

# Chem 728 Lecture Notes – Part 3 – Thermodynamics

The following are lecture notes for Chem 728 (by C. Martin, with minor modifications by L. Thompson). Much of the material is taken directly from the indicated references (old vH, some refs updated to new). This is not intended to replace the original references, but is made available solely for the convenience of students in the class.

**VH** = “Principles of Physical Biochemistry,” Kensal E. van Holde, W. Curtis Johnson, & P. Shing Ho, Prentice Hall, NJ, 1998 (ISBN 0-13-720459-0)

**CS** = “Biophysical Chemistry, Volumes I-III” Charles R. Cantor & Paul R. Schimmel, W. H. Freeman, NY, 1980 (ISBN 0-7167-1188-5, 0-7167-1190-7, 0-7167-1192-3)

<b>Review of Thermodynamics</b> .....	<b>1</b>
Definitions .....	1
<u>Thermodynamics</u> - describes distribution of <u>energy</u> within ( <u>states</u> of) a <u>system</u> .....	1
<u>System</u> - any subset of the universe having spatial boundaries .....	1
<u>State</u> of a system - defined only at equilibrium .....	1
<u>Energetics</u> .....	1
The First Law - “Energy is conserved” .....	2
$\Delta E = q + w$ $dE = dq + dw$ .....	2
For convenience, define enthalpy, $\Delta H = \Delta(E + PV)$ .....	2
Limiting conditions: .....	2
>> Review these ideas by following the reasoning of van Holde Fig 2.1 .....	2
The Molecular View - Statistical thermodynamics .....	2
Equilibrium is the most probable distribution of energies .....	2
This equation defines the <i>most probable</i> (ie. equilibrium) distribution for the system .....	3
Second Law of Thermodynamics .....	3
Define entropy: $S = k \ln W$ (an <i>extensive</i> property) .....	3
expanding our definition of entropy, .....	4
Define heat capacity, $C_p$ .....	5
Re-examine “equilibrium” .....	5
A more relevant term for biological systems - $\Delta G$ (constant T & P) .....	5
Back to Boltzmann .....	6
Brief Aside: .....	8
Introduction to MathCAD .....	8
Error Analysis .....	8
L'Hopital's Rule .....	8
Dimensional Analysis .....	8
Solution thermodynamics .....	9
Ideal Solutions .....	10
For a reaction: $aA + bB \leftrightarrow cC + dD$ .....	10
van't Hoff equations .....	11
<b>The Folding of Proteins and Nucleic Acids</b> .....	<b>13</b>
Quick review - the amino acids .....	13
The Forces that Govern Protein Folding .....	13
Dipole - Point Charge Interactions .....	14
van der Waals Interactions .....	14
Bond stretching .....	14
Torsional constraints - Rotations around rotatable bonds .....	14
The dipolar interaction significantly effects Ramachandran values for phi and psi .....	15
Effects of <i>unusual</i> amino acids on phi and psi .....	15
Covalent Interactions .....	15
“Non-bonding” Interactions .....	16
Hydrogen bonding .....	16
Water - the unusual solvent .....	16
Water structure - hydrophobic interactions .....	16

Denaturation by urea and guanidinium hydrochloride.....	16
Ionic interactions.....	16
Disulfide bonds.....	16
Dominant Forces in Protein Folding? - Ken Dill Review ( <i>Biochemistry</i> <b>29</b> , 7133, 1991) .....	17
<i>Cooperativity in formation of protein structures</i> .....	17
<i>Intro to hydrophobic interactions</i> .....	18
Summary: Definition of Hydrophobicity.....	20
<b>Analyses of Protein Folding.....</b>	<b>21</b>
Two-State Model for Protein Folding .....	21
Multi-State Model for Protein Folding .....	21
Calorimetry .....	22
Differential Scanning Calorimetry.....	24
<b>DNA Melting / Strand Association .....</b>	<b>25</b>
Thermodynamics - Two non- <i>self</i> -complementary, complementary strands .....	25
Thermodynamics - Two <i>self</i> -complementary strands.....	26
Generalities.....	26
Breslauer <i>et al.</i> .....	27

# Chem 728 Lecture Notes

The following are lecture notes for Chem 728 (by C. Martin Fall 1995, with minor modifications by L. Thompson). Much of the material is taken directly from the indicated references (old vH, some refs updated to new). This is not intended to replace the original references, but is made available solely for the convenience of students in the class.

## Review of Thermodynamics

See van Holde, Chapter 2

As chemists we're interested in molecules - the microscopic view... Thermodynamics takes a macroscopic view of the behavior of large numbers of molecules (and thus it can take advantage of statistical behavior) to deduce a great deal about a reaction without needing to know its microscopic details. This understanding can then provide important insight into the molecular level. We'll see how... >>example.

## Definitions

**Thermodynamics - describes distribution of energy within (states of) a system.**

**System - any subset of the universe having spatial boundaries**

Transfer of matter between system and universe:

open - transfer of matter allowed

closed - transfer of matter not allowed

Transfer of heat between system and universe

non-adiabatic - heat allowed (eg. eppendorf tube floating in a large water bath)

adiabatic - insulated, no transfer of heat (eg. (approx) eppendorf tube placed in styrofoam insulation)

**State of a system - defined only at equilibrium**

Specified by two of (T,P,V) plus the masses and identities of all chemical species

Properties of states of the system

Extensive - requires for their definition the full specification of the system including masses and identities of components

examples - volume, energy, entropy - double the number of molecules, double the volume

Intensive - only relative amounts of components need be known

examples - density, viscosity - double the number of molecules, no change in density

Changes between states

Reversible - path from initial to final state proceeds through a succession of "near-equilibrium" states

Irreversible - NOT above

State variables - path independent

This is quite important for calculations, particularly for complex processes: a simpler path connecting the same 2 states can be used to calculate path-independent quantities (typically all but q and w).

## Energetics

**q - Heat** □ energy supplied to the system as **heat**

energy transferred into or out of a system as a consequence of a temperature difference between the system and surroundings

heat absorbed by the system is *positive*

**w - Work** □ energy supplied to the system as **work**

any exchange of "**non-heat**" energy between the system and its surroundings. Examples:

volume change under constant pressure ( $P\Delta V$ )

changes in surface area under constant surface tension (2D analog of  $P\Delta V$ )

electrical work, etc.

work done *on* the system is *positive* - Beware: this convention is **NOT** universal! ie van Holde opposite

**E or U - Internal Energy** (extensive state function)

The energy *within* the system. For our purposes, only energy which can be modified by chemical processes (eg. we ignore nuclear energy).

examples - translational, vibrational, and rotational energy of the molecules, chemical bond energy, nonbonding interactions (eg. dipolar, ionic)

Typically defined *relative to* a standard state. Only *change* in E is usually considered.

**H - Enthalpy** (state function)

$H = E + PV$  Enthalpy is the sum of the internal energy plus the product of the system volume times the external pressure

### The First Law - "Energy is conserved"

$$\Delta E = q + w \quad dE = dq + dw$$

"The change in the internal energy of a system is the total heat *absorbed* by the system plus the total work *done on* the system by its surroundings"

Note that  $E$  is a state variable, but that  $q$  and  $w$  are not (some texts note this with different symbol for  $dq$ ,  $dw$ ).

If only  $P\Delta V$  work is possible then  $dw = -PdV$

$$dE = dq + dw = dq - PdV$$

Note: one must apply work **to** the system ( $dw > 0$ ) in order to compress the system ( $dV < 0$ )

### For convenience, define enthalpy, $\Delta H = \Delta(E + PV)$

$$dH = d(E + PV) = dE + PdV + VdP = dq + dw + PdV + VdP$$

If only  $P\Delta V$  work is possible then  $dw = -PdV$

$$dH = dq + VdP$$

### Limiting conditions:

At constant volume:  $dE = dq$  ( $\Delta E = q_v$ )

At constant pressure:  $dH = dq$  ( $\Delta H = q_p$ )

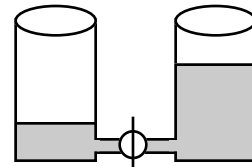
Since most biological processes occur under constant pressure (1 atm),  $\Delta H$  is the term of choice to measure the energy of the system. However, since volume changes in most biological processes are small,  $\Delta H \approx \Delta E$ , so the distinction is minor.

>> Review these ideas by following the reasoning of van Holde Fig 2.1

### The Molecular View - Statistical thermodynamics

From our point of view, a system is really just a collection of molecules. It is the individual actions and energies of the molecules that determine the properties of the system. Why then, do we often ignore the individual players? The answer is that the statistical behavior of very large numbers of well-behaved particles near equilibrium can be very accurately predicted, even though the exact behavior of any single particle cannot be predicted.

Intuitive example: consider two beakers containing a liquid and connected by a tube with a closed stopcock in between. You can readily predict what will happen when you open the stopcock. Fluid will flow (individual molecules will move from one container to the other) until the liquid levels are equal. It doesn't matter whether the fluid is water, honey, or benzene, the final state will be the same. Similarly, it doesn't matter what diameter the connecting tube is, or how long the connector is. Although the kinetics of approach to equilibrium may depend on these parameters, the final state will not. In contrast, you cannot predict *which* individual molecules will end up in which container (in fact the disposition of an individual molecule will vary with time, even at equilibrium).

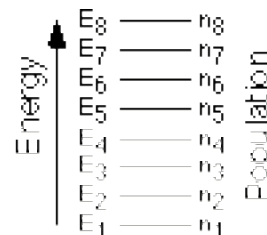


Thermodynamics is really concerned with how energy is *distributed* over a very large number of particles. In a gas or a liquid, different particles can have different energies. These energies are statistically distributed among the particles and the sum of their individual energies is the total energy of the system.

### Equilibrium is the most probable distribution of energies.

Using statistical mechanics it is possible to derive an equation known as the Boltzmann equation, which relates the relative (statistical) populations of non-degenerate states ("states" available to the particles) to their energy levels. For two such states,  $i$  and  $j$ , we have:

$$\frac{n_i}{n_j} = e^{-\frac{(E_i - E_j)}{k_B T}}$$



**This equation defines the *most probable* (ie. equilibrium) distribution for the system.**

The above assumed non-degenerate particle states. In other words, no two particle states may have the same energy. Since we are dealing with simple statistical distributions, if a given particle energy exists for two different particle states, then that energy will be more often represented in the statistical distribution than an energy represented by one state.

To include the general concept of degeneracy, the more general form of the Boltzmann equation is written:

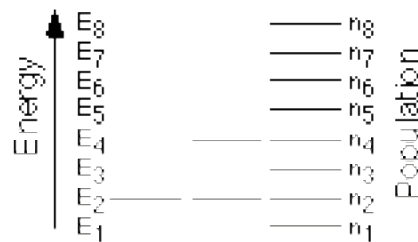
$$\frac{n_i}{n_j} = \frac{g_i}{g_j} e^{-\frac{(E_i - E_j)}{k_B T}}$$

where each energy level,  $E_i$ , contains  $g_i$  degenerate states, and  $n_i$  is the number of particles with energy  $E_i$ .

This explains the ratio of the populations of two energy levels.

To determine the population of one state relative to all possible states, we have

$$\frac{n_i}{N} = \frac{g_i e^{-\frac{(E_i - E_j)}{k_B T}}}{\sum_j g_j e^{-\frac{(E_j - E_j)}{k_B T}}} \quad (\text{N.B.: } j \text{ is simply an arbitrary reference state for the calculation})$$



The above equation is known in statistical mechanics as the molecular partition function. Remember well the basis that thermodynamics has in statistics.

