GCMB: Simple Models of Atoms & Proteins

Note Title   3/26/2007

Models for
   Individual atoms: Hard spheres & van der Waals radii
   Collections of atoms:
        Statistical Thermodynamics for Combinatorialists
             -Entropy   -Energy   -Equilibrium
          Lesson: Atomic properties by simple counting models
   Protein folding:
        HP folding model on 2D & 3D lattices
          Lesson: Folding is hard, even for simple models

Simple Models for individual atoms

0. First, consider how difficult it was to explain proportions in
   chemical reactions without the model of atoms & molecules.
   1 mol = 6.022×10^{23} molecules (= 12 grams of carbon C_{12})
   1 dalton = $\frac{1}{12}$ weight of a carbon C_{12}

1. Born-Oppenheimer approximation: electrons are fast
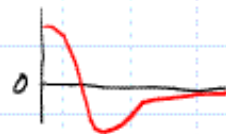   partially decouples motions of nucleus & electrons
   a. electron motion depends only on position of nucleus.
   b. nuclear motion sees a smeared distribution of electrons.

2. Atoms are spheres that weakly attract at a distance
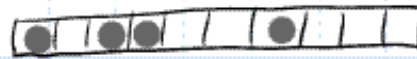   and strongly repel up close.
   van der Waals sphere radii
   Lennard-Jones potential $\frac{A}{r^{12}} - \frac{B}{r^6}$ r = interatom distance
                                                        A, B depend on atom type

The Simplest Model for Collections of Atoms

Lattice Models:

a grid in which each cell may be occupied by an atom.

Fix # cells V and # atoms N.

A "state" will be a set of assignments of atoms to cells.
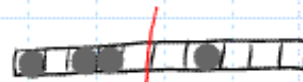
e.g. all atoms left of center

or evenly balanced

Which is more likely?

Lattice gas models show mixing

With 4 atoms and 10 slots,

What proportion of models have all atoms
to the left of center?

$$\frac{\text{nchoosek}(5, 4)}{\text{nchoosek}(10, 4)} \qquad 0.02380952$$

What proportion mix half-and-half?

$$\frac{\text{nchoosek}(5, 2)^2}{\text{nchoosek}(10, 4)} \qquad 0.47619048$$

Gas pressure at constant temperature

Fix N atoms but let the #cells V grow;

The number of states $W = \binom{V}{N} \sim V^N$ increase.

Consider $\frac{d}{dV} \log W \sim N \frac{d}{dV} \log V = \frac{N}{V}$

(Ideal gas law pressure $P = \frac{N}{V} kT$, so at fixed temp, counting gives pressure, up to a constant.)

Die Mathematiker sind eine Art Franzosen:
redet man zu ihnen, so übersetzen sie es in ihre
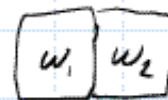Sprache, und dann ist es alsobald ganz etwas anders.
　　　　　　Goethe, Maximen und Reflexionen 1279

Entropy: $\log W$

Entropy $S = k \log W$
　　　　　　$\llcorner$ # of states
　　　$\llcorner$ Boltzmanns constant

Entropy is "extensive"
　When we put two systems together,
　the number of states multiply,
　so the entropies sum.

$\boxed{W_1}\ \boxed{W_2}$

$S_1 + S_2$

2nd law of thermodynamics: systems maximize entropy.

Boltzmann distribution: Max Entropy with Observations

Roll N t-sided dice $(\varepsilon_1, \varepsilon_2, \dots \varepsilon_t)$ and get the average $\bar{\varepsilon}$.
What probabilities $(p_1, p_2 \dots p_t)$ are consistent with this average?

We want to maximize entropy $S = -\sum_i p_i \log p_i$
We know $\sum_i p_i = 1$ and $\sum p_i \varepsilon_i = \bar{\varepsilon}$

Lagrange
multipliers

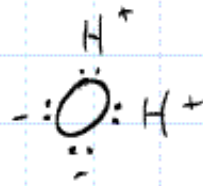$\therefore \sum dp_i = 0$ and $\sum \varepsilon_i dp_i = 0$

