

# Understanding Protein Structure and Function: Gateway Concepts of Physics and Mathematics

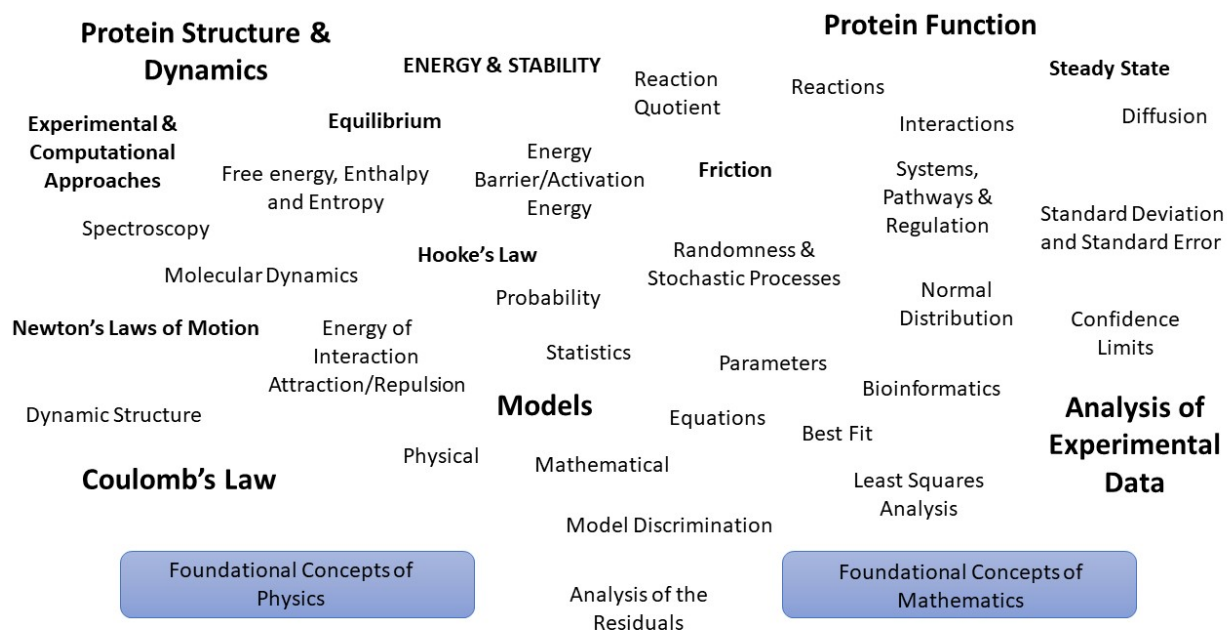
*Ellis Bell*

*Laboratory for Structural Biology, Biophysics and Bioinformatics*

*Department of Chemistry & Biochemistry*

*University of San Diego*

*San Diego, CA, 92110, USA*



## Synopsis

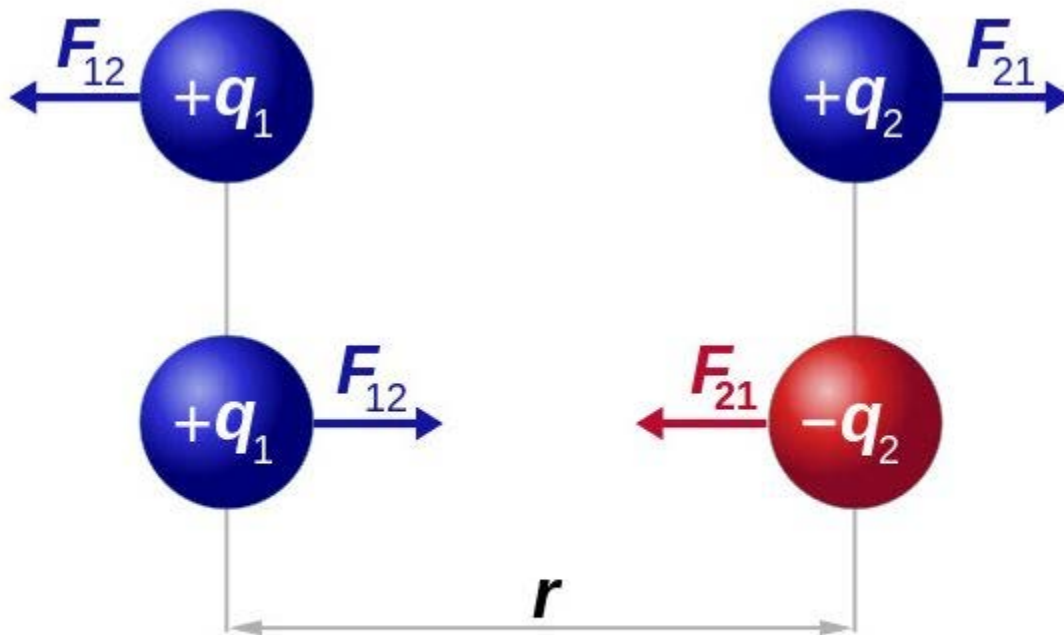
The relationship between structure and function is one of the foundational interdisciplinary concepts of science, and together with an understanding of the roles that models (both physical and mathematical) play in science requires foundational knowledge of both physics and mathematics. In addition, many experimental and computational approaches depend upon these concepts to give mechanistic insight into protein structure function relationships. Finally, establishing that the correct model is being considered to give detailed understanding comes from quantitative mathematical analysis of experimental data.

