Lecture 3: Boltzmann Distribution

I. Overview

At equilibrium the statistical probability (P_i) that any system (population of molecules) is in some state, i, depends on the energy (E_i) of the system. This is expressed in the Boltzmann probability formula:

$$P_i = e^{-Ei/kT}/Z \tag{1}$$

where T is the absolute temperature (in Kelvin), k is Boltzmann's constant, Z is the partition function (discussed later) and, again, E_i is the total energy of the system, comprising all the molecular translational, rotational and vibrational energy, plus the energy of noncovalent interactions between the molecules in the system. In this form the expression is oversimplified because it assumes that there is only one way to achieve each energy state.

If each energy state may have more than one configuration the states are **degenerate**. In this case g_i is the number of ways in which that total system energy may be achieved and is a measure of the entropy of the system. The above equation becomes:

$$P_i = g_i \cdot e^{-Ei/kT}/Z \tag{2}$$

Expression 2 describes a Boltzmann distribution of states and it means that the probability of a system's state is dependent on *both* **entropy** and **energy**. As we will see this result is a re-statement of the well-known thermodynamic expression for **free energy**, G:

$$G = E - TS$$

where E is the potential energy and S is the entropy of a system.

Boltzmann also defined entropy and assigned an energy value to the entropy of a system. His definition of entropy is,

$$S = k \log_e W$$

Where k is the Boltzmann energy conversion factor and W is the *Wahrscheinlichkeit* of the system. We will explore this relationship below.

Where do these Boltzmann equations come from? In fact, they can be derived from first principals without reference to any models or descriptions of molecules. This derivation from first principals is exactly what Boltzmann did in 1877 and it is one of the reasons his work is so famous. We will now explore Boltzmann's expression for entropy, then point out the highlights of a derivation of Boltzmann's probability expression; details of

the derivation are found in Appendix A. An alternate derivation that is somewhat simpler is given in Appendix B. A conceptual description of Boltzmann's constant is given in Appendix C. Finally, a calculation of protein fluctuations is presented in Appendix D.

II. Entropy
An understanding of entropy is central to understanding Boltzmann's theory and how it relates to the Second Law of thermodynamics. Most laws of physics are assertions of equality between two things. For example, Newton's second law,

$$a = F/m$$

where a is the acceleration of a body of mass m subject to force F. Another example is Einstein's law of mass-energy equivalence,

$$E = mc^2$$

where the energy of a system, E, is equal to its mass, m, times the square of the speed of light, c, asserting that energy and mass are interconvertible. In contrast, the Second Law of thermodynamics is an *inequality*, asserting that the entropy of a system is always greater at later times than it is at earlier times. Entropy can be thought of (as an oversimplification) as the randomness, or disorder, of a system. So the Second Law states that systems are getting more random all the time.

State counting
Let's consider a concrete entropy example that will lead us to some concepts in Boltzmann's theory. Assume we have a rectangular container of paint with two red and two blue circular paint particles as shown in the figure below. Assume we can't see the color of the individual paint particles (circles) but we can see the average color of a "region" of paint with two particles (squares). If a region has two red particles we see it as pink, two blue particles we see as blue and if the region has one red and one blue particle we see it as purple. There are two possible positions for particles in each region and the individual particles can be distributed in each rectangle of paint as shown. Each rectangle is a possible configuration of the paint molecules.

Initial condition   After mixing = **16** ways
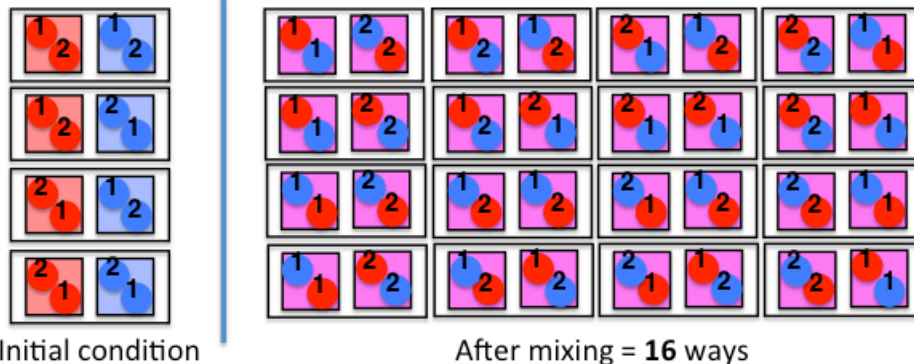
We start out with the condition on the left just after adding the red and blue paint to a rectangle (red and blue regions) and after mixing we end up with the condition on the right (purple regions).

How many different paint molecule configurations can each condition be composed of? For the initial condition there are two ways to have a red region and two ways to have a companion blue region in a rectangle so the total number of configurations for each color is 2x2 = 4 and these are all illustrated. For the final mixed condition there are eight ways to have the first purple region and eight ways to have the second purple region in a rectangle so there are 4x4=16 configurations.

After mixing there are many more ways to achieve the purple condition compared to the initial red/blue condition. The purple condition has more ways to exist and therefore we say it has more entropy than the initial conditions of red and blue. We identified the higher entropy condition by counting the possible configurations or states of the paint molecules. The condition with the most ways of existing (mixed) will be the equilibrium condition.

We also introduced the concept of measuring only what we can see which is an average quantity. We can't see the color of individual molecules but we can see (and measure the color or hue) of a "region" of paint. Another point to be made here is that the difference between low entropy at the start and high entropy in the final condition is *large* (4 versus 16). The magnitude of this difference increases exponentially as the number of configurations increases. When we deal with Avogadro's number of molecules the differences between even two adjoining entropy states becomes astronomical. *Both the concept of considering only what we can measure and the large differences in the number of configurations between entropy states will become important below*.

[As a final caveat it should be stated that the entropy obtained by state counting in this example is not related to energies of the particles; therefore, it is formally different than thermodynamic entropy. But for didactic purposes the above example is an intuitive way to introduce the concepts which are the same.]

Configuration Space
Now we increase the complexity and consider a case that allows us to introduce the concept of multidimensional configuration space. In the illustration below we have three particles at different positions in three dimensions. We can say that each particle has three degrees of freedom, x position, y position and z position. To describe this collection of three particles we require 3 * 3 pieces of information. All the possible configurations of our three particles are contained in what is called the *configuration space* (*C*) of the system and in this case it has nine degrees of freedom total and is nine dimensional (3 * 3).

For a system of d degrees of freedom, the configuration space would be a d-dimensional space. For example, if a system consisted of q particles $p_1, p_2, ... p_q$ (each without any internal degrees of freedom such as vibrational levels) then the configurational space would have $3q$ dimensions. This is because each individual molecule requires three coordinates (x, y, z) to determine its position, so there are $3q$ coordinates overall. A single point p in configuration space defines the locations of all of $p_1, p_2, ... p_q$. In the illustration below two different points (*p*) in the configuration space (*C*) of the three

particles are shown.
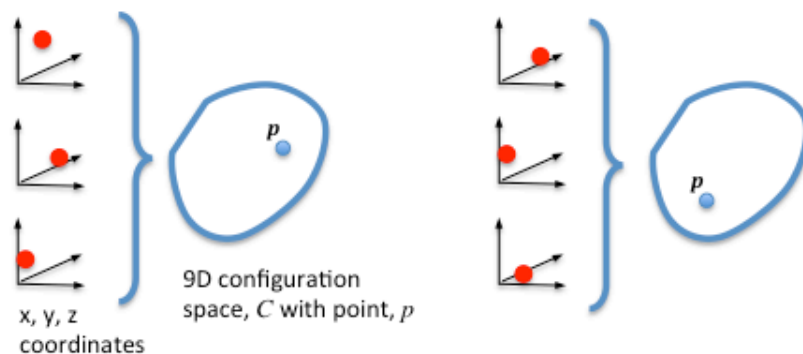
Of course we shouldn't think of the 9D configuration space as an extension of our familiar 3D world. Rather, configuration space is only a mathematical construct. It simply is not possible for us to intuit its shape. Configuration space can be massively dimensional because the dimensionality depends partly on the number of particles. If we are dealing with Avogadro's number of particles the dimensionality is truly large. Furthermore, as the dimensionality increases the difference between adjoining entropy states become even greater than the counting example above.

Also, since the positional coordinates are continuous rather than discrete as above in the paint example, we can't simply count states to calculate the entropies. The size of the configuration space is as large as the number of decimal points by which we specify the positions of the particles. In other words, it can be astronomical. We will have to find a different method to calculate the number of possible states.

Phase Space
Now, one more increase in complexity to introduce the concept of *phase space*. Let's say that each of the three particles above is moving and each has a certain velocity. In this case we refer to the *phase space* (**P**) of the system rather than the configuration space (*C*). Phase space includes both position and velocity (or more properly momentum) information. It has even more dimensions than configuration space! But in the end we can say that the state of our system is described by the location of **p** within **P**.