## Second Law of Thermodynamics

How do we know what *direction* a reaction will take? A reaction will go towards the *most probable* state (overall distribution of energies) for the system. From statistics again, we can look at the two distributions below for three degenerate energy levels and predict which is more likely.



There is only one way to distribute the particles to get state A. There are a lot more ways of distributing particles to get state B ( $11!/(4! 4! 3!)=11,550$ ). If A and B have equal energies, B will occur much more often. Another way of saying this is that *there is more randomness in distribution B* than in distribution A. Given this, we need to have a way to express “randomness.”

From probability theory, we know that the number of ways of arranging  $N$  distinguishable particles with  $n_1$  in one group,  $n_2$  in another group, etc. is

$$W = \frac{N!}{n_1! n_2! n_3! \dots n_n!} \quad \text{where } N = n_1 + n_2 + n_3 + \dots + n_n$$

Define entropy:  $S = k \ln W$  (an *extensive property*)

where  $k$  is the Boltzmann constant,  $k_B$  (from here forward, we will drop the subscript “B” except when it is necessary for clarity), and  $W$  is the number of ways in which the individual particle states can be distributed within a particular energy level.

Note that  $S$  is *extensive*, such that for a system composed of two parts, 1 and 2, the total entropy is given by:

$$S = S_1 + S_2 = k \ln W_1 + k \ln W_2 = k \ln(W_1 W_2)$$

in other words,  $W = W_1 W_2$ , which is expected from statistics: if the number of ways of distributing particles in state 1 is  $W_1$  and the number of ways for part 2 is  $W_2$ , then the number of ways of distributing particles in the combined state is simply  $W_1 W_2$ .

Similarly, we can ask about the change in entropy of a system initially characterized by a distribution  $W_0$  going to a distribution  $W_f$

$$\Delta S = S_f - S_o = k \ln W_f - k \ln W_o = k \ln \frac{W_f}{W_o}$$

expanding our definition of entropy,

$$\text{from } W = \frac{N!}{n_1! n_2! n_3! \dots n_n!} \quad \text{we expand} \quad \ln W = \ln(N!) - \ln(n_1!) - \ln(n_2!) - \ln(n_3!) \dots - \ln(n_n!)$$

but note that for large n (Stirling's approximation):  $\ln(n!) \approx n \ln(n) - n$

$$\ln W = (N \ln(N) - N) - (n_1 \ln n_1 - n_1) - (n_2 \ln n_2 - n_2) \dots - (n_n \ln n_n - n_n)$$

$$\ln W = N \ln N - n_1 \ln n_1 - n_2 \ln n_2 \dots - n_n \ln n_n - N + n_1 + n_2 \dots + n_n$$

$$\ln W = N \ln N - n_1 \ln n_1 - n_2 \ln n_2 \dots - n_n \ln n_n$$

(remember that  $N = n_1 + n_2 + n_3 + \dots + n_n$ ). Finally we can summarize:

$$\ln W = N \ln N - \sum_i n_i \ln n_i$$

Now let's look at a change of the system in which we redistribute some particles to different energy levels, and we'll do it in infinitesimally small steps. From before,  $\Delta S = k \Delta \ln W$ , so that:

$$dS = kd(\ln W) = kd \left( N \ln N - \sum_i n_i \ln n_i \right)$$

$$= kd(N \ln N) - k \sum_i d(n_i \ln n_i)$$

$$= 0 - k \sum_i \left[ d(n_i) \ln n_i + n_i d(\ln n_i) \right] = -k \sum_i \left[ dn_i \ln n_i + n_i \frac{dn_i}{n_i} \right] = -k \sum_i dn_i \ln n_i - k \sum_i dn_i$$

The summation in the last term represents the sum of all population changes in all levels. Since the N particles are restricted to the levels 1 through n (i.e., particles can only redistribute among the levels), the sum of all changes is simply 0.

If this causes you trouble, imagine that if you move 5 particles from level 3 to level 6. For this transfer  $dn_3 = -5$  and  $dn_6 = +5$ . As for this example, the sum of all transfers is simply 0.

$$\text{Therefore} \quad dS = -k \sum_i \ln n_i dn_i$$

$$\text{remembering the Boltzmann distribution, } \frac{n_i}{n_j} = e^{-\frac{(\epsilon_i - \epsilon_j)}{k_B T}}$$

$$\ln n_i = \ln n_j - \frac{(\epsilon_i - \epsilon_j)}{k_B T} = \ln n_j - \frac{\epsilon_i}{kT}$$

Placing this into the equation for dS above

$$dS = -k \sum_i \left[ \ln n_j - \frac{\epsilon_i}{kT} \right] dn_i = -k \ln n_j \sum_i dn_i + k \sum_i \frac{\epsilon_i}{kT} dn_i$$

$$dS = 0 + k \sum_i \frac{\epsilon_i}{kT} dn_i = \frac{1}{T} \sum_i \epsilon_i dn_i$$

Let's think again about the summation of  $\Delta \epsilon$ : we are talking about redistributing particles between various levels. The summation totals the changes in energies, weighted by the number of particles making each change. If *more particles move to higher energies* than move to lower energies, the system must *absorb heat* (and vice versa). Hence, the summation is simply the amount of heat absorbed by the system,  $Q_{\text{rev}}$ .

$$dS = \frac{1}{T} \sum_i \mu_i dn_i = \frac{dq_{rev}}{T} \quad \text{or} \quad \Delta S = \int \frac{dq_{rev}}{T} = \frac{q_{rev}}{T}$$

In general then, we can directly calculate S for a process *if* we can define  $q_{rev}$  along a *reversible* path from state A to state B. You will remember that since S is a state function, we can exploit this result even if the actual path taken in an experiment is not reversible. A simple example using the expansion of an ideal gas is given in van Holde (pp. 80-81).

This is then expanded to include mixing two solutions. Our intuition says that the entropy of the final mixture will be greater than the sum of the entropies of the two pure components (each particle has more choices available to it).

Again intuition tells us that the melting of a crystalline solid or the unfolding of a protein (but see important discussion below) will lead to an increase in entropy.

$$\Delta S = \frac{q_{rev}}{T} = \frac{\Delta H_{melting}}{T_{melting}}$$

In general, heating of a sample will lead to an increase in entropy, and this can be seen from the Boltzmann analysis above. Heating a sample must lead to increased population of higher particle energy levels. This results in a larger distribution of particle energies, and therefore to increased entropy.

### Define heat capacity, $C_p$

The change in the internal heat of a system as a function of temperature (and at constant pressure) can be expressed as:

$$\left[ \frac{\partial q_{rev}}{\partial T} \right]_p = C_p$$

Remembering that  $dS = \frac{dq_{rev}}{T}$

$$dS = \frac{C_p dT}{T} \quad \Delta S = \int_{T_1}^{T_2} \frac{C_p dT}{T}$$

Assuming (and this is an assumption) that the heat capacity is not temperature dependent over the temperature range from  $T_1$  to  $T_2$ , then

$$\Delta S = C_p \int_{T_1}^{T_2} \frac{dT}{T} = C_p \ln \left[ \frac{T_2}{T_1} \right]$$

### Re-examine "equilibrium"

From above we have  $\partial q_{rev} = T\partial S$ , and returning to our definition of E we have

$$dE = dq_{rev} + dw \\ = TdS - PdV \quad \text{at constant pressure.}$$

For a reversible process at equilibrium, the internal energy E and the system volume V are constant, so that dS must be zero. If we reversibly perturb the system in an infinitesimally small manner, dS around equilibrium is zero. This means that S must be in a local maximum or local minimum.

From our previous statements about most probable states, we conclude that an isolated system (E and V constant), will be at equilibrium only when the entropy is maximum (minimum entropy would correspond to few ways of arranging the particles).

Finally state the Second Law: An isolated system will approach a state of maximum randomness - directionality.

### A more relevant term for biological systems - $\Delta G$ (constant T & P)

We rarely encounter isolated systems at constant volume. Instead most of life's processes occur under constant temperature (eg. physiological 37°C or room

temperature  $\approx 25^\circ\text{C}$ ) and constant pressure (1 atm). We introduce the Gibbs free energy, as a function of T and P

$$G = H - TS$$

$$H = E + PV$$

$$dG = dH - d(TS) = dE + PdV + VdP - SdT - TdS$$

for a reversible process  $dE = TdS - PdV$ , so that

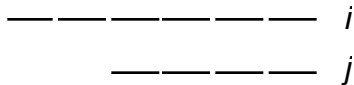
$$dG_{rev} = VdP - SdT$$

If we consider **equilibrium at constant temperature and pressure**, then  $dG = 0$

Just as entropy reaches a maximum at equilibrium, G reaches a minimum.

Also note the more familiar expression at constant temperature  $dG = -H - TdS$

## Back to Boltzmann



Consider two energy levels,  $i$  and  $j$ ,

$$\text{From before we had } \frac{n_i}{n_j} = \frac{g_i}{g_j} e^{-\frac{\epsilon_i - \epsilon_j}{k_B T}}$$

But  $g_i$  is the degeneracy, or number of ways of attaining energy  $\epsilon_i$ , which is  $W_i$

$$\frac{n_i}{n_j} = \frac{W_i}{W_j} e^{-\frac{\epsilon_i - \epsilon_j}{k_B T}} = e^{\ln \frac{W_i}{W_j} - \frac{\epsilon_i - \epsilon_j}{k_B T}} = e^{\ln W_i - \ln W_j - \frac{\epsilon_i - \epsilon_j}{k_B T}} = e^{\frac{k_B T \ln W_i - k_B T \ln W_j - \epsilon_i + \epsilon_j}{k_B T}}$$

If we move from particle energy ( $\Delta\epsilon$ ) to a molar energy ( $\Delta E$ ), also replacing  $k$  by  $R$ , and then approximate  $\Delta E \approx \Delta H$ , we have

$$\frac{n_i}{n_j} = e^{\frac{TS_i - TS_j - H_i + H_j}{RT}} = e^{\frac{H_j - H_i - T(S_j - S_i)}{RT}} = e^{-\frac{\Delta G}{RT}}$$

You will of course notice that the left side is just the equilibrium constant for a reaction going from state  $j$  to state  $i$ . This then leads us to a familiar expression

$$\frac{n_i}{n_j} = K = e^{-\frac{\Delta G}{RT}} \quad \text{or} \quad \ln K = -\frac{\Delta G}{RT}$$

**Aside:** The text (van Holde) first talks about the unfolding of a protein as necessarily having an associated increase in entropy, as a result of the increased motional freedom of the polypeptide and the side chains. This is what most people would come to, looking at the protein as an isolated system. At the end of the chapter, van Holde discusses the real situation - protein unfolding may actually have an *unfavorable* entropic component.

**Why?** The answer is that it is unrealistic to treat an isolated protein as a system unto itself. In fact, in solution a protein is interacting substantially with the solvent (water) and components of the solvent. Indeed, it is these interactions which lead to stable folding of proteins. When we draw the line for our system, we must include these interactions.

A charged or polar group on the surface of a protein will "want to" interact with a charge or polar group in solution. Even if a direct interaction with a salt ion is formed, water will ultimately be involved in forming interactions with the polar group(s). Hence, the exposure of a charged or polar group on the surface of the protein serves to orient (restrict the motion of) water molecules - reducing their entropy. Of course, water molecules will also interact with themselves, but the interactions are transient and more varied (large entropy). Even more so, exposure of a nonpolar group to solvent forces a preferred orientation on associated solvent molecules (water) reducing the entropy of the system.



