

# Background and Reference Material

---

## Probability and Statistics

### Probability Distribution

$P(X)$  is a probability distribution for observing a value  $X$  in a data set of multiple observations. It can describe either a discrete ( $i = 1$  to  $N$ ) data set or continuous function. For the continuous distribution,  $P(X)$  is known as a probability density, and  $P(X) dX$  is the probability of observing a value between  $X$  and  $X + dX$  within a large sample. All probability distributions are normalized such that

$$\text{Discrete} \\ \sum_{i=1}^N P_i = 1$$

$$\text{Continuous} \\ \int P(X) dX = 1$$

The *mean* or average value of  $X$  for this distribution is

$$\langle X \rangle = \frac{1}{N} \sum_{i=1}^N X_i = \sum_{i=1}^N P_i X_i \qquad \int X P(X) dX$$

The *root mean square* value of the distribution is the square root of the mean square value of  $x$ :

$$X_{rms} = \langle X^2 \rangle^{1/2}$$

The mean is the first *moment* of the distribution, the mean square value is the second moment, and the  $n^{\text{th}}$  moment of  $P(x)$  is

$$\langle X^n \rangle = \frac{1}{N} \sum_{i=1}^N X_i^n = \sum_{i=1}^N P_i X_i^n \qquad \int X^n P(X) dX$$

The width of the distribution function about its mean is quantified through the standard deviation  $\sigma$  or variance  $\sigma^2$ :

$$\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2$$

### Gaussian Distribution

The Gaussian normal distribution is a continuous distribution function commonly used to describe a large number of independent samples from a random distribution.

$$P(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(X - \langle X \rangle)^2}{2\sigma^2}\right]$$

The distribution function is completely described by its mean and variance. All higher moments of this distribution are zero. The most probable value, or peak, of the distribution is the mean, and the width is given by the standard deviation. The probability density with  $\pm 1\sigma$  of the mean is 68.2%

and within  $\pm 2\sigma$  is 95.4%. The full-width of  $P(X)$  at half the maximum height (FWHM) is  $\sqrt{8 \ln 2} \sigma = 2.355\sigma$ .

## Binomial and Poisson Distributions

A binomial distribution is a discrete probability distribution that relates to the number of times of observing  $N$  successful events from a series of  $M$  independent trials on a binary variable (yes/no; up/down). For instance the number of times of observing  $N$  copies of the amino acid tyrosine within a protein sequence of length  $M$ . Alternatively we can say that it gives the probability that  $N$  random events happen during a given period of time,  $M$ . Like the number of barrier crossings (jumps) during a time window in a two-state trajectory. Consider the number of ways of putting  $N$  indistinguishable objects into  $M$  equivalent boxes:

$$\Omega = \frac{M!}{N!(M-N)!} \equiv \binom{M}{N}$$

(Note: In the limit  $M \gg N$ ,  $\Omega \approx M^N/N!$ ). The choices made here are binary, since any given box can either be empty or occupied.  $\Omega$  is also known as the binomial coefficient. However, unlike a coin toss, a binary variable need not have equal probability for the two outcomes. We can say that we know the probability that a box is occupied is  $p_o$ . Then the probability that a box is empty is  $p_e = 1-p_o$ . The probability that in a given realization with  $M$  boxes that  $N$  are occupied and  $M-N$  are empty is  $p_o^N p_e^{M-N}$ . Then multiplying the probability for a given realization and the number of possible realizations, the probability of observing  $N$  of  $M$  boxes occupied is

$$P(N, M) = \binom{M}{N} p_o^N p_e^{M-N} = \frac{M!}{N!(M-N)!} p_o^N (1-p_o)^{M-N}$$

This is the binomial distribution. The average number of occupied boxes is  $\langle N \rangle = N p_o$ . A Gaussian distribution emerges from the binomial distribution in the limit that the number of trials becomes large ( $M \rightarrow \infty$ ). The mean value of the distribution is  $N p_o$  and standard deviation is  $N p_o (1-p_o)$ .