## Coulomb's Law

Interaction between charges, the charges,  $q$ , may be full, formal charges, +, -, or partial charges,  $\delta+$ ,  $\delta-$ , and from the equation below, the force between them will be attractive, opposite charges, or repulsive, like charges. The value of  $F$  being either negative, for opposite charges, or positive, for like charges. The separation between the charges is represented by  $r$ .

$$F = kq_1q_2/r^2$$

Coulomb's law is a vector equation and includes the fact that the force acts along the line joining the charges.



$$F_{12} = F_{21} = k \frac{q_1 q_2}{r^2}$$

Particularly in a biological setting you have to consider what is between the charges: with more polar compounds in effect shielding the charges and decreasing the force between them. This is represented by the Dielectric Constant,  $D$ , of the medium. In a biological setting the Dielectric Constant ranges from 80 for water to an estimated 4-8 in the hydrophobic interior of a protein, and hence has a significant effect on the force between two charges. Where the dielectric constant must be taken into account, Coulomb's Law is represented by:

$$F = kq_1q_2/D.r^2$$

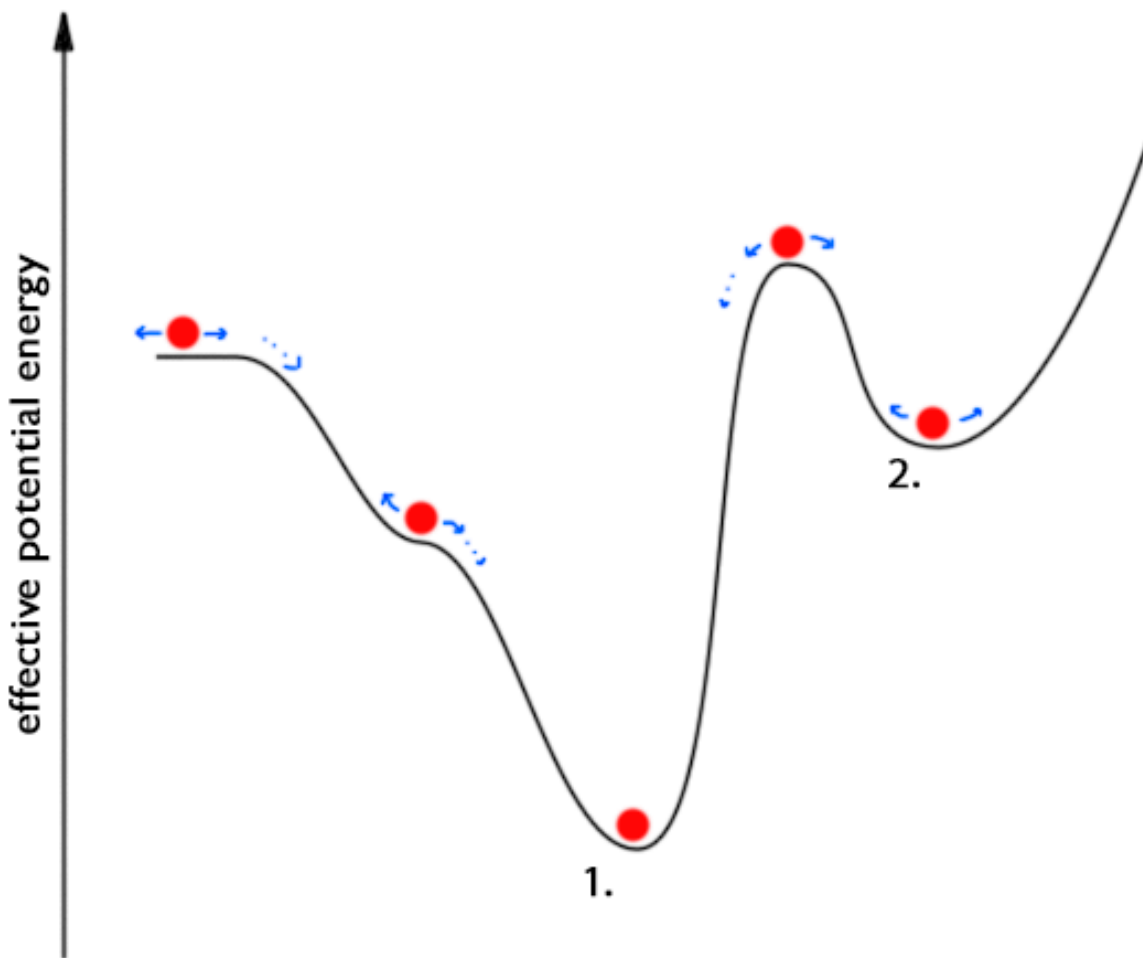
## Energy and Stability

Steric energy of a molecule arises from specific interactions within the molecule. These interactions include the stretching or compressing of bonds beyond their equilibrium lengths and angles, torsional effects of twisting about single bonds, the Van der Waals attractions or repulsions of atoms that come close together, and the electrostatic interactions between partial charges in a molecule due to polar bonds