When a protein folds up, many of these interactions are satisfied internally (a polar side chain might interact with a backbone carbonyl). As before the entropy associated with each group decreases, but it was already low since the two portions of the molecule are ultimately covalently linked. Since this decrease in entropy is less than the entropy loss associated with solvent caging, the total entropy (disorder) of the system can actually be higher in the system containing a folded protein. We will discuss this concept at greater length in future lectures.

**Brief Aside:****Introduction to MathCAD**

When you first run MathCAD you will be placed onto an empty document.

You can simply click anywhere with the mouse to type an equation.

To define a variable, use :=

eg. a:=3.0 YVal:=a/5 (this will then convert to a more familiar form)

Note that expressions on a page are evaluated left to right and top to bottom, so that the second expression above must be located to the right of or below the left expression, otherwise "a" will be undefined.

To check the value of a variable, use =

eg. after the above, type YVal= the screen will respond with YVal = 0.6

Any time that the "insertion point" is in a variable, typing will do as you expect.

When you want to "operate on" a variable, you often have to "back up" and select the whole variable first. To do this type either the "space bar" or the "up arrow key" while the insert point is anywhere in the name of that variable (you may have to click to get it there). You will see the single selection mark change to a rectangle surrounding the variable.

If you type "space bar" or "up arrow key" more than once, more variables will be enclosed in your rectangle for selection.

To delete a selection, you must type "Command-X" (or choose "Cut" from the edit menu).

For some strange reason, the "delete key" will not delete.

**Error Analysis**

For an observable  $f(x,y,z)$  which is a function of variables  $x$ ,  $y$ , and  $z$ , each with inherent error  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$ , respectively

$$\Delta f(x,y,z) = \sqrt{\left(\frac{\partial f(x,y,z)}{\partial x} \Delta x\right)^2 + \left(\frac{\partial f(x,y,z)}{\partial y} \Delta y\right)^2 + \left(\frac{\partial f(x,y,z)}{\partial z} \Delta z\right)^2}$$

**L'Hopital's Rule**

What to do in the limit as a function goes to either  $0/0$  or  $\pm\infty/\pm\infty$ .

Put simply:  $\lim_{x \rightarrow c} \frac{f(x)}{g(x)} \left[ \frac{0}{0} \text{ or } \frac{\infty}{\infty} \right] = \lim_{x \rightarrow c} \frac{\frac{df(x)}{dx}}{\frac{dg(x)}{dx}}$

**Dimensional Analysis**

When doing derivations, it is very easy to drop a term or part of a term. An easy way to find such errors is to remember that most, if not all, parameters in a function have **units**. Consequently, it doesn't make sense to add an expression in units of concentration to an expression in units of concentration per time. So at the end of a derivation, make sure that in any sum or difference expression, all additive terms have the same units. If they don't, then you can be certain that you made an error. To find the error, repeat the dimensional analysis on each preceding expression in your derivation.

An obvious corollary of this is to always make sure that the final result has the correct units. If you're solving for velocity, and the expression you end up with doesn't have time in the denominator, then you are in trouble.

### Solution thermodynamics

Definition: Solution - a single phase system containing more than one *independently variable* substance (component). At equilibrium, the solution can be described completely by specification of the components and knowledge of their interactions. Extensive and intensive variables depend on the composition.

Definition: Partial molar/specific quantities - for any extensive property X, the corresponding partial molar quantity is

$$\bar{X}_i = \left[ \frac{\partial X}{\partial n_i} \right]_{T,P,n_{j \neq i}}$$

In other words, it is the differential change in the extensive property X associated with the differential change in the molar amount of species *i*, holding all else constant. It may very well be a function of the other parameters (T,P, the amounts and identities of other components).

Partial *specific* quantities are the same, but are with respect to the differential change in *mass* of the component.

As an example, if you add volume *x* of a solute to volume *y* of a solution, the resulting volume is not necessarily *x+y*. The volume change is governed by the partial specific volume of the solute in that solution.

Note that the observable, the total extensive quantity, is sum of the products of all partial specific quantities multiplied by the number of moles of each:

$$X = \sum_{i=1}^n n_i \bar{X}_i \quad \text{or} \quad dX = \sum_{i=1}^n \bar{X}_i dn_i$$

One extensive property that deserves expansion is the free energy,  $\bar{G}_i$ , the partial molar free energy (also known as the chemical potential,  $\mu_i$ ).

$$\bar{G}_i = \left[ \frac{\partial G}{\partial n_i} \right]_{T,P,n_{j \neq i}} \quad G = \sum_{i=1}^n n_i \bar{G}_i$$

Just as we can look at the dependence of an extensive variable on the molar amount of a species, we can also look at its dependence on other solution parameters. In general:

$$dG = \left[ \frac{\partial G}{\partial T} \right]_{P,n_i} dT + \left[ \frac{\partial G}{\partial P} \right]_{T,n_i} dP + \sum_{i=1}^n \left[ \frac{\partial G}{\partial n_i} \right]_{T,P,n_{j \neq i}} dn_i$$

from before we had:  $dG = SdT + VdP$   
which is now more generally:

$$dG = SdT + VdP + \sum_{i=1}^n \left[ \frac{\partial G}{\partial n_i} \right]_{T,P,n_{j \neq i}} dn_i = SdT + VdP + \sum_{i=1}^n \bar{G}_i dn_i$$

Since we often restrict ourselves to constant T and P, then  $dG = \sum_{i=1}^n \bar{G}_i dn_i$

An important result of this can be seen, backing up to the more general  $G = \sum_{i=1}^n n_i \bar{G}_i$

we have  $dG = \sum_{i=1}^n n_i d\bar{G}_i + \sum_{i=1}^n \bar{G}_i dn_i$

but since at constant temperature and pressure  $dG = \sum_{i=1}^n \bar{G}_i dn_i$  then  $\sum_{i=1}^n n_i d\bar{G}_i = 0$

The simple result of this is that for an n-component system,  $\mu_i$  for only n-1 components are independent (Gibbs-Duhem).

Also, the differential free energy of a solute in a system of two phases (1 and 2) at equilibrium and at constant T and P is

$$dG = \sum_{i=1}^2 \mu_i dn_i = \mu_1 dn_1 + \mu_2 dn_2 = 0$$

If we transfer a small amount of the solute from phase 1 to phase 2,  $dn_1 = -dn_2$ , so that  $\mu_1 dn_1 + \mu_2 dn_2 = 0 \implies \mu_1 = \mu_2$

In other words, for a solute at equilibrium between two phases, the chemical potential for that species must be the same in both phases.

Since  $\mu_i$  is related to concentration,  $c_i$ , this will lead to rules for equilibria.

### Ideal Solutions

An ideal solution is one in which all components obey Raoult's law, relating vapor pressure to mole fraction,  $X_i$ .

Alternatively viewed, an ideal solution is one for which the enthalpy of mixing is 0 and the only source of energy change is due to changes in the *entropy* of mixing. It implies a lack of interactions between solutes, in exactly the same way that the ideal gas law ignores molecule-molecule interactions.

From above, we had  $S = k \ln W = k \ln \frac{N_0!}{N_1! N_2! \dots N_n!}$

Using Stirling's approximation as before,

$$\Delta S_{\text{mixing}} = -k \sum_{i=1}^n N_i \ln \frac{N_i}{N_0} = -k \sum_{i=1}^n N_i \ln X_i \quad \text{where } X_i = \frac{N_i}{N_0}$$

$$\text{Therefore } \Delta S_{\text{mixing}} = -R \sum_{i=1}^n n_i \ln X_i$$

$$\Delta G_{\text{mixing}} = \Delta H_{\text{mixing}} - T \Delta S_{\text{mixing}} = 0 - T \Delta S_{\text{mixing}} = RT \sum_{i=1}^n n_i \ln X_i$$

$$\text{but } G = \sum_{i=1}^n n_i \mu_i$$

$$\text{If we define } \Delta G_{\text{mixing}} = G_{\text{so}} - \sum_{i=1}^n G_i^{\text{pure}} = \sum_{i=1}^n n_i \mu_i - \sum_{i=1}^n n_i \mu_i^0 = \sum_{i=1}^n n_i (\mu_i - \mu_i^0)$$

Combining  $\Delta G_{\text{mixing}}$  from the two approaches

$$\Delta G_{\text{mixing}} = \sum_{i=1}^n n_i (\mu_i - \mu_i^0) = RT \sum_{i=1}^n n_i \ln X_i$$

therefore

$$\mu_i - \mu_i^0 = RT \ln X_i$$

van Holde then discusses that for dilute solutions,  $X_i \approx C_i$ , so that for most purposes:

$$\mu_i = \mu_i^0 + RT \ln C_i$$

**For a reaction:  $aA + bB \leftrightarrow cC + dD$**

$$\Delta G = G_{\text{final}} - G_{\text{initial}} \quad \text{but } G = \sum_{i=1}^n n_i \mu_i$$

$$\Delta G = c\mu_C + d\mu_D - a\mu_A - b\mu_B$$

$$\Delta G = (c\mu_C^0 + cRT \ln C_C) + (d\mu_D^0 + dRT \ln C_D) - (a\mu_A^0 + aRT \ln C_A) - (b\mu_B^0 + bRT \ln C_B)$$

$$\Delta G = (c\mu_C^0 + d\mu_D^0 - a\mu_A^0 - b\mu_B^0) + (cRT \ln C_C + dRT \ln C_D - aRT \ln C_A - bRT \ln C_B)$$

$$\Delta G = \Delta G^0 + RT \ln \frac{C_C^c C_D^d}{C_A^a C_B^b} \quad \text{a classic!}$$

At equilibrium  $\Delta G = 0$ , so

another classic!

$$\text{Also } \Delta G^\circ = \Delta H^\circ - T\Delta S^\circ = -RT \ln K_{eq} \text{ so } \ln K_{eq} = \frac{\Delta H^\circ}{RT} + \frac{\Delta S^\circ}{R}$$

If  $\Delta H^\circ$  and  $\Delta S^\circ$  are independent of temperature (often NOT for protein interactions, but often is OK over a narrow temperature range), then the temperature dependence of  $K_{eq}$  can yield  $\Delta H^\circ$  and  $\Delta S^\circ$ .

### van't Hoff equations

In calorimetry, one can directly measure heats associated with certain types of reactions.

Let's continue to look at the temperature dependence of some thermodynamic parameters.

$$\text{Starting with the expression: } \ln K_{eq} = \frac{\Delta H^\circ}{RT} + \frac{\Delta S^\circ}{R}$$

$$\text{We can derive the differential expression: } \left[ \frac{\partial \ln K_{eq}}{\partial (1/T)} \right]_p = \frac{\Delta H^\circ}{R}$$

$$\text{or alternatively, } \left[ \frac{\partial \ln K_{eq}}{\partial T} \right]_p = \frac{\Delta H^\circ}{RT^2}$$

These are various forms of the van't Hoff relation.