A Poisson distribution emerges in the limit that the number of trials ( $M$ ) becomes large and the probability of observing an occupied box is small,  $p_o \ll 1$ .

$$P(N) = \frac{\langle N \rangle^N e^{-\langle N \rangle}}{N!}$$

The mean of the Poisson distribution is  $\langle N \rangle$  and the standard deviation  $\sigma = \langle N \rangle^{1/2}$ . Fluctuations in  $N$  scale as  $\sigma$ , and the fluctuations relative to the mean as  $\sigma / \langle N \rangle = \langle N \rangle^{-1/2}$ . Then we see that fluctuations are most apparent to the observer for small  $N$  with the biggest contrast for  $N = 1$ .<sup>1</sup>

---

1. M. B. Jackson, *Molecular and Cellular Biophysics*. (Cambridge University Press, Cambridge, 2006), Ch. 12.

## Constant and Units

$$k_B = 1.381 \times 10^{-23} \text{ J K}^{-1}$$

$$N_A = 6.02205 \times 10^{23}$$

$$R = k_B N_A = 8.314 \text{ J mol}^{-1} \text{ K}^{-1}$$

$$e = 1.60219 \times 10^{-19} \text{ C}$$

$$4\pi\epsilon_0 = 1.112650 \times 10^{-10} \text{ C}^2 \text{ s}^2 \text{ kg}^{-1} \text{ m}^{-1}$$

$$1 \text{ V} \rightarrow 1 \text{ eV} = 1.60 \times 10^{-19} \text{ J}$$

$$1 \text{ atm} = 101352 \text{ N m}^{-2}$$

$$1 \text{ L} = 10^{-3} \text{ m}^3$$

### *Practical Units and Identities for Biophysical Purposes*

$$1 \text{ m} = 10^9 \text{ nm} = 10^{10} \text{ \AA} = 10^{12} \text{ pm}$$

$$1 \text{ nm} = 10^{-9} \text{ m}$$

$$1 \text{ pN} = 10^{-12} \text{ N}$$

$$1 \text{ aJ} = 10^{-18} \text{ J} \quad \text{zJ} = 10^{-21} \text{ J}$$

$$\frac{e^2}{4\pi\epsilon_0} = 230 \text{ pN nm}^2 = 230 \text{ zJ nm}$$

### **At 300 K**

$$k_B T = 4.1 \text{ pN nm} = 2.5 \text{ kJ/mol} = 0.60 \text{ kcal/mol}$$

$$\frac{1}{4\pi\epsilon_0} \frac{e^2}{k_B T} = 56 \text{ nm}$$

$$\frac{k_B T}{e} = 25 \text{ mV}$$

# Thermodynamics

## First Law

$$dU = \dot{d}q + \dot{d}w$$

$w$ : Work performed by the surroundings on the system.

Mechanical work: linear displacement, stretching surface against surface tension, and volume expansion against an external pressure.

$$dw = -\bar{f}_{\text{ext}} \cdot d\bar{r} + \gamma da - p_{\text{ext}} dV$$

Electrical work:  $dw = -q \bar{E} \cdot d\bar{r}$

Magnetic work:  $dw = -M \cdot dB$  if field is external to the system or  $dw = B \cdot dM$  if the field is internal to the system.

$q$ : Heat added to the system.

$C$ : Heat capacity links heat and temperature. At constant pressure:  $\dot{d}q_p = C_p dT$

## Second Law

$$dS = \frac{\dot{d}q_{\text{rev}}}{T}$$

## State Functions

Internal Energy,  $U$ :  $U = q + w$   
 $dU = TdS - pdV + \mu dN$       Natural variables:  $N, S, V$