$$E_{\text{steric}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{str-bend}} + E_{\text{oop}} + E_{\text{tor}} + E_{\text{VdW}} + E_{\text{qq}}$$

Bond stretching, bending, stretch-bend, out of plane, and torsion interactions are bonded interactions because the atoms involved must be directly bonded or bonded to a common atom. The Van der Waals and electrostatic (qq) interactions are between non-bonded atoms.

As a result, in general for a molecule, High Energy = Low Stability

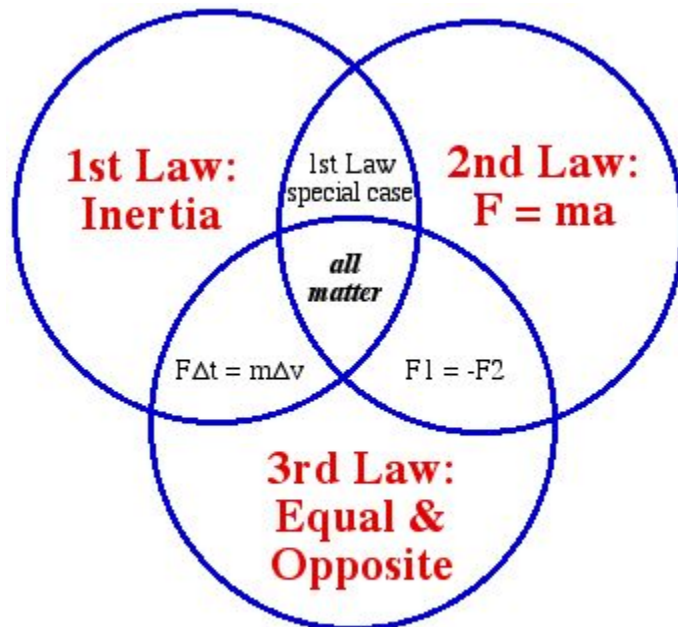


## Newton's Laws of Motion

First Law: an object will remain at rest or in uniform motion in a straight line unless acted upon by an external force

Second Law:  $F=ma$

Third Law: for every external force that acts on an object there is a force of equal magnitude but opposite direction which acts back on the object which exerted that external force



## Newton's Laws

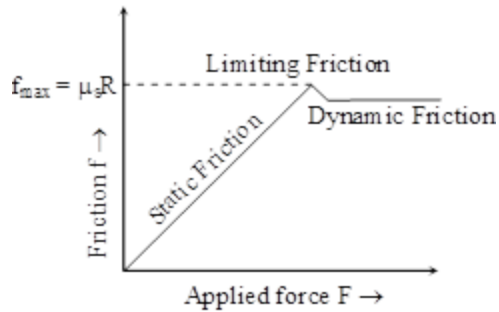
### Friction

For a moving object, Frictional resistance is usually proportional to the "normal force" and designated by  $N$ . (the force perpendicular or "normal" to the surfaces). Friction is independent of the area of contact and the coefficient of static friction is slightly greater than that of kinetic friction. In general, kinetic friction is independent of velocity. Friction is proportional to the roughness of the surfaces in contact.

