_0 / V)$. Hence the relationship between $C = \ln c(t)$ and t is in theory linear. We can estimate the slope $m = -F/V$ from concentration measurements and linear regression formulae. In practice, because of the time lag in reaching the mixing chamber and incomplete mixing, the actual experimental curve shows exponential decay only after a certain time t_0, we therefore perform a linear regression on C and t only after this time.

Exercise: Estimate the central blood volume V for the data given in above example using the integral formula for V given by

$$V = \frac{F\int_0^{\infty} tc(t)dt}{\int_0^{\infty} c(t)dt}$$

Solution: We have already seen $\int_0^{\infty} c(t)dt \approx 37.8$ and $F \approx 0.1323$

Applying Simpson's formula to $\int_0^{\infty} tc(t)dt$ gives

$$\int_0^{\infty} tc(t)dt \approx \int_0^{12} tc(t)dt \approx 190.933$$

Hence from the integral, volume is $V \approx (0.1323)(190.933)/37.8 = 0.67$ liters.

Continuous Infusion into a Compartment:

As illustrated in figure below, tracer is continuously infused into the tub at the known rate of l mg/min. The amount of tracer in the tub now satisfies $\dot{x} = -\lambda x + l$ where $\lambda = F/V$.

It follows that $x(t) = ae^{-\lambda t} + (l/\lambda)$ and so the concentration is given by $c(t) = a'e^{-\lambda t} + l/(V\lambda)$,

where $a' = c(0) - l/(V\lambda)$. Hence $c_\infty = \lim_{t\to\infty} c(t) = l/(V\lambda)$ and so $a' = c(0) - c_\infty$.

i.e.,

$$c_\infty - c(t) = (c_\infty - c(0))e^{-\lambda t} \text{ and } c_\infty = l/(V\lambda) \text{ and } \lambda = F/V$$

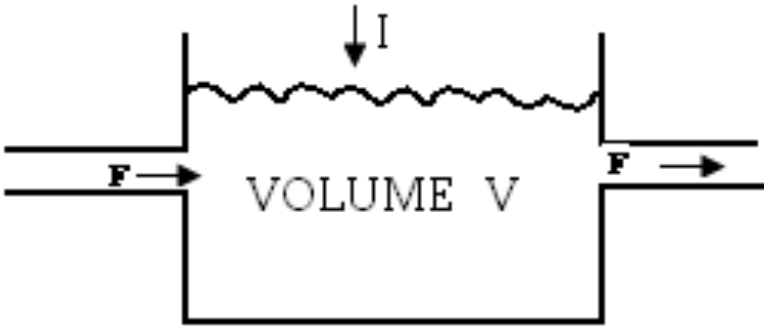


Figure: Constant infusion into a single compartment

An illustration of the model is provided by one of the standard glucose tolerance tests. To measure carbohydrate metabolism in a subject, glucose is infused into the blood stream at the known rate of $l\ mg/\min$. Blood sugar concentrations are measured at times 0, $t_1, t_2, \cdots, t_n$ until the steady state concentration $c_\infty = l/(V\lambda)$ mg is reached. With these measurements, the turnover rate $\lambda = F/V$ can be estimated as follows. If we choose $Y = \ln[c_\infty - c(t)]$, then $Y = \ln[c_\infty - c(0)] - \lambda t$. Hence if we find the regression line $Y = mt + b$, then

the turnover rate is just –m. The volume V can then be computed from the formula $V = l/(c_\infty \lambda)$.

Example: In preparation for the glucose tolerance test, a subject fasts for two days and begins the test with glucose concentration level c(0) = 85 mg/deciliter. Glucose is then infused intravenously at the rate of 300 mg/min. The steady state concentration is $c_\infty = 139.2$ mg/dl, and concentration samples are shown in table below.

Time t(min)	0	10	20	30	40	50
c(t) (mg/dl)	85.0	105.4	120.1	127.3	131.9	135.4
$Y = \ln(c_\infty - c)$	3.9927	3.5205	2.9497	2.4765	1.9879	1.5686

The regression of Y versus t is $Y = -0.0491\,t + 3.9773$. Hence the turnover rate is $\lambda = 0.049/\min$. and $V = l/(c_\infty \lambda) = 43.89$ deciliters of about 5.4 liters. This turnover rate is in the normal range.

Fick's Principle:

As illustrated in figure below, a dye of concentration c_1 flows into a tank and is removed at the constant rate of R mg/min.

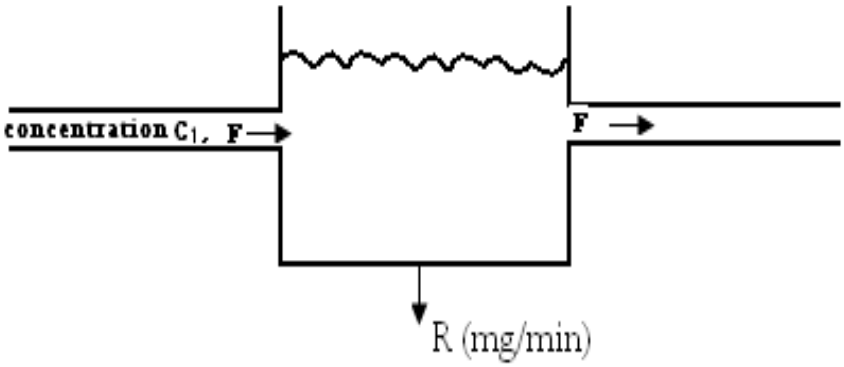


Figure: Constant removal rate from a single compartment

If $x(t)$ is the amount of dye in the tank at the time t, then

$$\dot{x} = -\frac{F}{V}x + (Fc_1 - R)$$

whose solution is of the form

$$x(t) = ce^{-(F/V)t} + \frac{c_1 F - R}{F} V$$

Hence $c_2 = \lim_{t\to\infty} c(t) = (c_1 F - R)/F$ and so we obtain Fick's principle

$$F = \frac{R}{c_1 - c_2}$$

This formulation can be used to measure flow rates through organs. Note however that the computation of F depends on our knowing the rate of removal R by the organ. Now in order to measure hepatic blood flow, if F_1 denote the flow rate through the liver and $F_1 + F_2$ the cardiac output, then, as shown in figure below, we have the dynamics

$$\dot{x}_1 = -(F_1 + F_2)c_1(t) + (F_1 + F_2)c_3(t) + I$$

$$\dot{x}_2 = F_1 c_1(t) - F_1 c_2(t) - R$$

$$\dot{x}_3 = F_2 c_1(t) + F_1 c_2(t) - (F_1 + F_2)c_3(t)$$

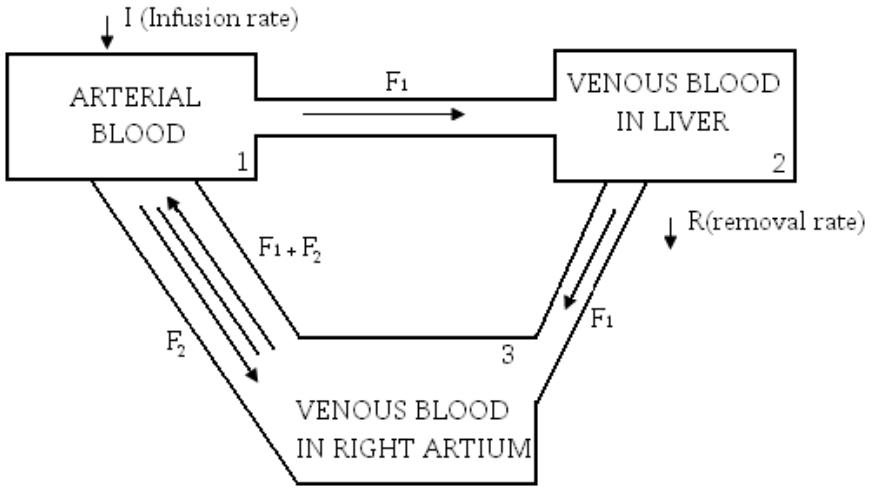


Figure: A compartment model for hepatic blood flow

Note that we have used the formula $\sum_{j\neq i} F_{ij} = \sum_{j\neq i} F_{ji}$ in labeling the flow rates.

The transfer matrix is of the form

$$A=\begin{bmatrix} -(a+b) & 0 & d \\ a & -c & 0 \\ b & c & -d \end{bmatrix}$$

and it is not hard to show that $\lambda = 0$ is an eigenvalue of multiplicity one while the other two eigenvalues both satisfy $\mathrm{Re}(\lambda) < 0$. Hence if $X_e(t)$ is the complementary solution, $\lim_{t\to\infty} X_e(t)$ exists. Therefore the long-range solution is dependent on the form of the particular solution $X_p(t)$. Since $\lambda = 0$ is an eigenvalue of multiplicity one, particular solutions are of the form $X_p(t) = C_1 + C_2(t)$. For $C_2 \neq 0$, no steady state solution will exist. We therefore need to determine conditions, which will insure that $\hat{X} = \lim_{t\to\infty} X(t)$ exists. If $\hat{X} = [\hat{x}_1, \hat{x}_2, \hat{x}_3]$ is a constant solution, let $\hat{c}_1 = \hat{x}_1 / V_1$, $\hat{c}_2 = \hat{x}_2 / V_2$ and $\hat{c}_3 = \hat{x}_3 / V_3$. It follows that

$$0 = -(F_1 + F_2)\hat{c}_1 + (F_1 + F_2)\hat{c}_3 + I$$

$$0 = F_1\hat{c}_1 - F_1\hat{c}_2 - R$$

$$0 = F_2\hat{c}_1 + F_1\hat{c}_2 - (F_1 + F_2)\hat{c}_3$$

Hence $I = (F_1 + F_2)\hat{c}_1 + (F_1 + F_2)\hat{c}_3 = (F_1 + F_2)(\hat{c}_1 - \hat{c}_3)$, $R = F_1\hat{c}_1 - F_1\hat{c}_2$ and $(F_1 + F_2)\hat{c}_3 = F_2\hat{c}_1 + F_1\hat{c}_2$.

It follows that $I = (F_1 + F_2)\hat{c}_1 - F_2\hat{c}_1 F_1\hat{c}_2 = F_1(\hat{c}_1 - \hat{c}_2) = R$. Hence we have concluded that $I = R$ is the necessary condition for $X = \lim X(t)$ to exist. Conversely, when $I = R$, it can be shown that the constant particular solution can be constructed.

When $I = R$, the hepatic blood flow rate can be computed from the Fick's formula

$$F_1 = \frac{I}{\hat{c}_1 - \hat{c}_2}$$

When $I \neq R$, the long-range solution will be of the form $C_1 + C_2 t$, with $C_2 \neq 0$. The long-range concentration levels will then change at a steady state.

To use the model, we use ICG, which is cleared from plasma, exclusively by the liver and converted into bile. A catheter is inserted into a vein in the arm. After an initial load dose of 12 mg, dye is infused at a constant rate of I = 0.25 mg/min. For a normally functioning liver, equilibrium should be reached in about 20 minutes. Minor adjustments in I may be needed to insure equilibrium has been reached. Note that if $I \neq R$, we should notice steady increases or decreases in concentration levels. Finally we measure concentrations from blood samples taken from an artery and a hepatic vein. For example if $\hat{c}_1$ (arterial blood concentration) = 0.7 mg/liter and $\hat{c}_2$ (hepatic blood concentration) = 0. 05 mg/liter, then

$F_1 = (0.5\ mg/\min)/(0.7-0.05)mg/liter = 0.77$ liter/min.

Elementary Pharmacokinetics:

Compartment models in pharmacy have proven extremely useful and surprisingly accurate in predicting drug concentration levels in organs and tissues and in estimating rates at which the drug is eliminated from the body. Important clinical applications have resulted from such models.

Example: A drug is quickly injected into a bloodstream and rapidly distributes itself throughout the body. We will assume that the drug does not collect in organs and tissues and is eliminated through urine only.

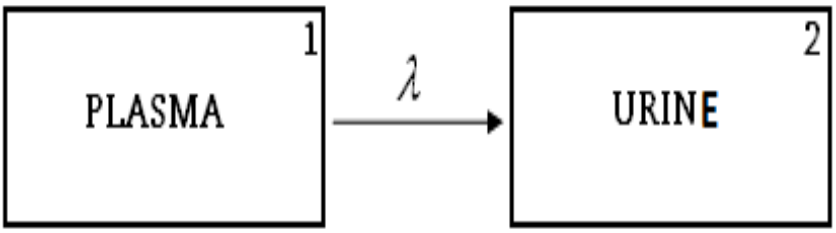


Figure: Drug elimination through urine

From Figure above, $x_1(t) = x_0 e^{-\lambda t}$ and the total amount of drug collected in the urine $x_2(t)$ satisfies

$$\dot{x}_2(t) = \lambda x_1(t) = \lambda x_0 e^{-\lambda t}$$

Hence assuming $x_2(0)=0,$ we have

$$x_2(t)=x_0\left(1-e^{-\lambda t}\right)$$

Letting $X=\ln\left(x_0-x_2(t)\right),$ we have $X=\ln x_0-\lambda t$. Hence, given urinary excretion data $x_2^*(t_1),\cdots,x_2^*(t_n)$, we can find the regression line of X versus t, and estimate the turnover rate λ. This technique is called sigma-minus method.

Example: (First Order Absorption): A fixed dose of x_0 mg of a drug is quickly injected into a bloodstream and makes its way to an organ where the drug is gradually destroyed or transformed. One such example is radioactive iodine-131 and the thyroid gland as shown in Figure below.

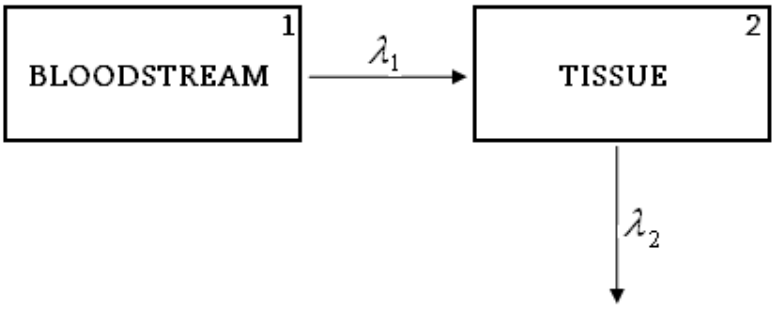


Figure: First order absorption

The drug leaves the bloodstream at the rate $I(t)=x_0\lambda_1 e^{-\lambda_1 t}$, the function serves as a driving function for the second compartment. Hence

$$\dot{x}_2=-\lambda_2 x_2+I(t)$$

and when $\lambda_1\neq\lambda_2$ a particular solution is given by

$$x_p(t)=\frac{x_0\lambda_1}{\lambda_2-\lambda_1}e^{-\lambda_1 t}$$

Assuming $x_2(0)=0,$ the final solution is given by

$$x_2(t)=\frac{-x_0\lambda_1}{\lambda_2-\lambda_1}\left[e^{-\lambda_2 t}-e^{-\lambda_1 t}\right]$$

In many cases too high a drug concentration level in the tissue can be extremely dangerous. For example, high concentration of Iodine-131 in the thyroid can produce cancerous tumors and destroy thyroid tissue. It is important then to find the maximum value of $x_2(t)$. Setting $\dot{x}_2(t) = 0$ in the expression of x_2, we can see that the maximum will occur at time

$$T = \frac{1}{\lambda_2 - \lambda_1} \ln(\lambda_2 / \lambda_1)$$

and the maximum value is $x_0(\lambda_2 / \lambda_1) e^{-\lambda_1 T}$. Setting $\dot{x}_2 = 0$ in the differential equation $\dot{x}_2 = -\lambda_2 x_2 + I(t)$, we have

For most drugs, λ_1 is significantly larger than λ_2. This has the important consequences that for large t,

$$c_2(t) = \frac{x_2(t)}{V_2} \approx \frac{-x_0 \lambda_1}{V_2(\lambda_2 - \lambda_1)} e^{-\lambda_2 (t)}$$

and so $c_2 = \ln c_2(t) \approx -\lambda_2 t + b$ for some constant b. Hence the key turnover rate can be found by collecting dye concentration data $c_2^*(t_1), c_2^*(t_2), \cdots, c_2^*(t_n)$ from the tissue and by finding the regression line of $c_2 = \ln c_2(t)$ versus t for large t. Thus the curve fitting technique, called the method of exponential peeling.

Suppose that our estimate for λ_2 is 0.04 and λ_1, estimated from plasma concentration samples, is about 0.10 per hour. The maximum amount of drug in the tissue occurs at time T = 15.27 hours. If the maximum amount of drug in the tissue should be no more than 20 mg, then the load dose x_0 must satisfy

$$x_0 \leq 20(\lambda_2 / \lambda_1) e^{\lambda_1 T} = 36.8\, mg$$

Parameter estimation in Two-Compartment Models:

A general principle in constructing a compartment model for a physical system is to select the model with the fewest number of compartments that adequately describes the data available on the system. It is quite remarkable that in physiological and pharmaceutical applications, models with two or three

compartments often do an excellent job in describing the kinetics. Figure shown below is a two-compartment model that occurs frequently in drug kinetics.

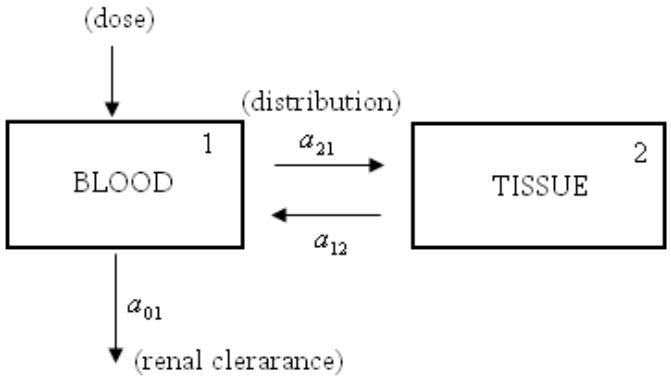


Figure: A common two-compartment model for tracer kinetics

In many cases, it is possible to obtain tracer concentration measurements from one compartment only (typically from blood to urine). We wish therefore to investigate under what conditions these measurements are sufficient to obtain estimates for the turnover rates a_{ij}, the volume V_j and the flow rates F_{ij}. As we will see, measurements from one compartment are rarely adequate when the number of compartments exceeds two. Although a four or five compartment model may appear to be reasonable in view of physiology, the model may be totally useless from clinical point of view unless the technology is present for sampling more compartments.

Shown in figure below is the bath-tub representation of the general two-compartment model. We will suppose that tracer makes its way into the system through the first compartment only so that $X(0)=[x_0,0]$ and $I(t)=[I_1(t),0]$.

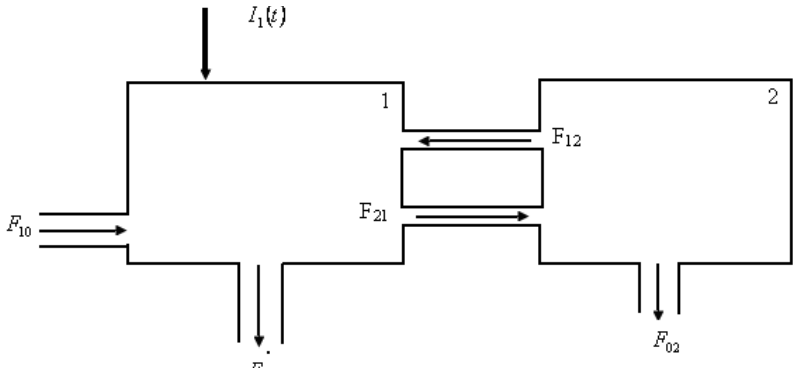


Figure: Infusion into the first compartment

Hence

$$\dot{x}_1 = -(F_{21} + F_{01})c_1(t) + F_{12}c_2(t) + I_1(t)$$
$$\dot{x}_2 = F_{21}c_1(t) - (F_{12} + F_{02})c_2(t)$$

or, in terms of $X = [x_1, x_2]$ alone,

$$\dot{X} = \begin{bmatrix} -(a_{21} + a_{01}) & a_{12} \\ a_{21} & -(a_{12} + a_{02}) \end{bmatrix} X + \begin{bmatrix} I_1(t) \\ 0 \end{bmatrix}$$

where $a_{ij} = F_{ij} / V_j$, and the original diagram takes the simple form as shown in figure below.

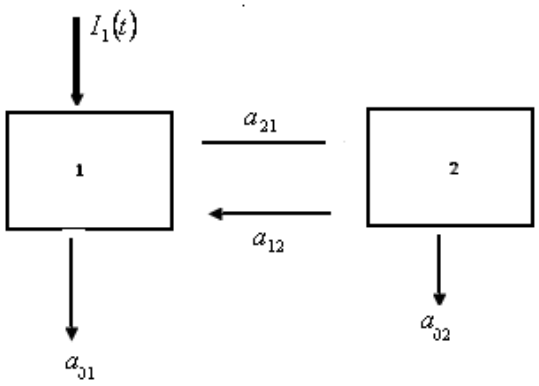


Figure: Simplified two compartment model with transfer coefficient

The Homogeneous Case: We first analyze the tracer is rapidly injected into the first compartment. Hence $X(0) = [x_0, 0]$ but $I_1(t) \equiv 0$.

Let $a = a_{21} + a_{01}$, $b = a_{12}$, $c = a_{21}$ and $d = a_{12} + a_{02}$. The homogeneous solution $X_e(t)$ is given in terms of the eigenvalues and eigenvectors of the matrix.

$$A = \begin{bmatrix} -a & b \\ c & -d \end{bmatrix}$$

The characteristic polynomial of A is $\lambda^2 + (a+d)\lambda + (ad - bc)$ with discriminant

$$D = (a+d)^2 - 4(ad - bc) = (a-d)^2 + 4bc$$

Hence $D \geq 0$ and so, both eigenvalues are real. Also note that $ad - bc = a_{21}a_{02} + a_{01}a_{12} + a_{01}a_{02} \geq 0$ and $a + d = a_{12} + a_{02} + a_{21} + a_{01} > 0$, thus using the quadratic formula, we have established the following theorem.

Theorem: In the general two-compartment model, both eigenvalues of the transfer matrix A are real and non-positive.