We can also predict the temperature dependence of the equilibrium constant over a finite change in temperature:

$$\ln K_{T_2} = \frac{\Delta G_{T_2}^\circ}{RT_2} \text{ leads to } \ln K_{T_2} - \ln K_{T_1} = \frac{\Delta G_{T_2}^\circ}{RT_2} - \frac{\Delta G_{T_1}^\circ}{RT_1}$$

$$\text{or } \ln K_{T_2} = \ln K_{T_1} + \frac{\Delta G_{T_2}^\circ}{RT_2} - \frac{\Delta G_{T_1}^\circ}{RT_1}$$

We can also express this in terms of  $\Delta H$  and  $\Delta S$

$$\ln K_{T_2} = \ln K_{T_1} + \frac{\Delta H_{T_2}^\circ - T_2 \Delta S_{T_2}^\circ}{RT_2} - \frac{\Delta H_{T_1}^\circ - T_1 \Delta S_{T_1}^\circ}{RT_1}$$

$$\ln K_{T_2} = \ln K_{T_1} + \frac{(\Delta H_{T_2}^\circ + T_2 \Delta S_{T_2}^\circ)}{RT_2} - \frac{(\Delta H_{T_1}^\circ + T_1 \Delta S_{T_1}^\circ)}{RT_1}$$

$$\ln K_{T_2} = \ln K_{T_1} + \frac{\Delta H_{T_1}^\circ}{RT_1} - \frac{\Delta H_{T_2}^\circ}{RT_2} + \frac{\Delta S_{T_2}^\circ - \Delta S_{T_1}^\circ}{R}$$

If we are trying to expand this over a temperature range which is narrow enough that we can assume  $\Delta H$  and  $\Delta S$  are temperature independent, then we can simplify this to:

$$\ln K_{T_2} = \ln K_{T_1} + \frac{\Delta H_T^\circ}{R} \left[ \frac{1}{T_1} - \frac{1}{T_2} \right] = \ln K_{T_1} + \frac{\Delta H_T^\circ}{R} \left[ \frac{T_2 - T_1}{T_1 T_2} \right]$$

If  $\Delta H < 0$  (exothermic reaction) then increasing temperature lowers  $K$ , pushing the reaction towards reactants. This is precisely what your intuition should tell you (*think LeChatelier*). The opposite is true for endothermic reactions. Alternatively, we can come up with a more general expression. Remembering our definition of Gibbs free energy:  $G = H - TS$  then  $\frac{G}{T} = \frac{H}{T} - S$ . Differentiating in  $T$  yields

$$\left[ \frac{\partial (G/T)}{\partial T} \right]_p = \frac{1}{T} \left[ \frac{\partial H}{\partial T} \right]_p - \frac{1}{T^2} H + \left[ \frac{\partial S}{\partial T} \right]_p$$

Finally, remember that *at equilibrium*  $\left[ C_p^o \right] = \left[ \frac{\partial \left[ H^o \right]}{\partial T} \right]_p = T \left[ \frac{\partial \left[ S^o \right]}{\partial T} \right]_p$

$$\frac{\partial \left[ \frac{G^o}{T} \right]}{\partial T} = \frac{1}{T} \left[ C_p^o \right] - \frac{H^o}{T^2} - \frac{1}{T} \left[ C_p^o \right]$$

$$\left[ \frac{\partial \left( \frac{G^o}{T} \right)}{\partial T} \right]_p = \frac{H^o}{T^2} \quad (\text{Gibbs-Helmholtz equation})$$

The integrated form of this expression is:

$$\int_{T_1}^{T_2} \left[ \frac{\partial \left( \frac{G^o}{T} \right)}{\partial T} \right]_p dT = \int_{T_1}^{T_2} \frac{H^o}{T^2} dT$$

$$\frac{G_{T_2}^o}{T_2} - \frac{G_{T_1}^o}{T_1} = \int_{T_1}^{T_2} \frac{H_T^o}{T^2} dT$$

Now, if  $\Delta H$  is independent of temperature, this will eventually lead to the expression above for the temperature dependence of  $\ln(K)$  - try it!

For the unfolding of a protein, we can assume a simple "two-state" model (in other words, the protein is either completely folded or "completely" unfolded, but intermediates do not accumulate significantly). You might think that this is an unreasonable assumption, however the folding of a protein is a highly cooperative process, and one that is often poised on the edge of stability, such that quite often we can safely make this assumption.

The following represent thermodynamic data for the unfolding of ribonuclease at 30°C.

In other words, for the reaction: F  $\leftrightarrow$  U

Note the large  $\Delta C_p^o$ .

pH	$\Delta G^o$ (kcal/mole)	$\Delta H^o$ (kcal/mole)	$T\Delta S^o$ (kcal/mole)	$\Delta S^o$ (cal deg <sup>-1</sup> mole <sup>-1</sup> )	$\Delta C_p^o$ (cal deg <sup>-1</sup> mole <sup>-1</sup> )
1.1 3	-1.09	60.3	61.2	202	2072
2.5 0	0.91	57.2	56.3	186	1985
3.1 5	3.09	53.0	50.0	165	1987

Note the large  $\Delta C_p^o$ . For comparison, heat capacities for pure liquids are:

water: 18 cal deg<sup>-1</sup> mole<sup>-1</sup>      benzene: 32 cal deg<sup>-1</sup> mole<sup>-1</sup>

Note that the protein is unstable at pH=1.13, but becomes stable at pH=2.50. Note also that  $\Delta H^o$  and  $\Delta S^o$  are large relative to  $\Delta G^o$ . The large  $\Delta C_p^o$  indicates that  $\Delta H^o$  and  $\Delta S^o$  will show a strong temperature dependence.

The following data represent thermodynamic parameters for the unfolding of a number of different proteins. Again, note the large  $\Delta C_p^o$ .

## The Folding of Proteins and Nucleic Acids

### Quick review - the amino acids

Nonpolar - low solubility in water

Always: Ala, Val, Ile, Leu. Almost always: Phe, Trp, Met

Charged - high solubility in water, must interact strongly with water

Glu, Asp, Arg, Lys

Strongly Polar - need to interact with water

Asn, Gln, Ser, Thr

Ambiguous

Cys, His - pK<sub>a</sub>'s near 7.0. Clearly polar when charged, fairly nonpolar when neutral

Tyr - has a H-bond donor, but is otherwise very nonpolar. Indeed it is normally found in nonpolar niches in protein structures - rarely at the surface.

Special

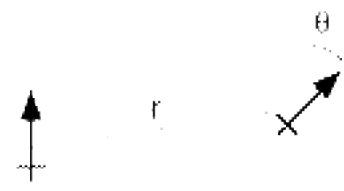
Gly - nonpolar. More freedom to rotate than others (wider range of phi and psi angles).

Pro - nonpolar. Less freedom to rotate than others (phi fixed, psi limited - see below).

## The Forces that Govern Protein Folding

### Coulombic Charge - Charge Interaction

### Dipolar Interaction

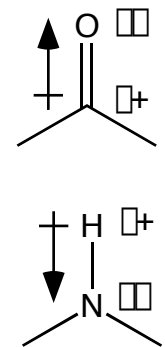
$$E_{dipolar} = \frac{1}{4\pi\epsilon_0} \frac{\mu_A \cdot \mu_B}{r^3} - \frac{3(\mu_A \cdot \vec{r})(\mu_B \cdot \vec{r})}{4\pi\epsilon_0 r^5}$$


where  $\mu_A$  and  $\mu_B$  are point dipoles (vectors) separated by a distance  $r$ . Remember that carbonyl C=O and amide N-H bonds have an uneven distribution of electrons - that is, each has a dipole moment associated with it. Note from above that the dipolar interaction depends on both the *distance* and the *angle* between the dipoles.

Point charge approximation. One can model the dipolar interaction using a point dipole approximation as above, or alternatively, since the distance between dipoles is short, it may be better to view each dipole as two separated charges. This is shown in the lower figure at right. We can now view the interactions as simple coulombic charge-charge interactions, according to:

$$E_{dipolar} = \sum_{i,j} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}^2}$$

This is a common approach in “molecular dynamics/mechanics” calculations. Each atom in the entire protein is assigned a partial charge and the coulombic interactions are summed. Again note that an important parameter in this estimation is the dielectric ( $\epsilon$ ) of the intervening medium. Values from 2-5 or a bit higher are realistic. ( $\epsilon$  for water is 80!). In this case, there is little between the charges and so a low dielectric is reasonable. For dipoles which are separated a greater distance in the protein, The effective dielectric of the intervening protein will be somewhat higher. However, estimates are difficult to make and the value will be different in different parts of the protein.



## Dipole - Point Charge Interactions

You can readily imagine interactions between a dipole moment from one part of the protein structure and a point charge nearby (a charged amino acid for example). This intermediate case is again well approximated by the full point charge picture presented above.

## van der Waals Interactions

this interaction is a result of dipolar interactions between transient *induced* dipole moments within atoms. Electrons in any given atom may at any point in time be non-uniformly distributed about the nucleus (although *on average* they are symmetrically displaced). This non-uniform distribution results in a momentary dipole. The resulting dipole can then induce a dipole in a neighboring atom, of opposite direction. This leads to an attractive force between the atoms. However, at shorter distances  $e^-e^-$  repulsion becomes greater than this attractive force, so that the atoms repel.

How important are van der Waals interactions to protein structure?

A single van der Waals interaction is only a few hundred calories, however, there are many such contacts in a single protein. So you can imagine that the energies could add up significantly.

However, the folded and unfolded forms of the protein *both* have a large number of van der Waals contacts (the latter has many with water), so that the *net* stabilization of protein structure imparted by van der Waals interactions may be small.

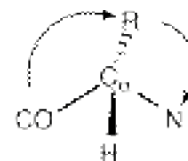
## Bond stretching

Energies are too big to worry about. In other words, the energy cost to significantly lengthening (or shortening) a covalent bond in the final (average) protein structure is so high that such considerations can be ignored. Covalent bonds are considered to be constant in length and set at classic values for the bond involved.

## Torsional constraints - Rotations around rotatable bonds

Aside - Stereochemistry at the  $C_\alpha$  position

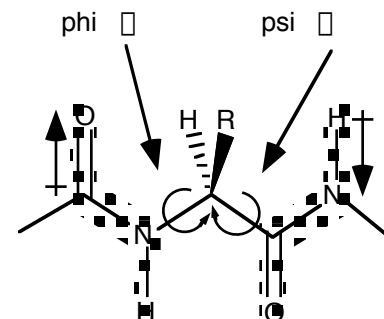
The stereochemistry at the  $C_\alpha$  position can be easily remembered by use of the little corny mnemonic "CORN." With the (little) hydrogen pointing towards you, proceeding in a clockwise direction, the other three substituents should spell out "CORN" as shown at right (another contribution from Jane Richardson).



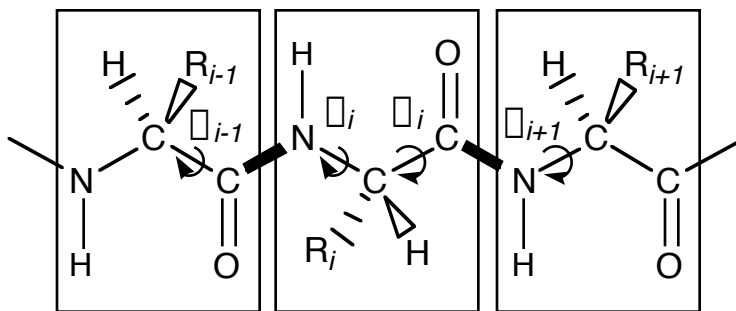
Rotation around some bonds is also too high in energy cost. For example, the peptide bond CO-NH has partial double bond character. For our purposes, then it can be assumed that rotation of the peptide bond from the angle at which the  $\pi$  system is planar costs too much energy to occur with any frequency.

However, rotation around the other two bonds in the polypeptide chain does occur at low energy. Consequently, variation in these angles will contribute to differences in protein structure. (see C&S pp 254-257 for a detailed explanation of the sign conventions).

$\phi$  ranges from  $-180^\circ$  to  $180^\circ$  with  $90^\circ$  corresponding to a *cis* configuration. An all *trans* backbone (if it adopted the simple planar structure above) would correspond to  $\phi$  and  $\psi = 180^\circ$ .







For a protein  $\alpha$ -helix,  $\phi = -57^\circ$  and  $\psi = +47^\circ$  for every amino acid unit in the helix.