The frictional resistance force may then be written:  $f_{\text{friction}} = \mu N$

$\mu_k$  = coefficient of kinetic friction

$\mu_s$  = coefficient of static friction



## Hooke's Law

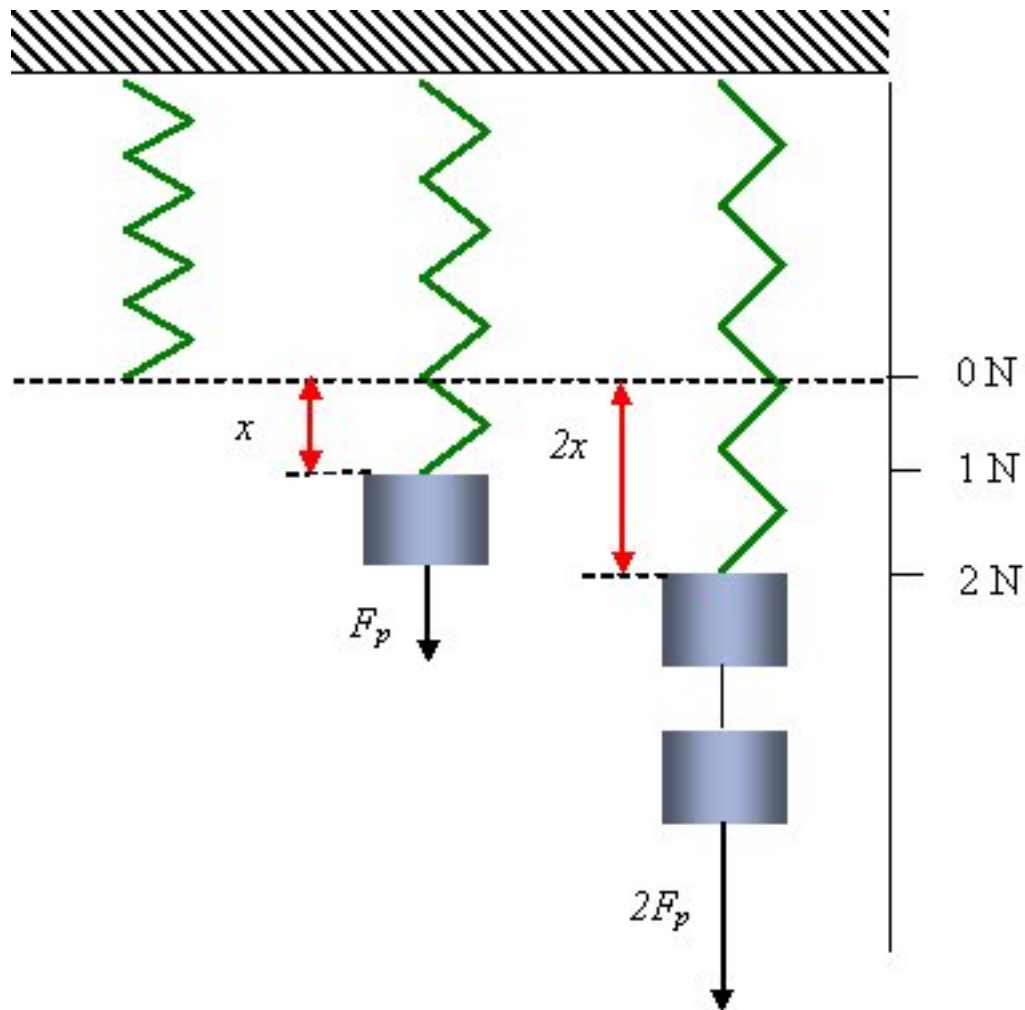
An elastic object, such as a spring, at equilibrium has a defined length. If the spring is either stretched or compressed, the change in length is called its extension and has either a +ve value (stretched) or a -ve value, (compressed). The extension of an elastic object is directly proportional to the force (F) applied to it, related by Hooke's Law

$$F = k.e$$

Where F is the force in Newtons, k is the Spring Constant in N/meter, (the greater the value of k, the stiffer the spring), and e is the extension (compression) in meters.

The Potential Energy stored in a spring,  $U_{el}(x) = \frac{1}{2} k x^2$ , (where  $x = e$ ). For a Harmonic Oscillator, the frequency of the oscillation, f, is related to k by the relationship

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{m}}$$



And the wavenumber,  $\nu$  is

$$\frac{\pi c}{2} \frac{k}{\mu}$$

Where  $\mu = \frac{m_1 m_2}{(m_1 + m_2)}$ , and  $c$  is the speed of light.

Stronger bonds have larger values of  $k$ , and give faster vibrations. Bonds to lighter atoms have faster vibrations than bonds to heavy atoms.

