CEM 882

Lecture Notes 2

Wellesley

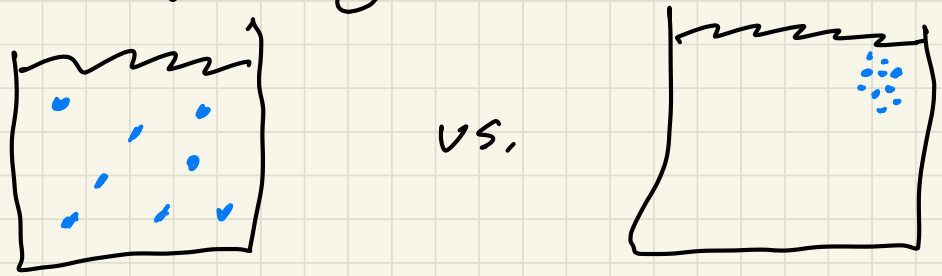
Thermodynamics / Statistical Mechanics are the same field. Thermodynamics was developed in the 19th century before a clear understanding of atoms and molecules and was an effort to understand how energy could be interchanged between heat (temperature changes) and mechanical motion in engines and machines ⇒ coupled with the industrial revolution in Western Europe. Statistical mechanics was developed in the late 19th century and an effort to understand how a large number

of molecules behaved based on understanding of the states of single molecules and statistics. Both approaches yield similar results but the statistical mechanics approach is easier to follow.

The guiding principle is that the properties of large numbers of molecules reflects the most probable distribution of the molecules subject to constraints like fixed temperature and pressure. The most probable distribution is

the one which can be achieved by the largest number of independent ways.

Equal distribution of solute molecules more probable than concentrated in one region of solution



vs.

Start with a simple model system. Protein can have backbone that is either helical (coil) or strand (—). Start with assumption that the two states for a single molecule are equal energy so that

state A

state B

$\mathbb{P}(A) = \mathbb{P}(B) = 0.5$ $\quad \mathbb{P} \equiv$ probability

Consider two molecules ⇒ four possibilities

Molecule 1	Molecule 2	
A	A	} all equal probability
A	B	
B	A	
B	B	

We can only measure the total numbers of A and B molecules

$A_n B_{N-n}$ ($n \equiv$ # of A, $N \equiv$ total # of molecules)

so there are three possible measurements (distributions)

Distribution	$\mathbb{P}$	
A_2B_0	0.25	
A_1B_1	0.5	⇐ more likely to have all states with some $\mathbb{P}$
A_0B_2	0.25	

For three molecules, use Pascal's Triangle

$\mathbb{P}(A_3) = 0.125$

$\mathbb{P}(A_2B_1) = \mathbb{P}(A_1B_2) = 0.375$

$\mathbb{P}(B_3) = 0.125$ $\Sigma\mathbb{P} = 1$

For N molecules,

normalization ⇓

# of independent ways of achieving A_nB_{N-n} ⇑

factorial non-negative integers

$$\mathbb{P}(A_nB_{N-n}) = (0.5)^N \frac{N!}{n!(N-n)!}$$

$N!$ by definition ⇓

$N! = (1)(2)(3)\ldots(N)$

$1! = 1$ $2! = 2$ $3! = 6$ $4! = 24$ $0! = 1$

If $\mathbb{P}(A) \neq \mathbb{P}(B)$ (true if $E_A \neq E_B$),

$\mathbb{P}(A) = p$ $\mathbb{P}(B) = 1-p$ ⇐ $\mathbb{P}(A) + \mathbb{P}(B) = 1$

$$\mathbb{P}(A_nB_{N-n}) = p^n(1-p)^{N-n} \frac{N!}{n!(N-n)!}$$

⇐ binomial distribution

For $\mathbb{P}(A) = 0.9$ $\mathbb{P}(B) = 0.1$ $N = 3$

$\mathbb{P}(A_3) = 0.729$

$\mathbb{P}(A_2B_1) = 0.243$

$\mathbb{P}(A_1B_2) = 0.027$

$\mathbb{P}(B_3) = 0.001$

($\approx \frac{1}{4}$)