Ramachandran plots show that only certain ranges (and combinations of ranges) are allowed for these angles. This is a result of steric considerations of the functional groups. An important distinction is made for the amino acid glycine (Gly): Due to the very small size of its functional group (hydrogen), a much wider variety of angles is energetically reasonable. (see C&S p. 259 for illustration). This has important implications for protein folding.

### Sterics effects possible phi / psi angles

Of course, the major restriction on phi and psi angles arises due to the bulk of the groups attached to a given set of peptide bonds. Only certain ranges of angles allow low energy steric interactions.

### Staggered vs. eclipsed

As we have all seen in organic chemistry classes, bonds prefer to rotate such that their attached bonds are "eclipsed." This is most simple explained in terms of sterics... Suffice it to say that this effects preferred values of phi and psi.

### The dipolar interaction significantly effects Ramachandran values for phi and psi

The picture at right shows that the dipole moments associated with the peptide bond place energetic restrictions on the angles which can be adopted by phi and psi. Adjacent peptide bonds would like to interact favorably.

### Effects of *unusual* amino acids on phi and psi

Gly has a simple hydrogen as its "functional group" (R above). Consequently the steric restrictions are much less for Gly and it is allowed a much wider range of phi/psi angles (see C&S).

Pro is also unusual in that the normal C-N bond is tied up in a 5-membered ring. Consequently the

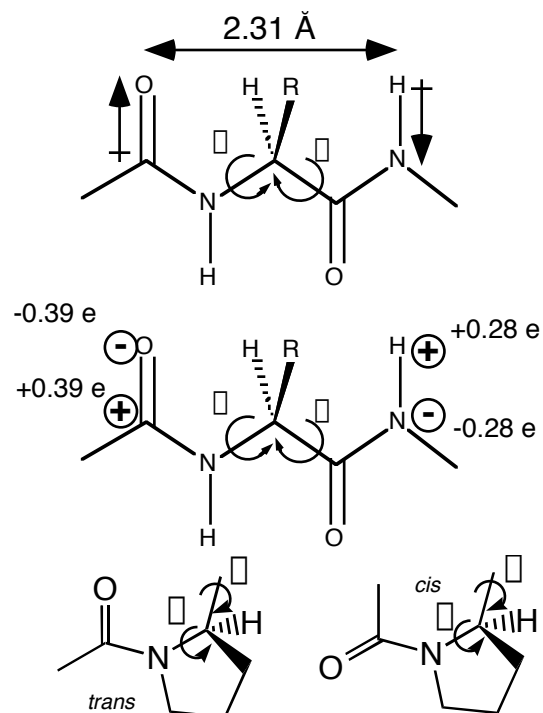
phi angle is rigidly fixed near  $-60^\circ$ . The psi angle is not rigidly fixed, but is significantly more restricted than in any other amino acid. The angles for psi which result in energy minima are  $-55^\circ$  (compact chain) and  $145^\circ$  (extended chain).

In addition, the psi angle at the amino acid  $i-1$  is more restricted and constrained to different values than in other contexts (the NH "upstream" of it is now N-R). So most amino acids are very sensitive to the presence of Pro at  $i+1$  (the exception is Gly, which already has a wide range of angles available to it and has little steric clash to present to the N-R group).

f

### Covalent Interactions

As for bond stretching and amide bond rotation (but much more strongly), the disruption of a covalent bond requires a very high energy cost. Thus, covalent bonds are not considered as variable.



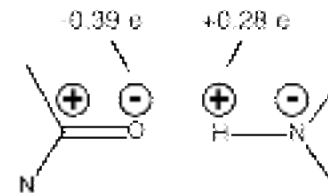
## “Non-bonding” Interactions

A variety of non-bonding interactions contribute to protein stability, and in fact, it is these interactions, together with torsion angle changes which drive the formation of secondary and tertiary structure.

London dispersion energy -

## Hydrogen bonding

One way of viewing hydrogen bonding is simple electrostatics. We could view it as a charge-charge attraction as shown at right. This is probably overly simplistic and does not include orbital overlap, which results in an angular dependence to the interaction.



## Water - the unusual solvent

Compound	Melting point (K)	Boiling point (K)
H <sub>2</sub> O	273	373
H <sub>2</sub> S	190	211
CH <sub>3</sub> (CO)OH	290	391
CH <sub>3</sub> (CO)CH <sub>3</sub>	178	330
CH <sub>3</sub> CH <sub>2</sub> OH	156	351
CH <sub>3</sub> CH <sub>2</sub> CH <sub>3</sub>	83	231
CH <sub>3</sub> OCH <sub>3</sub>	135	249
CH <sub>3</sub> NH <sub>2</sub>	181	267
CH <sub>3</sub> CH <sub>3</sub>	101	185

## Water structure - hydrophobic interactions

C&S Part I (pp 279-89) discuss hydrophobic interactions and water structure.

It's not as simple as they imply (see Dill below), but the concepts are interesting.

Water forms “cage-like lattices” around hydrophobic molecules

Water is indeed unique in its properties. Even in its liquid state, it adopts transient local “crystalline” structures (*microscopic icebergs*). **Cage like lattices** can form which can sequester small hydrophobic molecules. This results in enthalpic stabilization. But formation of these cage-like polyhedrons severely restricts the entropy of the water molecules - ie. entropically destabilized..

Hence there is a large entropic cost to the solvation of nonpolar molecules -- the hydrophobic effect is typically viewed as entropically driven.

Kauzman measured  $\Delta H$  and  $\Delta S$  for the transfer of a number of different molecules from nonpolar solvents to water (p. 285).  $\Delta H$  is negative in all cases (the nonpolar molecules prefer water because the ice like structures create hydrogen bonds), but  $\Delta S$  is negative as well (the nonpolar molecules decrease the entropy of the system in going to water and so prefer the non polar solvent). At room temperature,  $\Delta G$  strongly favors transfer to the non polar solvent (entropically driven).

## Denaturation by urea and guanidinium hydrochloride

Various hydrophobic or nonpolar side chains prefer 8M urea to water. (that is  $\Delta G_{\text{transfer to water}}^{\text{gua}}$  is negative (similar arguments).

## Ionic interactions

They occur in proteins. But are thought to be a minor contributor to stability (see below).

Like charges - repulsive - enthalpy unfavorable

Opposite charges - attractive - enthalpy favorable

A charge in aqueous solution requires substantial solvation. This decreases the entropy of the water.

An ion pair requires much less.

## Disulfide bonds

Significant entropic destabilization of the unfolded structure.

Recent studies suggest that disulfide bonds don't direct protein folding, but stabilize structures which already form. They stabilize the structure not so much by lowering the energy of the folded state, but by entropically raising the energy of the unfolded state.

## Dominant Forces in Protein Folding? - Ken Dill Review (*Biochemistry* 29, 7133, 1991)

### *Electrostatic interactions*

*Long-range* - if the net charge of a protein is non zero (ie the protein is on either side of its isoelectric point), then there will be a net repulsive interaction between unbalanced like charges. This forms the basis for the general trend that a protein most stable at its isoelectric point.

*Specific charge interactions* - attractive ion pairing (salt bridges, ionic bonds). Not too sensitive to bulk dielectric, since the interaction is short range - little intervening medium. Can effect unfolded state as well as folded state.

Most salt bridge interactions occur at the surface of the protein. The ion pair is not completely self-neutralizing and so requires solvation (cost to bury an ion pair is on the order of  $7\text{ kcal/mole}$ ).

Rarely found in the interior of proteins. Any solvent effects which stabilize/destabilize ion pairs at the surface of the protein would have similar stabilizing/destabilizing effects on the ions in the unfolded state (little net effect). In fact recent studies suggest that hydration is better in the unfolded state (ion is more accessible).

Often associated with a change in partial molal volume. Isolated charges require extensive hydration. The water packs tightly around the charge - electrostriction.

Clearly not the dominant force in protein folding - protein stability is not a steep function around the  $pI$ , nor is stability strongly dependent on salt (dielectric of the medium).

Evolutionary studies show that surface charges are not highly conserved. Mutagenesis studies have confirmed that surface charges can usually be mutated with little perturbation of stability.

Too few to be a major net structural force ( $1\text{--}3\text{ kcal/mole/ion pair}$ ,  $\approx 5$  ion pairs / 150 aa's).

### *Hydrogen bonding and van der Waals interactions*

*H-bonding*: discussed previously. View it either as 1) a local charge-charge attraction (with partial charges) or 2) as a bonding interaction involving orbital overlap.

*van der Waals*: Like H-bonding, it is short range and will have similar effects on energetics. Hard to model and understand the way we do H-bonding.

*Helix-coil transitions*: Balance between entropy and enthalpy. At low temperatures enthalpy dominates and the helix is favored. At high temperatures, entropy dominates and the random coil is favored. A sharp transition occurs. In general increased net charge of the protein should disfavor helix formation and favor the more expanded random coil. Solvents which H-bond well compete for the intrahelical H-bonding and so destabilize the helix.

Similar arguments for  $\beta$ -sheets.

Kauzmann argued early on (and Dill agrees) that although H-bonding interactions are important to details of structure, they cannot drive protein folding.

The reason is that the H-bond interactions with water in the unfolded state are quite strong. The net change is very small and could go either way.

### **Cooperativity in formation of protein structures**

The first H-bond in a unit of secondary structure typically costs energy due to loss of configurational entropy. But once that price is paid, subsequent H-bonds forms without the energetic expense. For this reason, helices become more stable and transitions sharpen with increasing length.

*N-methyl-acetamide* ( $\text{CH}_3(\text{CO})(\text{NH})\text{CH}_3$ ) is a popular model to study effects on hydrogen bonding interactions (mimics the peptide backbone). At  $25^\circ\text{C}$ , dimerization is disfavored. Loss of translational entropy?

Additionally, consider entropic effects in the formation of multiple H-bonds. In the pairing of a dipeptide, two bonds can form in a single dimer. When only one forms, the two molecules still maintain a rotational degree of freedom about that bond. When the second bond forms, that rotation is completely lost - entropy decreases. Difficult to separate the effects.

Dill's conclusion: H-bonding alone will not drive protein folding, but if other processes favor folding, then H-bonds will be favorable within the structure.