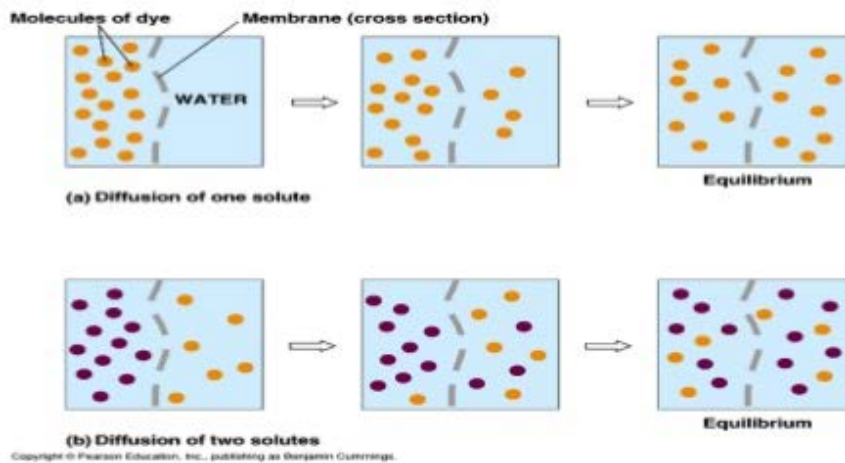
### Concept of Diffusion

Diffusion is the thermal motion of all (liquid and gas) molecules at temperatures above absolute zero. Diffusion rate is a function of only temperature, and is not affected by concentration.

Brownian motion is observed in molecules that are so large that they are not driven by their own thermal energy but by collisions with solvent particles. They move at random because they frequently collide. Molecular diffusion is relevant only on length scales between nanometer and millimeter. On larger length scales, transport in liquids and gases is normally due to another transport phenomenon, convection

## + Define Diffusion

The movement of molecules from a area in which they are highly concentrated to a area in which they are less concentrated.



## Free Energy, Enthalpy and Entropy

Table 4. Three Thermodynamic Quantities

Quantity	Symbol	Measures	Units
Enthalpy	$H$	Heat	Energy
Entropy	$S$	Disorder	Energy/K
Free energy	$G$	Reactivity	Energy

Spontaneous changes are ones in which the free energy (G) of a system decreases ( $\Delta G$  is negative). Heat energy is also called enthalpy (H). When heat is released, the change in the enthalpy for the system that is releasing the heat decreases, whereas when heat is absorbed, the change in the enthalpy increases. While a decrease in the enthalpy makes a process more spontaneous (favorable), the change in enthalpy alone cannot be used to predict whether an overall change is spontaneous. There is another factor that must be considered and that is the entropy (S). Entropy is a measure of disorder; when a system become more disordered, the change in entropy is positive. When a change in entropy is positive, it makes the change more spontaneous (favorable). In general reactions are favorable if  $\Delta H_o < 0$  and  $\Delta S_o > 0$ , or unfavorable if  $\Delta H_o > 0$  and  $\Delta S_o < 0$ ,

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

It is critical to appreciate the difference between  $\Delta G^\circ$  and  $\Delta G$ .  $\Delta G^\circ$  indicates standard conditions and M concentrations of reactants and products at equilibrium since  $\Delta G^\circ = -RT \ln K_{\text{equilibrium}}$   **$\Delta G^\circ$  tells you the Equilibrium position attained if the reaction occurs- it does not tell you whether the reaction will occur, or in which direction the reaction will proceed.**

**$\Delta G$ , on the other hand is the free energy under the conditions at hand, with**

$$\Delta G = -RT \ln Q$$

Where Q, the so-called Reaction Quotient = [Products]/[Reactants]

**$\Delta G^\circ$  and  $\Delta G$  are related by the equation**

$$\Delta G = \Delta G^\circ + R T \ln Q.$$

If  $K_{\text{equilibrium}} > Q$ , the **reaction** will proceed forward, converting reactants into products. If  $K_{\text{equilibrium}} < Q$ , the **reaction** will proceed in the reverse direction, converting products into reactants. If  $Q = K_{\text{equilibrium}}$  then the system is already at **equilibrium and  $\Delta G = 0$ .**