Because the size of a molecular system's phase space (**P**) is so large and continuous we have to define a reasonable sub-volume to be concerned with. Our collection of molecules will have measurable average quantities like conformations, energies, etc. So we divide up **P** into sub-volumes that "look the same" to the *limits we can measure* these parameters. Two points (**p**) that "look the same" are contained in the same sub-volume of **P** and have the same macroscopic state (with different microscopic states or values of x, y, z, and momentum). The size of this sub-volume is proportional to the number of particles in it, or the probability of finding a particle in it, or the number of states it represents, or its *multiplicity* - all terms which in German can be summed up by Wahrscheinlichkeit (**W**).

Finally, here is the key statement: The entropy *S* of a molecular state represented by a sub-volume in phase space *P* is defined by the following expression,

$$S = k \log_e W$$

where *W* is the probability (Wahrscheinlichkeit) of finding a system in the macroscopic state of that sub-volume and *k* is Boltzmann's constant. This was the crucial insight that Boltzmann had and it has proven very durable and useful. It follows that the entropy of a system is low if it occupies a volume in the appropriate phase space which is small compared to the full volume accessible to it. (See Appendix C for a derivation of the origin of Boltzmann's constant, *k*).

Other terms in use describe points *p* as *microstates* in the sub-volume *macrostate*. The number of microstates that are consistent with a given macrostate is called the *multiplicity Ω* of that macrostate. Each macrostate "looks the same" (has an average observable property) but the different microstates are (usually) still *distinguishable* by the individual positions and momenta of the molecules.

We still need a method to calculate the size of the sub-volume so we can determine differences in entropies as our system moves around in phase space. What we want is a measure of the sub-volume that contains *p*'s that look the same. An important feature of phase space is that it has a *natural measure* so that we can take volumes in phase space to be just dimensionless numbers. This is especially important when we want to compare high-dimensional volume measures with each other. Usually a two dimensional measure will have a smaller value than a similar three dimensional measure. But in the abstract multi-dimensional world of phase space we can compare volumes with different dimensions. The key to this conundrum is that phase space volume measures are essentially normalized to Planck's constant and somehow this makes the measure dimensionless! (Don't ask).

What does all this have to do with the Second Law? Well, suppose that we let the system evolve with time. You will remember that I made a big deal about the fact that adjacent entropy regions in phase space have astronomical differences in entropy. So any adjoining space that has less entropy is *astronomically* smaller and any adjoining space that has more entropy is *astronomically* larger. Which space do you think is more probable for the particle to move into with time based on the relative target space sizes? The larger one of course.

As time evolves particles will inexorably move into larger sub-volumes and therefore have larger entropies. Finally, a particle may enter the *largest* sub-volume in *P* and that stage is called thermal equilibrium. And since the size of the sub-volume is at a maximum, the *entropy* as defined by Boltzmann's equation, is also at a maximum. *In other words, with time the system evolves toward maximum entropy*. This is the Second Law of thermodynamics and Boltzmann defined it by considering little particles and using logic!

III. The Boltzmann probability formula
Derivation of the Boltzmann probability expression evolves from the logical conclusion that given an ensemble of many systems, the equilibrium distribution of systems in that ensemble is the most probable distribution. Sounds obvious, right? Another way to say it is that the most probable distribution will be that which happens the most often. So the derivation begins by mathematically defining the distribution that has the largest statistical weight, W. (Notice the parallels here with the above discussion of entropy).

As an example consider the number of ways of distributing an ensemble of 3 systems (x, y, z) in different states given the constraints that:
1. the number of systems ($\sum n_i$) is constant
2. the energy (E) comes in discrete units (0, 1, 2, 3)
3. the total ensemble energy, $\sum n_i E_i = 3$
4. there is only one way for the energy to be achieved in each state (i.e. the states are not degenerate, $g_i = 1$).

The available states will be n_1, n_2, n_3, n_4 with respective energies (E) of 0, 1, 2 and 3 units (the ground state with E=0 is always include). The only available ways of distributing the systems are indicated in the **table of states** below where i, j, and k are three different distributions of the ensemble.

Energy	State	Dist. i	Dist. j	Dist. k
$E = 0$	n_1	0	1	2
$E = 1$	n_2	3	1	0
$E = 2$	n_3	0	1	0
$E = 3$	n_4	0	0	1
$\sum n_i E_i$		3	3	3

In distribution i there are 3 systems (x, y and z) in state n_2 and 0 in other states; this is equivalent to the systems being evenly distributed in energy and since there is no degeneracy they are all identical. In distribution j, there is 1 system each in state n_1, n_2, and n_3, all with different energies. In distribution k, 2 systems in state n_1 and 1 system in state n_4. In all configurations the average system energy is 1 and the total energy is 3.

The important question is: **Which of these distributions (i, j or k) is the most probable?**

For each of these distributions there are *apparently* 3! ways to distribute the systems in the available states. For example, consider the three systems x, y and z in configuration j. We can choose any one of the three systems to be in state n_1 (3 choices). Assume we will choose y (n_{1y}). Then we have x or z to choose from for state n_2 (2 choices); we choose x (n_{2x}). Then we have only one choice left for state n_3 (n_{3z}). There are 3 systems to choose from to put the first system in a particular state, 2 systems to choose from to place the second system and only 1 system to choose to place the third (3x2x1 = 3!). This is true since each system in distribution j has a **distinguishable** energy.

However, consider distribution k; if there are 2 systems in state n_1, these systems have the

same energy state and therefore are indistinguishable. There are not as many distinguishable ways to distribute the systems. In distribution k the two systems in n_1 only count once because they are identical. Formally we would have to divide our 3! ways in distribution k by 2! to correct for the over counting that occurred in state n_1. Calculation of the actual weights for this example may make this concept clearer.

The general expression for calculating the weight of a configuration is

$$W = N! / n_1! \, n_2! \, \dots \, n_X! \qquad\qquad (3)$$

where N is the number of systems and x is the total number of states. So the final expression for distribution k with 2 systems in state n_1 and 1 system in state n_4 is

$$W_k = 3! \,/\, 2!0!0!1! = 3$$
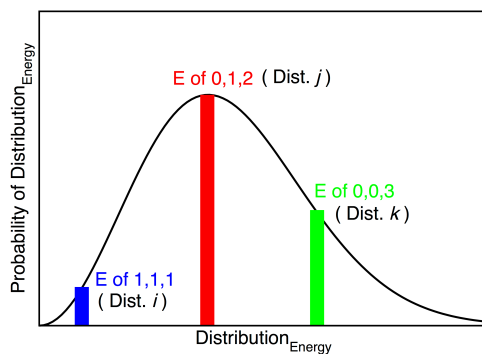
since $0! = 1$ by definition, and for distribution j it is

$$W_j = 3! \,/\, 1!1!1!0! = 6$$

and for distribution i it is

$$W_i = 3! \,/\, 0!3!0!0 = 1$$

In this case distribution j has the largest statistical weight (W) and is the configuration that is most probable.

The three distributions of energy states and their respective probabilities can be visualized as follows:


The goal of the Boltzmann derivation is to find an expression for W in terms of the distribution of energies (E) in the various states, $n_1, n_2, n_3 \dots n_i$. The distribution with the largest W will be the most likely distribution.

As a starting point we know one relationship of W to n. The derivative of W with respect to n will be zero when W is at a maximum

$$\partial W/\partial n = 0, \quad (W = \text{max}) \tag{4a}$$

and we could evaluate our final expression at this condition to insure that W is at a maximum. The following figure illustrates that at equilibrium, the most probable distribution of the states n_i is the peak of the curve and by definition this is where W is at a maximum. At this peak the instantaneous change in W with change in n is zero (at the absolute peak where slope is zero, an infinitesimal change in n does not change W).
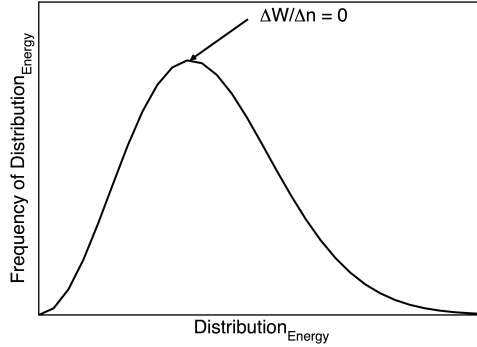
It is more convenient to find the maximum of the natural logarithm of W rather than W so we will use,

$$\partial \ln W/\partial n = 0, \quad (W = \text{max}) \tag{4b}$$

We will arbitrarily introduce two new constant terms to this expression,

$$\partial(\ln W - \alpha\sum n_i - \beta\sum n_i E_i) / \partial n = 0 \tag{5a}$$

These new terms ($\sum n_i$ and $\sum n_i E_i$) with constants α and β are equal to zero in this expression because the number of states ($\sum n_i$) is constant and therefore the derivative of these terms with respect to n equals zero; the values of α and βE_i can be anything because they are multiplied by a zero term in this expression. (The mathematical justification for doing this is explained in Appendix A under Lagrange multipliers). Equation 5a then becomes,

$$\partial \ln W/\partial n - \alpha - \beta E_i = 0 \tag{5b}$$

or

$$\partial \ln W/\partial n = \alpha + \beta E_i \tag{5c}$$

At this point the constants α and β are undefined, but we will come back to them.