Lagrange
multipliers
$$\frac{\partial S}{\partial p_i} - \alpha \left(\frac{\partial \sum p_i}{\partial p_i} \right) - \beta \left(\frac{\partial \sum \varepsilon_i p_i}{\partial p_i} \right) = 0 \quad \text{for } 1 \le i \le t$$

$$-1 - \log p_i^* - \alpha - \beta \varepsilon_i = 0 \qquad \text{so } p_i^* = \exp(-1 - \alpha - \beta \varepsilon_i)$$

$$\boxed{p_i^* = \frac{p_i^*}{\sum_i p_i^*} = \frac{e^{-\beta \varepsilon_i}}{\sum_i e^{-\beta \varepsilon_i}} \qquad \text{Boltzmann distribution}}$$
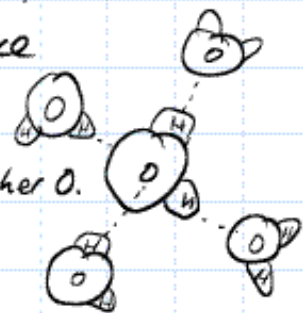
Entropy of ice

Giauque & Stout (1936) did calorimetry (measuring heat
of reactions) for water from freezing (15°K) to boiling.
They calculated the entropy of ice $S_1 = 44.28$ cal/Kmol, which
was smaller than for ideal gas: $S_2 = 45.10$ cal/Kmol
Pauling suggested that ice contributes $S_i = 0.81$ cal/Kmol
due to states of hydrogens in the lattice

H^+

$-:\overset{..}{\underset{..}{O}}: H^+$

Each O coordinates with 4 others.
Each H has 2 positions - close to one or the other O.
Each O needs 2close + 2far $\binom{4}{2} = 6$ choices
out of the possible $2^4 = 16$.
Estimated # states $W \leq 2^{2N} \left(\frac{6}{16}\right)^N$, so $S = k\log W \sim Nk \log(3/2)$
$= 0.81$ cal/Kmol

Energy by counting

Suppose that N particles have quantized energy levels. If the total energy is E, how many ways can the energy be distributed? $\quad \binom{N+E}{N} \sim (N+E)^N$

What is the probability that a particle receives αE energy?

for $\alpha \gg \frac{1}{N}$ roughly

ratio $\quad \dfrac{\binom{N-1 + (1-\alpha)E}{N-1}}{\binom{N+E}{N}} \sim \left(\dfrac{N+(1-\alpha)E}{N+E}\right)^N$

Given systems with N_A, N_B particles and E_A, E_B energy, what happens when you bring them together?

separate $\quad \binom{N_A + E_A}{N_A}\binom{N_B + E_B}{N_B} \qquad$ together $\quad \binom{N_A + N_B + E_A + E_B}{N_A + N_B}$

Energy flow

Given systems with N_A, N_B particles and E_A, E_B energy, what happens when you bring them together?

total states:

separate $\quad \binom{N_A + E_A}{N_A}\binom{N_B + E_B}{N_B}$

together $\quad \binom{N_A + N_B + E_A + E_B}{N_A + N_B} \qquad$ greater!

if you again separate the particles into N_A & N_B

energy divides proportionally: $\quad \dfrac{N_A(E_A + E_B)}{N_A + N_B} \qquad \dfrac{N_B(E_A + E_B)}{N_A + N_B}$

to maximize the number of states.

Equilibrium

to be completed from Dill's book, chapter 6.

Table 8.1 Fundamental equations and their natural variables.

	Function	Extremum at Equilibrium	Fundamental Equation	Definition
energy	$U(S,V,N)$	minimum	$dU = TdS - pdV + \sum_j \mu_i dN_i$	
entropy	$S(U,V,N)$	maximum	$dS = \left(\frac{1}{T}\right)dU + \left(\frac{p}{T}\right)dV - \sum_j \left(\frac{\mu_i}{T}\right)dN_j$	
enthalpy	$H(S,p,N)$	minimum	$dH = TdS + Vdp + \sum_j \mu_j dN_j$	$H = U + pV$
Helmholtz energy $A = arbeit$	$F(T,V,N)$	minimum	$dF = -SdT - pdV + \sum_j \mu_j dN_j$	$F = U - TS$
Gibbs free energy	$G(T,p,N)$	minimum	$dG = -SdT + Vdp + \sum_j \mu_j dN_j$	$G = H - TS$

From Dill & Bromberg

$S =$ entropy $V =$ volume $T =$ temperature $p =$ pressure $N_j = $ # atoms of type j

$\mu_j =$ chem potential ''

An example from D.11

The next example illustrates how to get Maxwell relations when types of work other than pV changes are involved.