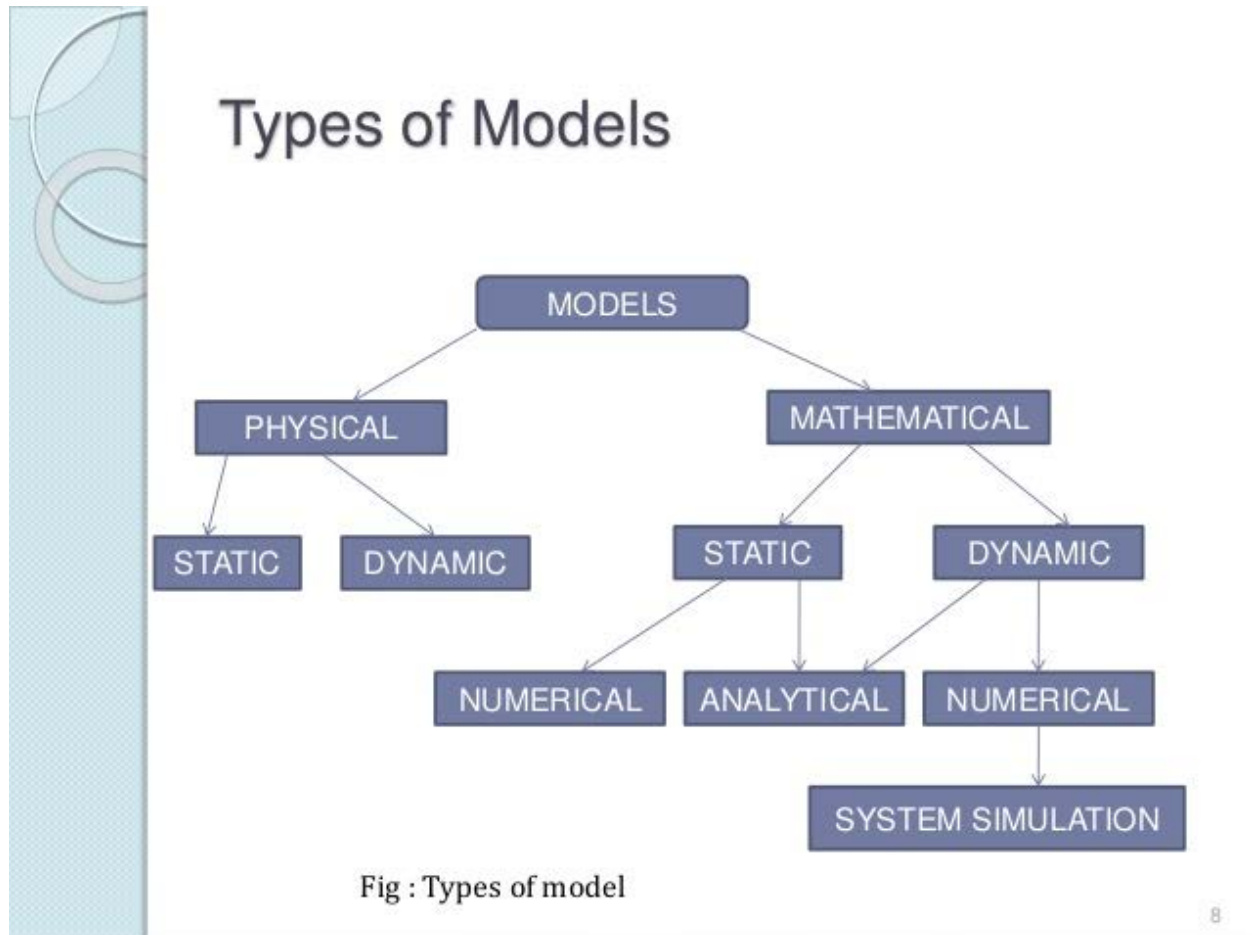
**Neither  $\Delta G$  or  $\Delta G^\circ$  tell you anything about how fast the reaction will proceed- That requires knowledge of the Activation Energy of the reaction.**

## Models

Models play important roles in science and come in three types: Physical, Mathematical and Conceptual and usually involve three components, i) Information input, ii) an Information processor, and iii) Information output (often some type of prediction). Physical and Conceptual models (sometimes referred to as Mental Models) are often qualitative while Mathematical ( and

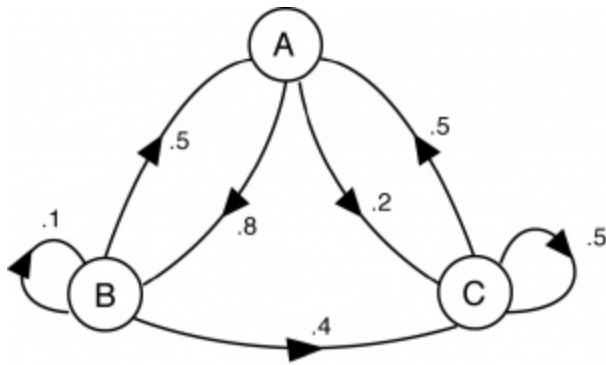


Statistical ) Models are quantitative. All Models have limitations but are useful in generating avenues to “test” the model, and, as a result, are often changed based upon new evidence. The most useful models are the simplest model that accommodates the available evidence