From logical extensions of the above definition of W in equation (3) it can also be shown that that the derivative is equal to the negative log of the ratio n/N (see Appendix A)

$$\partial \ln W/\partial n = -\ln(n_i/N) \tag{6}$$

Combining these two last expressions (equations 5c and 6) and changing signs we get

$$\ln(\boldsymbol{n}_i/\boldsymbol{N}) = -\boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{E}_i \tag{7}$$

or

$$n_i/N = e^{-\alpha}e^{-\beta E_i} \tag{8}$$

Since $\boldsymbol{n}_i/\boldsymbol{N}$ represents the probability $\boldsymbol{P}_i$ of finding a particular system in state i and

$$e^{-\alpha} = \frac{1}{e^{\alpha}}$$

then,

$$P_i = \frac{e^{-\beta E_i}}{e^{\alpha}} \tag{9}$$

Since $\boldsymbol{P}$ is a probability, the denominator in equation (9) must be equal to the sum of individual state probabilities and therefore

$$e^{\alpha} = \sum_i e^{-\beta E_i} \tag{10}$$

This quantity, the sum over all states, is known as the **partition function** and is frequently designated $\boldsymbol{Z}$,

$$Z = \sum_j e^{-\beta E_j} \tag{11}$$

The partition function is an important parameter of a system. If you know the partition function of a system you know everything there is to know about that system at equilibrium because you know *all* the energy states.

Finally, we have

$$P_i = \frac{e^{-\beta E_i}}{Z} \tag{12}$$

As we prove in Appendix A the constant $\boldsymbol{\beta}$ equals $\boldsymbol{1/kT}$ where $\boldsymbol{k}$ is Boltzmann's constant and $\boldsymbol{T}$ is temperature. So we may write

$$P_i = \frac{e^{-E_i/kT}}{Z} \tag{13}$$

This expression implies that, since low energy states are more likely, all systems would eventually exist at the lowest energy state, the ground state. Why doesn't this happen?

The reason is that real systems are degenerate; **the higher the energy of a state the more ways there are to distribute this energy**. Think of a protein that has a specific energy: There may be multiple combinations of bond angles, internal vibrations, and interactions with solvent that give the same total energy. So the more appropriate expression for real systems includes the degeneracy term, **g**

$$P_i = g_i * e^{-E_i/kT}/Z \qquad (14)$$

and here the partition function would also include the degeneracy term because it is the sum of all possible numerators,

$$Z = \sum_j g_i e^{-\beta E_j} \qquad (15)$$

The negative energy term and the positive degeneracy term have opposite influences on the probability of a state. Consider the following figure:
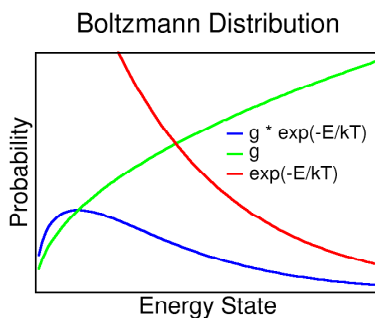
The two opposing terms result in a probability distribution of states that is approximately Gaussian (admittedly very skewed in this example). **This distribution represents the Boltzmann distribution and the above figure summarizes important concepts for this course.**

Another way to conceptualize the interplay of energy and entropy is shown in the following figure. The vertical axis represents potential energy of a protein and the horizontal (radial area) represents the number of possible ways to achieve each energy state. This plot is called an **energy landscape funnel**. One can imagine a protein folding by following a path from the rim of the funnel (where many possible conformations are possible) down to the energy minimum at the bottom where only one conformation is found.
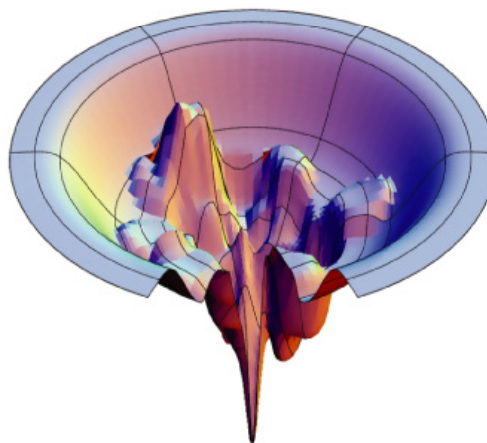
IV. Thermodynamics

Thermodynamics is the description of energy conversions at a level that deals with large collections of molecules - it helps us understand the average characteristics of energy utilization at the level of macrostates that can be observed rather than microstates that we cannot observe. The Boltzmann probability formula is also a restatement of the familiar thermodynamic equation defining Gibbs free energy

$$\Delta G° = \Delta H° - T\Delta S° = -RT\ln K \tag{16}$$

To arrive at this relationship from Boltzmann's point of view we first put the Boltzmann formula in terms of molar energy instead of molecular energy. To do this we multiply by Boltzmann's constant by Avogadro's number.

$$R = N_A k \tag{17}$$

From this relationship, you can see that the gas constant, R, is a molar Boltzmann's constant. Now we rearrange,

$$N_A = R/k \tag{18}$$

and convert the Boltzmann exponential term ($-E_i/kT$) to a molar quantity,

$$(-E_i/kT) * (N_A/N_A) = -N_A(E_i/RT) \tag{19}$$

We define E_A as the energy per mole for state A to give a molar Boltzmann exponential term,

$$-N_A(E_i/RT) = -E_A/RT \tag{20}$$

Nature usually operates under conditions of constant pressure therefore we use enthalpy (H_A) as the energy term (E_A is appropriate for constant volume). The enthalpy of a system is

$$H_A = E_A + PV \tag{21}$$

The pressure-volume term (PV) accounts for work done on or by the surroundings.

Thus the Boltzmann formula in terms of molar energy at constant pressure (and adding back the degenercy term) is

$$P_A = g_A \cdot e^{-H_A/RT}/Z \tag{22}$$

The system could also exist in a different state (B), with a different probability

$$P_B = g_B \cdot e^{-H_B/RT}/Z \tag{23}$$

and we assume the two states may exchange energy and interconvert.

$$A \longleftrightarrow B \tag{24}$$

The relative probability of finding the system in either state is equal to the equilibrium constant for the interconversion. As written above the equilibrium constant (K) is

$$K = [B]/[A] = P_B/P_A \tag{25}$$

Substituting the Boltzmann probability formula into this expression for the equilibrium constant gives

$$K = g_B e^{-H_B/RT}/Z \, / \, g_A e^{-H_A/RT}/Z \tag{26}$$

since Z includes both states A and B it cancels out.

$$K = g_B e^{-H_B/RT} \, / \, g_A e^{-H_A/RT} \tag{27}$$

Expand the exponential

$$\ln K = \ln g_B + \ln(e^{-H_B/RT}) - \ln g_A - \ln(e^{-H_A/RT}) \tag{28}$$

$$\ln K = \ln g_B - H_B/RT - \ln g_A + H_A/RT \tag{29}$$

Multiply both sides by RT and cancel

$$RT\ln K = RT\ln g_B - H_B - RT\ln g_A + H_A \tag{30}$$

Change signs

$$-RT\ln K = H_B - H_A - RT\ln g_B + RT\ln g_A \tag{31}$$

We define the standard state molar enthalpy difference between the two states as

$$\Delta H^\circ = H_B - H_A \tag{32}$$

giving

$$-RT\ln K = \Delta H^\circ - RT\ln g_B + RT\ln g_A \tag{33}$$

We change the sign and combine the natural log terms

$$-RT\ln K = \Delta H^\circ - RT\ln g_B - RT*(-\ln g_A) \tag{34}$$

$$-RT\ln K = \Delta H^\circ - RT\ln(g_B/g_A) \tag{35}$$

Now we interpret the g_B/g_A term. Boltzmann's famous equation relating entropy and probability is

$$S = k \ln W \tag{36}$$

or on a molar basis in the standard state

$$S° = R \ln W \tag{37}$$

Now as we discussed above W is equal to the number of distinguishable ways the system can be distributed among available states

$$W_A = g_A \tag{38}$$

and so we can write

$$\Delta S° = R(\ln W_B - \ln W_A) = R\ln(W_B/W_A) = R\ln(g_B/g_A) \tag{39}$$

Finally using this definition of entropy in the equation above (equation 34) we obtain

$$-RT\ln K = \Delta H° - T\Delta S° \tag{40}$$

or

$$\Delta G° = \Delta H° - T\Delta S° \tag{41}$$

In other words, the Gibbs free energy change ($\Delta G°$) of the interconversion ($A <\!\!-\!\!-\!\!-\!\!> B$) is related to the relative probability (P_B/P_A) of finding the system in state B compared to A. If $\Delta G°$ is negative, $P_B/P_A > 1$ and the system is likely to be in state B.