Statistics from known protein structures:

C=O groups:

89% are H-bonded  
43% to water  
11% to side chains  
46% to main chain

NH groups:

88% are H-bonded  
21% to water  
11% to side chains  
68% to main chain

If H-bonding were dominant, then

*Solvents which are good H-bonders should destabilize protein structure, those that are bad H-bonders should have little effect. Observation: 1% SDS can denature proteins, despite its low concentration and the fact that it does not destabilize helices.*

*Tetraalkylammonium salts increasing denature proteins as the length of the alkyl group increases (shouldn't effect H-bonding).*

*Alcohols are more hydrophobic than water and they stabilize helices, yet they denature proteins.*

*Mutagenesis: H-bonds effect stability, but the magnitude and direction of the effect varies significantly.*

*"Local" interactions - "Intrinsic" properties (note his definition of "intrinsic")*

*Must know the differential effect on native vs. unfolded state.*

*Define local interactions as those occurring between an amino acid and another amino acid 3-4 distant in the primary sequence*

*At 20°C, the propensity for a sequence to form a helix is generally quite small.*

*Additional stability occurs if helix formation leads to burial of a nonpolar surface.*

*Helices are further stabilized by charge compensation at the ends.*

*Helices can be stabilized by salt bridges and other side chain interactions.*

*Bottom line: context can be at least as important as "intrinsic" properties. Early on, people took sequences which were known to form helices in proteins and were somewhat surprised to find that they did not form helices when isolated in solution. Should be no surprise now.*

*Helices in globular proteins are on average short (6-12 residues) - much shorter than should be stable.*

*Conversely, the longer a helix, the more stable it should be, yet we don't see lots of long helices in solution (think about it - why?).*

*Context "rules" for formation of helices in proteins (note that "rules" is in quotes...)*

*Charges are distributed to stabilize the helix dipole moment (or the end carboxyl oxygen and amino proton).*

*Ends of helices are often at the surface (same reason as above)*

*Helices can pack in anti-aligned pairs to stabilize dipole moments (questionable importance)*

*Other problems with "intrinsic" properties*

*Cannot predict sheet structures, which are inherently non-local.*

*Intrinsic properties have shown a 64% average success rate in predicting structure*

*He goes through an analysis which says that this means local (intrinsic) factors contribute 15-30% of the total information required to predict a structure.*

*Hydrophobic effect - nonpolar solvation*

*Nonpolar solvents denature proteins by lowering the energy of the unfolded state*

*Evidence:*

- 1) Temp dependence of the free energy of folding follows that of transfer of nonpolar model compounds from water to nonpolar media*
- 2) Crystal structures confirm sequestering of nonpolar groups*
- 3) Dependence of stability on the nature of the salt species in solution follows the Hofmeister series, suggesting hydrophobic interactions*
- 4) Mutagenesis and other studies suggest that stability is proportional to the oil-water partitioning of the amino acid*
- 5) Hydrophobicity of core amino acids is evolutionarily conserved*
- 6) Computer simulations of incorrectly folded proteins shows that poor interior / exterior distribution of hydrophobic residues is a major factor in instability.*

### **Intro to hydrophobic interactions**

*Note first that mixing of simple solutions is entropically driven. Molecules want to "spread out" and obtain the highest possible translational entropy (dispersion). However repulsion between molecules counters this driving force somewhat. When two components A and B are relatively insoluble in each other (strong repulsive forces between them), an enthalpic term opposes the entropic term. With increasing temperature, the entropic term dominates and the solute is eventually allowed to mix. Or rather, solubility goes up.*

*However, entropic terms can also oppose mixing. In water, nonpolar molecules are harbored in water "shells," the polyhedron we discussed before. The dissolution of a nonpolar molecule in the water, leads to extensive ordering of this solvation shell, and hence to a decrease in entropy of the system.*

*Narrow definition of hydrophobic interactions: Dill describes the hydrophobic interaction very narrowly as the transfer of a nonpolar solute to an aqueous solution only when the mixing is "opposed by an excess entropy."*

*"Excess entropy" is the entropy associated with transfer of a solute which is not derived from the simple (statistical) entropy of mixing.*

*EMBED "Equation" "Word Object1" \\* mergeformat*

*Another property of the solvation of nonpolar solutes in water is that the transfer to water is characterized by an enthalpy with a strong temperature dependence (in other words, the heat capacity for transfer is large and positive).*

Remember that  $\Delta H = \int C_p \Delta T$

Now we have (assuming heat capacity independent of temperature):

$\Delta H = \int_{T_h}^{T_s} C_p dT = C_p (T_s - T_h)$

$\Delta S = \int_{T_h}^{T_s} \frac{C_p}{T} dT = C_p \ln \frac{T_s}{T_h}$

Therefore, assuming that the heat capacity is independent of temperature:

$$\Delta G_T = \Delta H_T - T \Delta S_T = \Delta H_{T_h} + \Delta C_p (T - T_h) - T \Delta S_{T_s} - T \Delta C_p \ln \frac{T}{T_s}$$

$$= \Delta H_{T_h} - T \Delta S_{T_s} + \Delta C_p \left[ \frac{T}{T_h} - T \ln \frac{T}{T_s} \right]$$

This means that  $\Delta G$  does not necessarily steadily decrease with temperature as would be expected for  $\Delta G = \Delta H - T \Delta S$  if  $\Delta H$  and  $\Delta S$  were temperature independent. The diagram at right shows that  $\Delta G$  (for the unfolding of the protein) may reach a maximum at a particular temperature. Indeed, a similar plot can be drawn for the transfer of benzene to water, and benzene is least soluble in water near room temperature.

If protein folding were driven by this process, then we might expect proteins to exhibit a temperature at which the hydrophobic effect is maximal and the protein is therefore maximally stable (ie the protein would tend to unfold more readily at higher temperatures and at lower temperatures). This may be somewhat counterintuitive, but once you accept hydrophobicity as a major determinant of protein stability and the fact that the hydrophobic effect weakens at temperatures below its maximum, then you can see why some proteins are unstable at low temperatures - "cold denaturation." Of course for many proteins, the low end of stability may be below the freezing point of water and so cold denaturation is never observed.

This also means that the hydrophobic effect is entropy driven at lower temperatures, but enthalpy driven at higher temperatures! Why? Going back to the iceberg model for nonpolar solute solvation, at low temperatures the "cages" are well-formed and water molecules get to form lots of hydrogen bonds ( $\Delta H$  happy), but there are few conformations available that optimally satisfy  $\Delta H$ , so entropy is low ( $\Delta S$  unhappy). At higher temperatures the opposite becomes true -- the clathrate structures break down (or rather they flex, "bending" the H-bonds), increasing entropy ( $\Delta S$  happy), but at the cost of lost (or weaker) H-bonds ( $\Delta H$  unhappy).

**Bottom line:** hydrophobic effects are characterized by a large heat capacity. Protein unfolding is typically characterized by a large heat capacity. Therefore protein unfolding must be accompanied (and by dominated by) the solvation (exposure) of nonpolar surfaces.

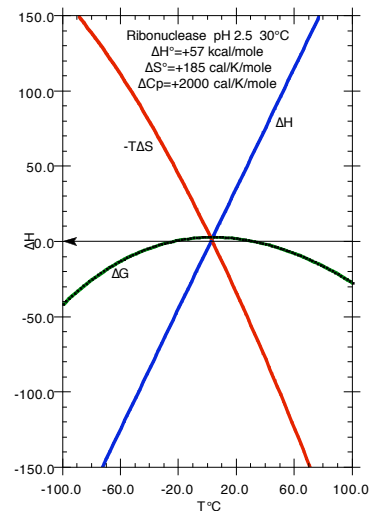
### Residual enthalpy and entropy of unfolding

There is an additional positive entropy and enthalpy of unfolding, not accounted for by predictions from nonpolar solvation. From where do these arise? Residual enthalpy seems to increase with polar content - one proposal: folded protein has more H-bonding contacts. As for entropy, unfolding should result in increased entropy for the protein (bond angles - configurational entropy).

### Site-directed mutagenesis studies - single amino acid substitutions

Various studies have now been done which involve the substitution of an interior amino acid in a protein by various other amino acids. If the system were simple, we could assume that the interior is represented by ethanol or a similar solvent, and that unfolding of the enzyme is a simple transfer from that solvent to water. To the extent that this is true, a plot of  $\Delta \Delta G$  for each mutant as a function of  $\Delta G_{tr}$  for that amino acid should yield a slope of 1. The non-ideality of this slope could arise from a number of factors:

- 1) "deformability" of the cavity in which the amino acid sits. How well does the protein adjust its local configuration to accommodate the substitution
- 2) specific interactions of each amino acid with the cavity, including enthalpy and entropy
- 3) how well does the denatured form of the protein mimic pure aqueous medium?
- 4) how exposed to water is the amino acid in the folded form (we assumed not at all)?



If the rest of the protein stabilizes a "cavity" into which the amino acid can fit, then the transfer of the side chain into that cavity will be less energetically costly than transferring the same side chain into the model solvent -- in the latter case, there is an energetic cost to create the cavity.

Conversely, if the cavity is rigid and one tries to insert a larger amino acid by substitution, then the energetic cost would be higher than expected (slope > 1). These kinds of analyses have been done on a number of proteins, with slopes ranging from 1 to 4.

### **Summary: Definition of Hydrophobicity**

Transfer of a nonpolar solute from non aqueous media to aqueous media when:

1) transfer is energetically strongly disfavored

**and**

2) transfer is associated with a large increase in heat capacity

### **Opposing Forces**

If hydrophobicity were the only force involved, proteins would be much more stable than they are. What else is there?

Our discussion of hydrophobicity dealt with entropic effects from ordering water, but we ignored the large decrease in the entropy of the polypeptide associated with folding. We return to the entropy associated with rotation about the phi and psi bonds, and about the bonds in the amino acid side chains. He calls this "configurational" entropy. He also refers to this as "local" entropy.

Another aspect of the entropy loss in folding is what he calls "nonlocal" entropy. This has to do with the number of possible ways a polymer chain can fold onto itself. The hydrophobic driving force wants to minimize the exposure of nonpolar groups - there is only one (or maybe a few) configuration which minimizes that exposure. Consequently, on protein folding the chain folding is reduced from a very large number of possibilities down to 1 or a few.

We come back to things like disulfide linkages. These covalent linkages reduce the nonlocal entropy of the unfolded state -- there are now many fewer ways of folding the polymer chain in solution.

Similarly, he argues that carbohydrates added to proteins can stabilize them by restricting the number of possible states in the unfolded form of the protein.

### **Internal architecture of proteins**

To some extent, the internal architecture of proteins is dictated by this need to pack in the one form which will optimize burial of hydrophobic groups. But of course, we can also now see an important role for the H-bonding which is so readily observable. Which is dominant? Hard to say. Both are important. The upshot of this entire article is that although H-bonds are easy to see and to talk about, when all energies are added up, the hydrophobic effect is energetically the much larger force -- not the only force, but the more dominant one...

## Analyses of Protein Folding

### Two-State Model for Protein Folding

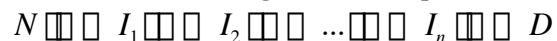
The basic model:  $N \rightleftharpoons D$

Most protein unfolding studies assume a two-state model for denaturation. The protein can exist in only two discrete states: native (N) and denatured (D).

### Multi-State Model for Protein Folding

Effect of multiple states on observables - what if unfolding is NOT two-state?

Assume the following scheme for protein unfolding:



where N refers to native and D to denatured.