### Randomness and Stochastic Processes

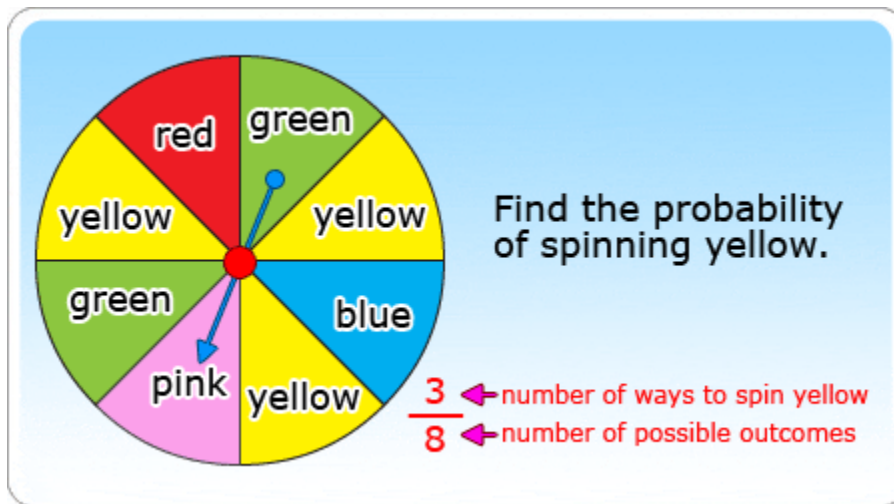
Stochastic vs Deterministic Processes. A stochastic process,  $X(t)$  or  $X_t$ , is a collection of random variables indexed by time,  $t$ . Discrete or continuous time Markov Chains. A process is deterministic if its future is completely determined by its present and past. A **Markov chain** is a stochastic process, but unlike a general stochastic process, **Markov chains** must be "memory-less" and the probability of future actions are independent of the preceding steps that gave rise to the current state.



Markov graph of transition probabilities between states A, B and C

### Probability

Probability Models- 3 components: 1] sample space (set whose elements are the “outcomes” or “sample points” 2] class of “events” (all subsets of the sample space) & 3] Probability Measure (assignment of a nonnegative number to each outcome, with the restriction that these numbers must sum to one over the sample space).



The probability of an event is the sum of the probabilities of the outcomes comprising that event

### Deriving Equations

Equations can be developed based upon a theoretical model (eg chemical and enzyme kinetics) or on experimental observations (eg Coulomb’s Law)