When Boltzmann demonstrated that macroscopic thermodynamic relationships could be derived from consideration of particle statistics it was revolutionary.

In the above we assumed that the relationship between the probabilities of *microstates* and their respective Boltzmann terms,

$$P_i = g_i{*}e^{-E_i/kT}/Z \tag{14}$$

would scale seamlessly to the relationship between the probabilities of *macrostates* and their respective Boltzmann energy terms,

$$P_A = g_A.e^{-H_A/RT}/Z \tag{22}$$

In general, the probability of the macrostate A is given by the sum of probabilities of all the microstates of the system that make up the observed state A.

$$P_A = \sum_{i_A} P_i = \frac{1}{Z} \sum_{i_A} g_i e^{-E_i/kT} \tag{42}$$

where the sum is over only those microstates that have observed characteristics of macrostate A. This idea that the Boltzmann weight of macrostate A is the sum of its

constituent microstates is illustrated in the following figure that lists all the states in this hypothetical system.

Macrostate	Microstate	Weight
A	1	$g_1 e^{-E_1/kT}$
	2	$g_2 e^{-E_2/kT}$
	.	.
	.	.
	n1	$g_{n1} e^{-E_{n1}/kT}$
B	n1+1	$g_{n1+1} e^{-E_{n1+1}/kT}$
	n1+2	$g_{n1+2} e^{-E_{n1+2}/kT}$
	.	.
	n2	$g_{n2} e^{-E_{n2}/kT}$
C	n2+1	$g_{n2+1} e^{-E_{n2+1}/kT}$
	n2+2	$g_{n2+2} e^{-E_{n2+1}/kT}$
	.	.
	n3	$g_{n3} e^{-E_{n3}/kT}$

Using the basic relation between the partition function and the free energy, $G = -kT\ln Z$, we can also express the probability of macrostate A as

$$P_A = \frac{1}{Z} e^{-G_A/RT} \tag{43}$$

where Z is now the sum of all the Boltzmann terms for individual *macrostates*,

$$G_A = -kT \ln\left(\sum_{iA} e^{-E_{iA}/RT}\right)$$

This is illustrated in the figure at right where the complete list of macrostates in the system and their respective weights is listed. Here G_A is the Gibbs free energy of the macrostate A. The free energies appearing in the weights are computed by doing a restricted sum over all the microstates contributing to the macrostate.

Macrostate	Weight
A	$e^{-G_A/RT}$
B	$e^{-G_B/RT}$
C	$e^{-G_C/RT}$

V. Second Moments and Fluctuations of protein energies.

Now we will see how a logical extension of the probability distribution concept allows us to say something about the behavior of individual molecules. The question we address: *What is the distribution of energy fluctuations in a population of protein molecules at room temperature?*

The following example is taken from a 1976 paper (Cooper PNAS 73(8):2740). A more complete description of the derivation is in Appendix D.

Folded proteins are, on average, in their free energy minimum state. We can think of the equilibrium between an unfolded and folded protein as follows,

$$U \leftrightarrow F \tag{44}$$

The free energy change upon folding for proteins is approximately,

$$\Delta G_F^0 = -10 \, kcal/mol \tag{45}$$

where the subscript F indicates in the direction of folding; the negative energy indicates that folding is a favorable reaction. We call this the **energy of stabilization** and most proteins have an energy of stabilization between -5 and -10 kcal/mol. Because the energy distribution is reasonably sharp most protein molecules stay within this range (or at least in the negative ΔG_F range) and are stable in the ensemble at room temperature. The figure below shows the situation schematically.
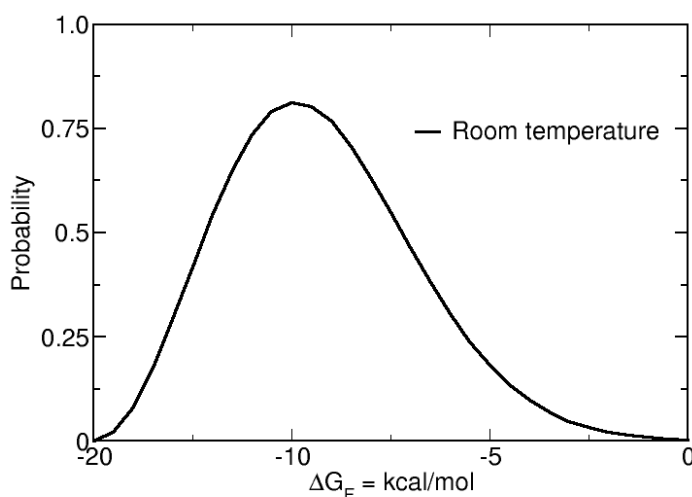
In other words, the *bulk* property is one of stability; most protein molecules in the ensemble have a negative ΔG_F. Individual protein molecules are very small systems, however, and in these cases statistical fluctuations in thermodynamic properties are of greater importance as shown below. The curve above approaches zero asymptotically on the right side and although the probability of a positive ΔG_F is low, it is finite.

The mean potential energy of a system such as a population of molecules at constant pressure is equal to the enthalpy of a state, H_i, times its probability, P_i, summed over all states

$$<H> = \sum H_i P_i = \sum (H_i e^{-\beta H_i}/Z) \tag{46}$$

The width of the potential energy distribution for a protein at constant pressure is proportional to the mean-square deviation of the enthalpy

$$< \Delta H^2 > = <(H_i - <H>)^2 >$$

The mean square deviation gives an approximation of the probability distribution. It is a measure of the fluctuation we are interested in. We can write this as a function of individual deviations and their respective probabilities.

$$< \Delta H^2 > = \sum \Delta H_i^2 P_i \tag{47}$$

In Appendix D we derive an expression which relates the mean-square deviation of the enthalpy to the heat capacity, C_p, of a protein

$$<\Delta H^2 > = kT^2 C_p \tag{48}$$

We can measure the heat capacity (C_p) of a protein (using calorimetry) and since the temperature (T) and Boltzmann's constant (k) are known values we can calculate the fluctuations of energy for a single protein. For a typical protein of MW 25,000:

mass	25,000 g/mol / 6.022×10^{23} molecule/mol = 4.15×10^{-20} g/molecule
C_p	0.32 cal g^{-1} Kelvin^{-1}
k	0.33 x 10^{-23} cal Kelvin^{-1}
T	298 °Kelvin

Therefore, the root mean square internal energy fluctuation (ΔH_{rms}) for an individual protein molecule is on the order of

$$\Delta H_{rms} = sqrt(<\Delta H_i^2>) = sqrt(kT^2 C_p) \tag{49}$$

Substituting known values and correcting for room temperature and the size of a typical protein molecule

$$= sqrt[(0.33 \times 10^{-23} \text{ cal K}^{-1}) \times (298^2 KK) \times (0.32 \text{ cal g}^{-1} \text{ K}^{-1}) \times (4.2 \times 10^{-20} \text{ g})]$$

$$= 6.3 \times 10^{-20} \text{ cal/molecule}$$

and to relate this to the more familiar energy units of kcal/mole,

$$= 6.3 \times 10^{-20} \text{ cal/molecule} \times 6.022 \times 10^{23} \text{ molecules/mol}$$

$$= 37938 \text{ cal mol}^{-1}$$

$$= \textbf{38 kcal mol}^{-1} \text{ (!)}$$

Remember that these are fluctuations within a single molecule and are uncorrelated with fluctuations in other molecules in the ensemble. (The number of 38 kcal/mol would be experienced only if the entire mole of proteins fluctuated the same degree at the same time!) The frequency of an individual molecule sampling this fluctuation is low (it may take years for a protein molecule to sample this extreme state) – but the point is that it does occur with a non-zero probability.

The value of 38 kcal mol^{-1} compares with the typical unfolding energy of proteins of approximately 5-10 kcal mol^{-1}. The energy fluctuation experienced (infrequently) by an individual protein is **many times** the energy necessary to unfold the protein! *Thus, individual proteins sample very large energetic fluctuations on the order of complete denaturation*. In a test tube of μmolar protein a small fraction of the proteins are undergoing complete denaturation at any instant in time. Many of these will re-fold, but because the re-folding may not be 100%, overall the bulk protein will usually denature with time.