⇐ still significant probability of having 1 strand out of 3 molecules

For the $P(A) = P(B) = 0.5$ case, as $N \uparrow$, the probability distribution becomes more peaked at $n_A \approx n_B$

For $N = 10$

$P(A_5B_5) = 0.246$

$P(A_3B_7) = P(A_7B_3) = 0.118$

$P(A_1B_9) = P(A_9B_1) = 0.001$

For larger N, an approximate analytical form for $P(A_nB_{N-n})$

$$\ln\{P(A_nB_{N-n})\} = -N\ln 2 + \ln(N!) - \ln(n!) - \ln\{(N-n)!\}$$

Define $n = N/2 + s = \frac{N+2s}{2}$

s: "excess" number of molecules in state A

$$N - n = \frac{N}{2} - s = \frac{N-2s}{2}$$

$$\ln(P(A_nB_{N-n})\} = -N\ln 2 + \ln(N!) - \ln\left\{\left[\frac{N+2s}{2}\right]!\right\} - \ln\left\{\left[\frac{N-2s}{2}\right]!\right\}$$

Stirling's Approximation for large N

$$\ln(N!) \approx \frac{\ln(2\pi)}{2} + (N+1/2)\ln(N) - N$$

Taylor Series to approximate any function $f(z)$ near $z=a$

$$f(z) \approx f(a) + \left(\frac{\delta f}{\delta z}\right)_{z=a}(z-a) + \frac{1}{2}\left(\frac{\delta^2 f}{\delta z^2}\right)_{z=a}(z-a)^2 + \dots$$

Approximate $f(z) = \ln(z)$ with $a=1$

$$\ln(1+x) \approx 0 + x + \frac{x^2}{2} + \dots$$

x: largest term for small x

$$\ln\{P(A_n B_{N-n})\} \approx \ln\{P(s)\} \approx \ln\left(\frac{2}{\pi N}\right) - \frac{2s^2}{N}$$

$$P(s) = \sqrt{\frac{2}{\pi N}}\, e^{-2s^2/N}$$ ⇐ Gaussian distribution

$$\sigma = \frac{\sqrt{N}}{2}$$ ⇐ standard deviation

$$P(s) = \sqrt{\frac{1}{2\pi\sigma^2}}\, e^{-s^2/2\sigma^2}$$

average value $\langle s \rangle = \int_{-\infty}^{+\infty} s\, P(s)\, ds = 0$

width of Gaussian distribution $\langle s^2 \rangle = \sigma^2$ mean-squared deviation