Given a microscopic observable for each state  $y_i$  then the overall macroscopic observable  $y$  is given by:

$$y = f_N y_N + f_D y_D + \sum_{i=1}^n f_i y_i$$

where  $f_i$  is the mole fraction of proteins in state  $i$ . Then  $f_N + f_D + \sum_{i=1}^n f_i = 1$

The observable can then be expressed as:

$$y = f_D y_D + \sum_{i=1}^n f_i y_i + f_N y_N$$

$$y = y_N \left( f_D y_N + f_D y_D + \sum_{i=1}^n (f_i y_i - f_i y_N) \right)$$

If we measure our observable as before, assuming a two-state unfolding process, we will calculate an *apparent* fractional unfolding:

$$f_{app} = \frac{y - y_N}{y_D - y_N} = \frac{y_N \left( f_D y_N + f_D y_D + \sum_{i=1}^n (f_i y_i - f_i y_N) \right) - y_N}{y_D - y_N}$$

$$f_{app} = \frac{f_D (y_D - y_N) + \sum_{i=1}^n (f_i y_i - f_i y_N)}{y_D - y_N}$$

$$f_{app} = f_D + \sum_{i=1}^n f_i \frac{y_i - y_N}{y_D - y_N}$$

$$f_{app} = f_D + \sum_{i=1}^n f_i d_i \quad \text{where} \quad d_i = \frac{y_i - y_N}{y_D - y_N}$$

from before, the apparent equilibrium constant for unfolding is:  $K_{app} = \frac{[D]_{app}}{[N]_{app}} = \frac{f_{app}}{1 - f_{app}}$

We can also talk about (but maybe not measure) the true equilibrium constant:  $K_D = \frac{D}{N} = \frac{f_D}{f_N}$

We can also talk about each microscopic equilibrium constant:  $K_i = \frac{I_i}{N} = \frac{f_i}{f_N}$

$$\frac{f_{app}}{f_N} = \frac{f_D}{f_N} + \sum_{i=1}^n \frac{f_i}{f_N} d_i = K_D + \sum_{i=1}^n K_i d_i = K_D \left( 1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i \right)$$

$$\frac{1 \square f_{app}}{f_N} = \frac{1 \square f_D + \sum_{i=1}^n f_i d_i}{f_N} = \frac{1 \square f_D}{f_N} \square \sum_{i=1}^n \frac{f_i}{f_N} d_i$$

$$\frac{1 \square f_{app}}{f_N} = \frac{f_N + \sum_{i=1}^n f_i}{f_N} \square \sum_{i=1}^n \frac{f_i}{f_N} d_i = 1 + \sum_{i=1}^n \frac{f_i}{f_N} \square \frac{f_i}{f_N} d_i = 1 + \sum_{i=1}^n \frac{f_i}{f_N} (1 \square d_i)$$

$$\frac{1 \square f_{app}}{f_N} = 1 + \sum_{i=1}^n [K_i (1 \square d_i)]$$

$$K_{app} = \frac{f_{app}}{1 \square f_{app}} = \frac{\frac{f_{app}}{f_N}}{\frac{1 \square f_{app}}{f_N}} = \frac{K_D \square 1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i}{1 + \sum_{i=1}^n [K_i (1 \square d_i)]} = K_D \frac{1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i}{1 + \sum_{i=1}^n [K_i (1 \square d_i)]}$$

Finally,

We have not yet said anything about how  $y_i$  depends on  $i$ . But let's look at two extremes, with one simple assumption.

The assumption: the observables for the intermediate states ( $y_i$ ) lie between  $y_N$  and  $y_D$ .

Case 1: intermediate states have properties similar to the denatured state  
then  $y_i \approx y_D$  and  $d_i \approx 1$

$$K_{app}(d_i \square 1) = K_D \frac{1 + \sum_{i=1}^n \frac{K_i}{K_D}}{1 + 0} = K_D + \sum_{i=1}^n K_i$$

Case 2: intermediate states have properties similar to the native state  
then  $y_i \approx y_N$  and  $d_i \approx 0$

$$K_{app}(d_i \square 0) = K_D \frac{1}{1 + \sum_{i=1}^n K_i}$$

In any case, only when  $K_i = 0$  (no intermediates), then  $K_{app} = K_D$ .

In general, if intermediates exist, then  $K_{app}$  will depend on how  $d_i$  depends on the intermediate states. One way to test for intermediates is to measure  $K_{app}$  with different observables. If the two values of  $K_{app}$  are *not* the same, then  $K_{app}$  is not a true  $K_D$  - the process is not two state.

(Note from the extreme cases above, that the converse is not necessarily true).

## Calorimetry

Similarly, we can worry about how the van't Hoff enthalpy that one has measured by the previous approaches is effected by the existence of intermediates.

Remember that



$$\Delta H = RT^2 \frac{\partial(\ln K)}{\partial T}$$

$$\Delta H_{app} = RT^2 \frac{\partial}{\partial T} \ln K_D \frac{1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i}{1 + \sum_{i=1}^n [K_i(1 + d_i)]}$$

$$\Delta H_{app} = RT^2 \frac{\partial}{\partial T} \ln K_D + \ln \left[ \frac{1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i}{1 + \sum_{i=1}^n [K_i(1 + d_i)]} \right]$$

$$\Delta H_{app} = RT^2 \frac{\partial \ln K_D}{\partial T} + RT^2 \frac{\partial}{\partial T} \ln \left[ \frac{1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i}{1 + \sum_{i=1}^n [K_i(1 + d_i)]} \right]$$

$$\Delta H_{app} = \Delta H_D + RT^2 \frac{1}{1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i} \frac{\partial}{\partial T} \sum_{i=1}^n \frac{K_i}{K_D} d_i - RT^2 \frac{1}{1 + \sum_{i=1}^n [K_i(1 + d_i)]} \frac{\partial}{\partial T} \sum_{i=1}^n [K_i(1 + d_i)]$$

$$\Delta H_{app} = \Delta H_D + RT^2 \frac{1}{1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i} \sum_{i=1}^n \frac{\partial K_i}{\partial T} \frac{K_i}{K_D} d_i - RT^2 \frac{1}{1 + \sum_{i=1}^n [K_i(1 + d_i)]} \sum_{i=1}^n \frac{\partial K_i}{\partial T} (1 + d_i)$$

Remember that:  $\Delta H = RT^2 \frac{\partial(\ln K)}{\partial T} = RT^2 \frac{1}{K} \frac{\partial K}{\partial T}$  so that  $\frac{\partial K}{\partial T} = \frac{K}{RT^2} \Delta H$

Then

$$\Delta H_{app} = \Delta H_D + \frac{1}{1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i} \sum_{i=1}^n \frac{\Delta H_i K_i}{K_D} d_i - \frac{1}{1 + \sum_{i=1}^n [K_i(1 + d_i)]} \sum_{i=1}^n [\Delta H_i K_i (1 + d_i)]$$

$$\Delta H_{app} = \Delta H_D + \frac{\sum_{i=1}^n (\Delta H_i - \Delta H_D) \frac{K_i}{K_D} d_i}{1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i} - \frac{\sum_{i=1}^n [\Delta H_i K_i (1 + d_i)]}{1 + \sum_{i=1}^n [K_i(1 + d_i)]}$$

agai

$$\Delta H_{app} = \Delta H_D + \frac{\sum_{i=1}^n \frac{(\Delta H_i - \Delta H_D) K_i}{K_D} d_i}{1 + \sum_{i=1}^n \frac{K_i}{K_D} d_i} - \frac{\sum_{i=1}^n \frac{\Delta H_i}{\Delta H_D} [K_i(1 + d_i)]}{1 + \sum_{i=1}^n [K_i(1 + d_i)]}$$

n, as expected if  $K_i$  are all zero (two state process), then  $\Delta H_{app} = \Delta H_D$ .

Again, however, if all of the  $d_i$  are zero but  $K_i$  are not, we have:

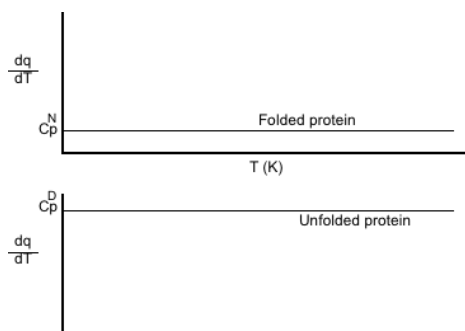
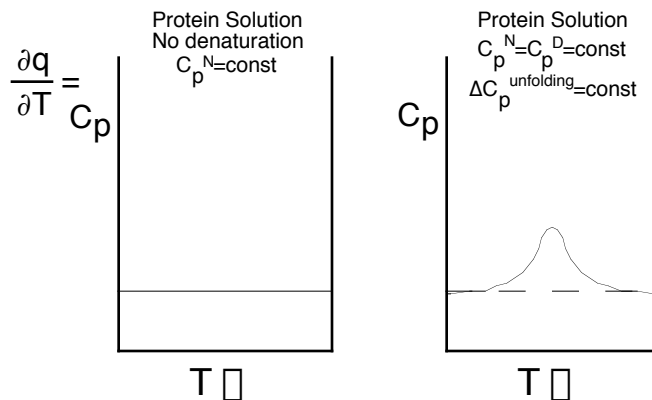
$$\Delta H_{app} = \Delta H_D + \frac{\sum_{i=1}^n \frac{\Delta H_i}{\Delta H_D} K_i}{1 + \sum_{i=1}^n K_i}$$

And, again  $\Delta H_{app} = \Delta H_D$ .

## Differential Scanning Calorimetry

### The method

The plots at right show a measurement of the heat capacity of a solution of protein. The plot to the left just shows the DSC scan for a solution of constant heat capacity, with no reaction proceeding. The plot to its right shows the scan for a solution with constant solution heat capacity, but for which a reaction with non-zero heat capacity is proceeding with increasing temperature. At the beginning of the reaction, the total heat capacity of the solution rises above that for the solution with no reaction, reaching a maximum at the midpoint of the reaction, and decreasing as the reaction goes to completion. This is the simplest protein unfolding DSC trace.

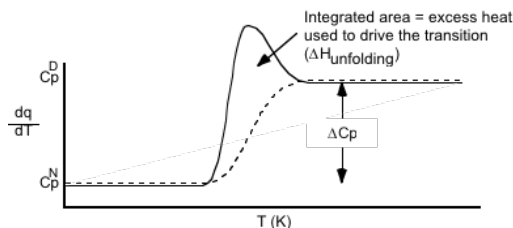
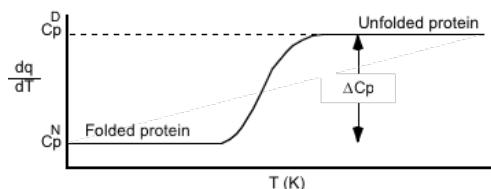


The traces at left show first two different solutions in which no reaction is occurring. As we discussed previously, the solution with denatured protein typically has a higher heat capacity than that with native protein. If the unfolding reaction had zero  $\Delta H$  associated with it, we could simply use the change in heat capacity of the solution as a function of the percent unfolded to monitor unfolding in the same way that we used CD spectra or other techniques to monitor unfolding. An observable changes with protein unfolding.

This is shown in the panel at right. Note that we can express the curve as follows:

$$\frac{dq}{dT}(\text{measured}) = \left( \chi_D C_p^D + \chi_N C_p^N \right) [\text{protein}]$$

Note that the curve at right is not typical. As the reaction proceeds, as we have seen, we must include the heat of the reaction.



The panel at left shows the addition of  $\Delta H^\circ$  for the reaction.

$$\frac{dq}{dT}(\text{measured}) = \chi_D C_p^D + \chi_N C_p^N + \chi \left[ \frac{\Delta H}{T} \right] [\text{protein}]$$

Derive an expression for total measured heat capacity as a function of the extent of the reaction (which is a function of temperature).

**Assume heat capacity independent of temperature!!**

**Assume that the heat capacity of a mixture of native and denatured protein is simply the mole fraction weighted average of the two heat capacities.**

Begin with the following parameters:

$C_p^D$  = molar heat capacity of a solution of denatured protein

$C_p^N$  = molar heat capacity of a solution of native protein

We also have given:

$\Delta H_{T_0}^\circ$  = molar enthalpy for unfolding of the protein at  $T = T_0$

$\Delta S_{T_0}^\circ$  = molar entropy for unfolding of the protein at  $T = T_0$

The total heat capacity for the solution is:

$$\begin{aligned} & \frac{dq}{dT}(\text{measured}) \\ &= (\text{heat from heat capacity of denatured protein solution}) \\ &+ (\text{heat from heat capacity of native protein solution}) \\ &+ (\text{heat from that fraction of protein which denatured during } \Delta T) \end{aligned}$$

$$\frac{dq}{dT}(\text{measured}) = f_D C_p^D + (1-f) C_p^N + f \frac{dH}{dT}[\text{protein}]$$

where we now remember that the enthalpy ( $\Delta H$ ) is temperature dependent.