Enthalpy,  $H$ :  $H \equiv U + pV$   
 $dH = TdS + Vdp + \mu dN$       Natural variables:  $N, S, p$

$$\Delta H(T_2) = \Delta H(T_1) + \int_{T_1}^{T_2} \Delta C_p dT$$

Helmholtz free energy,  $A$  or  $F$ :  $A = U - TS$   
 $dA = -pdV - SdT + \mu dN$       Natural variables:  $N, V, T$

more generally:  $dA = -SdT + w_{\text{rev}}$

where  $w_{\text{rev}} = -pdV + \sum_i \mu_i dN_i + \bar{f} \cdot d\bar{R} + \gamma da + \Phi dq$

Gibbs free energy,  $G$ :  $G = H - TS = A + pV$   
 $dG = -SdT + Vdp + \mu dN$       Natural variables:  $N, p, T$   
more generally:  $dG = -SdT + w_{\text{rev}}$

$$\text{where } w_{rev} = V dp + \sum_i \mu_i dN_i + \bar{f} \cdot d\bar{R} + \gamma da + \Phi dq$$

For a reversible process at constant T and p,  $\Delta G$  is equal to the non-pV work done on the system by the surroundings. Generally:  $(dG)_{T,p} \leq w_{\text{non-pV}}$ .

$$\text{Chemical potential: } \mu_i = -\left(\frac{\partial G}{\partial N_i}\right)_{T,p,\{N_{j \neq i}\}} = -\left(\frac{\partial A}{\partial N_i}\right)_{T,V,\{N_{j \neq i}\}}$$

$$\text{Entropy: } S = -\left(\frac{\partial G}{\partial T}\right)_p = -\left(\frac{\partial A}{\partial T}\right)_V$$

### If you know the free energy, you know everything!

$$V = \left(\frac{\partial G}{\partial p}\right)_T \quad p = -\left(\frac{\partial A}{\partial V}\right)_T$$

$$H = G + TS \quad \Rightarrow \quad H = G - T\left(\frac{\partial G}{\partial T}\right)_p$$

$$U = H - pV \quad \Rightarrow \quad U = G - T\left(\frac{\partial G}{\partial T}\right)_p - p\left(\frac{\partial G}{\partial p}\right)_T$$

$$A = U - TS \quad \Rightarrow \quad A = G - p\left(\frac{\partial G}{\partial p}\right)_T$$

$$C_p = T\left(\frac{\partial S}{\partial T}\right)_p \quad \Rightarrow \quad C_p = -T\left(\frac{\partial^2 G}{\partial T^2}\right)_p \quad C_V = \left(\frac{\partial U}{\partial T}\right)_{V,N}$$

### Spontaneous Processes

Conditions that determine the direction of spontaneous change and the conditions for equilibrium

$$(dG)_{p,T,N} \leq 0 \quad (dA)_{V,T,N} \leq 0 \quad (dS)_{H,p} \geq 0$$

$$(dU)_{V,S,N} \leq 0 \quad (dH)_{p,S,N} \leq 0$$

### Chemical Equilibria

$$\Delta G_{\text{rxn}}^0 = \Delta H_{\text{rxn}}^0 - T\Delta S_{\text{rxn}}^0 \quad \Delta G_{\text{rxn}}^0 = \sum_i^{\text{products}} N_i \mu_i^0 - \sum_j^{\text{reactants}} N_j \mu_j^0$$

$$K_{\text{eq}} = \exp(-\Delta G_{\text{rxn}}^0 / RT)$$

## Statistical Thermodynamics

The partition functions play a central role in statistical mechanics. All the thermodynamic functions can be calculated from them.

### Microcanonical Ensemble (N,V,E)

- All microstates are equally probable:  
 $P = 1/\Omega$  where  $\Omega$  is the number of degenerate microstates.  
 $S = k \ln \Omega$  Boltzmann entropy in terms of degeneracy

### Canonical Ensemble (N,V,T)

- Average energy  $\langle E \rangle$  is fixed.