Let $\lambda_1 = -m_1$ and $\lambda_2 = -m_2$ be the two eigenvalues, where m_1 and $m_2 \geq 0$. There is the possibility that $\lambda_1 = \lambda_2$ but this will occur only when $a = d$ and $bc = 0$. If for example $a_{12} = b = 0$, we would then need $a_{02} = a_{21} + a_{01}$. Since this case is highly unlikely in applications, we will assume that two eigenvalues are distinct. The solution of $\dot{X} = AX$ are now in the form

$$X(t) = E_1 e^{-m_1 t} + E_2 e^{-m_2 t}$$

where $E_1 = [c_1 \; c_2]$ and $E_2 = [d_1 \; d_2]$ are eigenvectors corresponding to λ_1 and λ_2 respectively. From $AE_1 = \lambda_1 E_1$, $AE_2 = \lambda_2 E_2$ and $X(0) = [x_0, 0]$, we obtain the equations

$$(a - m_1)c_1 = bc_2$$

$$(d - m_1)c_2 = cc_2$$

$$(a - m_2)d_1 = bd_2$$

$$(d - m_2)d_2 = cd_1$$

$$c_1 + d_1 = x_0$$

$$c_2 + d_2 = 0$$

In addition, since

$$\lambda^2 + (a + d)\lambda + (ad - bc) = (\lambda + m_1)(\lambda + m_2)$$

We have

$$m_1 m_2 = ad - bc$$

$$m_1 + m_2 = a + d$$

Adding above equations, we obtain

$$(d - m_1)c_2 + (d - m_2)d_2 = cx_0$$

Since $c_2 + d_2 = 0$, it follows that $c_2 = cx_0/(m_2 - m_1)$ and $d_2 = -c_2$. Therefore, $c_1 = (d - m_1)c_2/c = x_0(d - m_1)(m_2 - m_1)$.

Finally, we have $d_1 = (d - m_2)d_2/c = x_0(m_2 - d)/(m_2 - m_1)$.

The Non-homogeneous Case:

In clinical applications it is often desirable to maintain a constant therapeutic level of particular drug in a patient. To accomplish this, a drug is administered at a constant rate lf I mg/min through intravenous infusion into a blood stream. Thus, the technique involved can be used to solve

$$\dot{X} = AX + \begin{bmatrix} I \\ 0 \end{bmatrix}$$

and find the steady state solution $\hat{X}$. Assuming that both eigenvalues of A are negative (which will occur when a_{01} and $a_{12} > 0$), the constant particular solution X_p is given by

$$X(p) = -A^{-1}\begin{bmatrix} I \\ 0 \end{bmatrix}$$

$$= \frac{1}{ad - bc}\begin{bmatrix} d & b \\ c & a \end{bmatrix}\begin{bmatrix} I \\ 0 \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} Id \\ Ic \end{bmatrix}$$

Since $\lim_{t\to\infty} X_c(t) = 0$, the steady state solution $\hat{X}$ is just X_p.

Returning to the special case illustrated above in the beginning of this section, we have $a_{02} = 0$ and so ad-bc = $a_{12}a_{01}$, d = a_{12}, and c = a_{21}. The steady state solution is now given by

$$\hat{x}_1 = I/a_{01} \text{ and } \hat{x}_2 = (a_{21}/a_{12})\hat{x}_1$$

The steady state concentration in the bloodstream $\hat{c}_1$ is therefore given by

$$\hat{c}_1 = I/(a_{01}V_1)$$

***

Unit-IV

Biochemical Reactions and *Population Genetics*

Introduction

The understanding of cell function has been greatly facilitated through quantitative study of the growth of cells. By the growth of a colony of unicellular microorganisms is usually meant (and we shall mean here) changes in their number, rather than the change in size of the individual organism. A common method for cells to reproduce and increase their number is binary fission, in which a cell divides into two cells. The cell number in a culture can be determined in a variety of ways, such as the determination of the cell count per unit volume, the cell mass, or the biochemical activity of the cells.

This unit is based on the compartment models in physiology, Bath-tub model; discrete and continuous transfer; Stewart-Hamilton theory of cardiac output; Elementary pharmacokinetics etc. Some insight into how modeling is done to describe physiological functions. Two phenomena were considered in physiology (1) blood flow in arteries (2) the process of transportation of oxygen molecules to the tissues.

Cell Growth

The cell is a basic unit of structure and function in all-living things, except viruses. Cells are intensively investigated in the field of microbiology, which is the study of microorganisms too small to be seen by the naked eye, or less than 0.1 mm in diameter. Because of their small size, microorganisms are usually studied as aggregates in a culture, a colony of microorganism's growth in a medium in a laboratory.

Exponential Growth or Decay

The essential role of a useful mathematical theory in the sciences is to predict the future. That is to say, it attempts to answer the question: Given the state of a system at some initial time, what is the state of the system at some future time? The answer of often provided as the solution to differential equation.

Consider the growth of a population of N organisms. We shall think of population, for definiteness, as representing the number of bacteria in a colony,

although it could equally well represent the number of people in the country. In either case, we can assume that the population is so large that the change of a few members is infinitesimal. Therefore, we shall consider the number N to be a continuous variable that changes with time.

During a small time interval Δt subsequent to t, on what does the change in N depends? It is reasonable to suppose that this change ΔN is proportional to N as well as to Δt, or

$$\Delta N \propto N \Delta t \qquad \ldots(a)$$

The proportionality of the change ΔN to N is an example of the law of mass action. Many relations and processes in biology, chemistry, physics, and so on, obey such a law, at least over some range of variation. According to this law, if the population N were doubled, the change ΔN would also be doubled. Thus, the relation (a), can be therefore written as

$$\frac{dN}{dt} = kN \qquad \ldots(b)$$

The solution of the equation (b) is given by

$$N = N_0 e^{kt} \qquad \ldots(c)$$

where N_0 is the initial size of the population at t = 0.

The solution is said to be unique as no other function satisfies the differential equation (b).

The equation (b) must first instance be dimensionally correct to be at all meaningful. If we let the bracket [] denote the dimensions of the quantity contained within them, we have

$$\left[\frac{dN}{dt}\right] = \left[\frac{N}{t}\right] \qquad \ldots(d)$$

From equation (b), the dimensions of $\frac{dN}{dt}$ must be same as [k N], and therefore [k] = [1/t], or k has the dimensions of reciprocal time. We can also recognize this fact from equation (c), we see that the fractional growth rate at any time $\frac{1}{N}\frac{dN}{dt}$ is a constant, and this constant is k. It is also called specific growth rate.

Suppose now that k is positive, in case of bacteria population, the time interval taken for doubling the size of the population is called doubling time. If there are no cell deaths, the doubling time is the same as the mean generation time, which

is the mean life-time of a single cell. Let T denote the doubling time, then at t = T, $N = 2N_0$, substituting this relationship into equation (c) leads to $N_0 = N_0 e^{kT}$, or

$$k = \frac{\log 2}{T} = \frac{0.6931}{T} \qquad \ldots(e)$$

Thus, k is inversely related to the doubling time. Equation (c) has also been called the law of Malthus, since it was Malthus who called attention to the dangers of geometric growth characterized by positive k.

An important physical difference occurs when the parameter k is negative. For a population, k then represents a rate of decline, or death rate. Similarly if N represents the mass of a chemical species, and k is negative, then k represents a disappearance rate. More generally, it is called a decay rate. The time it takes for the species to decay to ½ of its initial value is called the half-life and is denoted by T½.

The use of radioactive substance as tracers has become widespread in biology and medicine. Because the radioactivity of atoms is easily measured, the incorporation of radioactive atoms into living organism is readily traced. Although the decay of a particular radioactive atom is a statistical event subject to the laws of probability, a macroscopic amount of radioactive material made up of very large number of atoms behaves in a quite predictable way. Thus, if N is the amount of radioactive material, its decay in time is governed by (b) with k negative.

Most of the neutrons produced in the atmosphere by cosmic rays are eventually absorbed by nuclei of nitrogen by the reaction

$$^{14}N + n \rightarrow {}^{14}C + {}^{1}H.$$

The ^{14}C atoms, which are radioactive, are formed high in the atmosphere (10 – 15 km altitude) and rapidly become oxidized to $^{14}CO_2$. This then mixes with natural CO_2 and participates in the same physical and biochemical processes. In particular, a steady state of exchange is set up between the atmosphere and organisms so that a certain proportion of the carbon in living plants and animals is in this radioactive form. This radioactivity amounts to 6.5 dis/min/g of carbon. When a plant or animal dies, the exchange ceases and the radioactive carbon remains embedded in the organism to decay according to its natural rate.

For ^{14}C , the half-life is 5730 years. These facts form the essential basis of radiocarbon dating. The basic mathematical theory of it is expressed by equation (6.3) with k negative.

Determination of Growth or Decay Rates

Cell populations that are described by equation (c) are said to be in "log phase" a terminology that appears unfortunate, since they are more correctly described as being in exponential growth phase, or simply exponential phase. The origin of the term log phase arises from the following considerations.

In biology it is usual case that a growth curve y(t) is experimentally determined. For example, we wish to know whether a bacterial population is in fact in exponential growth, and if so, what the growth constant k is. If we take logarithms of both sides of equation (c) (with y replacing N), we see that the relationship between y(t) and t can be described in linear fashion as

$$\log y = \log y_0 + kt \qquad \ldots(f)$$

In other words, the function logy, when plotted as a function if t, is described by a straight line with slope k and intercept log y_0. Therefore, if we make such a plot and find that, within experimental error, the resulting curve is in fact a straight line, then, we know that the population is indeed in exponential growth with a growth rate determined by the slope of the straight line. Because linear relationships are easy to comprehend and to visualize, it is often desirable to represent functional relationships in linear fashion.

Another important reason for using a logarithmic scale, that is to say, using logy instead of y as an ordinate, is that the values of y may be changing by orders of magnitude. Here we are conforming to the common vocabulary usage in the natural sciences for comparing two numbers (rather than two functions). One number is larger than the other by an order of magnitude if it is approximately 10 times as large, by two orders of magnitude if it is approximately 100 times as large, and so forth. When this is the case, it is obviously wiser to use a logarithmic scale rather than a linear scale for graphical purposes.

The use of $\log_{10} y$ for the ordinate instead of y is facilitated by the use of semilog paper. This is a kind of graph paper in which the abscissa utilizes a linear scale, equal intervals on the abscissa representing equal increments of the value of the abscissa variable. On the other hand, the ordinate utilizes a logarithmic scale; equal intervals on the ordinate scale represent equal intervals

of $\log_{10} y$. However, the values of y, the argument of the logarithm function, are assigned to the ordinate scale, rather than the values of the logarithm itself. Therefore, if y is measured, the necessity of taking $\log_{10} y$ in making a plot is obviated. We try to make this clear in figure below by assigning ordinate values to both y and $\log_{10} y$.

Recall that the relationship between natural logarithms and logarithms to the base 10 is given by $\log y = \log_{10} y.\log 10$. If then represents a quantity in exponential growth as in equation (f) , we may write

$$\log_{10} y = \log_{10} y_0 + (\log 10)^{-1} kt,$$

$$\log 10 = 2.3026, \frac{1}{\log 10} = 0.4343 \qquad \ldots(g)$$

and the curve is displayed in the Figure.

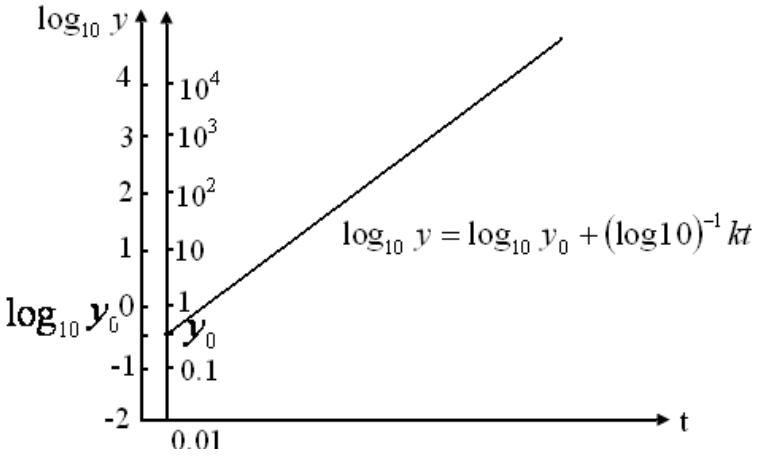


Figure: A semilog plot of the function $y = y_0 e^{kt}$

Assume that y has been determined experimentally as a function of the time t and plotted on semilog paper. Let the result be a straight line as shown in figure above. We wish to determine the values of y_0 and k. According to equation (f), the y intercept of the curve is $\log_{10} y_0$ (on the $\log_{10} y$ scale), and therefore the value of y_0 may be read off the curve directly from the y scale. The slope of the curve is given by

$$k\left(\frac{1}{\log 10}\right)=\frac{\Delta \log_{10} y}{\Delta t} \qquad \ldots(h)$$

Now let t_1 and t_2 are the values of the abscissa at any two points on the curve, and let y_1 and y_2 be the corresponding values of y. Then, to determine k by means of (h), we required the logarithm to the base 10 of y_1 and y_2. Thus,

$$k=\log 10\frac{\log_{10} y_2-\log_{10} y_1}{t_2-t_1} \qquad \ldots(i)$$

In the above example, convenient choices for y_1 and y_2 are 10 and 1, respectively, or any two successive powers of 10, where $\Delta \log_{10} y=1$, and $k=\log 10/\Delta t$.

The Method of Least Squares

A biologist who wishes to represent some experimental data in explicit quantified form is confronted with the problem of experimental error. For example, suppose he has reason to suspect that the relationship between two observed variables x and y is linear. Because of experimental error the observations will not all fall on a straight line. The question therefore, arises as to which straight-line fits the data best.

A criterion as to what is best was given by Gauss and is widely utilized. We assume at the outset that the values of x are given and ascribe all the error in measurement to y. (Many times, such an assumption is completely arbitrary, and in a second examination, we could equally well invert the roles of x and y.) Gauss criterion of the best line is the one, which minimizes the error, which is defined as the sum of the squares of the deviation of the observed values of y from the line. Let the equation of the line be $y=mx+b$, and let the observed values of y be denoted by y_i^* at the point x_i. The index i runs from 1 to n if there are n measurements altogether; then the error E is given by the expression as shown below (see Figure)

$$E=\sum_{i=1}^{n}\left[y(x_i)-y_i^*\right]^2 \qquad \ldots(j)$$

where $y(x_i)$ is the value of y at the point x_i. The error defined in this manner is seen to depend only on the absolute magnitude of the deviation of a given y_i^* from $y(x_i)$, and not on whether the point y_i^* falls above or below the line.

Substituting $y(x_i) = mx_i + b$, the above equation becomes

$$E = \sum_{i=1}^{n} \left(mx_i + b - y_i^* \right)^2 \qquad \ldots(k)$$

The quantities x_i and y_i^* are given numbers so that E is a function that depends on m and b, whose values we wish to choose in such a way that E is minimized.

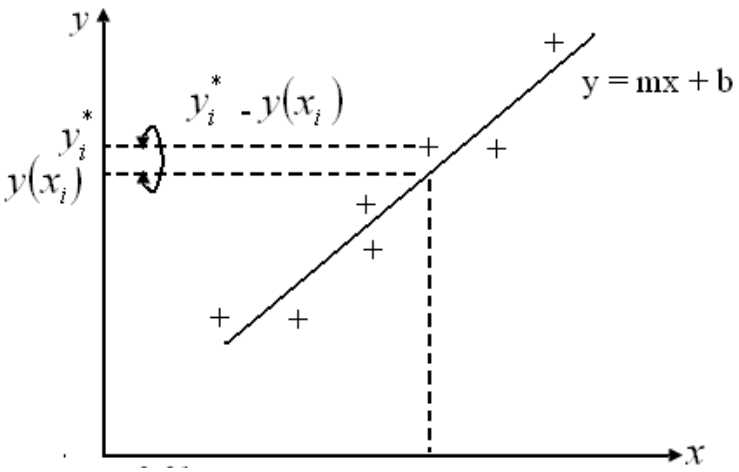


Figure: A set of experimental observations, indicated by the points +, to which the linear relationship y = mx + b has been theoretically assigned.

In other words, E is a function to be minimized with respect to the two variables m and b. The conditions for this are that the partial derivatives of E = E (m,b) with respect to m and b are set equal to zero, or

$$\frac{\partial E}{\partial m} = 2\left[m\sum_{i=1}^{n} (x_i)^2 + b\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i^* x_i \right] = 0$$

$$\frac{\partial E}{\partial b} = 2\left[m\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} 1 - \sum_{i=1}^{n} y_i^* \right] = 0 \qquad \ldots (l)$$

Equations (l) are two linear inhomogeneous equations for the quantities m and b. Their solution is

$$m = \frac{n\sum_{i=1}^{n} y_i^* x_i - \sum_{i=1}^{n} y_i^* \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} (x_i)^2 - \left(\sum_{i=1}^{n} x_i \right)^2},$$

$$b = \frac{\sum_{i=1}^{n} y_i^* \sum_{i=1}^{n} (x_i)^2 - \sum_{i=1}^{n} y_i^* x_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} (x_i)^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = n^{-1}\left(\sum_{i=1}^{n} y_i^* - n\sum_{i=1}^{n} x_i\right) \quad \ldots \text{(m)}$$

These values of m and b do minimize E because $E \to \infty$ as either |m| or $|b| \to \infty$. The above-illustrated procedure of fitting observations to an assumed functional form is called the method of least squares. The procedure does not depend on how many observations n have been made, as long as n is greater than the number of unknown parameters.

Such questions as, how good is the fit, what is the variance of this set of data as compared to another set, and so forth, are more properly statistical questions which can be answered only by appealing to the theory of probability.

Nutrient Uptake by a cell

As another example of a biological process governed by an exponential law, we shall consider the transport of the nutrient into some bacterial cell. Bacteria are microorganisms, very small single living cells, with a diameter usually not 3 μm. Reproduction is usually accomplished asexually by fission. Bacteria are termed procaryotic cells because they have a primitive type of nucleus that lacks a membrane. Escherichiae constitute a type of bacteria that has the shape of a short rod and is frequently motile. Escherichia coli, or E. coli for short, is a normal inhabitant of the intestinal tract of man and other animals. In its normal environment, It is nonpathogenic. Escherichia coli have been more thoroughly investigated than any other bacteria. Its length is about 2 μm and its mass is about $5 \times 10^{-13} g$. Many varieties or strains of E. coli exist.

Consider the mutant bacterial strain E/ coli ML 32,400, which is galactose-negative, that is, incapable of growth on galactose. Galactose is one of the simple sugars, widely distributed in nature and over 100 in number, that are a source of carbon and energy for cells. Some other such abundant simple sugars that are used as energy sources for growing microorganism are glucose, sucrose, maltose, arabinose, lactose and xylose. The parent strain E. coli ML 30 does utilize galactose as a source of carbon. The cause of the inability to do this in the mutant strain is known to be the lack of an enzyme, galactokinase, which is necessary to metabolize galactose. Testing cells of this mutant will, however, take up large amounts of galactose when it is present in the cell medium.

The uptake of galactose by the strain can be readily observed by utilizing galactose that is labeled with ^{14}C in the growth medium. Assume that the galactose concentration in the medium is constant, and let c denote the radioactive galactose concentration inside the cells. Initially, because the strain is galactose-negative, the concentration c is zero. We assume also that the rate of entry of galactose into the E. coli at any moment depends only on the difference between the concentration in the interior c and the final concentration $\bar{c}$, which is found to be much larger than the external concentration in the medium. Such an assumption no doubt represents a simplified view of the transport mechanism, although it leads to an accurate representation of observations. It follows from the assumption that $c = c(t)$ satisfies the equation

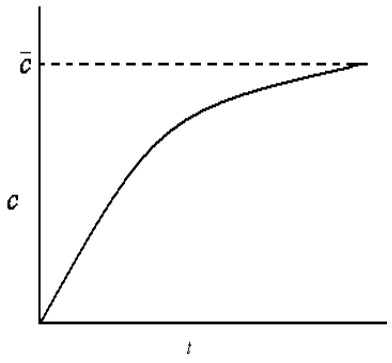
$$\frac{dc}{dt} = k(\bar{c} - c), \qquad \ldots(n)$$

where k is a positive constant and $c(0) = 0$. This is a slightly more complicated than our previous one. Before we solve it, we ask first what the equilibrium values of c is, or, when does $dc/dt = 0$? Clearly this occurs when influx $k\bar{c}$ equals to out flux kc, or $c = \bar{c}$. This form of equation (n) suggests that we should investigate the quantity $(c - \bar{c})$. It is readily seen that equation (n) can be rewritten as

$$\frac{d(c - \bar{c})}{dt} = -k(c - \bar{c}).$$

The solution of this equation is

$$(c - \bar{c}) = (c - \bar{c})e^{-kt} = -\bar{c}e^{-kt} \qquad \ldots(o)$$



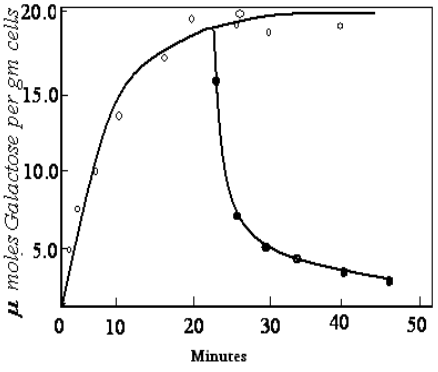


Figure: Theoretical dependence of the concentration c as a function of the time: Figure (b): The observed uptake of radioactive ^{14}C .

The subscript 0 on the right hand side means that we are to evaluate the quantity in parenthesis at t = 0. The solution for c may therefore be written, by transposing $\bar{c}$ in equation (o), as

$$c = \bar{c}\left[1 - e^{-kt}\right], \qquad \ldots(p)$$

which expresses the fact that c approaches its equilibrium value asymptotically. The curve based on equation (p) is illustrated in Figure.

For purposes of qualitative comparison, the observed intracellular galactose concentration is shown in Figure. In the quoted experiment, some of the cells, which had achieved the maximum or saturation value of intracellular concentration, C were introduced suddenly at the time $t_0 = 21$min (indicated by an arrow in Figure) to a non-radioactive ^{12}C galactose medium. The intracellular radioactive galactose then disappeared as shown by the decaying branch of the curve in Figure. When representing such circumstances on the basis of the simple theory presented here in, equation (n) no longer contains the influx term k $\bar{c}$ and its solution becomes $c = \bar{c}e^{k(t\ t_0)}$ for $t \quad t_0$. We see from Figure that the decaying branch does have the qualitative features of such a decaying exponential.

How would we obtain the rate constant k from data represented by the curve shown in Figure? If we plot c verses t on semilog paper, we would find that the resulting curve is not a straight line. However, for large times, we would see that the curve is a horizontal straight line because then e^{-kt} is negligible and $\log \approx \log \bar{c}$.