What happens to the shape of the above protein energy stability curve with increasing temperature? We know from equation 48 that the width of the curve will increase. The maximum energies sampled will also increase. The following figure illustrates schematically how the Boltzmann distribution of protein energies changes at high temperatures. Notice that a significant proportion of the population has a positive ΔG_F at higher temperatures. This means that the "folding" reaction is not favorable and as individual molecules fluctuate into the high energy unfolded conformation they will not re-fold – the protein denatures.
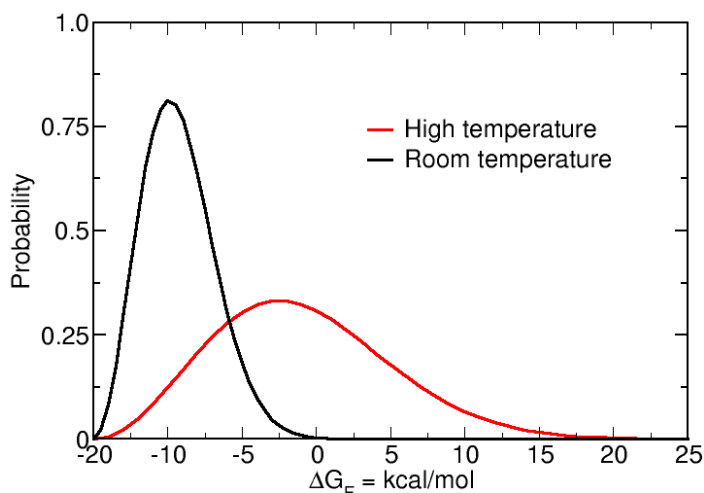
Keywords. You should be able to define or explain each of these terms with a short phrase, mathematical expression or sentence.

Boltzmann distribution formula (law), components
Boltzmann's constant, relation to gas constant
configuration space, C
degeneracy
enthalpy
entropy
free energy
macrostate
microstate
multiplicity, Ω
phase space, P
potential energy
protein energy fluctuations
protein energy of stabilization
state counting
Wahrscheinlichkeit
Derive $\Delta G° = \Delta H° - T\Delta S°$ from $P_i = (g_i * e^{-Ei/kT})/Z$

Appendix A: Derivation of Boltzmann's equations.

We will derive the expression describing a Boltzmann distribution from first principles in excruciating detail. Parts of this derivation may be found in many reviews and textbooks. These treatments usually skip over critical steps. One frequently comes to the phrase, "It can be shown that …". For students with extensive math background such a phrase might be a stimulation to go figure out how it can be shown. For others (like myself) this phrase is frustrating. Here we *show* how one expression can be restated as another, what assumptions are necessary and what math tricks are used.

You are not responsible for this level of detail. The reason for including the detailed derivation here is simply to have a reference in one place for all the steps. I wanted it for myself and thought it might be useful to some of you. There is another derivation that is simpler in Appendix B below.

Consider a "system", which in our case is a single protein molecule immersed in and interacting with a "bath" which may be the solvent. The microstate of the system at any one time may be described in terms of extensive physical quantities such as total energy, volume, momentum, extent of hydration, etc. Because of molecular interactions and collisions with the bath, the microstate of the system and the associated physical parameters will change with time. *We want the mean values of these parameters and statistical variations of these values*. With these values in hand we may state the probability that a system will be in any particular microstate. Rather than taking the time average of a single system it is advantageous to use the concept of a thermodynamic ensemble and take an instantaneous average over a large number of identical systems.

Imagine the ensemble depicted in Figure 1 which consists of N systems and M baths surrounded by an isolating wall. This is equivalent the canonical (NVT) ensemble. N and M are arbitrarily large and may be thought of as protein molecules (systems) immersed in respective regions of the solvent (baths). It is likely that each system is slightly different than the others and therefore there will be a distribution of states.
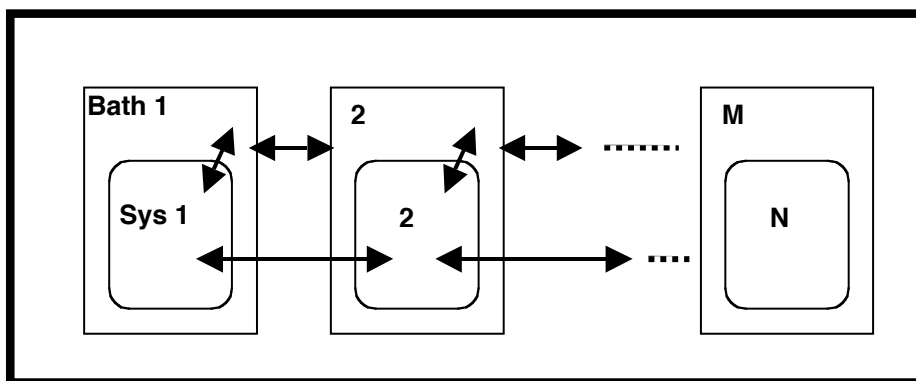

Figure 1. Ensemble of N systems and M baths contained with an isolated enclosure.

Any particular state, i, of a system or bath will be defined by the set of quantities where,

for example, E_i = energy of the system in state i, V_i = volume of system in state i, etc. All the systems and baths in the ensemble can exchange energy and other extensive quantities by molecular collisions (arrows, Figure 1, above).

We may arbitrarily assign n_1 systems to system state 1, n_2 to state 2…n_i to state i and m_1 baths to bath state 1, m_2 to state 2…m_i to state i. Assume for now that g=1, i.e. the states are not degenerate and all states have one system. The total numbers of systems and baths is constant and arbitrarily large:

$$\sum n_i = N \qquad \text{and} \qquad \sum m_i = M \qquad \qquad (1)$$

and any changes in the distributions due to exchange must obey:

$$\sum dn_i = 0 \qquad \text{and} \qquad \sum dm_i = 0 \qquad \qquad (2)$$

i.e. if one state loses energy another gains the equivalent energy because the extensive quantities are conserved, thus the total quantity E (for example, energy) in the ensemble is

$$E(tot) = \sum n_i E(system) + \sum m_i E(bath) \qquad \qquad (3)$$

We will derive the case for the extensive parameter of energy (E_i) and for the systems only. The same derivation holds for other extensive quantities (volume, momentum, hydration, etc.) and for the baths as well. The equivalent expressions to equations 1 and 3 are then

$$\sum n_i = N \qquad \qquad (1a)$$

$$E_{tot}(systems) = \sum n_i E_i(system) \qquad \qquad (3a)$$

Equations 1a and 3a represent **constraints** on the ways that states can be distributed amongst the systems in the ensemble.

The distribution of states is defined by how many systems are in state n_1, n_2…n_i. **At equilibrium the distribution of states is the most probable distribution - this is the central dogma of statistical thermodynamics.** The most probable distribution will be that which happens the most often; that is, it has the largest statistical weight. The statistical weight (W) is proportional to the number of ways of arranging N systems across the available states n_1, n_2, … in our ensemble.

As an example consider the number of ways of distributing an ensemble of 3 systems (**x**, **y**, **z**) in different states given the constraints that:
1. the number of systems ($\sum n_i$) is constant
2. the energy comes in discrete units (0, 1, 2, 3)
3. the average system energy $<E_i>$ = 1; the total ensemble energy $\sum n_i E_i = 3$
4. there is only one way for the energy to be achieved in each state (i.e. the states are not degenerate, $g_i = 1$).

The available states will be n_1, n_2, n_3, n_4 with respective energies (**E**) of 0, 1, 2 and 3 units. The only available ways of distributing the systems are indicated in the table below where i, j, and k are three different distributions of the ensemble.

Energy	State	Dist. i	Dist. j	Dist. k
E = 0	n_1	0	1	2
E = 1	n_2	3	1	0
E = 2	n_3	0	1	0
E = 3	n_4	0	0	1
$\sum n_i E_i$		3	3	3

In distribution i there are 3 systems (**x**, **y** and **z**) in state n_2 and 0 in other states; this is equivalent to the systems being evenly distributed in energy and since there is no degeneracy they are all identical. In distribution j, there is 1 system each in states n_1, n_2, and n_3, all with different energies. In distribution k, 2 systems in state n_1 and 1 system in state n_4. In all configurations the average system energy is 1 and the total energy is 3.

For each of these distributions there are *apparently* 3! ways to distribute the systems in the available states. For example, consider the three systems **x**, **y** and **z** in configuration j. We can choose any one of the three systems to be in state n_1 (3 choices). Assume we will choose **y** (n_{1y}). Then we have **x** or **z** to choose from for state n_2 (2 choices); we choose **x** (n_{2x}). Then we have only one choice left for state n_3 (n_{3z}). There are 3 systems to choose from to put the first system in a particular state, 2 systems to choose from to place the second system and only 1 system to choose to place the third (3x2x1 = 3!). This is true since each system in distribution j has a **distinguishable** energy.

However, consider distribution k; if there are 2 systems in state n_1, these systems have the same energy state and therefore are indistinguishable. There are not as many distinguishable ways to distribute the systems. In distribution k the two systems in n_1 only count once. Formally we would have to divide our 3! ways in distribution k by 2! to correct for the over counting that occurred in state n_1. Calculation of the actual weights for this example may make this concept clearer.
The general expression for calculating the weight of a configuration is

$$W = N! / n_1! \, n_2! \, \ldots \, n_X! \qquad\qquad (4)$$

where **N** is the number of systems and **X** is the total number of states. So the final expression for distribution k with 2 systems in state n_1 and 1 system in state n_4 is

$$W_k = 3! / 2!0!0!1! = 3$$

since $0! = 1$ by definition, and for distribution j it is

$$W_j = 3! / 1!1!1!0! = 6$$

and for distribution i it is

$$W_i = 3! / 0!3!0!0 = 1$$

In this case distribution j has the largest weight (W) and is the configuration that is most probable.