#### Canonical Partition Function: Q

$$Q = \sum_i e^{-E_i/kT}$$
$$= \int e^{-E/kT} dE$$

For one particle in  $n$  dimensions:

$$Q = \frac{1}{h^n} \iint e^{-H(q,p)/kT} dq^n dp^n$$

The classical partition function for  $N$  non-interacting particles in 3D:

$$Q = \frac{1}{h^{3N}} \int d^{3N} p \int d^{3N} q e^{-H(p^{3N}, q^{3N})/kT}$$
$$q^{3N} = \{q^3, q_2^3 \dots q_N^3\} \text{ where } q_i^3 = \{x_i, y_i, z_i\}$$

If the kinetic and potential energy terms in the Hamiltonian are separable as  $T(p^{3N}) + V(q^{3N})$ , then

$$Q = \left( \int d^{3N} q e^{-V(q^{3N})/kT} \right) \underbrace{\left( \frac{1}{h^{3N}} \int d^{3N} p e^{-T(p^{3N})/kT} \right)}_{\lambda^{-3N}}$$

where  $\lambda = (h^2 / 2\pi mkT)^{1/2}$ . So

$$Q = \frac{1}{\lambda^{3N}} \left( \int d^{3N} q e^{-V(q^{3N})/kT} \right)$$

### Microstate Probabilities

$$P_i(E_i) = \frac{e^{-E_i/kT}}{Q}$$

or describing the probability of occupying microstates of an energy  $E_j$ , which have a distinguishable degeneracy of  $g(E_j)$ :

$$P_j(E_j) = \frac{g(E_j)e^{-E_j/kT}}{Q}$$

or as a probability density:

$$P(r) = \frac{e^{-E(r)/kT}}{Q}$$

### Internal Energy

$$U = \langle E \rangle = \sum_j P_j E_j = \frac{1}{Q} \sum_j E_j e^{-E_j/kT}$$

### Helmholtz Free Energy

$Q(N, T, V)$  leads to  $A(N, T, V)$ . Using  $dA = -pdV - SdT + \mu dN$ :

$$A = U - TS = U + T \left( \frac{\partial A}{\partial T} \right)_{V, N}$$

$$\Rightarrow A = -kT \ln Q$$



All the other functions follow from  $Q$ ,  $U$  and  $A$

$$\begin{aligned}
 -\frac{U}{T^2} &= \frac{1}{T} \left( \frac{\partial A}{\partial T} \right)_{V,N} & -\frac{A}{T^2} &= \left( \frac{\partial(A/T)}{\partial T} \right)_{V,N} = k \left( \frac{\partial \ln Q}{\partial T} \right)_{V,N} \\
 U &= -kT^2 \left( \frac{\partial \ln Q}{\partial T} \right)_{V,N} \\
 C_V &= \left( \frac{\partial U}{\partial T} \right)_{V,N} \\
 S &= -\frac{A}{T} + \frac{U}{T} = k \ln Q + kT \left( \frac{\partial \ln Q}{\partial T} \right)_{V,N} \\
 C_p &= T \left( \frac{\partial S}{\partial T} \right)_p \\
 p &= - \left( \frac{\partial A}{\partial V} \right)_{T,N} = kT \left( \frac{\partial \ln Q}{\partial V} \right)_{T,N} \\
 \mu &= - \left( \frac{\partial A}{\partial N} \right)_{T,V} = -kT \left( \frac{\partial \ln Q}{\partial N} \right)_{T,V}
 \end{aligned}$$

## Entropy

Entropy in terms of microstate probabilities and degeneracies

$$\frac{S}{k} = \frac{U - A}{kT} = \sum_i P_i \frac{E_i}{kT} + \ln Q \quad \text{But } \frac{E_i}{kT} = -\ln e^{-E_i/kT}, \text{ so}$$

$$\frac{S}{k} = -\sum_i P_i (\ln e^{-E_i/kT}) + \ln Q$$

$$= -\sum_i P_i \ln \left( \frac{e^{-E_i/kT}}{Q} \right)$$

$$= -\sum_i P_i \ln P_i$$

$$S = -k \sum_i P_i \ln P_i$$

Gibbs equation:  $S$  in terms of microstate probabilities

## Ensemble Averages

Other internal variables ( $X$ ) can be statistically described by

$$\langle X \rangle = \sum_{i=1}^N P_i X_i \quad P_i(E_i) = \frac{e^{-E_i/kT}}{Q}$$