In other words, the asymptote for the semilog plot is $\log \bar{c}$. Knowing $\bar{c}$, we can then calculate $(c - \bar{c})$ and make a semilog plot of the later quantity verses time. Such a plot permits us to determine k in the usual way.

Inhomogeneous Differential Equation

In a common way, we have just solved the differential equation

$$\frac{dy}{dt} + ay = b \qquad \ldots(q)$$

where a and b are constants. Such an equation is called an inhomogeneous differential equation, because of the presence in it of the inhomogeneous term b, which does not depend on y or its derivative. If b were equal to zero, the differential would be called homogeneous. What is the solution to this equation

if a and b are functions of t? The answer is that it equals the general solution of the homogeneous equation (obtaining by setting b equal to zero) plus any particular solution of the inhomogeneous equation.

To understand this last remark, we introduce some terminology from the theory of differential equation A differential equation is said to be ordinary when it contains a single independent variable, as in equation (b) and equation (q). The solution to the nth order ordinary differential equation contains n arbitrary constants, and is called the general solution. If particular values to the arbitrary are assigned, the resulting solution is called a particular solution.

In the solution (p) of equation (n), $c = \bar{c}$ is a particular solution, and de^{-kt} with d constant is the general solution to the homogeneous equation. We choose d equal to $\bar{c}$ in order to satisfy the initial condition. When a and b are time dependent, the solution to eqn. (q), subject to the initial condition $y(0) = y_0$, is

$$y(t) = \exp\left[-\int_0^t a(x)dx\right]\left\{y_0 + \int_0^t b(s)\exp\left[\int_0^s a(x)dx\right]ds\right\} \qquad \ldots(r)$$

The solution is seen to consist of two terms. The first term derives from the general solution of the homogeneous equation, and the second term is a particular solution of the inhomogeneous equation.

Growth of the Microbial Colony

We know that a colony of bacteria or other microorganisms contained in a nutrient medium will not grow indefinitely. There are many possible reasons for this, to wit, lack of space, lack of oxygen, disappearance of nutrients, the appearance of toxic substances, or changes in ion concentration in the medium, especially pH. We attempt to represent microbial growth in a more realistic way by assuming that the population N(t) at any time t is represented by the differential equation

$$\frac{dN}{dt} = kN - \beta N^2 \;\; ; \; k > 0, \;\; \beta > 0 \qquad \ldots(s)$$

The last term on the right hand side of the equation is always negative, and we can see intuitively that N can never grow indefinitely. As it attempts to do so, $\frac{dN}{dt}$ must ultimately become negative because N^2 is of a higher order than N.

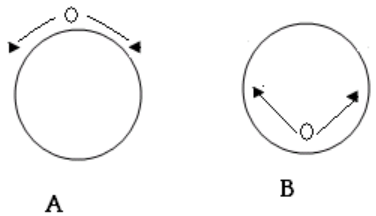
We emphasize here that the term kN represents the net growth rate, or the excess of the birth rate over the death rate. if the limiting factor in growth is he

appearance of toxic substances, say, we can provide a plausibility argument for its representation by the term $-\beta N^2$ as follows. A given cell detects the cumulative toxic effect on all N cells, if the toxic material diffuses freely throughout the intracellular medium. Thus, the toxic effect on a given cell is proportional to N. The toxic effect on all N cells is N times the effect on one cell and hence proportional to N.N. In other words, the number of intracellular interactions of N cells is of the order of N^2. In any case, the representation of growth process of the colony of microorganisms by equation (s) is a pure assumption whose validity is tested by the comparison between the deduction of the theory and experience.

Before solving (s) we obtain valuable information about the solution by inquiring about the stationary state when $dN/dt = 0$. Then

$$kN - \beta N^2 = 0$$

This has two solutions, $N = 0$ and $N = N_e = k/\beta$. These two solutions are called the stationary states or equilibrium states of the system. We can characterize the equilibrium states as being stable or unstable, by examining the behavior of these states in response to a small perturbation. That is to say, we shall consider the behaviour of these states when N suffers a small displacement away from its equilibrium value. This method of analysis is called the method of small perturbations. The mechanical origin of this idea is illustrated by the consideration of tiny ball balanced on a cylindrical shell in either of the two equilibrium positions as shown in figure below:



In case A, the ball is unstable because a small displacement from equilibrium grows larger and larger. In case B, the ball is stable, because the ball will return to its equilibrium position after a small displacement.

For equation (s), we saw that there are two equilibrium states and their stability properties must be investigated separately. First we examine the neighborhood of the equilibrium value $N = 0$. To do so, we expand the function f(N), the right

hand side of equation (s) in a Taylor series about the value $N = 0$ and neglect all terms higher than the linear term. Thus,

$$f(N) = kN - \beta N^2 = f(0) + (N-0)f'(0) + \cdots = kN + \cdots. \qquad \ldots(t)$$

Because we are only inquiring about the behavior N in a small neighborhood of $N = 0$, we replace f(N) in (s) by the expression (t). Then equation (s) becomes

$$\frac{dN}{dt} = kN \qquad \ldots(u)$$

in the neighborhood of $N = 0$. The solution is $N = ce^{kt}$, which shows exponentially away from $N = 0$. Because k is positive, the equilibrium value $N = 0$ is said to be unstable.

In the neighborhood of $N = N_e$,

$$f(N) = f(N_e) + (N - N_e)f'(N_e) + \cdots = -k(N - N_e) + \cdots. \qquad \ldots(v)$$

Then (s) becomes

$$\frac{dN}{dt} = -k(N - N_e) \qquad \ldots(w)$$

in the neighborhood of $N = N_e$. The solution to (w) is already known to us to be

$$N = N_e + ce^{-kt} \qquad \ldots(x)$$

which appears N_e as t increases. Because the coefficient of t in eqn. (x) is negative, the equilibrium value N_e is said to be stable. Thus, we see that the sign of the coefficient of t determines the stability in the exponential term representing the displacement of N from its equilibrium value.

Let us return to the solution of equation (s). We shall solve it by the method of separation of variables. In this method, we try to write the equation so that a function of N appears on one side of the equation, and a function of t on the other side. If this can be done, the method is applicable. Thus, equation (s) can be rewritten as

$$\frac{dN}{dt} = kN\frac{(N_e - N)}{N_e}$$

On solving this equation and making use of initial conditions, we get

$$N = \frac{N_0 N_e}{N_0 + (N_e - N_0)e^{-kt}} \qquad \ldots(y)$$

This equation is called the logistic law of growth. Usually $N_0 < N_e$, in which case N never exceeds N_e and has characteristic S-shaped appearance as seen in unit-I of this course. However, it is conceivable that $N_0 > N_e$. The N is never less than N_e and has the appearance of a simple decay to equilibrium.

We note that when $N_e e^{-kt} >> N_0$, the term $N_e e^{-kt}$ in the denominator is dominant or much larger than the other terms in the denominator, and (y) reduces to the Malthusian growth law,

$$N = N_e e^{-kt} + \cdots \qquad \ldots(z)$$

The dots on the right-hand side signify that we have neglected terms of lower order than proceeding term. After a sufficiently long time, however, the opposite extreme prevails when $N_e e^{-kt} << N_0$, then the term N_0 in the denominator is dominant and the solution can be expressed approximately as

$$N = \frac{N_e}{1+\left((N_e - N_0)/N_0\right)e^{-kt}} = N_e\left[1 - \frac{N_e - N_0}{N_0}e^{-kt} + \cdots\right]$$

$$= N_e - \frac{N_e}{N_0}(N_e - N_0)e^{-kt} + \cdots. \qquad \ldots(a1)$$

This equation indicates that the approach to the equilibrium value N_e is exponential.

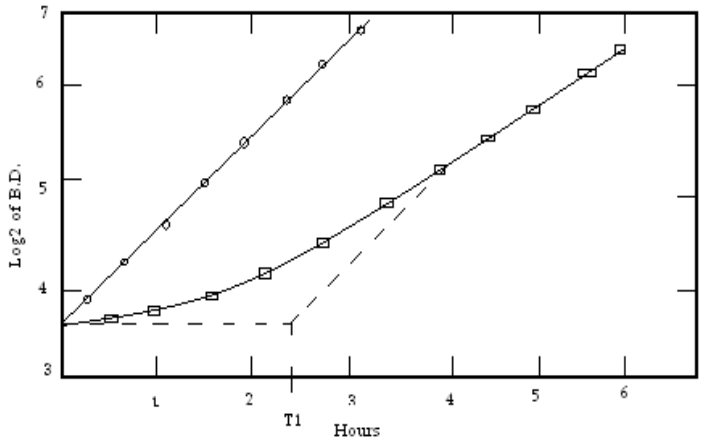


Figure: The growth of E-coli in synthetic media with glucose (circles) and xylose (rectangles) as organic nutrient source.

A bacterial colony grown in a nutrient medium or broth usually displays a growth curve given by the logistic law, provided the initial inoculum N_0 is

taken from an exponentially grown colony. If it is not, there is usually an initial phase of growth called the **lag phase** during which the growth of population is retarded.

Figure below displays the temporal growth of the colony of E-coli that was grown into two different media; one containing glucose and other xylose as an essential organic nutrient source. The cells utilized to initiate the growth of the colonies where taken from a culture maintained on an arabinose medium. It is seen that the cells transferred to the xylose medium exhibit a lag phase of several hours duration, before a rate of growth is achieved which is the same as that in the glucose medium. This lag phase is believed to be due to biochemical adjustments in the cells, in particular the formation of a specific enzyme system, which must be made in order for the cells to be capable of growing in the presence of the nutrient xylose.

When the population of colony is at or very close to its equilibrium or stationary value N_e, it is said to be in its stationary phase of growth. Also, after a long time, the population usually decays to zero from its equilibrium value N_e. Obviously, new growth and death conditions have intervened, such as exhaustion of a necessary nutrient in the medium, say, oxygen or the accumulation of a toxic metabolic product, which are not represented by equation (s).

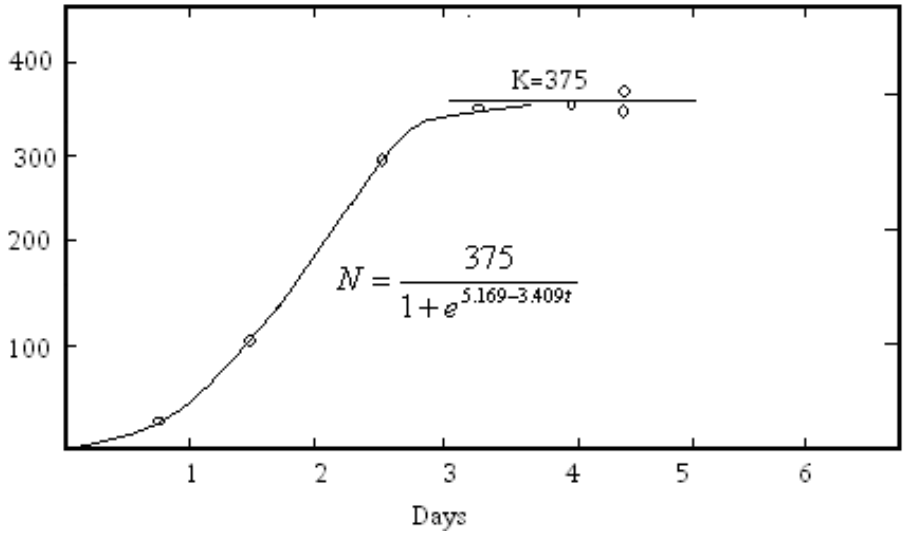


Figure: Growth of Paramecium caudatum in a medium of fixed volume

Figure below shows the growth in time of a population of the infusorian paramecium caudatum, and illustrates the logistic growth law. At t = 0 cells were placed in individual test tubes containing 0.5 cm^3 of a nutrient medium.

The average data of 63 subsequent population counts are displayed in the figure. These data points were then fitted by least squares to the logistic law, the parameter values as shown. Infusoria constitute a class of protozones having hair like protrusions called cilia, useful for locomotion and food ingestion. Protozoa are single celled animals, the lowest division of the animal kingdom. They are microorganisms that are more complex and generally larger than the bacteria. There complexity derives from the possession of many organs like structures called organelles, such as cilia, gullet, contractile vacuole. Because protozoa possess well-defined nuclear membranes, they are classified as eukaryotic cells. A typical protozone volume is of the order of $10^4 \mu m^3$ in contrast to a typical volume of a prokaryotic cell, which is of the order of $1 \mu m^3$.

Growth in a Chemostat

In any investigation of the underlying cellular processes governing growth, it is usually desirable to maintain a bacterial culture under study in a steady state of exponential growth. The previously described method of growth of a microorganism in a container with a fixed amount of nutrient medium is called the batch culture technique. Because of the lag phase of growth, and the stationary phase i.e., eventually achieved, there is really only an intermediate interval between these two phases during which the growth of the colony can be accurately described as being in steady exponential growth.

An alternate solution to this experimental problem of maintaining cells in a steady state of growth is the introduction of this continuous culture technique for growing microorganisms. In this method, a growth chamber of fixed volume V contains a nutrient liquid medium. The chamber has a feeding system whereby the nutrient liquid is supplied at a volume rate per unit time Q, and an overflow provision to maintain the volume of the chamber constant. In addition, the culture is mixed continuously, and aerated continuously, if necessary as shown in Figure below.

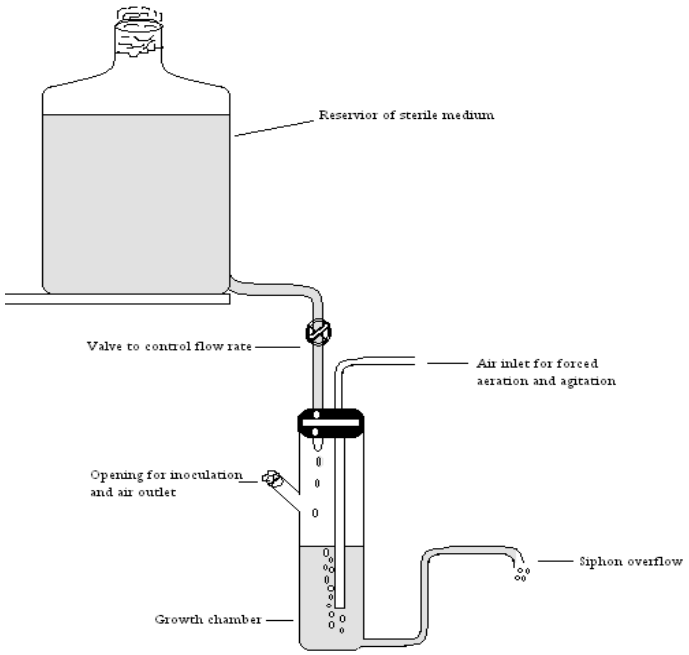


Figure: Diagram of Chemostat, a device for growing bacteria by continuous culture.

Because of the overflow and the mixing, cells are washed out continuously at a rate, which is proportional to the total number of cells in the chamber. Under these conditions, the population N in the chamber obeys the equation

$$\frac{dN}{dt} = kN - qN \qquad \ldots(b2)$$

where the first term on the right hand side represents the growth rate of the population in the undisturbed culture under the given conditions when the cell density is not too great, and the second term represents the disappearance rate of cells from the culture due to dilution and overflow. Here q is called the dilution rate and is defined as q = Q/V. It represents the fractional rate of flow of medium into or out of the growth chamber per unit time.

It can be seen from the equation (b2) that if q is greater than k, the culture is being diluted out too quickly, and the population will decline to zero, if, however, q is less than k, the density of microorganisms in the chamber will increase. Monitoring the cell density continuously, by optical methods, say, and

using the information so obtained to regulate the dilution rate q can maintain the total population in the chamber maintained at a desired level N_0. When the cell density in the chamber is monitored by turbidity measurements, the continuous flow device is called turbidostat.

A different principle for controlling the population is utilized in the continuous culture device called the chemostat. Here the nutrient medium is composed of an excess of all required nutrients except one, the limiting growth factor, the concentration of which can be varied. An example of limiting growth factor that is essential for the growth of almost all bacteria is an energy source containing carbon, such as a simple sugar. Clearly, if the concentration c of the limiting growth factor is zero, no growth can occur, while if it appears in excess, the growth rate is maximal. In general the growth rate k (in the absence of outflow) will depend on c, and the governing growth equation (b2) is more properly written as

$$\frac{dN}{dt} = k(c)N - qN \qquad \ldots(c3)$$

The dependence of k on the concentration of a typical limiting growth factor has the form shown in Figure below:

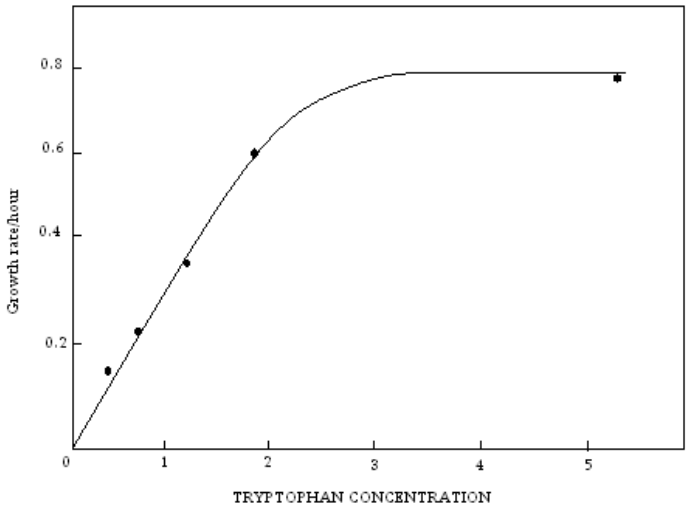


Figure: The growth rate of the tryptophan requiring E-coli B/1, t is shown as a function of the concentration of tryptophan

In the figure the growth rate k is plotted as a function of the concentration of the nutrient tryptophan, for a strain of the bacterium E-coli B/1, t. This strain requires the essential amino acid tryptophan for its growth. Amino acids, which number about twenty, are the fundamental building blocks of all proteins. The mathematical form of k proposed to be

$$k(c) = k_m \frac{c}{K + c}$$

Law of mass action

It states that the rate of any chemical reaction is proportional to the product of the masses of the reacting substances, with each mass raised to a power equal to the coefficient that occur in the chemical equation.

In chemistry, the law of mass action is a mathematical model that explains and predicts behaviors of solutions in dynamical equilibrium. Two aspects are involved in the initial formulation of the law.

i) The equilibrium aspect, concerning the composition of a reaction mixture at equilibrium and

ii) The kinetic aspect concerning the rate equations for elementary reactions.

Both aspects stem from the research performed by Cato M. Gulberg and Peter Waag between 1864-1879. The equilibrium constants were derived by using kinetic data and by the rate equation, which they have proposed. Gulberg and Waag also recognized the chemical equilibrium in a dynamical process in which rates of reactions for the forward and backward reactions must be equal at chemical equilibrium.

In order to derive the expression of the equilibrium constant appealing to kinetics, the expression of the rate equation must be used.

Applications

1. Mathematical ecology: The Lotka-Voltera equation describes dynamics of the predator-prey system. The rate of predation upon the prey is assumed to be proportional to the rate at which the predator and prey meet.
2. In mathematical epidemiology, the law of mass action forms the basis of the compartment model of disease spread in mathematical epidemiology, in which population of humans, animals, or other individuals is divided into categories of susceptible, infected and recovered (Immune), for example SIR model.

Reaction properties:

Mathematical formulation

Given a chemical equation of the form

$$aA + bB + ... \rightarrow xX + yY + ...$$

The forward rate is equal to $k_1[A]^a[B]^b .....$

and backward rate is equal to $k_{-1}[X]^x[Y]^y .....$

In equilibrium

$$k_1[A]^a[B]^b ..... = k_{-1}[X]^x[Y]^y .....$$

or

$$k_{eq} = \frac{k_1}{k_{-1}} = \frac{[X]^x[Y]^y .....}{[A]^a[B]^b .....}$$

For a chemical reaction

$$A + B \rightarrow C$$

The law of mass action gives the rate as

$$\frac{d}{dt}[N_c] = k[N_A][N_B] \qquad \ldots(1)$$

where the brackets denote the concentrations and the reverse reaction is assumed to be negligible.

Denote the initial values with a subscript and define

$$N = N_c$$

$$N_A = [N_A]_0$$

$$N_B = [N_B]_0$$

Since the amounts remaining of A and B after N molecules of C have been formed as $N_A \quad N$ and $N_B \quad N$. Therefore, from equation (1), we have

$$\frac{dN}{dt} = k(N_A \quad N)(N_B \quad N)$$

$$= k\left[N_A N_B \quad (N_A + N_B)N + N^2\right]$$

This is a non-linear differential equation and its solution can be sought as

$$\frac{dN}{N_A N_B \quad (N_A + N_B)N + N^2} = \quad k\,dt \qquad \ldots(2)$$

This integral is of the form

$$\int \frac{dx}{a+bx+cx^2} = \frac{1}{\sqrt{q}} \ln\left(\frac{2cx+b-\sqrt{q}}{2cx+b+\sqrt{q}}\right)$$

for $q<0$, where $a = N_A N_B$, $b = -(N_A + N_B)$, $c = 1$ and the discriminant q is given by

$$\begin{aligned} q = 4ac - b^2 &= 4N_A N_B - (N_A + N_B)^2 \\ &= -(N_A^2 - 2N_A N_B + N_B^2) \\ &= -(N_A - N_B)^2 < 0 \end{aligned}$$

which implies $\sqrt{q} = N_A - N_B$

Now the solution on the basis of (2)

$$\frac{1}{N_A - N_B} \ln\left(\frac{2.1.N + [-(N_A + N_B)] - (N_A - N_B)}{2.1.N + [-(N_A + N_B)] + (N_A - N_B)}\right) = kt + c$$

where c is the constant of integration.

Simplifying, we get $$\frac{1}{N_A - N_B} \ln\left(\frac{N - N_A}{N - N_B}\right) = kt + c$$

To find the value of constant c, we use the initial condition that $N(t=0) = 0$,

Therefore, $$\frac{1}{N_A - N_B} \ln\left(\frac{N_A}{N_B}\right) = c$$

Thus, $$\frac{N - N_A}{N - N_B} = e^{(N_A - N_B)c} . e^{kt(N_A - N_B)}$$

or $$\frac{N - N_A}{N - N_B} = \left(\frac{N_A}{N_B}\right) e^{kt(N_A - N_B)}$$

Solving for N, we have

$$N = N_A \frac{e^{kt(N_A - N_B)} - 1}{\frac{N_A}{N_B} e^{kt(N_A - N_B)} - 1}$$

which tends to zero as t tends to zero; and tends to N_B as t tends to infinity.