We wish to find the energy distribution of n that leads to a maximum W. To accomplish this we require an expression that relates W to the distribution of energy states with the constraints that the number of systems is constant and the total energy is 3.

In more mathematical terms we want an expression that gives the derivative of W with respect to the different states, $\mathbf{n_1} \ldots \mathbf{n_3}$ and to find the maximum of this expression. For the derivation to follow it will be more convenient to find the maximum of the natural logarithm of W rather than W. We start by expanding and removing the factorials in equation 4. The expansion is

$$\ln W = \ln(N!/n_1! \, n_2! \ldots n_i!)$$
$$\ln W = \ln N! - \ln (n_1! \, n_2! \ldots n_i!)$$
$$\ln W = \ln N! - (\ln n_1! + \ln n_2! + \ldots \ln n_i!) \qquad (5)$$

where N is the total number of systems. And then we use Stirling's approximation for large numbers to eliminate the factorial which will make further derivation easier. Stirling's approximation is

$$\ln x! = x \ln x - x$$

so

$$\ln W = (N \ln N - N) - \Sigma_i[n_i \, \ln(n_i) - n_i] \qquad (6)$$

Separate the sum terms

$$\ln W = N \ln N - N - \Sigma_i[n_i \, \ln(n_i)] + \Sigma_i \, n_i$$

and since $\mathbf{N = \Sigma n_i}$

$$\ln W = (N \ln N) - \Sigma_i[n_i \, \ln(n_i)] \qquad (7)$$

When the distributions within the ensemble change states (e.g. $\mathbf{n_i \rightarrow n_i + dn_i}$) the function $\mathbf{lnW}$ changes to $\mathbf{lnW + dlnW}$ where

$$d\ln W = \Sigma_i[\partial\ln W/\partial n]dn_i$$

At the maximum $\mathbf{W}$, $\mathbf{dlnW} = 0$. This is the case we wish to define. That is we wish to arrive at an expression that describes the distribution of systems among states, $\mathbf{n_i}$ for the most probable case.

When the change in **lnW** is evaluated with respect to **n**

$$\partial lnW/\partial n = \partial(N\ lnN\ )/\partial n - \sum_j [\partial(n_i\ ln\ n_i)\ /\partial n\ ] \qquad (8)$$

and at maximum **W**

$$\partial lnW/\partial n = \partial(N\ lnN\ )/\partial n - \sum_j [\partial(n_i\ ln\ n_i)\ /\partial n\ ] = 0$$

Because states $\mathbf{n}_i$ are not all independent of one another (as one loses energy another must gain it), we cannot evaluate the differential as written. This dependence of states on each other is defined by the constraints that the total energy is a constant (equation 3a) and the number of states is also a constant (equation 1a). These constraints can actually help us out at this point in the derivation by allowing us to introduce terms that relate **W** to the energy of each state $\mathbf{E}_i$.

We can take dependent constraints into account using the method of undetermined multipliers developed by the French mathematician, Lagrange. The rule is that a constraint should be multiplied by a constant and then the product added to the main expression. The dependent variables become independent and we can evaluate the above expression. The values for the constants can be evaluated at the end of the calculation.

Remember from equations 1a and 3a

$$\sum n_i = N$$

$$\sum n_i E_i(system) = E_{tot}(systems)$$

and since these are constants

$$d(\sum n_i) = \sum dn_i = 0$$

$$d(\sum n_i E_i) = \sum dn_i E_i = 0$$

We can multiply these expressions by arbitrary constants and add them to anything since they are zero. Here we introduce two constants: α, for the constraint that any changes in the distribution of systems among states must sum to zero (number of systems does not change, equation 1a) and β for the constraint that any changes in the distribution of energies must sum to zero (total energy does not change with exchange of states, equation 3a). We add them as the negative value to help with the algebra later on.

$$dlnW = \sum_i [\partial lnW/\partial n]dn_i + \sum_i [-\alpha dn_i\ ] + \sum_i [-\beta E_i dn_i\ ]$$

$$dlnW = \sum_i [(\partial lnW/\partial n) - \alpha - \beta E_i\ ]dn_i \qquad (9)$$

All the n_i are now treated as independent. Hence the only way of satisfying **dlnW** = 0 is to require that for each *i*,

$$(\partial \ln W/\partial n) - \alpha - \beta E_i = 0 \tag{10}$$

Let us go back to equation (7) and remember that

$$\ln W = (N \ln N) - \sum_i[n_i \ln(n_i)]$$

and that differentiation with respect to $\mathbf{n}_i$ gives equation 8

$$\partial \ln W/\partial n = \partial(N \ln N)/\partial n - \sum_j [\partial(n_i \ln n_i)/\partial n]$$

We want to differentiate this equation taking each state with respect to the other states. In other words, $\mathbf{n_1}$ with respect to $\mathbf{n_1, n_2, n_3, n_4...n_N}$; $\mathbf{n_2}$ with respect to $\mathbf{n_1, n_2, n_3, n_4...n_N}$; etc. To indicate that the indices of the state $\mathbf{n}$ in the numerator are not the same as the indices of the $\mathbf{n}$ in the denominator we introduce the index *j* in the numerator and include the index *i* in the denominator

$$\partial \ln W/\partial n = \partial (N \ln N)/\partial n_i - \sum_j \partial[n_j \ln(n_j)]/\partial n_i \tag{11}$$

We now take the derivative of the first term on the right of equation (11). Using the product rule for differentiation

$$\partial (N \ln N)/\partial n = (\partial N/\partial n) \ln N + N(\partial \ln N/\partial n) \tag{12}$$

And remembering that

$$d\ln N = 1/N (dN)$$

We may write

$$\partial (N \ln N)/\partial n = (\partial N/\partial n_i) \ln N + N [1/N (\partial N/\partial n)]$$

$$\partial (N \ln N)/\partial n = (\partial N/\partial n) \ln N + \partial N/\partial n \tag{13}$$

and because $\mathbf{N = n_1 + n_2 + n_3 +...}$ and its derivative with respect to any of the $\mathbf{n}$'s is 1 we end up with

$$\partial (N \ln N)/\partial n = \ln N + 1 \tag{14}$$

The derivative of the second term on the right of equation (11) is (again using the product rule)

$$\sum_j \partial[n_j \ln(n_j)]/\partial n_i = \sum_j[(\partial n_j/\partial n_i)\ln n_j + n_j(\partial \ln n_j/\partial n_i)] \tag{15}$$

And using

$$d\ln n = 1/n \, (dn)$$

we may write

$$\sum_j \partial[n_j \ln(n_j)]/\partial n_i = \sum_j[(\partial n_j /\partial n_i)\ln n_j + (n_j / n_j)(\partial n_j /\partial n_i)] \qquad (16)$$

Now (n_j /n_i) is a delta function. For the cases where i does not $= j$, n_i is independent of n_j and so $\partial n_j/\partial n_i = 0$ and in these cases the term vanishes. For the case where $i = j$, $\partial n_j/\partial n_i = \partial n_j/\partial n_j = 1$. For one case in the summation then, $\ln n_j = \ln n_i$. Therefore

$$\sum_j \partial[n_j \ln(n_j)]/\partial n_i = [1*(\ln n_j) + (n_j / n_j)*1] \qquad (17)$$

And since $(n_j / n_j) =1$

$$\sum_j \partial[n_j \ln(n_j)]/\partial n_i = (\ln n_i + 1) \qquad (18)$$

Combining both terms we have

$$\partial \ln W/\partial n = \ln N +1 -(\ln n + 1) = -\ln(n_i/N) \qquad (19)$$

and substituting this definition into equation (10) above

$$-\ln(n_i/N) - \alpha - \beta E_i = 0$$

$$\ln(n_i/N) = -\alpha - \beta E_i \qquad (20)$$

and therefore

$$n_i /N = e^{-\alpha}e^{-\beta E_i} \qquad (21)$$

Since n_i/N represents the probability P_i of finding a particular system in state i with respect to its internal energy E

$$P_i = \frac{e^{-\beta E_i}}{e^{\alpha}} \qquad (22)$$

Thus we have the probability distribution function in terms of the arbitrary multipliers α and β. We can eliminate α by recognizing that if P_i is a probability of state i the sum of probabilities for all states, $i...N$, is unity

$$\sum_i P_i = 1$$

$$\sum_i \left(e^{-\beta E_i} / e^{\alpha} \right) = 1$$

$$\left(\frac{1}{e^{\alpha}} \right) \sum_i \left(e^{-\beta E_i} \right) = 1$$

$$\sum_i e^{-\beta E_i} = e^{\alpha}$$

so that

$$e^{\alpha} = \sum_i e^{-\beta E_i} \tag{23}$$

This quantity, the sum over all states, is known as the partition function and is frequently designated Z, and it is so important that the definition is repeated:

$$Z = \sum_i e^{-\beta E_i} \tag{24}$$

The adjective *partition* is used because the quantity is a description of how systems are partitioned (or distributed) among the available states. If you know the partition function you know everything there is to know about that system.