EXAMPLE 9.3 The thermodynamics of a rubber band. Is the retraction of a rubber band driven by a change in enthalpy or in entropy? The answer to this question will help us to construct a model for the microscopic behavior of polymeric materials in Chapter 29. Suppose you apply a quasi-static stretching force that increases the length ℓ of a rubber band. The force of retraction f exerted by the rubber band is equal and opposite to the applied stretching force. To deal with elastic forces when there is no particle exchange, we have $U = U(S, V, \ell)$ and

$$dU = TdS - pdV + fd\ell. \qquad (9.12)$$

We are interested in experiments at constant T and p, so we want the Gibbs free energy $dG = d(H - TS) = d(U + pV - TS)$. Substitute Equation (9.12) into this expression to get

$$dG = -SdT + Vdp - fd\ell. \qquad (9.13)$$

It follows from Equation (9.13) and the definition $G = H - TS$ that the force f can be defined in terms of enthalpic and entropic components,

$$f = \left(\frac{\partial G}{\partial \ell}\right)_{T,p} = \left(\frac{\partial H}{\partial \ell}\right)_{T,p} - T\left(\frac{\partial S}{\partial \ell}\right)_{T,p}. \qquad (9.14)$$

To get a Maxwell relationship for $(\partial S/\partial \ell)_{T,p}$, take the cross-derivatives in Equation (9.13):

$$\left(\frac{\partial S}{\partial \ell}\right)_{T,p} = -\left(\frac{\partial f}{\partial T}\right)_{p,\ell}. \qquad (9.15)$$

Equation (9.15) implies that you can get the entropic component of the force, $(\partial S/\partial \ell)_T$, from a very simple experiment. Hold the rubber band at a fixed stretched length ℓ (and constant pressure) and measure how the retractive force depends on the temperature (see Figure 9.1). The slope of that line, $(\partial f/\partial T)_\ell$, will give $-(\partial S/\partial \ell)_T$. The positive slope in Figure 9.1 indicates that the entropy decreases upon stretching.

How do you get the enthalpy component, $(\partial H/\partial \ell)_T$? Substituting Equation (9.15) into (9.14) and combining with Equation (9.15) gives

$$\left(\frac{\partial H}{\partial \ell}\right)_{T,p} = f - T\left(\frac{\partial f}{\partial T}\right)_{p,\ell}. \qquad (9.16)$$

which you can determine from the same experiment. Figure 9.1 shows that the retraction of rubber is stronger at higher temperatures. This observation, first made in 1806 by J. Gough, a British chemist, distinguishes rubber from
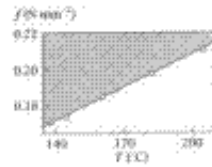

Figure 9.1 The retraction force f of a rubber polymer, amorphous polyethylene, held at constant length as a function of temperature T. The slope is $(\partial f/\partial T)_\ell$. Source: JE Mark, A Eisenberg, WW Graessley, L Mandelkern, and JL Koenig, Physical Properties of Polymers, 2nd edition, American Chemical Society, Washington, DC (1993). The data are from JE Mark, Macromolecules 11, 131 (1978).
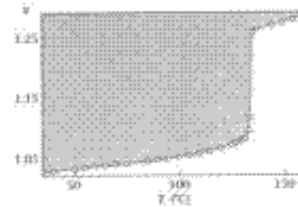
Figure 9.2 The specific volume v (cm³ g⁻¹) of polyethylene versus temperature T. According to Equation (9.17), the thermal expansion coefficient α is proportional to the slope $\partial v/\partial T$. At low temperatures, polyethylene is a hard crystalline glassy material. On melting at around 130 °C, the specific volume v increases sharply and α is large. Source: JE Mark, A Eisenberg, WW Graessley, L Mandelkern, and JL Koenig, Physical Properties of Polymers, 2nd edition, American Chemical Society, Washington, DC (1993). Data are from FA Quinn Jr and L Mandelkern, J Am Chem Soc 80, 3851 (1961).

metal. The entropy of rubber decreases on stretching, while the entropy of metal increases on stretching. Stretching metal increases the bonds, increasing the volume per atom, leading to translational disorder. Stretching rubber decreases the multiplicity of polymer conformations (see Example 2.4, page 34).

Changes in volume with temperature and pressure are readily measurable properties. These properties, in conjunction with the Maxwell relations, give useful insights into molecular behavior, as illustrated below.

Protein Folding

Anfinsen (1961) showed that a protein (ribonuclease)
   could be denatured (unfolded) and would return to function
   $\Rightarrow$ the amino acid sequence determines 3-d structure.