Thus, $N(0) = 0$ and $\lim_{t\to\infty} N(t) = N_B$.

Mathematical Model for Law of mass action

The law of mass action describes the rate at which chemicals interact in reactions. It is assumed that different chemical molecules come into contact by collision before reacting, and that the collision rate is directly proportional to

the number of molecules of each reacting species. Suppose that two chemicals A and B react to form a product chemical C, written as

$$A + B \xrightarrow{k} C$$

where k is the rate constant of the reaction.

For simplicity, we shall use the same symbol C, say, to refer to both the chemical C and its concentration. Therefor by the law of mass action

$$\frac{dC}{dt} = kAB$$

Similarly, the law of mass action of reacting concentrations A and B implies

$$\frac{dA}{dt} = -kAB \quad \text{and} \quad \frac{dB}{dt} = -kAB$$

The negative sign is due to the fact that the concentration of reacting substances decreases.

From the above equations, we have

$$\frac{d}{dt}(A + C) = 0 \quad \text{implies} \quad A + C = A_0$$

$$\frac{d}{dt}(B + C) = 0 \quad \text{implies} \quad B + C = B_0$$

where A_0 and B_0 are initial concentrations of the reactants and no product is present initially.

Therefore, Using law of mass action, we have

$$\frac{dC}{dt} = k(A_0 - C)(B_0 - C) \text{ with } C(0) = 0$$

which can be integrated by using the method of separation of variables. On solving the above non-linear differential equation, we have

$$C(t) = A_0 B_0 \frac{e^{kt(B_0 - A_0)} - 1}{B_0 e^{kt(B_0 - A_0)} - A_0}$$

which is complicated expression with the simple limits

$$\lim_{t\to\infty} C(t) = A_0 \text{ if } A_0 < B_0;$$

$$B_0 \text{ if } B_0 < A_0$$

The reaction stops after one of the reactants is depleted; and the final concentration of the product is equal to the initial concentration of the depleted reactant.

If we include the reverse reaction,

$$A + B \underset{k}{\overset{k_+}{\Leftrightarrow}} C$$

Then, the time derivative of the product is given by

$$\frac{dC}{dt} = k_+ AB \quad k\, C$$

Note that k_+ and k have different units. At equilibrium $\dot{C} = 0$, and using the conservation laws $A + C = A_0$ and $B + C = B_0$, we obtain

$$\left(A_0 \quad C\right)\left(B_0 \quad C\right) \quad \frac{k}{k_+} C = 0$$

The equilibrium constant is therefore given by

$$k_{eq} = \frac{k}{k_+}$$

which has units of concentration.

Therefore, at equilibrium, the concentration of the product is given by the solution of the quadratic equation.

$$C^2 \quad \left(A_0 + B_0 + k_{eq}\right) C + A_0 B_0 = 0$$

with the extra condition that $0 < C < \min\left(A_0, B_0\right)$

For example if $A_0 = B_0 = R_0$, then at equilibrium

$$C = R_0 \quad \frac{1}{2}\left(\sqrt{1 + \frac{4R_0}{k_{eq} \quad 1}}\right)$$

If $k_{eq} << R_0$, then A and B have high affinity, and the reaction proceeds mainly to C, with $C \to R_0$.

Exercise:

Write down the equations for $\dot{X}$ in reaction (i) and $\dot{X}$ and $\dot{X}$ in reaction (ii), Also calculate the behavior of chemicals

(i) $A + X \underset{k}{\overset{k_+}{\Leftrightarrow}} 2X$ (ii) $A + X \overset{k_1}{\to} 2X\,;\; X + Y \overset{k_2}{\to} 2Y\,;\; Y \overset{k_3}{\to} B$

Enzyme Kinetics

Virtually all-chemical reactions in the cell involve the direct participation of enzymes, which are proteins that act as catalysts. Proteins are the major constituents of cellular matter. They are complex nitrogenous compounds, macromolecules composed of combinations of the twenty amino acids. The

amino acids, which in combination are called peptides, are joined together in a linear fashion by peptide bonds. Hence a protein molecule is referred to as a polypeptide chain or simply as a polypeptide. The number of different proteins to be found in a bacterium of E. coli is estimated to be about 2×10^3 to 3×10^3, while for a human mammalian cell, the order of magnitude of this number is 10^6. A single protein molecule consists of anywhere from 50 to perhaps 10^4 amino acids molecules. Clearly, the diversity of proteins arises from the many possible ways of forming linear arrays or words (polypeptides) using twenty different letters (amino acids). The polypeptide chain of proteins are folded structures, which have two principal shapes, and are classified as being either filamentous proteins, also called fibrous proteins, or globular proteins. The latter are the most numerous.

The biological catalysts differ from all other catalysts known to chemistry in two essential ways. First they are exceptionally efficient under the mild conditions of the normal physiological state: aqueous medium, standard pressure, and physiological temperature. Second, they exhibit specificity, acting selectivity to find small molecular species called ligands. A ligand that is acted upon by an enzyme to form a product is called a substrate. A single enzyme molecule can transform 10^3 to 10^6 molecules of substrate per minute. That is why their catalytic function in a cell can be performed rapidly and why extremely small quantities of enzyme suffice to carry out cellular processes. Enzymes also act as regulators of biochemical processes. Here, we wish to examine some aspects of the kinetics of enzymatic reactions, by which is meant their reaction rates, the temporal dependence of the reactants and related properties.

The Michaelis Menten Theory

The Michaelis-Menten equation has been successful in describing most enzymatically controlled reactions (Michaelis and Menten, 1913) we shall describe here its theoretical basis. Consider the simplest case of an enzymatic reaction, that between an enzyme and a single substrate. The fundamental assumption of the theory of Michaelis and Menten is that the enzyme and the substrate react reversibly to form a complex initially. The complex subsequently breaks down to form the free enzyme plus one or more products. The reactions are represented schematically as follows:

$$S + E \leftrightarrow C \qquad \ldots(d4)$$

$$C \rightarrow E + P \qquad \ldots(e5)$$

Here E, S, C and P stands for enzyme, substrate, complex, and product, respectively. The arrows in (d4) and (e5) indicate the possible directions of the reactions. The second assumption of the theory is that the equilibrium is rapidly established among E, S, and C in accordance with (d4) and in disregard of the reaction (e5). This approximation may be expected to be valid provided the reactions of (d4) proceed very rapidly as compared to the reactions in (e5). We shall present here the mathematical representation of this theory.

In analytical chemistry, the concentration of reactant in a solution is often specified, rather than the absolute amount. When the concentration represents the solute mass in moles per kilogram of solute, it is called the molality and the unit is denoted by m. When the concentration is measured in moles per liter of solution, it is called the molarity, and the unit is denoted by M. We denote concentration of E, S, C and P by lower case letters naturally; the concentrations are time dependent functions.

In accordance with the law of mass action, the rate equations are assume to be the following

$$\frac{ds}{dt} = -k_{+1}es + k_{-1}c, \qquad \ldots(f6)$$

$$\frac{de}{dt} = -k_{+1}es + (k_{-1} + k_{+2})c, \qquad \ldots(g7)$$

$$\frac{dc}{dt} = k_{+1}es - (k_{-1} + k_{+2})c, \qquad \ldots(h8)$$

$$\frac{dp}{dt} = k_{+2}c. \qquad \ldots(i9)$$

Equation (f6) states that the change Δs in the concentration of substrate in the infinitesimal time interval $t + \Delta t$ consists of two parts. One part is proportional to e, s and Δt and is a result of the association or formation of complex from enzyme and substrate. The other part is proportional to c and Δt and is a consequence of the back reaction or dissociation of complex into enzyme and

substrate. These proportionality relations become equalities with the introduction of the positive proportionality constants k_+ and k_- respectively. The minus sign in front of the term $k_{+1}es$ then indicates that the change in s resulting from complex formation is a decrease, while the plus sign in front of $k_{-1}c$ indicates that the change in s resulting from the back reaction is an increase. Similar interpretations are applicable to equations (g7) - (h8).

Note that all growth terms are positive and all disappearance terms are negative, so that the rate constants $k_{\pm i}$ are positive quantities by definition. As in the Lotka-Voltera equations, a non-linear term, which in this case is proportional to es, appears. This term represents the collisions between enzyme and substrate molecules that lead to complex formation. Initially, it is assumed we start with enzyme and substrate only. Therefore,

$$(s,e,c,p)=(s_0,e_0,0,0) \text{ at } t=0 \qquad \ldots(j10)$$

where the subscript 0 denotes the initial value. The temporal dependence of the concentrations is now completely know, provided we can solve the equation system (f6)-(j10).

The equations can be simplified by recognizing that the amount of enzyme is always conserved existing either as free enzyme or as part of the complex. Mathematically, we observe, by adding equations (g7) and (h8) that $\frac{d}{dt}(e+c)=0$ and therefore $e(t) + c(t)$ is a constant. The constant is determined by setting $t=0$ when the values of e and c are known from (j10).

Thus,

$$e+c=e_0 \qquad \ldots(k11)$$

We utilize eqn.(g7) to eliminate e from equations (f6) and (c3) which then become

$$\frac{ds}{dt}=-k_{+1}e_0s+(K_{+1}s+k_{-1})c \qquad \ldots(l\text{-}12)$$

$$\frac{dc}{dt}=k_{+1}e_0s-(K_{+1}s+k_{-1}+k_{+2})c \qquad \ldots(m13)$$

This is a system of two non-linear differential equations. Clearly if we solve these two equations we can obtain p and e from equations (j10) and (k11) respectively. However, we are interested in the neighborhood of equilibrium states here; we wish to determine the velocity of the reaction, especially in the initial stages of the reaction. The initial velocity usually measures.

In order to simply the mathematical problem, we make use of the second assumption of the theory that a quasi steady state is established very rapidly, so that the concentration of the complex c is changing very slowly with time. Thus, we assume that

$$\frac{dc}{dt} = 0 \qquad \ldots(n14)$$

With this assumption, equation (m13) becomes algebraic equation that can be readily solved for c in terms of s, namely

$$c(t) = \frac{k_{+1} e_0 s(t)}{k_{+1} s(t) + k_{-1} + k_{+2}} \qquad \ldots(o15)$$

We now substitute (o15) into (m13) and obtain the equation

$$\frac{ds}{dt} = -k_{+2} c = \frac{k_{+1} k_{+2} e_0 s}{k_{+1} s + k_{-1} + k_{+2}} \qquad \ldots(p16)$$

This equation can be directly integrated to yield.

$$s + \frac{k_{-1} + k_{+2}}{k_{+1}} \log \frac{s}{s_0} = s_0 - k_{+2} e_0 t \qquad \ldots(q17)$$

in which the initial condition has been imposed in order to determine the constant of integration.

Biochemists commonly wish to determine the velocity of the reaction v, which is usually defined as either the rate of appearance of the product, or the rate of disappearance of the substrate. These two rates are not strictly equal. However, with the approximation of the theory, it is seen from equation (i9) and (p16) that these definitions are equivalent. Thus, the velocity of the reaction at any time t is

$$v(t) = \frac{dp}{dt} = \left| \frac{ds}{dt} \right| \qquad \ldots(r18)$$

Usually, reaction velocities are determined at $t = 0$. Denoting the initial value $v(0) = v_0$, we obtain from equations (p16) and (r18) the relation

$$v_0 = \frac{V s_0}{s_0 + K_m} \qquad \ldots(s19)$$

where V and K_m are defined as

$$V = k_{+2} e_0 \qquad \ldots(t20)$$

$$K_m = \frac{K_{-1} + k_{+2}}{k_{+1}} \qquad \ldots(u21)$$

Equation (s19), the main result of the theory, is called the *Michaelis-Menten equation*. The constant K_m is called the *Michaelis constant*. The curve of v_0 as a function of s_0 is a rectangular hyperbola, and approaches its maximum value V asymptotically as shown in figure below. Thus V is the maximum value of the velocity of the reaction, or, simply, v-max. Because k_{+2} is the only reaction rate appearing in V, the rate-limiting step in the reaction is that represented by equation (t20). However, the rate-limiting constituent is the enzyme.

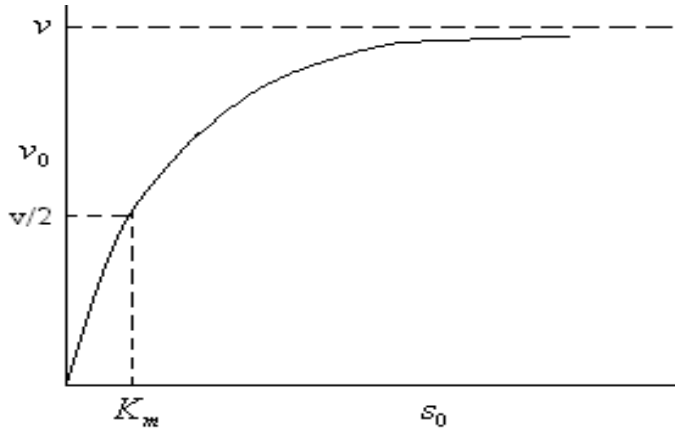


Figure: Illustrates the Michaelis-Menten equation

For large values of s_0, we see that the velocity of the reaction is approximately V and is independent of the substrate concentration. The reaction velocity is then said to saturate with respect to the substrate or to exhibit the phenomenon of saturation. The introduction into a chemical reaction of a reactant in excess amount so as to make the reaction independent of the concentration so that reactant is a common biochemical technique for reducing the number of independent variables associated with the given reaction.

If k_{+2} is small compared to k_{-1}, as the case for many reactions, it can be neglected in the equations (f6)-(i9). The system then has true equilibrium state, given by $ds/dt = de/dt = dc/dt = 0$, for which the inter-conversions among the molecular species leave the net number of molecules of a given species

unchanging with time. Such a dynamical equilibrium sate is also called a steady state. The equilibrium values of e, s and c, denoted with an over bar are found to be related by the equation

$$\frac{\bar{e}\bar{s}}{\bar{c}} = \frac{k_{-1}}{k_{+1}} \equiv K_d \approx K_m \qquad \ldots(v22)$$

Hence in such a case, K_m is approximately the equilibrium constant or dissociation constant K_d for the reaction (d4). As the equations (f6) - (i9) stand, the only true equilibrium state for $e_0 \neq 0$ is given by (e,s,c,p) = (e_0,0,0,s_0) and is not of interest. This state represents the state of total conversion of substrate to product. The back reaction $C \leftarrow E + P$ must exist on thermodynamics grounds, and complements the reaction (e5). Consequently, the complete conversion of substrate to product is never achieved. In practice, the rate constant for the back reaction is usually small. Therefore, the back reaction is usually neglected, as has been done here. When the back reaction is taken into account, the equilibrium constant relating the equilibrium concentrations of S and P can be expressed in terms of all the rate constants.

In experimental kinetics investigations, the velocity is measured for various values of the initial substrate s_0 and the curve is shown in figure above. The kinetic properties of the enzyme in interaction with a particular substrate are characterized by the two parameters V and K_m. The value of K_m is determined according to equation (u21) as the value of s_0 for which $v_0 = v/2$. A commonly employed equivalent representation of the data is the double reciprocal or Line-weaver-Burk plot, in which v_0^{-1} is plotted as a function of s_0^{-1}.

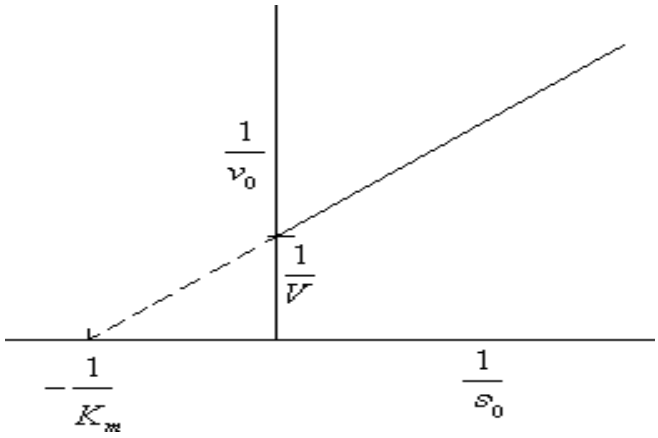


Figure: Lineweaver-Burk plot of Michaelis-Menten equations

According to eqn. (t20) this relationship is represented simply by a straight line,

$$\frac{1}{v_0} = \frac{1}{V} + \frac{K_m}{Vs_0} \qquad \ldots(w23)$$

The Figure below gives the Line-weaver-Burk plot of the Michaelis-Menten equations for an enzyme.

Early Time Behavior of Enzyme Reactions

If we examine more closely the solution we obtain for c(t), equation (o15), we observe that it does not satisfy the initial condition that c(0) = 0. In fact, the solution we have obtained is not valid in the neighborhood of t = 0. In spite of this fact, the inference regarding v(0) is in good quantitative agreement with many enzymatically controlled reactions. In other words, s(t) as given by eqn. (r18) represents the experimental facts satisfactorily during the initiation of an enzymatically controlled reaction, but c(t) as given by (o15) does not. In order to understand this seemingly paradoxical result, we shall solve the rate equations (n14)-(m13) in a more formal albeit approximate manner without making the restrictive assumption, equation (n14).

We shall utilize a perturbation procedure to obtain formal solution to equations (l-12)-(m13). To understand this procedure and the approximation implied by it, it is necessary to express our equations in non-dimensional form by introducing the non-dimensional variables $t' = k_{+1}e_0 t$, $s' = s/s_0$ and $e' = c/c_0$. Note that $k_{+1}e_0$, according to equations (l-12)-(m13), has the dimensions of t^{-1} so that t' is dimensionless. It is called the non-dimensional time. If now the equations (l-12) - (m13) are divided by $k_{+1}e_0 s_0$ and the non-dimensional variables are introduced, the equations take the form

$$\frac{ds'}{dt'} = -s' + \left(s' + \frac{k_{-1}}{k_{+1}s_0} \right) c' \qquad \ldots(x24)$$

$$\frac{e_0}{s_0}\frac{dc'}{dt'} = s' - \left(s' + \frac{k_{-1} + k_{+2}}{k_{+1}s_0} \right) c' \qquad \ldots(y25)$$

In this form we see that $\frac{e_0}{s_0}$ is one of three non-dimensional parameters characterizing the system, the other two being $\frac{k_{-1}s_0}{k_{+1}}$, and $\frac{(k_{-1} + k_{+2})s_0}{k_{+1}}$. In

experimental investigations of enzyme kinetics it is usually the case that $s_0 >> e_0$ and $e_0 / s_0 \sim 10^{-3}$ is not uncommon. We shall therefore treat the quantity $\frac{e_0}{s_0}$ as a very small quantity, and seek the solution of the above equation system in the form

$$s' = s'^{(0)} + \frac{e_0}{s_0} s'^{(1)} + \left(\frac{e_0}{s_0}\right)^2 s'^{(2)} + \cdots,$$

$$c' = c'^{(0)} + \frac{e_0}{s_0} c'^{(1)} + \left(\frac{e_0}{s_0}\right)^2 c'^{(2)} + \cdots. \qquad \ldots(z26)$$

Here $s'^{(j)} = s'^{(j)}(t')$ and $c'^{(j)} = c'^{(j)}(t')$, where j is a non-negative integer, are functions of the time to be determined. Either of these series is called an asymptotic expansion because it tends to the true solution as the small parameter $\frac{e_0}{s_0}$ tends to zero. The method of finding a solution to an equation system such as (x24)- (y25) by means of such a series is called a singular perturbation procedure. It is called a perturbation procedure because the solution is expanded as a power series in a small parameter. The procedure is singular when the small quantity or perturbation multiplies the highest derivative term appearing in the differential equations.

If we substitute equation (z26) into eqn. (y25) and neglect all terms that are of order $\frac{e_0}{s_0}$, we obtain two equations for $s'^{(0)}$ and $c'^{(0)}$. These equations are exactly the same as equations (l-12)-(m13) with dc/dt set equal to zero. Thus, the solution we found there for s(t) and c(t) constitute the first term in the expansion (z26). However, it is valid only for large values of t, and is called the outer solution.

To investigate the problem mathematically for early times, we cannot neglect the left-hand side of equation (y25) but must treat it on the footing as the right-hand side. To accomplish this, we introduce a new non-dimensional time $\tau = k_{+1} s_0 t = s_0 t' / e_0$ in the place of old one. Because the new non-dimensional variable is much larger than the old one by a factor s_0 / e_0. It is called the stretched variable. We introduce also new functions $\tilde{s}(\tau)$ and $\tilde{c}(\tau)$ by the

equations $\tilde{s}(\tau) = s'(\tau e_0 / s_0)$, and $\tilde{c}(\tau) = c'(\tau e_0 / s_0)$. In terms of these variables, equations (x24)-(y25) take the form

$$\frac{d\tilde{s}}{d\tau} = \frac{e_0}{s_0}\left[-\tilde{s} + \left(\tilde{s} + \frac{k_{-1}}{k_{+1}s_0}\right)\tilde{c}\right], \qquad \ldots(a53)$$

$$\frac{d\tilde{c}}{d\tau} = \tilde{s} - \left(\tilde{s} + \frac{k_{-1} + k_{+2}}{k_{+1}s_0}\right)\tilde{c} \qquad \ldots(a54)$$

It is seen that the small parameters e_0 / s_0 now multiplies a different term than the one it multiplied previously. Again, let us neglect terms of order e_0 / s_0. Equation (a53) becomes $\frac{d\tilde{s}}{d\tau} = 0$, with the solution $\tilde{s}$ = constant. After imposing the initial conditions, the solution becomes

$$\tilde{s} = 1 \text{ or } s = s_0 \qquad \ldots(a55)$$

Now we substitute this result into (a54), after which that equation is readily integrable. After imposing the initial conditions, we obtain the concentration of the complex, which, expressed in terms of the original dimensional variables, is

$$c = \bar{c}\left[1 - e^{-(k_{+1}s_0 + k_{-1} + k_{+2})t}\right],$$

$$\bar{c} = \frac{k_{+1}e_0 s_0}{k_{+1}s_0 + k_{-1} + k_{+2}} \qquad \ldots(a56)$$

According to (a56), the complex concentration c increases from its value 0 at t = 0 to final $\bar{c}$ which we the quasi-steady state value. At the same time according to (a55), the substrate concentration remains unchanged. Of course, these conditions cannot apply to the state of affairs at large times, because we know, for example that the substrate begins to get exhausted. The expressions given (a55) and (a56) satisfy the initial conditions and represent the solution for small times. Together they comprise the inner solution. The inner solution, valid for short times, plus the outer solution, valid for long times, comprise the total solution to the problem valid to the order e_0 / s_0.