Finally, we have

$$P_i = \frac{e^{-\beta E_i}}{Z} \tag{25}$$

which describes the Boltzmann distribution, i.e. the fraction of systems in state *i* with extensive quantity (such as energy) E_i. (Note that some texts use **q** instead of **Z** for the partition function).

We still want to relate β to some physical quantity. Remember that in the ensemble each β is the same for all systems. This suggests that the β's might be related to intensive thermodynamic quantities such as temperature, pressure or chemical potential: all of which must be the same for systems and baths in classical thermodynamic equilibrium. In fact we will now show that β is related to temperature. Since in this course we are more interested in conformational diversity (entropy) here we will start from an expression of entropy. (Note: An alternate derivation for the relationship of β to temperature may be obtained starting with an expression for the internal energy of a system.)

Boltzmann's famous equation relating entropy and probability is

$$S = k \ln W \tag{26}$$

where **k** is Boltzmann's constant. This brings us back to the above assumption that at equilibrium our ensemble has maximum **W** and is equivalent to the Second Law of Thermodynamics which states that at equilibrium the entropy of the ensemble (the entire system) is a maximum.

Recall from equation 4 that

$$W = N!/ n_1! \, n_2! \, \ldots \, n_i!$$

so the entropy per system (divide equation 26 by **N**) will be

$$S = k/N \ln[N!/ n_1! \, n_2! \, \ldots \, n_i!] \tag{27}$$

A second form of Sterling's approximation for large factorials is,

$$\lim_{n \to \infty} \frac{n!}{\sqrt{2\pi n}\left(\dfrac{n}{e}\right)^n} = 1$$

At large n the square root is relatively insignificant compared to the term raised to the power, n, and we can say that,

$$\left(\frac{n}{e}\right)^n = n!$$

So equation (27) becomes,

$$S = k/N \ln[(N/e)^N / (n_1/e)^{n1} (n_2/e)^{n2} \ldots (n_i/e)^{ni}]$$

$$S = k/N \ln[(N^N/e^N) / (n_1^{n1}/e^{n1}) (n_2^{n2}/e^{n2}) \ldots (n_i^{ni}/e^{ni})]$$

$$S = k/N \ln[N^N/e^N / \prod(n_i^{ni}/e^{ni})] \tag{28}$$

since $\mathbf{e^N = \prod e^{ni}}$ the denominators in the log term cancel

$$S = k/N \ln[N^N / \prod(n_i^{ni})]$$

since $\mathbf{N = \sum n_i}$, $\mathbf{N^N = \prod N^{ni}}$

$$S = k/N \ln[\prod(N^{ni} / n_i^{ni})]$$

$$S = k/N \ln[\prod(1/P_i)^{ni}]$$

or
$$S = k/N \ln[1 / P_1^{n_1} P_2^{n_i} \ldots Pi^{n_i}]$$

Change sign and expand log

$$S = k/N(-\sum[n_i \ln P_i])$$

$$S = k(-\sum[n_i/N \ln P_i])$$

and since $\mathbf{P_i = n_i/N}$

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

(Pause here to note that the above equation relates entropy to probabilities. If you study information theory you will see that the entropy of a variable, $\mathbf{X}$ is a measure of the uncertainty of the value of $\mathbf{X}$ and is defined as $\mathbf{-log\sum_{x \in X}[P_x \log P_x]}$)

$$S = -k \sum_i[P_i \ln (e^{-\beta E_i} / Z)]$$

$$S = -k \sum_i[P_i (-\beta E_i - \ln Z)]$$

$$S = -k \sum_i[-P_i \beta E_j i - P_i \ln Z)]$$

$$S = k\sum_i P_i \beta E_i + k\sum_i P_i \ln Z$$

and since $\sum_i P_i E_i = <E_i>$ and $\sum_i P_i = 1$,

$$S = k\beta<E_i> + k\ln Z$$

we arrive at

$$S = k\beta E + k\ln Z \qquad\qquad (29)$$

Partial differentiation of this expression with respect to the extensive quantity $\mathbf{E}$ gives

$$\partial S/\partial E = k\beta$$

or

$$\beta = 1/k \; \partial S/\partial E \qquad\qquad (30)$$

Recalling from classical thermodynamics that $\mathbf{dE = TdS - PdV}$, then at constant volume

$$\partial S/\partial E = 1/T$$

and we may write

$$\beta = 1/k \; \partial S/\partial E = 1/kT \tag{31}$$

which gives the most familiar form of the Boltzmann distribution

$$P_i = e^{-E/kT} / Z \tag{32}$$

In our derivation we assumed that all systems with the same energy were indistinguishable, i.e. $g_i = 1$. In reality most systems are degenerate, i.e. a system may have more than one configuration at a particular energy level. The appropriate expression is then written as

$$P_i = g_i e^{-E_i/kT}/Z \tag{33}$$

where **g** is the degeneracy of energy level *i*. In this case the **Z** would be written

$$Z = \sum g_j e^{-E_j/kT} \tag{34}$$

where again the subscript **j** indicates that the sum is over all energy levels.

At low **T**, $e^{-E/kT}$ is small therefore the number of states is low and at high temperature the number of states is large. At **T** = 0 only the ground state is populated and at **T** = ∞ all states are populated equally. In between these temperature extremes the population depends on E_i.

Appendix B: Alternate derivation of Boltzmann's equations.

The above derivation is the standard one found in most textbooks. However, another derivation was published in 1996 by Douglas Russell in the Journal of Chemical Education (Vol. 73 (4), p. 299-300) and this latter derivation may be easier to follow. The following is based on the article in the Journal of Chemical Education.

Consider the equation on Boltzmann's tombstone

$$S = k \ln W \qquad (1)$$

where **W** refers to entropy or the number of ways a system can achieve an energy state. And remembering the definition of Helmholtz Energy

$$A = E - TS \qquad (2)$$

By substitution

$$A = E - Tk \ln W \qquad (3)$$

we may evaluate **A** for each of the possible energy states from the example above in Appendix A.

Distribution	i	j	k
Number states	1	6	3
Entropy	0	$k \ln 6$	$k \ln 3$
Energy (A)	E - 0	E - Tk ln6	E - Tk ln3

Again, the lowest energy (most favorable) configuration will be j because the total energy (**A**) is lowest for this configuration.

Now extend the above to the more general problem with many microstates (**n_1, n_2, n_3, . . .**). The general expression for calculating the weight of a configuration is

$$W = N! / n_1! \, n_2! \, \dots \, n_X! \qquad (4)$$

where **N** is the total number of systems and **X** is the total number of states.

Substituting equation (4) into equation (3) and summing over all states we obtain

$$A = \sum n_i E_i - kT \ln N! + kT \sum \ln n_i! \qquad (5)$$

We now seek the distribution of **n_i** that gives the minimum value of **A**. Consider the effect on **A** of a transfer of one system from level 1 to level 2 to generate the configuration **n_1 -1, n_2 +1, n_3, ...** Only those terms in **n_1** and **n_2** will be affected so that we find

$$\Delta A = (n_1 -1)E_0 + (n_2 +1)E_1 + kT\ln(n_1 -1)! + kT\ln(n_2 +1)!$$
$$- n_1 E_0 - n_2 E_1 - kT\ln n_1! - kT\ln n_2! \tag{6}$$

Using the definition of factorials [**n! = n(n -1) . . . 2 x 1**] this expression can be rearranged to

$$\Delta A = (E_1 - E_0) + kT \ln (n_2 +1 / n_1) \tag{7}$$

Assuming that the number of systems in each state n_i is very large we can make the approximation

$$\Delta A = (E_1 - E_0) + kT \ln (n_2 / n_1) \tag{8}$$

This form would not be valid in the example above with only three systems but would be approximately true for large numbers.

We may derive a similar expression for the reverse transfer from $\mathbf{E_1}$ to $\mathbf{E_0}$ to produce the configuration $\mathbf{n_1 +1, n_2 -1, n_3, . . .}$ This would yield

$$\Delta A' = (E_0 - E_1) + kT \ln (n_1 / n_2) \tag{9}$$

This is just the negative of equation 8. If the original configuration was that which produced the minimum **A**, then no transfer of a system from any level to any other can produce a negative change in **A**. The only way in which both eq. 8 and eq. 9 can be non-negative is if both are equal to zero. Equating eqs. (8) or (9) to zero and rearranging produces

$$\frac{n_1}{n_2} = \frac{e^{-E_1 /kT}}{e^{-E_2 /kt}} \tag{10}$$

which is a restatement of the Boltzmann relationship.