For our purposes, we will see that translational, rotational, and conformational degrees of freedom are separable

$$Q = Q_{\text{trans}} Q_{\text{rot}} Q_{\text{conf}} \dots$$

## Grand Canonical Ensemble ( $\mu, V, T$ )

Average energy  $\langle E \rangle$  and average particle number  $\langle N_j \rangle$  are fixed. Thermodynamic quantities depend on microstates ( $i$ ) and particle type ( $j$ ).  $P_i(j)$  is the probability that a particle of type  $j$  will occupy a microstate of energy  $E_i$ :

$$P_i(j) = \frac{e^{-E_i(j)/k_B T} e^{\mu_j N(j)/k_B T}}{\Xi}$$
$$U = \langle E \rangle = \sum_j \sum_i P_i(j) E_i(j) \quad \text{and} \quad \langle N \rangle = \sum_j \sum_i P_i(j) N(j)$$

The grand canonical partition function is

$$\Xi = \sum_j Q(N_j, V, T) e^{\mu_j N(j)/k_B T}$$

## Stirling's Approximation

For large  $N$ :

$$N! \approx N \ln N - N$$
$$\text{or } N! \approx \left(\frac{N}{e}\right)^N \sqrt{2\pi N} \quad \rightarrow \quad N! \approx N \ln N - N + \frac{1}{2} \ln(2\pi N)$$

## First Law Revisited

Let's relate work and the action of a force to changes in statistical thermodynamic variables:<sup>2</sup>

The internal energy:

---

2. T. L. Hill, *An Introduction to Statistical Thermodynamics*. (Addison-Wesley, Reading, MA, 1960), pp. 11–13, 66–77.

$$U = \langle E \rangle = \sum_j P_j E_j$$

$$dU = d\langle E \rangle = \sum_j E_j dP_j + \sum_j P_j dE_j$$

Note the relationship between this expression and the First Law:

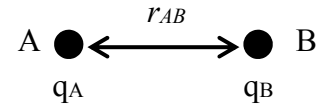
$$dU = \bar{d}q + \bar{d}w$$

$$\bar{d}q_{rev} = TdS \leftrightarrow \sum_j E_j dP_j$$

$$\bar{d}w \cong p dV \leftrightarrow \sum_j P_j dE_j$$

## Continuum Electrostatics

- The interaction of charges can be formulated through a Force, Field, or Potential.
- Consider the interaction between two ions  $A$  and  $B$ , separated by a distance  $r$ , with charges  $q_A$  and  $q_B$ .



$$\begin{aligned} \mathbf{r}_{AB} &= (\mathbf{r}_A - \mathbf{r}_B) \\ r_{AB} &= |\mathbf{r}_{AB}| \\ \hat{\mathbf{r}}_{AB} &= \mathbf{r}_{AB} / r \end{aligned}$$

### Force

Coulomb's Law gives the force that B exerts on A.

$$\bar{\mathbf{f}}_{AB} = -\frac{1}{4\pi\epsilon} \frac{q_A q_B}{r_{AB}^2} \hat{\mathbf{r}}_{AB}$$

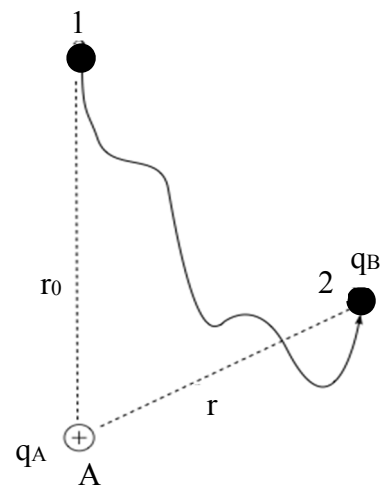
### Work

Electrical work comes from moving charges

$$dw = -\bar{\mathbf{f}} \cdot d\bar{\mathbf{r}}$$

As long as  $q$  and  $\epsilon$  are independent of  $r$ , and the process is reversible, then work only depends on  $r$ , and is independent of path. To move particle B from point 1 at a separation  $r_0$  to point 2 at a separation  $r$  requires the following work

$$w_{1 \rightarrow 2} = \frac{1}{4\pi\epsilon} q_A q_B \left( \frac{1}{r} - \frac{1}{r_0} \right)$$