How long does it take for c to increase from zero to its quasi-steady state value? The answer is the time it takes for the exponential term in (a56) to become negligible, for which the characteristic decay time τ_0 appearing to the exponent is given by the expression

$$\tau_0 = \left(k_{+1}s_0 + k_{-1} + k_{+2}\right)^{-1} \qquad \ldots(a57)$$

For enzymes, this time is of the order of a second or less. The reason this time is so short is that k_{+1} for enzymatic reactions is usually found to be in the range $10^6/M$ sec to $10^8/M$ sec.

We note that the limiting value of the outer solution as $t \to \infty$, $s(s,c) = (s_0, \bar{c})$, is the same as the limiting value of the outer solution at $t \to 0$ as is readily verified from equations (a51), (a53), (a55) and (a56). This fact indicates that the two solutions have a common time domain of validity, that is, the outer solution takes up where the inner solution leaves off. The term in the solution that is common to both solutions is called the overlap term. An alternative way of writing the total solution (up to given order e_0/s_0) is to write it as the sum of the inner solution and the outer solution, minus the overlap term. Thus, the total solution for s and c up to order e_0/s_0 can also be written in the form

$$s + K_m \log\frac{s}{s_0} = s_0 - Vt \qquad \ldots(a58)$$

$$c = e_0\left[\frac{s}{s+K_m} - \frac{s_0}{s_0+K_m}e^{-t/\tau_0}\right] \qquad \ldots(a59)$$

Graphs of s, c and p are shown schematically in figure below. Note that $\bar{c} \to e_0$ as $s_0 \to \infty$. Also, the total solution for s happens to be the same as the outer solution, to this order e_0/s_0. The total solution given by (a58)-(a59) is not subject to the paradox set forth at the beginning of this section.

We are now in a position to understand better what occurs in an enzymatically-controlled reaction. The reciprocal of the Michaelis constant is considered to be a quantitative measure of the affinity that the enzyme holds for a particular substrate.

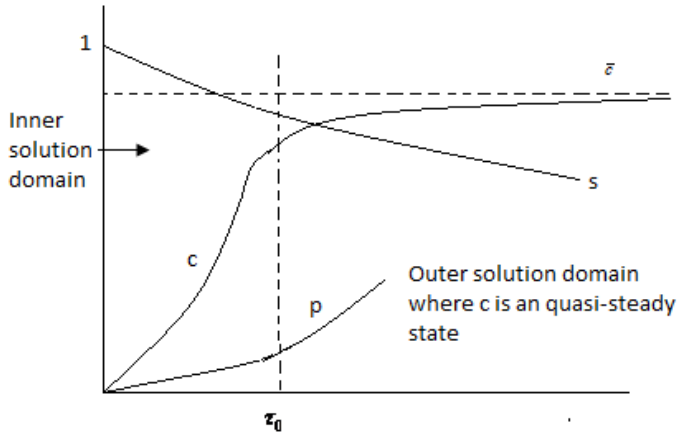


Figure: Schematic representation of the behavior with time of the concentration s, c and p and enzyme-substrate system

Enzyme-Substrate-Inhibitor System

The principal result of Michaelis-Menten theory, equation (a45) or equivalently equation (a49), did not really require the solution to any differential equations. The result followed formally from the underlying differential equations (a38)-(a39) by setting dc/dt equal to zero in (a39), denoting ds/dt in equation (l-12) by v_0 , and eliminating c between these two equations.

We shall now derive the result of applying Michaelis-Menten theory to the case in which two ligands compete for the same enzymatic site. Such a reaction is said to be fully competitive. The reactant that is singled out for measurement is called the substrate. The second reactant is called the inhibitor. Clearly, their roles may be reversed in subsequent studies. The reactions are represented schematically as

$$S + E \underset{k_{-1}}{\overset{k_{+1}}{\leftrightarrow}} C_1 \overset{k_{+2}}{\rightarrow} P_1 + E$$

$$I + E \underset{k_{-3}}{\overset{k_{+3}}{\leftrightarrow}} C_2 \overset{k_{+4}}{\rightarrow} P_2 + E \qquad \ldots(a60)$$

where I represents the inhibitor. Note that each allowed transition has been labeled with its association rate constant. Again, we have neglected the reversibility of the reaction that forms products in (a60). For the short times during which such reactions are usually investigated, the amount of product

formed is small, and the importance of the back reaction is thereby diminished. The kinetic equations corresponding to these reactions are the following:

$$\frac{ds}{dt} = -k_{+1}e_0 s + (k_{+1}s + k_{-1})c_1 + k_{+1}sc_2,$$

$$\frac{dc_1}{dt} = k_{+1}e_0 s - (k_{+1}s + k_{-1} + k_{+2})c_1 - k_{+1}sc_2,$$

$$\frac{di}{dt} = -k_{+3}e_0 i + (k_{+3}i + k_{-3})c_2 + k_{+3}ic_1, \qquad \ldots(a61)$$

$$\frac{dc_2}{dt} = k_{+3}e_0 i - (k_{+3}i + k_{-3} + k_{+4})c_2 - k_{+3}ic_1,$$

$$e = e_0 - c_1 - c_2$$

The concentrations are subject to the following initial conditions,

$$(s, c_1, i, c_2, e) = (s_0, 0, i_0, 0, e_0) \text{ at } t = 0. \qquad \ldots(a62)$$

In deriving equations (a61), we have followed the same procedure utilized in deriving equations earlier. We have omitted the equations representing the concentration of the products P_1 and P_2. Now assume that e_0/s_0 is very small, and that both k_{+1}/k_{+3} and s_0/i_0 behave like constants as e_0/s_0 tends to zero. We express these assumptions in mathematical notation as follows:

$$\frac{e_0}{s_0} << 1, \quad \frac{k_{+1}}{k_{+3}} \sim 1, \quad \frac{s_0}{i_0} \sim 1 \qquad \ldots(a63)$$

These assumptions are the mathematical requirements of the quasi-steady-state hypothesis.

Formally, in correspondence to the quasi-steady-state consideration of equations (l-12)-(m13) of the single substrate enzyme, we set dc_1/dt and dc_2/dt equal to zero in (a61). The two algebraic equations which result permit us to solve c_1 and c_2 in terms of s and i. Hence c_1 and c_2 can be eliminated from the differential equations for s and i.

Denoting the velocity of the reaction by $v(t) = |ds/dt|$, it follows from (a61) that, in the quasi-steady-state approximation,

$$\frac{1}{v(t)} = \frac{1}{V^s}\left[1 + \frac{K_m^S}{s(t)}\left(1 + \frac{i(t)}{K_m^i}\right)\right] \qquad \ldots(a64)$$

where $$K_m^S = \frac{k_{-1} + k_{+2}}{k_{+1}}, \; K_m^i = \frac{k_{-3} + k_{+4}}{k_{+3}}, \; V^S = k_{+2} e_0 \qquad \ldots(a65)$$

The result (a64) is called the time dependent Michaelis-Menten equation for an enzyme-substrate-inhibitor system. The differential equations for s(t) and i(t) in the quasi-steady state approximation can be readily solved, in the same manner as described earlier. The result is

$$s(t) - s_0 - \frac{V^s}{V^i}\left\{1 - \left[\frac{s(t)}{s_0}\right]^\delta\right\} + K_m^S \log\left[\frac{s(t)}{s_0}\right] = -V^S t,$$

$$\frac{i(t)}{i_0} = \left[\frac{s(t)}{s_0}\right]^\delta \qquad \ldots(a66)$$

where δ is defined as

$$\delta = \frac{V^i K_m^S}{V^S K_m^i} = \frac{k_{+4}(k_{-1} + k_{+2})}{k_{+2}(k_{-3} + k_{+4})}, \; V^i = k_{+4} e_0$$

To obtain the usual time-independent form of Michaelis-Menten equation, set t = 0 in equation (a64), and denote by v_0 the initial velocity of the reaction v(0). Then

$$\frac{1}{v_0} = \frac{1}{V^s}\left[1 + \frac{K_m^S}{s_0}\left(1 + \frac{i_0}{K_m^i}\right)\right] \qquad \ldots(a67)$$

The result is that a Lineweaver-Burk plot of v_0^{-1} versus s_0^{-1} is again a straight line, except that the slope depends on the particular value of i_0 .

Cooperative properties of Enzymes

The macromolecular structure of enzymes permits them to engage in a variety of interactions involving many kinds of bonds. In general, however, an enzyme catalyzes only a single reaction. This great specificity is believed to be a steric property; that is to say, it depends on the shape of the enzyme molecule in a local region. A local region of interaction or contact between the enzyme molecule and a reacting substrate molecule is called an active centre or active site.

Many enzymes consist of subunits, which are globular polypeptides. A protein consisting of many subunits is called an oligomer. The term is not usually applied to proteins consisting of more than about twenty subunits. The oligomer is called a dimer, trimer, tetramer, and so on, when the number of subunits is respectively 2, 3, 4, and so on. Subunits, which are identical, are protomers. A protomer obtained by dissociation of the oligomer may also be referred to as a monomer, or monomeric unit. In some enzymes, each subunit contains an active centre. We wish to discuss the kinetic behavior of enzymes with more than one active center.

We shall first consider the theory for an idealized protein which is an oligomer consisting of n identical protomers, each containing one active centre as shown in figure below.

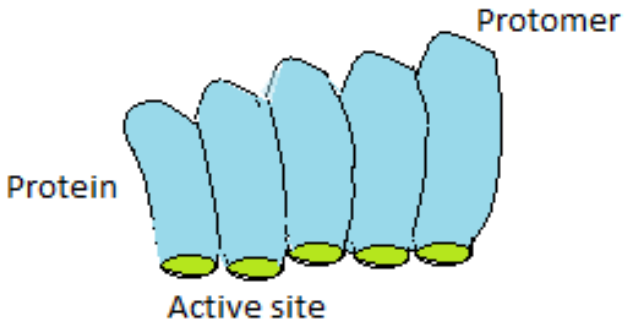


Figure: Schematic representation of an idealized protein consisting of several identical protomers, each containing one active site

The active sites are assumed to be independent of each other in their interaction with the molecule of a ligand, which we shall again call the substrate. Denote the substrate by S, and the complex of the protein combined with j ligand molecule by C_j , where j runs from 0 (meaning the protein is bare) to n (meaning the protein is fully reached with n substrate molecules). The individual reactions are represented as follows,

$$S + C_j \leftrightarrow C_{j+1}, \qquad j = 0,1,2,\cdots,n-1 \qquad \ldots(a68)$$

Denote the rate constant for binding the substrate ligand to a particular site of the protein as k_{+1} for association and k_{-1} for dissociation. We shall further assume that there is an abundance of substrate, and its concentration is varying so slowly that we may treat it as constant and set its value equal to s_0 .

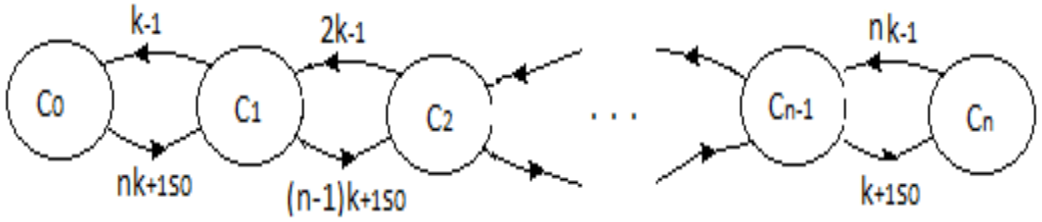


Figure: Representation by a graph of the system of reactions

As an alternative to the representation (a68) of the reactions, we shall introduce figure below, which has a one-one correspondence with the rate equations for the concentrations of C_j. Thus, an appropriately labeled circle represents a distinct protein state. Such a circle is called a node or vertex, and a reaction pathway connecting two states represented by a directed line segment or branch. Because of this correspondence, such diagrams have been studied in their own right and are called graphs.

Each vertex has a rate equation associated with it. To find it, choose a vertex and set the time rate of change of the concentration of material associated with that vertex equal to a sum of terms, one for each branch entering and one for each branch leaving the vertex. A term is the product of the branch label with the concentration of the node from which the branch emanates. The sign of the term is positive or negative, according as the arrowhead points towards or away from the chosen vertex.

Note that the rate constant in going from C_0 to C_1 is $nk_{+1}s_0$ because there are n unoccupied sites in the state C_0. It is $k_{+1}s_0$ in going from C_{n-1} to C_n because only one site is unoccupied. Similarly, in going from C_1 to C_0, there is only one site occupied, and the rate constant for dissociation is therefore k_{-1}. In going from C_n to C_{n-1}, there are n sites occupied by ligand molecules, and the rate constant is nk_{-1}. We see that the assumption of independence of binding sites leads to particularly simple interrelations among the rate constants for the various transitions.

By following the above prescription for inferring the rate equations from the graph, we obtain

$$\frac{dc_0}{dt} = -nk_{+1}s_0c_0 + k_{-1}c_1,$$

$$\frac{dc_j}{dt} = (n+1-j)k_{+1}s_0c_{j-1} - jk_{-1}c_j - (n-j)k_{+1}s_0c_j + (j+1)k_{-1}c_{j+1},$$

$$j = 1,2,\cdots,n-1$$

$$\frac{dc_n}{dt} = k_{+1}s_0c_{n-1} - nk_{-1}c_n. \qquad \ldots(a69)$$

We shall now make the steady state hypothesis, setting time derivatives in (n14) equal to zero and solve for the steady state or equilibrium values of c_j. Define the equilibrium constant K as

$$K = \frac{k_{-1}}{k_{+1}}. \qquad \ldots(a70)$$

Consider first the steady state condition $dc_n/dt = 0$. It follows that the equilibrium values c_n and c_{n-1} are related to the equation

$$c_n = \frac{1}{n}\frac{s_0}{K}c_{n-1} \qquad \ldots(a71)$$

Note that K has the dimensions of concentration, so that s_0/K is dimensionless. Going next to equations (a69) with j = n-1 and setting $dc_{n-1}/dt = 0$, we see that the terms on the right hand side vanish in pairs. Proceeding in this fashion, we find that

$$c_{n-1} = \frac{2}{n-1}\frac{s_0}{K}c_{n-2}$$

.....

$$c_2 = \frac{n-1}{2}\frac{s_0}{K}c_1 \qquad \ldots(a72)$$

$$c_1 = n\frac{s_0}{K}c_0$$

By combining the equations appearing in (a71) and (a72), we see that all the equilibrium values of c_j for $j \geq 1$ may be expressed in terms of c_0 in a regular fashion. Thus

$$c_2 = \frac{n(n-1)}{1.2}\left(\frac{s_0}{K}\right)^2 c_0$$

$$c_3 = \frac{n(n-1)(n-2)}{1.2.3}\left(\frac{s_0}{K}\right)^3 c_0$$

and so forth. In general

$$c_j = b_j^n x^j c_0, \qquad j = 0,1,2,\cdots,n \qquad \ldots(a73)$$

where the non-dimensional substrate concentration x is defined as

$$x = \frac{s_0}{K} \qquad \ldots(a74)$$

The quantity b_j^n is the binomial coefficient defined as

$$b_j^n = \frac{n!}{j!(n-j)!}, \quad j,n \text{ integers}, \ j \le n \qquad \ldots(a75)$$

The fraction of sites actually bound by the ligand is denoted by $Y(s_0)$ and is called the saturation function. It is defined by the equation

$$Y(s_0) = \frac{\sum_{j=1}^{n} jc_j}{n\sum_{j=0}^{n} c_j} \qquad \ldots(a76)$$

From equation (44) and using the value of y = 1 in the binomial theorem

$$(x+y)^n = \sum_{j=0}^{n} b_j^n x^{n-j} y^j$$

We have

$$\sum_{j=0}^{n} c_j = c_0 \sum_{j=0}^{n} b_j^n x^j = c_0(1+x)^n \qquad \ldots(a77)$$

Also, because the operations of differentiation and summation can be computed,

$$\sum_{j=1}^{n} jc_j = c_0 \sum_{j=1}^{n} jb_j^n x^j = c_0 x \frac{d}{dx}\sum_{j=0}^{n} b_j^n x^j = c_0 x(1+x)^n$$

Or

$$\sum_{j=1}^{n} jc_j == c_0 xn(1+x)^{n-1} \qquad \ldots(a78)$$

From the results (a77) and (a78), the saturation function Y defined by (a76) becomes

$$Y(s_0) = \frac{x}{1+x} = \frac{s_0}{K+s_0} \qquad \ldots(a79)$$

As is easily verified from equation (a76) with n =1, this is exactly the same result that would be obtained if the protein were a monomer!

By the same token, suppose that the protein is an enzyme; and that in addition to the reactions (a68), the reactions

$$C_i \rightarrow C_{i-1} + P\,, \quad i = 1,2,\cdots,n \qquad \ldots(a80)$$

where P is a product, are also occurring. Let k_{+2} be the rate constant for conversion of a ligand molecule to a product molecule at a particular site. The effect of the inclusion of (a80) on the rate equation (a69) for c_i is to replace k_{-1} wherever it appears by $(k_{-1}+k_{+2})$. Then, with the assumption of a quasi-steady state for the conversion of substrate to a product, the Michaelis constant $K_m = (k_{-1}+k_{+2})/k_{+1}$ replaces the equilibrium constant K in equation (a80) and those following it. The velocity of the reaction v_0 is given by the expression

$$v_0 = k_{+2}\sum_{j=1}^{n} jc_j = V\frac{\sum_{j=1}^{n} jc_j}{n\sum_{j=0}^{n} c_j} \qquad \ldots(a81)$$

where by definition, $V = nk_{+2}e_0$, and e_0 is the total initial enzyme concentration,

$$e_0 = \sum_{j=0}^{n} c_j\,. \qquad \ldots(a82)$$

Hence, from (a76) and (a79), with K_m replacing K,

$$v_0 = \frac{Vs_0}{K_m + s_0} \qquad \ldots(a83)$$

From this result we see that, as far as the reaction velocity with respect to a given substrate is concerned, an oligomeric enzyme that consists of an arbitrary number of independent identical protomeric units is indistinguishable from an enzyme that is a monomer. An oligomeric enzyme that possesses a reaction

velocity that is different from the monomeric result (a83) is said to be a cooperative system. Hence, the idealized oligomeric enzyme or protein described above can be called a non-cooperative system, or its behavior can be described as non-cooperative, it can be said to display zero cooperativity.

The Cooperative Dimer

We shall consider the simplest example of an oligomeric cooperative system, a cooperative dimer. Let each subunit of the dimer contain an active centre. The enzyme molecule, in interaction with a substrate, has three possible states: E, a free molecule; C_1, with one centre free and the other occupied by a substrate molecule; C_2, with both centres occupied by substrate molecules. The reactions are represented as

$$S + E \underset{k_{-1}}{\overset{k_{+1}}{\leftrightarrow}} C_1 \overset{k_{+2}}{\rightarrow} E + P$$

$$S + C_1 \underset{k_{-3}}{\overset{k_{+3}}{\leftrightarrow}} C_2 \overset{k_{+4}}{\rightarrow} E + P \qquad \ldots(a84)$$

Because the sites are no longer independent, there is no simple relationship between k_{+1} and k_{+3}, or between k_{-1} and k_{-3}, as there was for a non-cooperative system. We shall invoke the quasi-steady state hypothesis as in Michaelis-Menten theory. Hence, we treat the substrate concentration as constant and equal to its initial value s_0. The graph representation of these reactions is then given in the Figure.