Leventhal "paradox": a protein with 100 amino acids
   will have $> 3^{100}$ states; a random search would take
   the age of the universe, yet proteins do fold...

What in the sequence & environment helps proteins fold quickly?

Let's start with an extremely simple model.

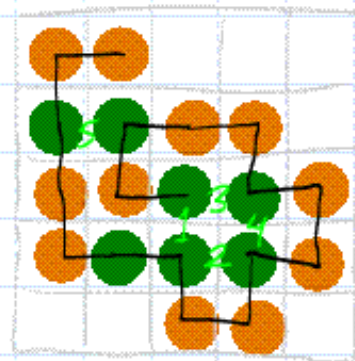Hydrophobicity (fear of water)

Recall that water is polar.

Water molecules will orient themselves around charged
amino acids, pulling these toward the surface
and burying the "hydrophobic" residues.

Hydrophilic   ↑   Arg Lys (end)
(polar or   Asp Glu His Asn Gln
charged)   Pro Tyr   } P olar
           Arg Lys (near backbone)

           Ala Thr Ser Gly   } H ydrophobic
Hydrophobic ↓   Val Ile Leu Met Phe Trp Cys

HP folding model   [Dill 85]

A protein is a string from $\{H, P\}^n$
that is embedded in a lattice
(2d or 3d grid)
The energy of an embedding is
   $-1 *$ # of pairs of Hs at distance 1
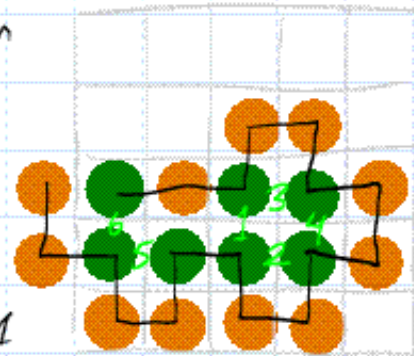   that are not adjacent in the string

energy −5
(can you do better?)

Find min energy embedding.

HP folding model

A protein is a string from $\{H, P\}^n$
that is embedded in a lattice
(2d or 3d grid)
The energy of an embedding is
  -1 · # of pairs of (H)s at distance 1
    that are not adjacent in the string

energy -6

Complexity: NP·complete in 2D & 3D
Algorithms: Exhaustive, Move heuristics, Search
What we can learn: Statistics of ensembles, core packing,
                   design for foldability?

HP folding is NP complete

2-D   Crescenzi, Goldman, Papadimitriou, Piccolboni, Yannakakis 98
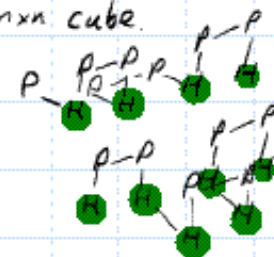   Reduction from Hamilton paths on special graphs w/ Trevisan codes

3-D   Berger, Leighton 98
      Reduction from Bin-Packing

Idea: perfect packing of n^3 Hs is an n×n×n cube.
      So $(PHP)^k$ must pack an edge
      and $(PHHP)^k$ must pack a path
                              on a face

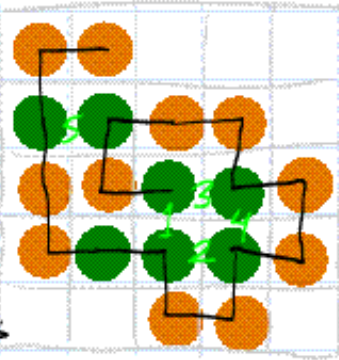All the action happens on the surface of a cube.

HP folding algorithms [Dill 85]

Algorithms:

Exhaustive search for small instances

≤30 is tractible

20 shows core/surface behavior best

   with such small numbers of residues

More clever: [YFTCSD95]

Hydrophobic zipper — start by pairing random near H-Hs

Constrained Hydrophobic Core Construction — assemble core with

       min surface area; also compute lower bounds

Move sets & Monte Carlo   eg [MMW] online

HP folding: what can we learn?

Simple models should   [S97]

1. have fast algorithms to allow
   - performing many runs to gather statistics
   - studying properties of ensembles

2. be coarse-grained to tell us about properties on
   - time scales of 1ms - 1sec
   - length scales of 10Å - 200Å

3. have many states (Levinthal) but unique min (Anfinsen)

4. clarify
   - cooperativity of folding
   - design for foldabilily