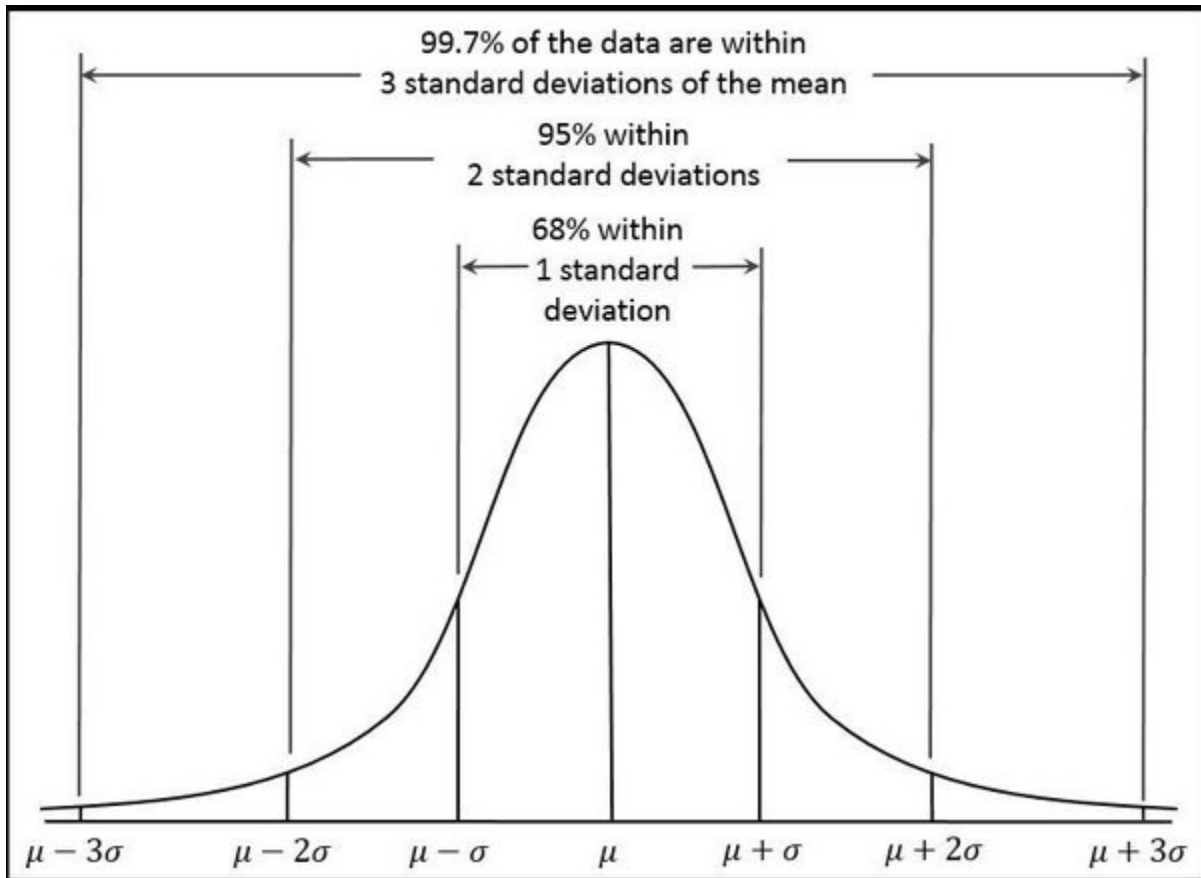
### Using Equations

Equations describe many processes in the molecular life sciences: linear-  $y=mx + c$ , hyperbolic-  $y=mx/(x+K)$ , exponential rise-  $y=a(1-e^{-kx})$ , exponential decay- $y=ae^{-kx}$

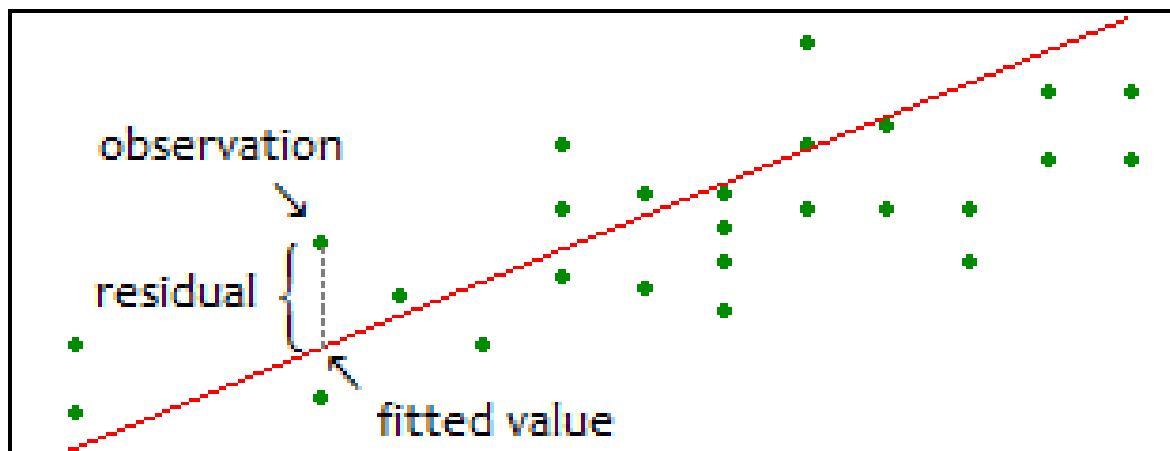
Sigmoidal- $y=y_0+a/(1+e^{-((x-x_0)/b)})$  and allow descriptor constants,  $m,K,k,a, b$  etc, to be calculated from experimental data. Equations also allow experimental data to be extrapolated to predict values for the observable,  $y$ , for inaccessible values of the independent variable,  $x$ .

### Populations, Averages, Normal Distributions and Standard Deviations

Reproducibility and error analysis: average,  $x = \Sigma xi/n$ . Sample standard deviation is the square root of  $\Sigma(xi-xav)^2/(n-1)$ . If you compare two normal distributions and take 95% confidence limits: ie the values laying between the 2.5% and 97.5% values, of each distribution and the values do not overlap, there is only 2.5% of the possible estimates of the higher number that could, in a normal distribution fall into the upper echelons of the normal distribution of the lower number: the numbers would be said to differ at the level of the 95% confidence limits, or at a p value of 0.025.



## Linear Regression and Residuals



Least squares analysis involves making an initial guess to the values of the parameters of the equation being fit and calculating using these initial guesses the values of  $y$  for the chosen values of  $x$ . The difference between theoretical and experimental data points is the “residual” and is negative or positive depending upon whether the calculated value is bigger or smaller than the actual experimental value. The residuals are squared and summed. The sum of the squares of the residuals is hence an estimate of the fit of the line to the data. The parameter estimates are then changed and a sum of squares of the new residuals calculated. The process is repeated until the sum of squares of the residuals is a minimum. The parameters giving the lowest sum of squares of the residuals are the best fit parameters. The residual is the difference between the actual data point and the computed best fit data point and can be negative or positive. The residuals calculated from these best fit parameters can be plotted against the value of  $x$ , and a random distribution of residuals = correct equation. A pattern in the residuals indicates that a different equation is more appropriate

Images from Google Image

Mindmap from Project Mindmap, <https://www.molecularlifesciences.org/>