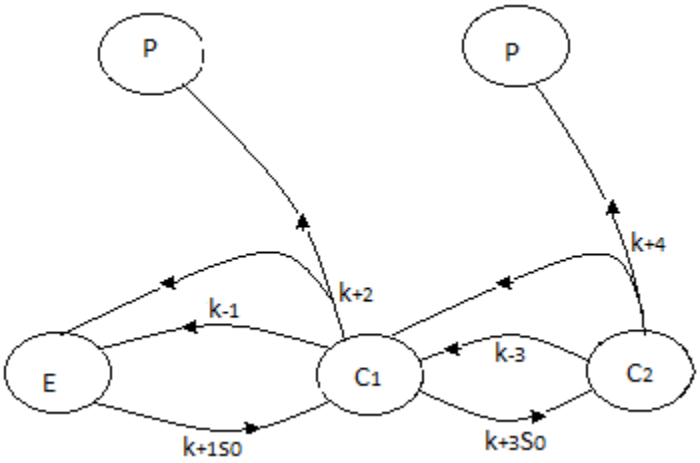


Figure: The dimer enzyme interacting with a substrate

By following the prescription for writing down the rate question from such a graph (see equation (a69)), we find that the rate equations for the concentrations e, c_1, c_2, and p are

$$\frac{de}{dt} = -k_{+1}s_0 e + (k_{-1} + k_{+2})c_1,$$

$$\frac{dc_1}{dt} = k_{+1}s_0 e - (k_{-1} + k_{+2} + k_{+3}s_0)c_1 + (k_{-3} + k_{+4})c_2, \qquad \ldots(a85)$$

$$\frac{dc_2}{dt} = k_{+3}s_0 c_1 - (k_{-3} + k_{+4})c_2,$$

$$\frac{dp}{dt} = k_{+2}c_1 + k_{+4}c_2. \qquad \ldots(a86)$$

In addition, because of the conservation of enzyme molecules,

$$e_0 = e + c_1 + c_2 \qquad \ldots(a87)$$

To apply Michaelis-Menten theory to this system, we impose the steady state condition on the enzyme and its complexes that is we set $de/dt = dc_1/dt = dc_2/dt = 0$. In addition, we denote dp/dt by v_0, the quasi-steady state initial reaction velocity. Equations (a85) become a set of three homogeneous linear algebraic equations for e, c_1 and c_2. Because of homogeneity, these three equations can only determine e, c_1 and c_2 up to a constant factor. We therefore replace one of the set (a85) by (a87), which contains the inhomogeneous term e_0. We therefore solve the steady-state values of e, c_1 and c_2 uniquely obtaining the result

$$e = e_0\left[1 + \frac{s_0}{K_m} + \frac{s_0^2}{K_m K'_m}\right]^{-1}, \qquad \ldots (a87a)$$

$$c_1 = \frac{e_0 s_0}{K_m}\left[1 + \frac{s_0}{K_m} + \frac{s_0^2}{K_m K'_m}\right]^{-1}, \qquad \ldots (a87b)$$

$$c_2 = \frac{e_0 s_0^2}{K_m K'_m}\left[1 + \frac{s_0}{K_m} + \frac{s_0^2}{K_m K'_m}\right]^{-1}, \qquad \ldots (a87c)$$

where

$$K_m = \frac{k_{-1} + k_{+2}}{k_{+1}}, \qquad \ldots (a87d)$$

$$K'_m = \frac{k_{-3} + k_{+4}}{k_{+3}} \qquad \ldots \text{(a87e)}$$

By substituting equation (a87b) and (a87c) into (a86), the reaction velocity v_0 is determined to be

$$v_0 = \frac{e_0 s_0 \left(k_{+2} K'_m + k_{+4} s_0\right)}{\left[K_m K'_m + K'_m s_0 + s_0^2\right]} \qquad \ldots\text{(a88)}$$

Let $V = K_{+2} e_0$, $K_{+4} e_0 = \alpha k + 2$, $K'_m = \beta K_m$ and $K_m / s_0 = \xi$. The equation (a88) can be written as

$$\frac{1}{v_0} = \frac{\beta\xi^2 + \beta\xi + 1}{V(\beta\xi + \alpha)} \qquad \ldots\text{(a89)}$$

Hence cooperativity is accounted for by the assignment of suitable values to the two parameters α and β. The parameter β is to some extent a relative measure of the binding of substrate to enzyme in the two states C_1 and C_2, while α is a relative measure of product information from the two states C_1 and C_2. The dependence of v_0 on the rate constants associated with the state C_2 appears above in these two parameters only. By differentiation of equation (a89) with respect to ξ, we compute

$$\frac{d}{d\xi}\left(\frac{1}{v_0}\right) = \frac{\beta}{V}\frac{\left(\beta\xi^2 + 2\alpha\xi + \alpha - 1\right)}{(\beta\xi + \alpha)^2}$$

$$\frac{d}{d\xi^2}\left(\frac{1}{v_0}\right) = \frac{2\beta}{V}\frac{\left(\alpha^2 + \beta(1-\alpha)\right)}{(\beta\xi + \alpha)^3}$$

In the last expression on the RHS, the denominator is always positive, and we see that the curvature depends on the sign of the quantity in the parenthesis in the numerator on the right hand side. Thus, if $\alpha < 1$, the curvature is always positive, but if $\alpha > 1$, the curvature can be positive, zero, or negative. When

$$\beta = \frac{\alpha^2}{\alpha - 1},\ \alpha > 1 \qquad \ldots\text{(a90)}$$

the curvature is zero. Substitution of (a90) into (a89) makes the denominator on the right hand side of equation (a89) a factor of the numerator, and reduces (a89) to the simple expression

$$\frac{1}{v_0} = \frac{\alpha\xi + 1}{\alpha V} \qquad \ldots\text{(a91)}$$

which is the Michaelis-Menten or zero cooperative result.

Allosteric Enzymes

Consider an enzyme that contains a ligand site, which is distinct and distant from the catalytically active site of the enzyme. A proposed mechanism of control of the catalytic activity of the enzyme is that ligand binding at the distant site affects the activity of the enzyme at the active site. The distant site is called allosteric site and the effect is called allosteric effect, because the ligand that binds at the distant site can be structurally unlike or allosteric with the substrate that binds at the active site. An enzyme exhibiting such an effect is called allosteric enzyme. By contrast, when substrate and inhibitor molecules compete for a single active site on the enzyme as in the enzyme substrate inhibitor system discussed earlier, the substrate and inhibitor molecules are structurally similar or isosteric.

The allosteric effect is presumed to arise because of a reversible conformational change in the enzyme: A change in the folding of the polypeptide chain, called an allosteric transition. The ligand that binds at the allosteric site is called an effector or modifier. If the modifier increases the activity or binding of the substrate, it is said to be activator. If the modifier decreases the activity of the substrate, it is called an inhibitor.

If the modifier molecule is the same as the ligand that binds at the active site, the allosteric effect is said to be homotropic. If the modifier and substrate molecules are different, the effect is said to be heterotropic. Most allosteric proteins are oligomers. Furthermore, allosteric enzymes exhibiting cooperative (homotrophic) effects almost invariably exhibit heterotropic effects as well.

Because allosteric enzymes are recognized by their anomalous kinetic behavior, in particular the appearance of a sigmoidal curve for $v_0(s_0)$, the descriptive term allosteric is often used to mean that the enzyme displays such a sigmoidal curve. Clearly, in the case when the ligands are identical, the allosteric effect is a mechanism by which an oligomer can achieve cooperative behavior. In other words, conformational change in the enzyme explains how distant active sites can interact during binding with substrate.

The theory assumes the following properties for the allosteric protein:

a) The protein is an oligomer made up of n identical protomers
b) Each protomer has only one active site that is able to combine with a given ligand.

c) The active sites associated with a given ligand are independent of each other in their interactions with the ligand.
d) The protomer exists in two states, each of which has a different affinity for the ligand.
e) The protein can exist only in either of two states, in which the subunits are all in one or the other of their two states.

The allosteric effect is a consequence of the fact that the protein can exist in two different states. This is presumed due to conformational change in the protein.

We denote the two states of the protein by R and T and the ligand molecule by S. The subscript j is attached to R or T will denote the presence of j ligands attached to the protein in either state. Thus, j can vary from 0 to n, and the subscript 0 represents the protein with no occupied sites. We now impose the steady state hypothesis, and assume that the protein complexes are in equilibrium with a large number of ligand molecules of concentration s_0, which is treated as constant. The possibility of transformation of ligand molecules into product molecules is initially ignored, so that a true steady state is assumed to exist. Furthermore, it is assumed that the states R_0 and T_0 can convert from one to the other and are in equilibrium, with an equilibrium constant L, the allosteric constant. Thus, the possible reactions are written as

$$\left.\begin{aligned} S + R_j &\leftrightarrow R_{j+1} \\ S + T_j &\leftrightarrow T_{j+1} \end{aligned}\right\} \quad j = 0,1,2,\cdots,n-1 \qquad \ldots(a92)$$

$$R_0 \leftrightarrow T_0$$

Let us denote the rate constant for binding the ligand molecule to a particular site of the protein in the R state by k_{+1} for association, and k_{-1} for dissociation. Similarly, we denote by k_{+3} and k_{-3} the corresponding rate constants for the binding of the ligand molecule to the protomer of the protein in the T state, the rate constants for transformation for the R_0 to the T_0 state and back are denoted by k_{+0} and k_{-0}, respectively.

The graph for the process is shown in figure below. We see that it is composed of two subgraphs representing the R and T systems, each consisting of n+1vertices, which are joined together via the interaction between the states R_0 and T_0. We remark that the theory is presented with the conformational change from the R state to the T state occurring only when there are no ligand molecules attached, that is $R_j \leftrightarrow T_j$ for $j \neq 0$ is not allowed.

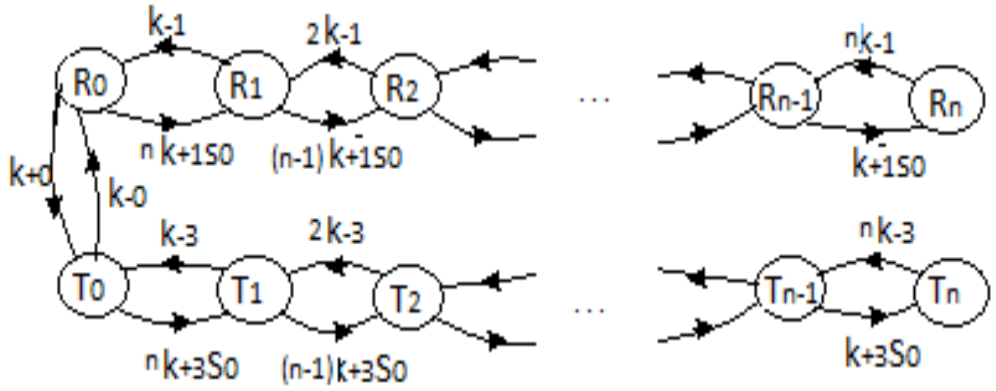


Figure: Graph of the allosteric reaction system

We can now write the rate equations corresponding to the graph, according to the prescription given before. Denoting concentrations by lower case letters, we obtain

$$\frac{dr_0}{dt} = -k_{+0}r_0 + k_{-0}t_0 - nk_{+1}s_0r_0 + k_{-1}r_1$$

$$\frac{dr_j}{dt} = (n+1-j)k_{+1}s_0r_{j-1} - jk_{-1}r_j - (n-j)k_{+1}s_0r_j + (j+1)k_{-1}r_{j+1},$$

$$\frac{dc_n}{dt} = k_{+1}s_0r_{n-1} - nk_{-1}r_n. \qquad \ldots(a93)$$

$$\frac{dt_0}{dt} = k_{+0}r_0 - k_{-0}t_0 - nk_{+3}s_0t_0 + k_{-3}t_1$$

$$\frac{dt_j}{dt} = (n+1-j)k_{+3}s_0t_{j-1} - jk_{-3}t_j - (n-j)k_{+3}s_0t_j + (j+1)k_{-3}t_{j+1},$$

$$\frac{dt_n}{dt} = k_{+3}s_0 t_{n-1} - nk_{-3}t_n \cdot \quad ; \qquad j = 1,2,\cdots,n-1$$

At equilibrium, we set the time derivative in (a93) equal to zero and solve for the equilibrium values of r_j and t_j. We define the equilibrium constants K_R, K_T and L as

$$K_R = \frac{k_{-1}}{k_{+1}},$$

$$K_T = \frac{k_{-3}}{k_{+3}}, \qquad \ldots(a94)$$

$$L = \frac{k_{+0}}{k_{-0}}$$

We recognize that either the R states or the T states by themselves constitute a non-cooperative system. In combining these two systems, only the transition between R_0 and T_0 has been added. Using the same procedure as done previously, we see that

$$r_j = b_j^n x^j r_0$$

$$t_j = b_j^n (cx)^j t_0$$

$$t_0 = Lr_0 \qquad \ldots(a95)$$

$$x \equiv \frac{s_0}{K_R}$$

$$c \equiv \frac{K_R}{K^T}$$

Suppose now that the allosteric transition can occur, regardless of the number of ligand molecules attached to the protein. Then equations (a92) are supplemented by the reactions

$$R_j \leftrightarrow T_j \quad ; j = 1,2,\cdots,n \qquad \ldots(a96)$$

and branches connecting R_j and Tj directly appear in the graph of figure above. Consider for definiteness the transition

$$R_1 \underset{k_{-5}}{\overset{k_{+5}}{\leftrightarrow}} T_1 \qquad \ldots(a96a)$$

in which we have assigned the rate constant k_{+5} for the forward reaction and k_{-5} for the backward reaction. Permitting this transition to occur would appear to have the consequence that the relationship between the equilibrium values of r_1 and t_1 would be altered, and the additional equilibrium constant $K \equiv k_{+5}/k_{-5}$ would appear in the theory, but that is not the case. The justification for this statement is a consequence of the thermodynamic principle of detailed balancing. This principle, applied to a chemical reaction in a state of equilibrium, states that the frequency of transitions from one molecular equilibrium state to another equilibrium state is equal to that in the reverse direction. It follows that, for a cyclic set of reactions in equilibrium, the product of the rate constants taken in a clockwise direction is equal to the product of the rate constants taken in a counterclockwise direction.

Now from the graph above with the transition given in (a96a) included, and consider the loops formed by the branches connecting the states R_0, T_0, R_1, and T_1. We infer from the above principle that $k_{+0}k_{+3}k_{-5}k_{-1} = k_{-0}k_{+1}k_{+5}k_{-3}$, or equivalently with the aid of equations (a94) and (a95) that

$$\frac{k_{+5}}{k_{-5}} = \frac{k_{+0}k_{-1}k_{+3}}{k_{-0}k_{+1}k_{-3}}, \quad K_5 = Lc \qquad \ldots(a97)$$

Similar considerations apply to the transitions $R_j \leftrightarrow T_j$ for $j > 1$.

Therefore the relationship between the equilibrium values of r_1 and t_1 given by equations (a95) as

$$\frac{t_1}{r_1} = Lc \qquad \ldots(a98)$$

is unaltered by the inclusion of the reaction (a96), and is the same as the equilibrium condition $t_1/r_1 = K_5$. The same considerations apply to the reactions (a96) for j >1, so that the equilibrium relation between r_j and t_j, which according to equations (a95) is

$$\frac{t_j}{r_j} = Lc^j \;; \quad j = 0,1,2,\cdots,n \qquad \ldots(a99)$$

is unaltered by the inclusion of the reaction (a96). Consequently, the above equilibrium relations have a more general significance than the simplified reaction scheme (a92) indicates, and the graph above could equally well have included the transitions (a96).

The saturation function, which is the fraction of all sites to which ligands are bound, is defined by the expression

$$Y(s_0) = \frac{\sum_{j=1}^{n} j(r_j + t_j)}{n\sum_{j=0}^{n}(r_j + t_j)} \qquad \ldots(a100)$$

By using (a95) in (a100) and proceeding as earlier, we find that

$$Y = \frac{Lcx(1+cx)^{n-1} + x(1+x)^{n-1}}{L(1+cx)^n + (1+x)^n} \qquad \ldots(a101)$$

The curve for Y(x) may be sigmoidal. If so, it is taken to the evidence for the allosteric behavior of the system.

When c =1 (the affinity of both states towards the ligand is the same), or when $L \to 0$ or $L \to \infty$, the protein exists exclusively in one state consisting of n identical, independent monomeric units, namely, a non-cooperative system. If the first two of these limiting cases, it is seen from (a101) that the saturation function reduces to the simplified expression

$$Y = \frac{x}{1+x} = \frac{s_0}{K_R + s_0} \qquad \ldots \text{ (a102)}$$

In the third case, for which $L \to \infty$, Y again reduces to the above expression with K_R replaced by K_T. Equation (a102) is exactly the same result that is obtained if the protein is a monomer, as we expect.

When c is negligible, (a101) reduces to the expression

$$Y = \frac{x(1+x)^{n-1}}{L + (1+x)^n} \qquad \ldots(a103)$$

Numerical calculations based on equation (a101) shows that Y(x) is sigmoidal when L assumes a large value. If L is large, it means that the bare protein prefers to exist in the T state than the R state. The sigmoidal character of Y is accentuated when c<<1.

Hemoglobin

Although hemoglobin is not an enzyme, it is good example of a cooperative system, and therefore serves as a model with which to test theoretical concepts of cooperative behavior.

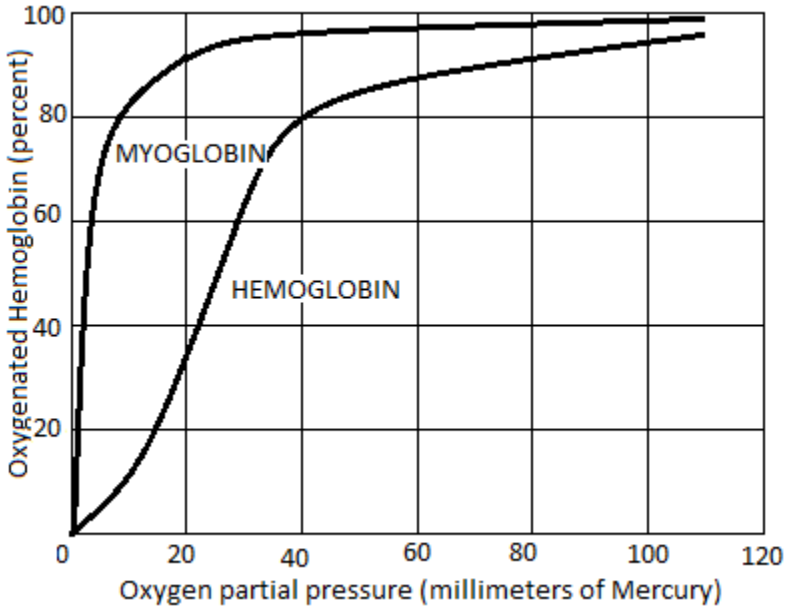


Figure: Hemoglobin and myoglobin verses oxygen partial pressure

More is known about myoglobin and hemoglobin, iron-bearing macro-molecules, than any other proteins. Myoglobin is found in muscle, and its function is to transport and store oxygen for subsequent use in an oxidative process. Hemoglobin exists normally in the blood, entirely within the red cell. The principal function of the red cell in fact is to contain hemoglobin. The chemical action of hemoglobin is to combine with oxygen in the lungs to form oxyhemoglobin. The latter gives up oxygen in the tissue to form hemoglobin again. Hemoglobin also combines with CO_2 in the tissues and releases it to the lungs. These properties are better understood by examining the saturation function representing the fractional oxygenation of hemoglobin as a function of

the partial pressure of oxygen in the air surrounding it. The saturation function is a directly observable quantity. Whereas the saturation curve for myglobin has the form of a rectangular hyperbola, the curve for hemoglobin is S-shaped as shown in Figure.

Thus the partial pressure of oxygen in the lungs where oxygen is taken up by hemoglobin is about 100mmHg. At this pressure, the hemoglobin is about 98% saturated, and its saturation is but slightly affected by changes in oxygen content of the air, because the saturation curve is relatively flat there. In the veins or tissue, the oxygen partial pressure is about 40mmHg, which is near the steep portion of the saturation curve, and represents a saturation value of about 75%. Furthermore, any abnormal demands for oxygen on the part of the tissue will deplete the oxygen partial pressure there to, say, 20mmHg, and the oxyhemoglobin will readily give up its oxygen, reducing its saturation to about 35%. At this value of the oxygen partial pressure, the myoglobin saturation value is about 90%. Hence, if the tissue is muscle, the oxygen will transfer itself from the hemoglobin to the myoglobin. In other words, the relative positions of the saturation curves illustrate the fact that, for a given partial pressure of oxygen the affinity of myoglobin for oxygen is very much greater than that of hemoglobin.

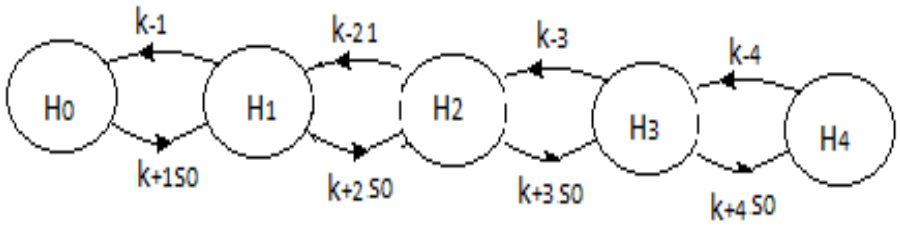


Figure: Graphical states of hemoglobin interacting with ligand of concentration s_0.

The hemoglobin molecule consists of a protein part called globin and four disk-shaped molecular rings called heme groups. Globin is composed of four polypeptide subunits, divided equally into two types of structures, called α and β. Each subunit has a heme group attached, and the latter is the active centre of the subunit. At the centre of each heme group is an iron atom. The iron atom is capable of combining with oxygen, which accounts for the functioning of hemoglobin. Myoglobin consists of only one polypeptide chain and one heme

group, and the molecular weight of myoglobin is about one fourth that of hemoglobin. The theoretical prediction of the observed shape of the hemoglobin saturation curve is an unsolved problem.

Perhaps the simplest theory of hemoglobin proposed is that it behaves like a cooperative tetramer. Thus, we assume the hemoglobin exists in five possible states H_j, where j runs from 0 to 4 and represents the number of oxygen molecules attached to the hemoglobin molecule. In other words, H_1 represents HbO_2, H_2 represents HbO_4, and so forth. The hemoglobin is assumed to be in equilibrium with oxygen, as shown by the graph given in figure below. In the graph the quantity, s_0 again represents the ligand molecule, which in the present instance is oxygen. Each transition between states possesses a different rate constant, as indicated by the branch values of the graph. In equilibrium, only the branch values will enter into the determination of the saturation function.

Let the equilibrium constants be defined as

$$K_j = \frac{k_{-j}}{k_{+j}}, \quad j = 1,2,3,4 \qquad \ldots(a104)$$

It is not necessary for us to write down the rate equations for H_j, as we can readily write down the equilibrium equations for the concentrations h_j of the hemoglobin complexes as follows

$$K_j = \frac{s_0 h_{j-1}}{h_j}, \qquad j = 1,2,3,4 \qquad \ldots(a105)$$

Therefore, the equilibrium concentrations are expressible in terms of h_0 as

$$h_1 = \frac{s_0 h_0}{K_1}, \; h_2 = \frac{{s_0}^2 h_0}{K_1 K_2}, \; h_3 = \frac{{s_0}^3 h_0}{K_1 K_2 K_3}, \; h_4 = \frac{{s_0}^4 h_0}{K_1 K_2 K_3 K_4} \qquad \ldots (a106)$$

The saturation function $Y(s_0)$ representing the fractional saturation of the available oxygen sites is defined earlier by the expression

$$Y(s_0)=\frac{\sum_{j=1}^{4} jh_j}{4\sum_{j=0}^{4} h_j} \qquad \ldots \text{(a107)}$$

Substituting equation (a106) into (a107) leads the final result

$$Y(s_0)=\frac{A_1 s_0 + 2A_2 {s_0}^2 + 3A_3 {s_0}^3 + 4A_4 {s_0}^4}{4\left(1 + A_1 s_0 + A_2 {s_0}^2 + A_3 {s_0}^3 + A_4 {s_0}^4\right)} \qquad \ldots(6.108)$$

where

$$\begin{aligned} A_1 &= K_1^{-1}, \\ A_2 &= (K_1 K_2)^{-1} \\ A_3 &= (K_1 K_2 K_3)^{-1}, \\ A_4 &= (K_1 K_2 K_3 K_4)^{-1} \end{aligned} \qquad \ldots(a109)$$

Since O_2 concentration is directly proportional to oxygen partial pressure p, s_0 in equation (a108) can be replaced by p so as to make a direct comparison with the experimental curve shown in Figure. The proportionality constant can be absorbed into the A_j, thus if $s_0 = \alpha p$, where p is the partial pressure of oxygen in the gas surrounding the hemoglobin solution and α is the constant of proportionality called the solvability coefficient. Then Y(p) is given by equation (a108) with s_0 replaced by p and

$$\begin{aligned} A_1 &= \alpha K_1^{-1}, \\ A_2 &= \alpha^2 (K_1 K_2)^{-1} \\ A_3 &= \alpha^3 (K_1 K_2 K_3)^{-1}, \\ A_4 &= \alpha^4 (K_1 K_2 K_3 K_4)^{-1} \end{aligned} \qquad \ldots(a110)$$

with four disposable constants to fit one curve, it is not surprising that the experimental curve can be adequately represented by a judicious choice of the

K_j. we emphasize that the above theory is a cooperative theory in that the binding the ligand molecule at one site facilitates or inhabits the binding of the ligand molecule at a second site and the K_j are not inter-related in a simple way.