Appendix C: The origin of Boltzmann's constant, *k*

Entropy as defined by *S*, is an extensive property of the state of a system. Entropy increases with the temperature of a system and it increases with the size of the system.

The increase in entropy of a system is manifested by an increase in the number of possible microstates. Therefore, there must be a relationship between the increase in entropy and the increase in the number of microstates. This relationship provides a fundamental link between classical thermodynamics and statistical thermodynamics.

Consider two systems that are exactly alike. If the entropy in each system is *S*, then the entropy of a larger system formed by the combination of the two is $S + S = 2S$.

The number of microstates does not combine in this manner. If *W* represents the total number of microstates in each system before combination, then the total number for the larger combined system is $W * W = W^2$. This is because for each microstate of one original system there are *W* microstates of the other one. The total number of microstates is obtained when all *W* microstates of the first system have each been multiplied by *W* microstates of the second system – i.e. the product of *W* values.

The only way that an increase in *S* can be directly related to an increase in *W* is for *S* to be proportional to the log of *W*. The constant of proportionality can be obtained if both the *S* and the *W* can be calculated – this can be done for the ideal gas. The proportionality constant in this case is found to be the gas constant divided by Avogadro's number and this is Boltzmann's constant, *k*.

Appendix D: The energetics of protein conformational fluctuations.

We want to relate Boltzmann-like probability distributions to fluctuations of a system such as a protein. We are interested in the fluctuations that individual proteins undergo and how we can estimate the magnitude and dynamics of these fluctuations. In this section we will derive an expression that allows us to estimate the fluctuations of an individual protein system from experimental measurements of an ensemble extensive parameter, heat capacity. As in Appendix A, each step of the derivation is explained at a level of detail that may be superfluous for those with a strong math background.

First a review of some definitions. The mean of any system quantity, $\mathbf{X}$ may be expressed in terms of its probability, $\mathbf{P}$, summed over all states

$$< X > = \sum X_i P_i = \sum (X_i e^{-\beta X_i}/Z) \tag{1}$$

and, in general, fluctuations about the mean are given by the **moments**, defined as

$$< \Delta X^n > = < (X - <X>)^n > \tag{2}$$

or

$$< \Delta X^n > = \sum (X_i - <X>)^n \, P_i; \quad n = 1, 2, 3\ldots \tag{3}$$

Note: $< \Delta X^n >$ is not the same as $< \Delta X >^n$

The ($\mathbf{n}$=0) moment is always zero. The first moment ($\mathbf{n}$=1) is the **mean**, the second moment ($\mathbf{n} = 2$)

$$< \Delta X^2 > = \sum \Delta X_i^2 P_i \tag{4}$$

is the **mean-square deviation** from the mean and it gives an approximation to the width of the probability distribution. It is a measure of the fluctuation that we are interested in. Higher moments depend on the shape of the distribution and we won't be concerned with these.

We will develop the form of our derivation using the first moment and the form of equation 1:

$$<E> = \sum E_i e^{-\beta E_i}/Z) \tag{5}$$

Now we use the mathematical equivalence

$$(d/da)(e^{ax}) = Xe^{ax} \text{ (p. 640, Atkins, 7th ed. for justification)},$$

and apply this to the numerator in the right hand sum of equation 5

$$<E> = \sum [(-\partial/\partial\beta) (e^{-\beta E_i})] (1/Z) \tag{6}$$

place the derivative in front of the sum

$$\langle E \rangle = -(\partial/\partial\beta) \sum [e^{-\beta E_i}] (1/Z) \tag{7}$$

and since $\sum[e^{-\beta E_i}] = Z$ (Appendix A, equation 24)

$$\langle E \rangle = -(\partial/\partial\beta)Z (1/Z) \tag{8}$$

or

$$\langle E \rangle = -1/Z (\partial Z/\partial\beta) \tag{9}$$

and using $d\ln x = 1/x(dx)$

$$\langle E \rangle = -\partial\ln Z / \partial\beta \tag{10}$$

A completely analogous derivation may be made for the second moment $\langle \Delta E^2 \rangle$

$$\langle \Delta E^2 \rangle = 1/Z (\partial^2 Z / \partial\beta^2) \tag{11}$$

We now have expressions for $\langle E \rangle$ and $\langle \Delta E^2 \rangle$ and from equation 2

$$\langle \Delta E^2 \rangle = \langle (E - \langle E \rangle)^2 \rangle$$

or

$$\langle \Delta E^2 \rangle = \langle E^2 \rangle - \langle E \rangle^2 \tag{12}$$

Substituting equations 10 and 11

$$\langle \Delta E^2 \rangle = 1/Z (\partial^2 Z/\partial\beta^2) - (-\partial\ln Z/\partial\beta)^2 \tag{13}$$

and the square loses the internal minus sign

$$\langle \Delta E^2 \rangle = 1/Z (\partial^2 Z/\partial\beta^2) - (\partial\ln Z/\partial\beta)^2 \tag{14}$$

and immediately below in the bracketed section we show how this expression may be simplified to

$$\langle \Delta E^2 \rangle = \partial^2\ln Z / \partial\beta^2 \tag{15}$$

[We work backwards from the answer to show that the right hand side of equation 15 is equivalent to that of equation 14,

$$\partial^2\ln Z / \partial\beta^2 = (\partial/\partial\beta) [\partial\ln Z/\partial\beta]$$

$$\partial^2\ln Z / \partial\beta^2 = (\partial/\partial\beta) [(1/Z)\partial Z/\partial\beta]$$

$$\partial^2\ln Z / \partial\beta^2 = (\partial Z^{-1}/\partial\beta) (\partial Z/\partial\beta) + [(1/Z)\partial^2 Z/\partial\beta^2]$$

$$\partial^2 lnZ \ / \ \partial\beta^2 = -\ (1/Z^2) \ (\partial Z/\partial\beta) \ (\partial Z/\partial\beta) + [(1/Z)\partial^2 Z/\partial\beta^2]$$

$$\partial^2 lnZ \ / \ \partial\beta^2 = -\ (1/Z^2) \ (\partial Z/\partial\beta)^2 + [(1/Z)\partial^2 Z/\partial\beta^2]$$

$$\partial^2 lnZ \ / \ \partial\beta^2 = -\ (\partial lnZ/\partial\beta)^2 + 1/Z \ (\partial^2 Z/\partial\beta^2)$$

where the right hand side is now the same as the right hand side of equation 14.]

Equation 15 can be expanded to

$$< \Delta E^2 > = (\partial/\partial\beta) \ (\partial lnZ/\partial\beta) \qquad\qquad\qquad (16)$$

and substituting equation 10

$$< \Delta E^2 > = (\partial/\partial\beta) \ (-<E>)$$

or

$$< \Delta E^2 > = -\partial<E>/\partial\beta \qquad\qquad\qquad\qquad (17)$$

We now have an expression which relates the mean square fluctuations of energy with the derivative of energy with respect to **β**. Recall from classical thermodynamics that the **heat capacity C$_v$** is defined as

$$C_v = \partial<E>/\partial T \qquad\qquad\qquad\qquad\qquad (18)$$

where the subscript **V** indicates constant volume. We want to relate the derivative of **<ΔE^2>** with respect to **T** instead of **β**, as in equation 17, in order to directly relate heat capacity with fluctuations. This is accomplished by multiplying by **(∂T/∂T)** and expanding

$$\partial<E>/\partial\beta = (\partial T/\partial\beta) \ (\partial<E>/\partial T) \qquad\qquad\qquad (19)$$

since **β=1/kT, T=1/kβ**

$$\partial<E>/\partial\beta = [\partial(k\beta)^{-1}/\partial\beta] \ (\partial<E>/\partial T)$$

$$\partial<E>/\partial\beta = [(-1/k) \ (1/\beta^2)] \ (\partial<E>/\partial T)$$

$$\partial<E>/\partial\beta = (-1/k) \ (k^2 T^2/1) \ (\partial<E>/\partial T)$$

$$\partial<E>/\partial\beta = -kT^2 \ (\partial<E>/\partial T)$$

or

$$-\partial<E>/\partial\beta = kT^2 \ (\partial<E>/\partial T) \qquad\qquad\qquad (20)$$

From equation 17 we know that the left side of equation 20 equals **< ΔE^2 >**

$$<\Delta E^2> = kT^2 \, (\partial <E>/\partial T) \tag{21}$$

and the right side is related to the heat capacity (see equation 18)

$$<\Delta E^2> = kT^2 \, C_v \tag{22}$$

We now have an expression to relate the mean standard fluctuation of energy at the system (protein) level to a measurable macroscopic quantity, the heat capacity.

As explained in Appendix A, a more relevant energy term for laboratory measurements is the enthalpy, $\mathbf{H = E + PV}$, for which the expression is

$$<\Delta H^2> = \, kT^2 \, C_p \tag{23}$$

where the subscript $\mathbf{P}$ indicates constant pressure.