First, let's determine the temperature dependence of  $\Delta H$ ,  $\Delta S$ ,  $\Delta G$ ,  $K_D$ , and  $f$ :

We know that:

$$dH = C_p dT \quad d \ln K = \frac{dH(T)}{R} \frac{dT}{T^2}$$

integrating from one temperature to another:

$$\int_{T_1}^{T_2} dH = \int_{T_1}^{T_2} C_p dT \quad H_{T_2} - H_{T_1} = C_p (T_2 - T_1) \quad H_T = H_{T_0} + C_p (T - T_0)$$

Similarly,

$$\int_{T_1}^{T_2} dS = \int_{T_1}^{T_2} C_p \frac{dT}{T} \quad S_{T_2} - S_{T_1} = C_p \ln \frac{T_2}{T_1} \quad S_T = S_{T_0} + C_p \ln \frac{T}{T_0}$$

and

$$G_T = H_T - T S_T$$

$$G_T = [H_{T_0} + C_p (T - T_0)] - T [S_{T_0} + C_p \ln \frac{T}{T_0}]$$

We can get the equilibrium constant,  $K$  from  $K = e^{-\Delta G/RT}$

So, we have the extent of the reaction given by  $f$  (mole fraction protein unfolded), which can be derived as:

$$K = \frac{[\text{unfolded}]}{[\text{folded}]} = \frac{f}{1-f} \quad \text{so that} \quad f = \frac{K}{1+K}$$

Finally,

$$\begin{aligned} \frac{dq}{dT}(\text{measured}) &= f C_p^D + (1-f) C_p^N + f \frac{dH}{dT}[\text{protein}] \\ &= f C_p^D + C_p^N - f C_p^N + f \frac{dH}{dT}[\text{protein}] \\ &= f (C_p^D - C_p^N) + C_p^N + f \frac{dH}{dT}[\text{protein}] \\ &= f C_p + C_p^N + f \frac{dH}{dT}[\text{protein}] \end{aligned}$$

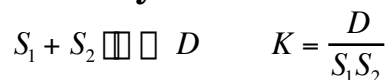
### DNA Melting / Strand Association

DNA duplex formation/melting is particularly amenable to simple thermodynamics.

$\Delta C_p = 0$  for this process. Hence not only are basic assumptions held up, but  $\Delta H^0$  and  $\Delta S^0$  are temperature independent.

Strand association appears to be dominated by "nucleation." It can usually be treated as a simple two-state process. Again, assumptions are valid.

### Thermodynamics - Two non-self-complementary, complementary strands



Assume that when fully melted,  $S_1 = S_2 = S_0$ .

At any temperature, the fraction associated can be represented as  $\alpha$ .

$$D = \alpha S_0 \quad S_1 = S_2 = (1 - \alpha) S_0$$

$$K = \frac{\alpha S_0}{(1 - \alpha)^2 S_0^2} = \frac{\alpha}{(1 - \alpha)^2} \frac{1}{S_0}$$

Remembering  $\Delta H_{vH} = RT^2 \frac{\partial \ln K}{\partial T}$

$$\Delta H_{vH} = RT^2 \frac{\partial}{\partial T} [\ln \alpha - 2 \ln(1 - \alpha) - \ln S_0]$$

$$\Delta H_{vH} = RT^2 \left[ \frac{1}{\alpha} + \frac{2}{1 - \alpha} \right] \frac{\partial \alpha}{\partial T} = RT^2 \frac{\alpha + 2(1 - \alpha)}{\alpha(1 - \alpha)} \frac{\partial \alpha}{\partial T} = RT^2 \frac{1 - \alpha}{\alpha(1 - \alpha)} \frac{\partial \alpha}{\partial T}$$

At the  $T_m$ ,  $\alpha = 1/2$ , such that

$$\Delta H_{vH} = RT_m^2 \frac{1 - 1/2}{(1/2)(1/2)} \frac{\partial \alpha}{\partial T} \bigg|_{T_m} = 6RT_m^2 \frac{\partial \alpha}{\partial T} \bigg|_{T_m} = 6R \frac{\partial \alpha}{\partial T} \bigg|_{T_m}$$

Plot  $\alpha$  vs.  $1/T$  and plot according to the above equation. This yields  $\Delta H_{vH}$ .

### Thermodynamics - Two self-complementary strands

as above, but now  $S_0$  = total conc of ss DNA

$$S_1 + S_2 = S_0 \quad S_1 = S_2 = \frac{1}{2}(1 - \alpha) S_0 \quad D = \frac{1}{2} \alpha S_0$$

$$K = \frac{\frac{1}{2} \alpha S_0}{\frac{1}{2} \alpha^2 (1 - \alpha)^2 S_0^2} = \frac{\alpha}{(1 - \alpha)^2} \frac{1}{2 S_0}$$

as before:

$$\Delta H_{vH} = RT^2 \frac{\partial}{\partial T} [\ln \alpha - 2 \ln(1 - \alpha) - \ln S_0 - \ln 2]$$

$$\Delta H_{vH} = RT^2 \left[ \frac{1}{\alpha} + \frac{2}{1 - \alpha} \right] \frac{\partial \alpha}{\partial T} = RT^2 \frac{\alpha + 2(1 - \alpha)}{\alpha(1 - \alpha)} \frac{\partial \alpha}{\partial T} = RT^2 \frac{1 - \alpha}{\alpha(1 - \alpha)} \frac{\partial \alpha}{\partial T}$$

Exactly as before.

### Generalities

Given that  $\Delta C_p = 0$ , we can very simply extrapolate from a set of data

The temperature dependence of the equilibrium constant is given by:  $\ln K = \frac{\Delta H^\circ}{R} \frac{1}{T} - \frac{1}{T_m}$

And from this we can readily get to  $\Delta G^\circ$  and  $\Delta S^\circ$ .

**Breslauer et al.**

From Breslauer et al., Proc. Natl. Acad. Sci. U.S.A. 83, 3746-3750 (1986).

All data for DNA in 1 M NaCl, pH 7, 25°C

Dinucleotide Step		$\Delta H^\circ$	$\Delta S^\circ$	$\Delta G^\circ$
AA/TT	TT/AA	9.1	24.0	1.9
AT/TA	-	8.6	23.9	1.5
TA/AT	-	6.0	16.9	0.9
CA/GT	TG/AC	5.8	12.9	1.9
GT/CA	AC/TG	6.5	17.3	1.3
CT/GA	AG/TC	7.8	20.8	1.6
GA/CT	TC/AG	5.6	13.5	1.6
CG/GC	-	11.9	27.8	3.6
GC/CG	-	11.1	26.7	3.1
GG/CC	CC/GG	11.0	26.6	3.1

 $\Delta H^\circ$  and  $\Delta G^\circ$  in kcal/mole;  $\Delta S^\circ$  in cal/K/moleFor non-self-complementary, but complementary duplexes:  $\Delta H_{\text{total}} = \approx 0 + \sum \Delta h_x$ For non-self-complementary, but complementary duplexes:  $\Delta G_{\text{total}} = -6 \text{ kcal/mole}^* + \sum \Delta g_x$ 

\*The free energy of duplex nucleation is estimated at -6 kcal/mole for non-self-complementary duplexes and -5 kcal/mole for self-complementary duplexes. Empirically determined?

Example: Duplex: CACTATA

$$\Delta H_{\text{total}} = 0 + 5.8 + 6.5 + 7.8 + 6.0 + 8.6 + 6.0 = 40.7 \text{ kcal / mole}$$

$$\Delta G_{\text{total}} = -6 + 1.9 + 1.3 + 1.6 + 0.9 + 1.5 + 0.9 = 2.1 \text{ kcal / mole}$$

By this accounting,  $\Delta S^\circ$  should be  $(\Delta H^\circ - \Delta G^\circ)/(298 \text{ K}) = 129.5 \text{ cal / K / mole}$ 

$$\sum \Delta s_x = \Delta s_{\text{init}} + 12.9 + 17.3 + 20.8 + 16.9 + 23.9 + 16.9 = \Delta s_{\text{init}} + 35.1 \text{ cal / K / mole}$$

Therefore  $\Delta s_{\text{init}} = 94.4 \text{ cal / K / mole}$ Also, this predicts:  $T_m = \Delta H/\Delta S = 314 \text{ K} (41^\circ\text{C})$ Can we use this to predict a  $T_m$ ?For equal concentrations of **self-complementary** strands:

$$S_1 + S_2 = S_o \quad S_1 = S_2 = \frac{1}{2}(1 - \alpha)S_o \quad D = \frac{1}{2}\alpha S_o \quad K = \frac{\alpha}{(1 - \alpha)^2} \frac{1}{2S_o}$$

At the midpoint of the melting transition,  $\alpha=1/2$

$$K_{T_m} = \frac{1/2}{(1 - 1/2)^2} \frac{1}{2S_o} = \frac{1}{S_o}$$

$$\alpha RT_m \ln K_{T_m} = \alpha RT_m \ln \frac{\alpha}{(1 - \alpha)^2} \frac{1}{2S_o} = \alpha H - \alpha T_m \alpha S \quad T_m = \frac{\alpha H}{\alpha S - \alpha R \ln \frac{\alpha}{(1 - \alpha)^2} \frac{1}{2S_o}}$$

For equal concentrations of **non-self-complementary, but complementary** strands:

$$D = \alpha S_o \quad S_1 = S_2 = (1 - \alpha)S_o \quad K = \frac{\alpha S_o}{(1 - \alpha)^2 S_o^2} = \frac{\alpha}{(1 - \alpha)^2} \frac{1}{S_o}$$

At the midpoint of the melting transition,  $\alpha=1/2$

$$K_{T_m} = \frac{1/2}{(1 - 1/2)^2} \frac{1}{S_o} = \frac{2}{S_o}$$

$$\Delta RT_m \ln K_{T_m} = \Delta RT_m \ln \frac{2}{S_0} = \Delta H - T_m \Delta S \quad T_m = \frac{\Delta H}{\Delta S - R \ln \frac{2}{S_0}}$$

$$T_m = \frac{\Delta H}{\Delta S - R \ln \frac{2}{S_0}} = \frac{40700 \text{ cal/mol}}{(129.5 \text{ cal/K/mol}) - (1.987 \text{ cal/K/mol}) \ln \frac{2}{S_0}}$$

$$\text{Assume } [S_0] = 20 \text{ nM, then } T_m = \frac{40700 \text{ cal/mol}}{(129.5 \text{ cal/K/mol}) - (1.987 \text{ cal/K/mol}) \ln \frac{2}{10^{-8}}}$$

**However**, remember our error analysis (**This needs to be corrected slightly** - see above):

$$T_m = \frac{\Delta H}{\Delta S} = \frac{T \Delta H}{\Delta H - \Delta G}$$

$$\Delta T_m^2 = \left| \frac{T}{\Delta H - \Delta G} - \frac{T \Delta H}{(\Delta H - \Delta G)^2} \right|^2 \Delta H^2 + \left| \frac{\Delta T \Delta H}{(\Delta H - \Delta G)^2} \right|^2 \Delta G^2$$

$$\Delta T_m^2 = \left| \frac{\Delta T \Delta G}{(\Delta H - \Delta G)^2} \right|^2 \Delta H^2 + \left| \frac{\Delta T \Delta H}{(\Delta H - \Delta G)^2} \right|^2 \Delta G^2$$

where  $\Delta \Delta H$  and  $\Delta \Delta G$  refer to the errors in our estimations of  $\Delta H$  and  $\Delta G$ , respectively. Assume that they are:  $\Delta \Delta H = 2 \text{ kcal/mole}$ ,  $\Delta \Delta G = 1 \text{ kcal/mole}$

Then

$$\Delta T_m = \left| \frac{(298)2.1}{(38.6)^2} \right| 2 + \left| \frac{(298)40.7}{(38.6)^2} \right| 1 = 0.84 + 8.1 = 9.0 \text{ K}$$