Population Genetics

To understand how population genetics came into existence, and to appreciate its intellectual significance, a brief excursion into the history of biology is necessary. Darwin's *Origin of Species*, published in 1859, propounded two main theses: firstly, that modern species were descended from common ancestors, and secondly that the process of natural selection was the major mechanism of evolutionary change. The first thesis quickly won acceptance in the scientific community, but the second did not. Many people found it difficult to accept that natural selection could play the explanatory role required of it by Darwin's theory. This situation—accepting that evolution had happened but doubting Darwin's account of what had caused it to happen—persisted well into the twentieth century.

Opposition to natural selection was understandable, for Darwin's theory, though compelling, contained a major lacuna: an account of the mechanism of inheritance. For evolution by natural selection to occur, it is necessary that parents should tend to resemble their offspring; otherwise, fitness-enhancing traits will have no tendency to spread through a population. (For example, if fast zebras leave more offspring then slow ones, this will only lead to evolutionary change if the offspring of fast zebras are themselves fast runners.) In the *Origin*, Darwin rested his argument on the observed fact that offspring *do* tend to resemble their parents—'the strong principle of inheritance'—while admitting that he did not know why this was. Darwin did later attempt an explicit theory of inheritance, based on hypothetical entities called 'gemmules', but it turned out to have no basis in fact.

Darwin was deeply troubled by not having a proper understanding of the inheritance mechanism, for it left him unable to rebut one of the most powerful objections to his overall theory. For a population to evolve by natural selection, the members of the population must vary—if all organisms are identical, no selection can occur. So for selection to gradually modify a population over a long period of time, in the manner suggested by Darwin, a continual supply of

variation is needed. This was the basis for Fleeming Jenkins' famous objection to Darwin, namely that the available variation would be used up too fast. Jenkins' reasoning assumed a 'blending' theory of inheritance, i.e. that an offspring's phenotypic traits are a 'blend' of those of its parents. (So for example, if a short and a tall organism mate, the height of the offspring will be intermediate between the two.) Jenkins argued that given blending inheritance, a sexually reproducing population would become phenotypically homogenous in just a few generations, far shorter than the number of generations needed for natural selection to produce complex adaptations.

Fortunately for Darwin's theory, inheritance does not actually work the way Jenkins thought. The type of inheritance that we call 'Mendelian', after Gregor Mendel, is 'particulate' rather than 'blending'—offspring inherit discrete hereditary particles (genes) from their parents, which means that sexual reproduction does not diminish the heritable variation present in the population. However, this realization took a long time to come, for two reasons. Firstly, Mendel's work was overlooked by the scientific community for forty years. Secondly, even after the rediscovery of Mendel's work at the turn of the twentieth century, it was widely believed that Darwinian evolution and Mendelian inheritance were incompatible. The early Mendelians did not accept that natural selection played an important role in evolution, so were not well placed to see that Mendel had given Darwin's theory the lifeline it needed. The synthesis of Darwinism and Mendelism, which marked the birth of modern population genetics, was achieved by a long and tortuous route.

The key ideas behind Mendel's theory of inheritance are straightforward. In his experimental work on pea plants, Mendel observed an unusual phenomenon. He began with two 'pure breeding' lines, one producing plants with round seeds, the other wrinkled seeds. He then crossed these to produce the first daughter generation (the F_1 generation). The F_1 plants all had round seeds—the wrinkled trait had disappeared from the population. Mendel then crossed the F_1 plants with each other to produce the F_2 generation. Strikingly, approximately one quarter of the F_2 plants had wrinkled seeds. So the wrinkled trait had made a comeback, skipping a generation.

These and similar observations were explained by Mendel as follows. He hypothesised that each plant contains a pair of 'factors' that together determine some aspect of its phenotype—in this case, seed shape. A plant inherits one

factor from each of its parents. Suppose that there is one factor for round seeds (R), another for wrinkled seeds (W). There are then three possible types of plant: RR, RW and WW. An RR plant will have round seeds, a WW plant wrinkled seeds. What about an RW plant? Mendel suggested that it would have round seeds—the R factor is 'dominant' over the W factor. The observations could then be easily explained. The initial pure-breeding lines were RR and WW. The F_1 plants were formed by RR × WW crosses, so were all of the RW type and thus had round seeds. The F2 plants were formed by RW × RW crosses, so contained a mixture of the RR, RW and WW types. If we assume that each RW parent transmits the R and W factors to its offspring with equal probability, then the F2 plants would contain RR, RW and WW in the ratio 1:2:1. (This assumption is known as *Mendel's First Law* or *The Law of Segregation*.) Since RR and RW both have round seeds, this explains why three quarters of the F2 plants had round seeds, one quarter wrinkled seeds.

Obviously, our modern understanding of heredity is vastly more sophisticated than Mendel's, but the key elements of Mendel's theory—discrete hereditary particles that come in different types, dominance and recessiveness, and the law of segregation—have turned out to be essentially correct. Mendel's 'factors' are the genes of modern population genetics, and the alternative forms that a factor can take (e.g. R versus W in the pea plant example) are known as the *alleles* of a gene. The law of segregation is explained by the fact that during gametogenesis, each gamete (sex cell) receives only one of each chromosome pair from its parent organism. Other aspects of Mendel's theory have been modified in the light of later discoveries. Mendel thought that most phenotypic traits were controlled by a single pair of factors, like seed shape in his pea plants, but it is now known that most traits are affected by many pairs of genes, not just one. Mendel believed that the pairs of factors responsible for different traits (e.g. seed shape and flower colour) segregated independently of each other, but we now know that this need not be so. Despite these points, Mendel's theory marks a turning point in our understanding of inheritance.

The rediscovery of Mendel's work in 1900 did not lead the scientific community to be converted to Mendelism overnight. The dominant approach to the study of heredity at the time was biometry, spearheaded by Karl Pearson in London, which involved statistical analysis of the phenotypic variation found in natural populations. Biometricians were mainly interested in continuously varying traits

such as height, rather than the 'discrete' traits such as seed shape that Mendel studied, and were generally believers in Darwinian gradualism. Opposed to the biometricians were the Mendelians, spearheaded by William Bateson, who emphasized discontinuous variation, and believed that major adaptive change could be produced by single mutational steps, rather than by cumulative natural selection à la Darwin. A heated controversy between the biometricians and the Mendelians ensued. As a result, Mendelian inheritance came to be associated with an anti-Darwinian view of evolution.

Population genetics as we know it today arose from the need to reconcile Mendel with Darwin, a need which became increasingly urgent as the empirical evidence for Mendelian inheritance began to pile up. The first significant milestone was R.A. Fisher's 1918 paper, 'The Correlation between Relatives on the Supposition of Mendelian Inheritance', which showed how the biometrical and Mendelian research traditions could be unified. Fisher demonstrated that if a given continuous trait, e.g. height, was affected by a large number of Mendelian factors, each of which made a small difference to the trait, then the trait would show an approximately normal distribution in a population. Since the Darwinian process was widely believed to work best on continuously varying traits, showing that the distribution of such traits was compatible with Mendelism was an important step towards reconciling Darwin with Mendel.

The full reconciliation was achieved in the 1920s and early 30s, thanks to the mathematical work of Fisher, Haldane and Wright. Each of these theorists developed formal models to explore how natural selection, and other evolutionary forces such as mutation, would modify the genetic composition of a Mendelian population over time. This work marked a major step forward in our understanding of evolution, for it enabled the consequences of various evolutionary hypotheses to be explored quantitatively rather than just qualitatively. Verbal arguments about what natural selection could or could not achieve, or about the patterns of genetic variation, to which it could give rise, were replaced with explicit mathematical arguments. The strategy of devising formal models to shed light on the process of evolution is still the dominant research methodology in contemporary population genetics.

There were important intellectual differences between Fisher, Haldane and Wright, some of which have left legacies on the subsequent development of the subject. One difference concerned their respective attitudes towards natural

selection. Fisher and Haldane were both strong Darwinians—they believed that natural selection was by far the most important factor affecting a population's genetic composition. Wright did not downplay the role of natural selection, but he believed that chance factors also played a crucial role in evolution, as did migration between the constituent populations of a species. A related difference is that Wright emphasized epistasis, or non-additive interactions between the genes within a single genome, to a much greater extent than Fisher or Haldane. Despite these differences, the work of all three was remarkably consonant in overall approach.

Deoxyribonucleic acid (DNA) - a large double stranded helical molecule, with rings made from the four base pairs adenine (A), cytosine(C), thymine (T), and guanine (G)-carries inherited genetic information. The ordering of the base pairs A, C, T and G determines the DNA sequence. A gene is particular DNA sequence that is the fundamental unit of heredity fpr a particular trait. Some species develop as diploids, carrying two copies of every gene, one from each parent, and some species develop as haploids, with only one copy, there are even species that develop as both haploids and diploids.

Consider the pea plant, which develops as a diploid, when we say there is a gene for pea colour, say, we mean there is a particular DNA sequence that may vary in a pea plant population, and that there are at least two subtypes, called alleles, where plants with two copies of the yellow colour allele have yellow peas, those with two copies of the green allele have green peas. A plant with two copies same allele is homozygous for that particular gene (or a homozygote), while a plant carrying two different alleles is heterozygous (or a heterozygote). For the pea colour gene, a plant carrying both a yellow and a green colour allele has yellow peas. We say that the green colour is a recessive trait (or the green colour allele is recessive) and the yellow colour is a dominant trait (or yellow colour allele is dominant).

The combination of alleles carried by the plant is called its genotype, while the actual trait (green or yellow peas) is called its phenotype. A gene that has more than one allele in a population is called polymorphic and we say the population has a polymorphism for that particular gene.

Allele frequencies in a population can change due to the influence of four primary evolutionary forces: natural selection, genetic drift, mutation and migration.

Haploid Genetics

We first consider the modeling of selection in a population of haploid organisms. Selection is modelled by fitness coefficients with different genotypes having different fitness's. We begin with a simple model that counts the number of individuals in the next generation and then show how this model can be reformulated in terms of allele frequencies and relative fitness coefficient.

Genotype	A	A
Number	n_A	n_a
Viability fitness	g_A	g_a
Fertility fitness	f_A	f_a

Table**: Haploid Genetics using population size, absolute viability and fertility fitness**

Table-8 above formulates the basic model. We assume that there are two alleles A and a for a particular haploid gene. These alleles are carried in the population by n_A and n_a individuals respectively. A fraction $g_A(g_a)$ of individuals carrying allele A(a) is assumed to survive to reproduction age and those that survive contribute $f_A(f_a)$ offspring to the next generation. These are of course average values, but under the assumption of an (almost) infinite population, the modelis deterministic. Accordingly with ${n_A}^{(i)}({n_a}^{(i)})$ representing the number of individuals carrying allele A(a) in the ith generation and formulating a discrete generation model we have

$$ {n_A}^{(i+1)} = f_A g_A {n_A}^{(i)} , \quad {n_a}^{(i+1)} = f_a g_a {n_a}^{(i)} \qquad \ldots (c1) $$

It is mathematically easier and more transparent to work with allele frequencies rather than individual numbers. We denote the frequency (or more accurately, proportion) of allele A(a) in the ith generation by $p_i(q_i)$ i.e.

$$p_i = \frac{n_A^{(i)}}{n_A^{(i+1)} + n_a^{(i+1)}}, \quad q_i = \frac{n_a^{(i)}}{n_A^{(i+1)} + n_a^{(i+1)}}$$

where evidently $p_i + q_i = 1$, now from table

$$n_A^{(i+1)} + n_a^{(i+1)} = f_A g_A n_A^{(i)} + f_a g_a n_a^{(i)} \qquad \ldots(c2)$$

so that dividing the first equation in (c1) by (c2) yields

$$p_{i+1} = \frac{f_A g_A n_A^{(i)}}{f_A g_A n_A^{(i)} + f_a g_a n_a^{(i)}}$$

$$= \frac{f_A g_A p_i}{f_A g_A p_i + f_a g_a q_i} = \frac{\frac{f_A g_A}{f_a g_a} \div p_i}{\frac{f_A g_A}{f_a g_a} \div p_i + q_i} \qquad \ldots(c3)$$

where the second equality comes from dividing the numerator and denominator by $n_A^{(i)} + n_a^{(i)}$ and the third equality from dividing the numerator and denominator by $f_A g_A$.

Similarly

$$q_{i+1} = \frac{q_i}{\frac{f_A g_A}{f_a g_a} \div p_i + q_i} \qquad \ldots(c4)$$

which could also be derived by using $q_{i+1} = 1 \quad p_{i+1}$.

We observe from the evolution equation for the allele frequencies, (c3 and c4) that only the relative fitness $\left(f_A g_A / f_a g_a\right)$ of the alleles matters.

Accordingly in these models we will consider only relative fitness, and we will arbitrarily set one-fitness to unity to simplify the algebra and make the final result more transparent.

Spread of a favoured allele

We consider a simple model for the spread of a favoured allele in Table -1 with s > 0.

Genotype	A	a
Frequency of gamete	P	Q
Relative fitness	$1+s$	1
Frequency after selection	$(1+s)p/w$	q/w
Normalization	$w=(1+s)p+q$	

Table1 : Haploid genetic model of the spread of a favoured allele.

Denoting

$$p'=(1+s)p/w=(1+s)p/1+sp \qquad \ldots(c5)$$

where we have used $(1+s)p+q=1+sp$, since $p+q=1$

Note that (c5) is same as (c3) with $p'=p_{i+1}$, $p=p_i$ and $(f_Ag_A/f_ag_a)=1+s$.

Fixed points of (c5) are determined from $p'=p$. We find two fixed points $p^*=0$, corresponding to a population in which allele A is absent; and $p^*=1$ corresponding to a population in which allele A is fixed. Intuitively, $p^*=0$ is unstable while $p^*=1$ is stable.

To illustrate how a stability analysis is performed analytically for a difference equation (instead of a differential equation).

Consider the general differential equation

$$p'=f(p) \qquad \ldots. (c6)$$

with $p=p^*$ a fixed point such that $p^*=f(p^*)$, we write $p=p^*+$

So that (c6) becomes

$$p^* + \ = f\left(p^* + \ \right) = f\left(p^*\right) + \ f'\left(p^*\right) + \ldots$$
$$= p^* + \ f\left(p^*\right) + \ldots$$

where $f'\left(p^*\right)$ denotes the derivative of f evaluated at p^*

Therefore, to leading order in ϵ

$$\left| \xi / \epsilon \right| = |f'(p^*)|$$

And the fixed point is stable provided that

$$|f'\left(p^*\right)|<1$$

For these haploid model

$$f\left(p\right) = \left(1+s\right)p/1+sp, \quad f'\left(p\right) = \left(1+s\right)/\left(1+sp\right)^2$$

So that $f'\left(p^*=0\right) = \left(1+s\right) > 1, \quad f'\left(p^*=1\right) = 1/\left(1+s\right) < 1$

Confirming that $p^* = 0$ is unstable while as $p^* = 1$ is unstable.

If the selection coefficient s is small, the model equation (c5) simplifies further we have

$$p = \left(1+s\right)p/1+sp = \left(1+s\right)p\left(1+sp\right)^{\ 1}$$
$$= \left(1+s\right)p\left\{1 \quad sp + o\left(s^2\right)\right\}$$
$$= p + \left(p \quad p^2\right) + o\left(s^2\right)$$

So that leading order in s is $p' \quad p = sp\left(1 \quad p\right)$

If $p' \quad p << 1$ which is valid for $s << 1$, we can approximate this difference equation by the differential equation

$$\frac{dp}{cn} = sp\left(1 \quad p\right)p' \quad p$$

which show that the frequency allele A satisfies the now very familiar logistic equation.

Although a polymorphism for this gene exists in the population as the new allele spreads, eventually A becomes fixed in the population and the polymorphism is lost.

Mutation Selection Balance

We consider a gene with two alleles: a wild type allele A and a mutant allele a. We view the mutant allele as a defective genotype, which confers on the carrier a lowered fitness $1 \quad s$ relative to the wild type. Although all mutant alleles may not have identical DNA sequences, we assume that they share in common the same phenotype of reduced fitness. We model the opposing effects of two evolutionary forces: natural selection, which favours the wildtype allele over the mutant allele a, and mutation, which confers a small probability u that allele A mutates to allele a in each newborn individual.
Schematically

$$A \underset{s}{\overset{u}{\Leftrightarrow}} a$$

where u represents mutation and s represents selection.

The model is shown in table.

Genotype	A	A
Frequency of gamete	P	Q
Relative fitness	1	$1 \quad s$
Frequency after selection	p/w	$(1 \quad s)q/w$
Frequency after mutation	$(1 \quad u)p/w$	$(1 \quad s)q+up/w$
Normalization		$w = p+(1 \quad s)q$

The equations for p and q in the next generation are

$$p' = \left(1 \quad u\right) p / w = \left(1 \quad u\right) p / \left(1 \quad s\left(1 \quad p\right)\right) \qquad \text{.... (c7)}$$

$$q' = \left(1 \quad s\right) q + up / w = \left(1 \quad s \quad u\right) q + u / \left(1 \quad sq\right) \qquad \text{.... (c8)}$$

where we have used $p + q = 1$ to eliminate q from the equation p' and p from the equation for q'. The equations for p' and q' are linearly dependent since $p' + q' = 1$ and we now solve only one of them.

Considering (c7) the fixed points determined from $p' = p$ are $p^* = 0$ for which the mutant allele a is fixed in the population and there is no polymorphism and the solution to

$$1 \quad s\left(1 \quad p^*\right) = 1 \quad u$$

which is $p^* = 1 \quad u / s$ and there is a polymorphism.

The stabilities of these two fixed points are determined by considering $p' = f\left(p\right)$ with $f\left(p\right)$ given by the R.H.S of (c7).

Taking the derivative of f

$$f\left(p\right) = \frac{\left(1 \quad u\right)\left(1 \quad s\right)}{1 \quad s\left(1 \quad p\right)^2}$$

so that

$$f\left(p^* = 0\right) = \left(1 \quad u\right) / \left(1 \quad s\right),\ f\left(p^* = 1 \quad u / s\right) = \left(1 \quad s\right) / \left(1 \quad u\right)$$

Applying the criterion $|f\left(p^*\right)| < 1$ for stability;

$p^* = 0$ is stable for $s < u$.

and $p^* = 1 \quad u / s$ is stable for $s > u$.

A polymorphism is therefore possible under mutation – selection balance when $s > u > 0$.

Diploid Genetics

Most sexually reproducing species are diploid. In particular, these species homosapeins is diploid with two exceptions: we are haploid at the gamete stage (sperm and unfertilized egg) and males are haploid for most genes on the un-matched X and Y sex chromosomes (females are XX and diploid). This later seemingly innocent fact is of great significance to males suffering from genetic diseases due to an X linked recessive mutation inherited from their mother. Females inheriting this mutation are most probably disease free because of the functional gene inherited from their father.

A polymorphism gene with alleles A and a can appear in a diploid gene as three distinct genotypes: AA, Aa, aa. Conventionally we denote A to be the wild type allele and a, the mutant allele. Table below represents the terminology of diploidy

Genotype	AA	Aa	aa
Referred to as	Wild type homozygote	Heterozygote	Mutant homozygote
Frequency	P	Q	R

As for haploid genetics, we will determine evolution equations for allele and / or genotype frequencies. To develop the appropriate definitions and relations we initially assume a population of size N (which we will later take to be infinite), and assume that the number of individuals with genotypes AA, Aa, aa are N_{AA}, N_{Aa}, N_{aa}, now $N = N_{AA} + N_{Aa} + N_{aa}$.

Define genotype frequencies P, Q and R as

$$P = \frac{N_{AA}}{N},\ Q = \frac{N_{Aa}}{N},\ R = \frac{N_{aa}}{N}$$

So that $P + Q + R = 1$.

It will also be useful to define allele frequencies. Let n_A and n_a be the number of alleles A and a in the population with $n = n_A + n_a$ the total number of alleles. Since the population is of size N and diploidy, $n = 2N$ and since each homozygote contains two identical alleles, and heterozygote contains one of each allele, $n_A = 2Nn_A = 2N_{AA} + N_{Aa}$ and $n_a = 2N_{aa} + N_{Aa}$.

Defining the allele frequencies p and q as previously

$$p = \frac{n_A}{n} = 2N_{AA} + N_{Aa} / 2N = P + \frac{1}{2}Q$$

and similarly

$$q = \frac{n_a}{n} = 2N_{aa} + N_{Aa} / 2N = R + \frac{1}{2}Q$$

with five frequencies P, Q, R, p, q and four constraints

$$P + Q + R = 1,\ p + q = 1,$$

$$p = P + \frac{1}{2}Q,\ q = R + \frac{1}{2}Q$$

How many independent frequencies are there? In fact, there are two because one of the four constraints is linearly dependent.

We may choose any two frequencies other than the choice $\{p, q\}$ as the linearly independent set. For instance one choice is $\{P, p\}$, then

$$q = 1 \quad p, Q = 2(p \quad P),\ R = 1 + P \quad 2p$$

Similarly, other choice is $\{P, Q\}$,

Then R = 1-P-Q, p = P + ½ Q, q = 1- P -1/2 Q .

Sexual reproduction

Diploid reproduction may be sexual and Asexual and sexual reproduction may be of varying types (e.g. random mating, various other types of assertive mating). The two simplest types to model exactly are random mating and

selfing. These mating systems are useful for contrasting the biology of both out reeding and in breeding.

Random Mating

The simplest type of sexual reproduction to model mathematically is random mating. Here, we assume a well–mixed population of individuals that have equal probability of mating with every other individual. We will determine the genotype frequencies of the zygotes (fertilized eggs) in terms of allele frequencies using two approaches:

(1) The gene pool approach

(2). The mating table approach.

The gene pool approach models sexual reproduction by assuming that males and females release their gametes into pools. Offspring genotypes are determined by randomly combining one gamete from the male pool and one gamete from the female pool. As the probability of a random gamete containing allele A or a is equal to the allele·s population frequency p or q respectively, the probability of an offspring being AA is p^2, of being Aa is 2pq (male A female a + female A male a) and of being aa is q^2.