Gaussian distribution

$N = 100$

$\sigma = 5$

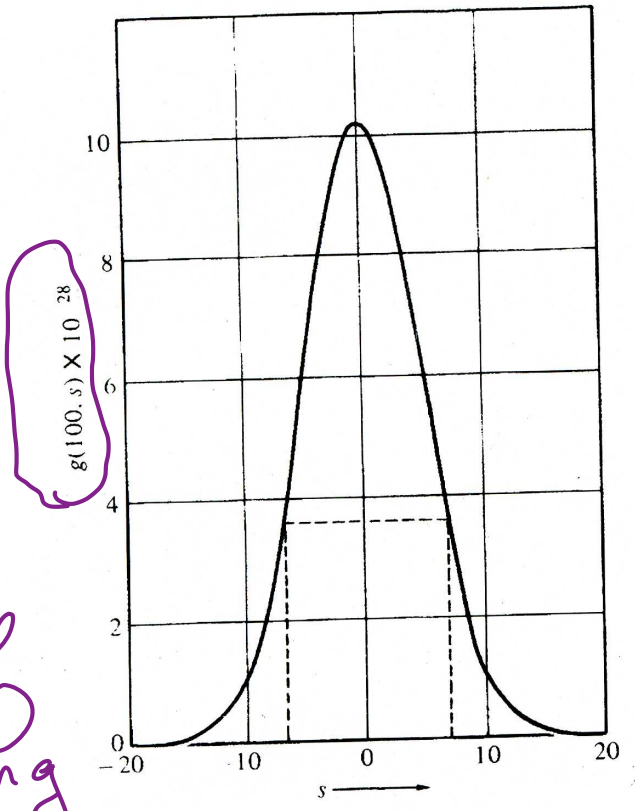


Figure 1.9 The Gaussian approximation to the binomial coefficients $g(100,s)$ plotted on a linear scale. On this scale it is not possible to distinguish on the drawing the approximation from the exact values over the range of s plotted. The entire range of s is from -50 to $+50$. The dashed lines are drawn from the points at $1/e$ of the maximum value of g.

number of ways of achieving distribution

All probable distributions have $|s| \leq 15$

$$P(0) = \frac{0.4}{\sigma}$$ $\Leftarrow$ $n_A = n_B = N/2$ most probable distribution

$$P\left(\frac{N}{2}\right) = \left(\frac{0.4}{\sigma}\right)\left(e^{-N/2}\right)$$ $\Leftarrow$ least probable distribution $n_A = N,\ n_B = 0$

$$\frac{P(N/2)}{P(0)} = e^{-N/2}$$ $\Leftarrow$ For $N = 10^{17}$ (0.15 μmole) $\approx 10^{-10^{16}} \approx 0$

Most probable distributions are for $|s| \leq \sigma$

$$\frac{P(\sigma)}{P(0)} \approx 0.6$$

For $N = 10^{17}$, $\sigma = 1.6 \times 10^{8} \ll 10^{17}$

$$\frac{\sigma}{N/2} = \frac{1}{\sqrt{N}}$$ $\Leftarrow \approx 3 \times 10^{-8}$ for $N = 10^{17}$

For large N, the most probable distributions are a very small subset of the possible distributions

It's also useful to think about the numbers of ways that a few distributions can be achieved ∝ Π

$W \equiv$ multiplicity

$N = 10^{17}$

all helical | 1 strand / all others helical | half-helical half-strand

S

W

The most common function with W is

$$S = k \ln W$$

S = entropy

k = Boltzmann's constant $= 1.381 \times 10^{-23} \frac{J}{K}$

$= \frac{R}{N_A}$ (R = ideal gas constant, N_A = Avogadro's number)

experimentally determined (not derivable)

For the helix/strand 2-state protein system

$$W(A_n B_{N-n}) = \frac{N!}{n!(N-n)!}$$

This equation is useful for deriving a formula for S based on p_A and p_B

$$p_A = \frac{n}{N} \qquad p_B = \frac{N-n}{N}$$

$$W = \frac{N!}{(N \times p_A)!\,(N \times p_B)!}$$

Stirling's approximation for large N

$$\ln(N!) \approx N \ln N - N$$

$$\ln W \approx N \ln N - (N \times p_A) \ln(N \times p_A) - (N \times p_B) \ln(N \times p_B)$$

$$\approx N(-p_A \ln p_A - p_B \ln p_B)$$

$$\ln W = N \sum_{j=1}^{t} -p_j \ln p_j$$

j: states of individual molecules
t: total states

Consider a multiplicity model for how a protein might attain its final folded structure

Each residue in the final folded structure has a single "correct" structure

Proteins synthesized in an organism or purified from a synthesis often don't have the folded structure

One simple model of folding is that the protein samples in time conformations until every residue has the correct conformation.

What is the overall W of a protein?

Worst case scenario

Two conformations per residue

$P_A = P_B = 0.5$

100 residue protein

$W \approx$

Takes $\approx 10^{-12}$ s to sample a protein structure

Maximum time to sample all structures

How might model be adjusted to estimate a more reasonable time for protein folding?