### Field, E

The electric field is a vector quantity that describes the action of charge B at point A is

$$\bar{\mathbf{E}}_{AB}(\mathbf{r}_A) = -\frac{1}{4\pi\epsilon} \frac{q_B}{r_{AB}^2} \hat{\mathbf{r}}_{AB}$$

$\hat{\mathbf{r}}_{AB}$  is a unit vector pointing from  $\mathbf{r}_B$  to  $\mathbf{r}_A$ .  $E$  is related to force that charged particle B exerts on a charged test particle at A through

$$\bar{\mathbf{f}}_A = q_A \bar{\mathbf{E}}_{AB}(\mathbf{r}_A)$$

More generally for the field exerted by multiple charged particles at point  $\mathbf{r}_A$  is the vector sum of the field from multiple charges ( $i$ ):

$$\bar{\mathbf{E}}(\mathbf{r}_A) = \sum_i \bar{\mathbf{E}}_{Ai}(\mathbf{r}_A) = -\frac{1}{4\pi\epsilon} \sum_i \frac{q_i}{r_{Ai}^2} \hat{\mathbf{r}}_{Ai}$$

where  $r_{Ai} = |\mathbf{r}_A - \mathbf{r}_i|$  and the unit vector  $\hat{r}_{Ai} = (\mathbf{r}_A - \mathbf{r}_i) / r_{Ai}$ . Alternatively for a continuum charge density  $\rho(r)$ ,

$$\bar{E}_A(\mathbf{r}_A) = -\frac{1}{4\pi\epsilon} \int \rho(\mathbf{r}) \frac{(\mathbf{r}_A - \mathbf{r})}{|\mathbf{r}_A - \mathbf{r}|^3} d\mathbf{r}$$

Electrical work comes from moving charges in an electrostatic field is

$$dw = -q \bar{E} \cdot d\bar{r}$$

## Electrostatic Potential, $\Phi$

The electrostatic potential  $\Phi$  is a scalar quantity defined in terms of the electric field through

$$\bar{E}_A = -\bar{\nabla}\Phi_A \quad \text{or} \quad \Phi = \int \bar{E} \cdot \bar{r} \, dr$$

The integral expression illustrates the relationship between the electrostatic potential at point A and the work needed to move a particle within that potential. The electrostatic potential at point A, which results from a point charge at B, is

$$\Phi_A = \frac{1}{4\pi\epsilon} \frac{q_B}{r_{AB}}$$

or for a continuum:

$$\Phi_A(\mathbf{r}_A) = \frac{1}{4\pi\epsilon} \int \frac{\rho(\mathbf{r})}{|\mathbf{r}_A - \mathbf{r}|} d\mathbf{r}$$

$\Phi$  is related to the potential energy for interaction between two charges as

$$U_{AB} = q_A \Phi_A = q_B \Phi_B = \frac{1}{4\pi\epsilon} \frac{q_A q_B}{r_{AB}}$$

This is the Coulomb interaction potential. More generally, for many charges, the interaction energy is

$$\begin{aligned} U_{AB} &= \frac{1}{2} \sum_i q_i \Phi(r_{Ai}) \\ &= \frac{1}{2} \int \Phi_A(\mathbf{r}_A) \rho(\mathbf{r}_A) d\mathbf{r}_A \end{aligned}$$

The factor of  $1/2$  is to assure that one does not double count the interaction between two particles.