Therefore, after a single generation of random mating, the genotype frequencies can be given in terms of the allele frequencies by

$$P = p^2, Q = 2pq, R = q^2$$

This is the celebrated Hardy-weinberg law. Notice that under the assumption of random mating, there is now only a single independent frequency, greatly simplifying the mathematical modeling,

For example, if p is taken as the independent frequency then

$$q = 1-p , \ P = p^2 , Q = 2p(1-p), R = (1-p)^2$$

Most modeling is done assuming random mating unless inbreeding influences the biology under study.

The second approach uses a mating Table-2. This approach to modeling sexual reproduction is more general and can be applied to other mating systems.

We explain this approach by considering the mating AA x Aa.The genotypes AA and Aa have frequencies P and Q respectively .The frequency of AA males mating with Aa females is PQ and is the same as AA females mating with Aa males, so the sum is 2PQ. Half of the offspring will be AA and half Aa, and the frequencies PQ are denoted under progeny frequency. The sums of all the progeny frequencies are given in the totals row, and the random mating results are recovered upon use of the relationship between the genotype and allele frequencies.

Mating	Frequency	AA	Aa	Aa
AA x AA	P^2	P^2	0	0
AA x Aa	2PQ	PQ	PQ	0
AA x aa	2PR	0	2PR	0
Aa x Aa	Q^2	¼ Q^2	½ Q^2	¼ Q^2
Aa x aa	2QR	0	QR	QR
aa x aa	R^2	0	0	R^2
Totals	$(P+Q+R)^2 = 1$	$(P+1/2Q)^2 = P^2$	$2(P+1/2Q)(R+1/2Q) = 2PQ$	$(R+1/2Q)^2 = Q^2$

Table-2: Random mating

Random Drift

Random genetic drift refers to the chance fluctuations in gene frequency that arise in finite populations; it can be thought of as a type of 'sampling error'. In many evolutionary models, the population is assumed to be infinite, or very

large, precisely in order to abstract away from chance fluctuations. But though mathematically convenient, this assumption is often unrealistic. In real life, chance factors will invariably play a role, particularly in small populations. The term 'random drift' is sometimes used in broad sense, to refer to any stochastic factors that affect gene frequencies in a population, including for example chance fluctuations in survival and mating success; and sometimes in a narrower sense, to refer to the random sampling of gametes to form the offspring generation (which arises because organisms produce many more gametes than will ever make it into a fertilized zygote). The broader sense is used here.

To understand the nature of random drift, consider a simple example. A population contains just ten organisms, five of type *A* and five of type *B*; the organisms reproduce asexually and beget offspring of the same type. Suppose that neither type is selectively superior to the other—both are equally well-adapted to the environment. However, this does not imply that the two types will produce identical numbers of offspring, for chance factors may play a role. For example, it is possible that all the type *B*s might die by accident before reproducing; in which case the frequency of *B* in the second generation will fall to zero. If so, then the decline of the *B* type (and thus the spread of the *A*type) is the result of random drift. Evolutionists are often interested in knowing whether a given gene frequency change is the result of drift, selection, or some combination of the two.

The label 'random drift' is slightly misleading. In saying that the spread of the *A* type is due to random drift, or chance, we do not mean that no cause can be found of its spread. In theory, we could presumably discover the complete causal story about why each organism in the population left exactly the number of offspring that it did. In ascribing the evolutionary change to random drift, we are not denying that there is such a causal story to be told. Rather, we mean that the spread of the *A* type was not due to its adaptive superiority over the *B* type. Put differently, the *A* and the *B*types had the same *expected* number of offspring, so were equally fit; but the *A* types had a greater*actual* number of offspring. In a finite population, actual reproductive output will almost always deviate from expectation, leading to evolutionary change.

An analogy with coin tossing can illuminate random drift. Suppose a fair coin is tossed ten times. The probability of heads on any one toss is ½, and so the *expected* frequency of heads in the sequence of ten is 50%. But the

probability of *actually* getting half heads and half tails is only 242/1024, or approximately 23.6%. So even though the coin is fair, we are unlikely to get equal proportions of heads and tails in a sequence of ten tosses; some deviation from expectation is more probable than not. In just the same way, even though the *A* and *B* types are equally fit in the example above, it is likely that some evolutionary change will occur. This analogy also illustrates the role of population size. If we tossed the coin a hundred times rather than ten, the proportion of heads would probably be very close to ½. In just the same way, the larger the population, the less important the effect of random drift on gene frequencies; in the infinite limit, drift has no effect.

Drift greatly complicates the task facing the population geneticist. In the example above, it is obviously impossible to *deduce* the composition of the population in the second generation from its composition in the first generation; at most, we can hope to deduce the probability distribution over all the possible compositions. So once drift is taken into account, no simple recurrence relation for gene frequencies, of the sort expressed in equation (1) above, can be derived. To analyse the evolutionary consequences of drift, population geneticists use a mathematical technique known as diffusion modelling, which is beyond the scope of this article; see Gillespie (2004) or Rice (2004) for good introductions. However, many of these consequences are quite intuitive, and can be understood without the mathematics.

One important effect of random drift is to decrease the degree of heterozygosity in a population over time. This happens because, given enough time, any finite population will eventually become homozygous through drift (though if the population is large, the approach to homozygosity will be slow.) It is easy to see why this is—for gene frequencies of 0 and 1 are 'absorbing boundaries', meaning that once the boundary is reached, there is no way back from it (apart from mutation). So eventually, a given allele will eventually become fixed in a population, or go extinct, the latter being the more likely fate. Indeed mathematical models show that a neutral allele arising by mutation has a very low probability of becoming fixed in a population; the larger the population, the lower the probability of fixation.

Another important effect of random drift is to cause the different subpopulations of a species to diverge genetically from each other, as the chance accumulation of alleles will probably proceed differently in each, particularly if

the alleles confer little selective advantage or disadvantage. By chance, one population may become fixed for allele A_1, while a second population becomes fixed for another allele A_2. This possibility is an important one, for if we ignore it, we may mistakenly conclude that the A_1 allele must have been advantageous in the environment of the first population, the A_2 allele in the environment of the second, i.e. that selection was responsible for the genetic differentiation. Such an explanation *might* be right, but it is not the only one—random drift provides an alternative.

The question of whether drift or selection plays a more important role in molecular evolution was much debated in the 1960s and 1970s; it lay at the heart of the heated controversy between 'selectionists' and 'neutralists' (see Dietrich 1994). The neutralist camp, headed by M. Kimura, argued that most molecular variants had no effect on phenotype, so were not subject to natural selection; random drift was the major determinant of their fate. Kimura argued that the apparently constant rate at which the amino acid sequences of proteins evolved, and the extent of genetic polymorphism observed in natural populations, could best be explained by the neutralist hypothesis (Kimura 1977, 1994). Selectionists countered that natural selection was also capable of explaining the molecular data. In recent years, the controversy has subsided somewhat, without a clear victory for either side. Most biologists believe that some molecular variants are indeed neutral, though fewer than were claimed by the original neutralists.

Spread of a favoured allele

We consider the spread of a favoured allele in a diploid population .The classic example – widely repeated in biology textbooks, as a modern example of natural selection – is the change in the frequencies of the dark and light phenotypes of the peppered moth during England's industrial revolution. The evolutionary story begins with the observation that population killed the light coloured lichen on trees during industrialization of cities. Light coloured pepper moths camouflage well on light coloured lichens but are exposed to birds on plain tree bark. On the other hand, dark coloured peper moths camouflage well on plain tree bark, but are exposed on light coloured lichens. Natural selection therefore favoured the light coloured allele in pre-industrilized England and dark coloured allele during industrialization. The dark coloured allele, presumably kept at low frequency by mutation – selection balance in pre-industrilized England.

Here we consider aa as the wildtype genotype and normalize its fitness to unity. The allele A is the mutant whose frequency increases in the population. In our example of a peppered moth, the aa phenotype is light coloured and the AA phenotype is dark coloured. The colour of the Aa phenotype depends on the relative dominance of A and a. Usually no pigment results in light colour and are a consequence of non-functioning pigment producing genes. One functioning pigment producing allele is usually sufficient to result in a dark coloured moth. With A a functioning pigment-producing allele and a the mutated non-functioning allele, a is most likely recessive, A is most likely dominant and the phenotype of Aa is most likely dark, so h ≈ 1.For the moment, though we leave h as a free parameter.

Genotype	AA	Aa	aa
Frequency of zygote	p^2	$2pq$	q^2
Relative fitness	$1+s$	$1+sh$	1
Frequency after selection	$(1+s)p^2/w$	$2(1+sh)pq/w$	q^2/w
Normalization	$w=(1+s)p^2+2(1+sh)pq+q^2$		

Table-3: A diploid genetic model of the spread of a favoured allele assuming random mating

We assume random mating, and this simplification is used to write the genotype frequencies as $P=p^2, Q=2pq, R=q^2$. Since $q=1\ \ p$ we reduce this problem to determining an equation for p' in terms of P. Using $p'=P_s+\frac{Q_s}{2}$, where $p'$is the A allele frequency in the next generation`s zygotes, and P_s and Q_s are the AA and Aa genotype frequencies, respectively, in the present generation after selection,

$$P'=(1+s)p^2+(1+sh)pq/w, \text{ where } q=1\ \ p$$

and

$$w=(1+s)p^2+2(1+sh)pq+q^2=1+s\left(p^2+2hpq\right) \qquad \ldots(c8)$$

After some algebra, the final evolution equation written solely in terms of p is

$$P'=(1+sh)p+s(1\quad h)p^2/1+2shp+s(1\quad 2h)p^2$$

The expected fixed points of this equation are $P^*=0$(unstable) and $P^*=1$(stable) where our assignment of stability assumes positive selection coefficients.

The evolution (1) in this form is not particularly illuminating. In general, a numerical solution would require specifying numerical values for s and h, as well as an initial value for p]. Here, to determine how the spread of A depends on the dominance coefficient h, we investigate analytically the increase of A assuming s ≪1.Taylor series expand the R.H.S. of (1) in powers of s, keeping terms to order s:

$$\begin{aligned} P' &= (1+sh)p+s(1\quad h)p^2/1+2shp+s(1\quad 2h)p^2 \\ &= p+s\ hp+(1\quad h)p^2\ /1+s\ 2hp+(1\quad 2h)p^2 \\ &= p+s\ hp+(1\quad h)p^2\quad 1\quad s\left(2hp+(1\quad 2h)p^2\right)+o\left(s^2\right) \qquad \ldots(c9) \\ &= p+sp\ h+1(1\quad 3h)p\quad (1\quad 2h)p^2\ +o\left(s^2\right) \end{aligned}$$

If s≪1 we expect a small change in allele frequency in each generation, so we can approximate

$p'\quad p=\frac{dp}{dn}$, where n denotes the generation number and $p=p(n)$. The approximate differential equation obtained from (c8) is

$$\frac{dp}{dn}=sp\ h+(1\quad 3h)p\quad (1\quad 2h)p^2 \qquad \ldots(c10)$$

If A is partially dominant so that h ≠ 0 (e.g.the heterozygous moth is darker than the homozygous mutant moth) then the solution to (c9) behaves similarly to the solution of logistic equation: p initially grows exponentially as $p(n)=p_0\exp(sha)$and asymptotes to one for large n. If A is recessive so that h =

0 (e.g. the heterozygous moth is as light coloured as the homozygous mutant moth), then (c9) reduces to

$$\frac{dp}{dn} = sp^2(1 \quad p) \quad \text{for } h = 0 \qquad \ldots(c11)$$

Our main interest is the initial growth of p when $p(0) = p_0 << 1$,

so that $\frac{dp}{dn} \quad sp^2$.

This differential equation may be integrated by separating variables to yield

$$p(n) = \frac{p_0}{1 \quad sp_0 n} \quad p_0(1 + sp_0 n)$$

The frequency of a recessive favoured allele increases only linearly across generations, a consequence of the heterozygote being hidden from natural selection. Most likely the peppered moth heterozygote is significantly darker than the light coloured homozygotes since the dark coloured moth rapidly increased in frequency over a short period of time.

Mutation – Selection balance

By virtue of self-knowledge, the species with the most known mutant phenotypes is homosapeins. There are thousands of known genetic diseases in humans, many of them caused by mutation of a single gene (called a monogenic disease)

Disease	Mutation	Symptoms
Thalassemia	Hemoglobin	Anemia
Sickle cell anemia	Hemoglobin	Anemia
Hemophilia	Blood clotting factor	Un-controlled bleeding
Cystic fibrosis	Chloride ion channel	Thick lung mucous
Tay-sachs disease	Hexosaminidase A enzyme	Nerve cell damage
Fragile X-syndrome	FMRI gene	Mental retardation

Hungtingtons disease	HD gene	Brain degeneration

Table-4: Lists seven common monogenic diseases

The first two diseases are maintained at significant frequencies in some human populations by heterosis. It is postulated that Taysahs disease, prevalent among ancestors of eastern European Jews, and cystic fibrosis may also have been maintained by heterosis acting in the past. (Note that cystic fibrosis gene was identified in 1989 by a Toronto group led by lap chee tsui, who later became president of the university of hong kong). The other disease genes listed may be maintained by mutation-selection balance.

Genotype	**AA**	**Aa**	**Aa**
Frequency of zygote	p^2	$2pq$	q^2
Relative fitness	1	$1 \quad sh$	$1 \quad s$
Frequency after selection	p^2 / w	$2(1 \quad sh)pq / w$	$(1 \quad s)q^2 / w$
Normalization	$w = p^2 + 2(1 \quad sh)pq + (1 \quad s)q^2$		

Table-5: Diploid genetic model of mutation selection balance in random mating

We further assume that mutations of type $A \to a$ occur in gamete production with frequency u, back mutation is neglected. The gametic frequency of A and a after selection but before mutation is given by $p^{\wedge} = P_s + Q_s/2$ and $q^{\wedge} = R_s + Q_s/2$.

And the gametic frequency of a after mutation is given by

$$q' = \left\{u\left(p^2 + (1 \quad sh)pq\right)\right\} + (1 \quad s)q^2 + (1 \quad sh)pq/w$$

where $$w = p^2 + 2(1+sh)pq + (1 \quad s)q^2$$

$$= 1 \quad sq(2hp+q)$$

Using $q = 1 \quad p$, we write the evolution equation for q' in terms of q alone. we obtain

$$q' = u + \{1 \quad u \quad sh(1+u)\}q \quad s\{1 \quad h(1+u)\}q^2/1 \quad 2shq \quad s(1 \quad 2h)q^2$$

to determine the equation solutions of (c8) we set $q^* = q' = q$ to obtain a cubic equation for q^*. Because of the neglect of back mutation in this model, one solution readily found is $q^* = 1$, in which all the A alleles have mutated to a. The $q^* = 1$ solution may be factored out of the cubic equation resulting in a quadratic equation with two solutions. Rather than show the exact result here, we determine equation solutions under two approximations

(i) $0 < u << h, s$ (ii) $0 = h < u < s$

First when $0 < u << h, s$ we look for a solution of the form $q^* = au + o(u^2)$ with a constant and Taylor's series expand in u {assuming s, $h = 0(u^0)$}. If such a solution exists, then (c8) will determine the unknown coefficient a.we have

$$au + o(u^2) = u + (1 \quad sh)au + o(u^2)/1 \quad 2shau + o(u^2)$$

$$= (1+a \quad sha)u + o(u^2)$$

and equating powers of u, we find $a = 1 + a \quad sha$

or $a = 1/sh,$

Therefore

$$q^* = u/sh + o(u^2) \text{ for } 0 < u << h, s$$

Second when $0 = h < u < s$, we substitute h = 0 directly in (c8)

$$q^* = u + (1 \quad u)q^* \quad sq^{*2}/1 \quad sq^{*2}$$

which we then write as a cubic equation for q^*

$$q^{*3} \quad q^{*2} \quad (u/s)q^* + (u/s) = 0$$

By factoring this cubic equation, we find

$$(q^* \quad 1)(q^{*2} \quad u/s) = 0$$

And the polymorphic equation solution is $q^* = \sqrt{u/s}$ for $0 = h < u < s$ because $q^* < 1$ only if $s > u.$

Note that this solution does not exist if $s < u.$

Table below summarizes the results for the equation frequencies of the genotypes at mutation –selection balance.

Genotype	AA	Aa	aa
Frequency $0 < u << s, h$	$1 + o(u)$	$2u/sh + o(u)$	$u^2/(sh)^2 + + o(u^3)$
Frequency $0 = h < u < s$	$1 + o(\sqrt{u})$	$2\sqrt{u/s} + o(u)$	u/s

Equation frequency of the genotypes at the diploid mutation selection balance

The first row of frequencies $0 < u << s, h$ corresponds to a dominant $(h = 1)$ or partially dominant $(u << h < 1)$ mutation, where the heterozygote is of reduced fitness and shows symptoms of the genetic disease. The second row of frequencies $0 = h < u < s$ corresponds to a recessive mutation where the heterozygote is symptom free. Notice that individuals carrying a dominant mutation are twice as prevalent in the population as individuals homozygous for a recessive mutation (with the same u and s).

A heterozygote carrying a dominant mutation most commonly arises either by direct mutation of allele A or by mating of a heterozygote with a wild type. The later is more common for $s << 1$, while the former must occur for $s = 1$.

Heterosis

Heterosis also called over dominance or heterozygote advantage occurs when the heterozygote has higher fitness than earlier homozygote. The best-known examples are sickling cell anemia and thalassemia, diseases that both effect hemoglobin, the oxygen carrier protein of red blood cells. The sickle cell mutations are most common in people of West African decent, while the thalassemia mutations are most common in people from the Mediterranean and Asia. In Hong Kong, the television stations occasionally play public service announcements concerning thalassemia. The heterozygote carrier of the sickle cell or thalassemia gene is healthy and resistant to malaria; the wild type homozygote is healthy, but susceptible to malaria; the mutant homozygote is sick with anemia. Table below presents the model of heterosis.

Genotype	AA	Aa	aa
Frequency of zygote	p^2	$2pq$	q^2
Relative fitness	$1 \quad s$	1	$1 \quad t$
Frequency after selection	$(1 \quad s)p^2 / w$	$2pq / w$	$(1 \quad t)q^2 / w$

Normalization	$w=(1 \quad s)p^2+2pq+(1 \quad t)q^2$

Table: A diploid genetic model of heterosis assuming random mating

Both homozygotes are of lower fitness than the heterozygote, whose relative fitness we arbitrarily set to unity.

Writing the equation p` we have

$$P'=\{(1 \quad s)p^2+pq\}/(1 \quad s)p^2 \quad tq^2$$
$$=\left(p \quad sp^2\right)/1 \quad t+2tp \quad (s+t)p^2$$

At equilibrium $p^*=p'=p,$ and we obtain a cubic equation for p^*.

$$(s+t)p^{*3} \quad (s+2t)p^{*2}+tp^*=0 \qquad \ldots \text{(c12)}$$

Evidently $p^*=0$ and $p^*=1$ are fixed points and (c9) can be factored as

$$p(1 \quad p)\{t \quad (s+t)p\}=0$$

The polymorphic solution is therefore

$$p^*=t/(s+t), \qquad q^*=s/(s+t)$$

valid when $s,t>0$. Since the value of q^* can be large, recessive mutations that cause disease, yet are highly prevalent in a population, are suspected to provide some benefit to the heterozygote.

Frequency-Dependent Selection

A polymorphism may also result from frequency-dependent selection .A well known model of frequency–dependent selection is the hawk-dove game. Most commonly, frequency – dependent selection is studied using game theory, and following john Maynard Smith, one looks for an evolutionary stable strategy (ESS). Here we study frequency–dependent selection using a population genetics model.

Consider two phenotypes: Hawk and Dove, with no mating between different phenotypes (for e.g. different phenotypes may correspond to different species, such as Hawks and Doves), we describe the Hawk – Dove game as follows

(i) When the Hawk meets Dove, Hawk gets the resource and Dove retreats before injury;
(ii) When two Hawks meet, they engage in an escalating fight, seriously risking injury and
(iii) When two Doves meet, they shear the resource.

Player/opponent	H	D
H	$E_{HH} = 2$	$E_{HD} = 2$
D	$E_{DH} = 0$	$E_{DD} = 1$

The Hawk – Dove game is modelled by a payoff matrix

The payoffs are played to the player (first column) when playing against the opponent (first row).

The player in the first column receives the payoff when playing the opponent in the first row. For instance Hawk playing Dove gets the payoff E_{HD} .The numerical values are commonly chosen such that $E_{HH} < E_{DH} < E_{DD} < E_{HD}$, i.e. Hawk playing Dove does better than Dove plying Dove does better than Dove playing Hawk does better than Hawk playing Hawk.

Frequency - dependent selection occurs because the fitness of Hawk or Dove depends on the frequency of Hawks and Doves in the population. For e.g. a Hawk in a population of Doves does well, but a Hawk in a population of Hawks does poorly

We model the Hawk –Dove game using a haploid genetic model, with p and q the population frequencies of Hawks and Doves, respectively. We assume that the time a player phenotype plays an opponent phenotype is proportional to the population frequency of the opponent phenotype. For e.g. if the population is composed of ¼ Hawks and ¾ Doves, the Hawk or Dove plays Hawk ¼ of the time and Dove ¾ of the time.

The fitness parameter K is arbitrarily, but assumed to be large and positive so that the fitness of Hawk or Dove is always positive. A negative fitness in this haploid model has no, meaning:

From table below, the equation for $p' = \left(K + pE_{HH} + qE_{HD}\right)/w$

with $w = \left(K + pE_{HH} + qE_{HD}\right)p + \left(K + pE_{DH} + qE_{DD}\right)q$

The haploid genetic model assuming frequency - dependent selection is formulated in table below

Genotype	H	D
Frequency of zygote	P	Q
Relative fitness	$K + pE_{HH} + qE_{HD}$	$K + pE_{DH} + qE_{DD}$
Frequency after selection	$\left(K + pE_{HH} + qE_{HD}\right)/w$	$\left(K + pE_{DH} + qE_{DD}\right)/w$
Normalization	$w = \left(K + pE_{HH} + qE_{HD}\right)p + \left(K + pE_{DH} + qE_{DD}\right)q$	

Haploid genetic model for frequency – selection in the Hawk-Dove game

Reference Books:

1. J. D. Murray, Mathematical Biology
2. S. I. Rubinow, Introduction to Mathematical Biology
3. J.N. Kapur, Mathematical Modeling
4. J. R. Chasnov, Lecture Notes on Mathematical Biology
5. www.wikipedia.org

***