

MCB100A/Chem130
MidTerm Exam 2
April 4, 2013

Name _____ Student ID _____

True/False (2 points each).

1. T The Boltzmann constant, k_bT sets the energy scale for observing energy microstates
2. T Atoms with favorable electronic configurations gain stability by forming covalent bonds
3. F Conformational entropy favors the folded state over the unfolded state
4. F QM potential energies are tractable for calculating the heat capacity of a protein
5. F The highest potential energy is the most probable energy state at constant NVT
6. T Statistical entropy can be applied to isothermal ideal gas expansion and ideal DNA pulling
7. F The natural log of multiplicity, $\ln(W)$, is not extensive
8. F The work done in an equilibrium process is less than work done in a non-equilibrium process
9. T Chemical potential for ideal gas is only dependent on entropy changes with molecule numbers
10. T The heat capacity at constant pressure is greater than the heat capacity at constant temperature
11. T C_v is inversely proportional to the variance of the Gaussian distribution for energy at constant T
12. F The units of entropy are kJ/mol
13. T The Helmholtz free energy is available energy to do mechanical or chemical work
14. T The chemical potential, μ , is the tendency of system to realize concentration changes.
15. F A reaction will go forward if the reaction quotient, Q is greater than equilibrium constant K
16. T Henderson-Hasselbach relates the pH to the acid dissociation constant K_a
17. F The standard enthalpy and standard entropy are always independent of temperature, pressure
18. F The chemical potential is related to the mechanical (expansion) work of Gibbs free energy
19. F If protein A has a larger partition function than protein B then B has the higher heat capacity
20. T Temperature signifies how multiplicity of bath grows with energy

Multiple Choice (5 points each).

21. Molecular interactions are classified as long-ranged when the power law exponent of r^{-n} is

- (a) $n < 3$ (b) $n = 3$
(c) $n < 2$ (d) $n > 4$

22. What is the value of kT

- (a) ~ 2.5 kJ/mol (b) ~ 210 cm $^{-1}$
(c) ~ 0.6 kcal/mol (d) all of the above

23. What is the probability of observing system with energy E , at constant N, V, T

- (a) $\exp(-\beta E,^2)/Q$ (b) $1/Q$
(c) $\exp(-\beta E,)/Q$ (d) all of the above

24. If a covalent bond vibrational excitation is ~ 25 kJ/mole, is it significantly populated at 298K?

- (a) yes (b) no
(c) maybe (d) all of the above

25. For an isolated system (constant N, V, E) all energy microstates

- (a) have different Boltzmann probabilities (b) are equally probable
(c) have different multiplicities (d) all of the above

26. Temperature signifies how multiplicity of bath (surroundings) grows with energy is embodied in

- (a) $(dE/dS)_{N,V}$ (b) $(dW/dE)_{N,V}$
(c) $(dE/dT)_{N,V}$ (d) all of the above

27. The condition for the multiplicity to be at an extremum or maximum for large N is

- (a) $dW/dN=0$ (b) $d(\ln W)/dN=0$
(c) both (a) and (b) (d) none of the above

28. Which state function(s) predict spontaneous change?

- (a) $dS > 0$ (b) $dA < 0$
(c) $dG < 0$ (d) all of the above

29. Direction of spontaneous change is when dN particles move from regions of

- (a) low chemical potential to high (b) equal chemical potential
(c) high chemical potential to low (d) all of the above

30. The extent of a chemical reaction with large and negative values of ΔG° correspond to

- (a) $K_{eq} \gg 1$ (b) $K_{eq} \sim 1$
(c) $K_{eq} \ll 1$ (d) none of the above

Short Problems (15 points each)

31. The basis of proton NMR is that the hydrogen atom has a magnetic moment, so that in a magnetic field it can populate two states: spin up ($S\uparrow$) and spin down ($S\downarrow$). The energy difference in a 300 MHz NMR spectrometer is 3.5×10^{-2} kJ/mol. For $N=100$ hydrogen atoms and $T = 300$ K

(a) Compute the relative population difference $|N_{S\uparrow} - N_{S\downarrow}| / (N_{S\uparrow} + N_{S\downarrow})$

$$Q = e^{0/kT} + e^{-0.035/2.5} = 1 + 0.9861$$

$$P(N_{S\uparrow}) = \frac{e^0}{1.9861} \quad P(N_{S\downarrow}) = \frac{e^{-0.035/2.5}}{1.9861}$$

$$N_{S\uparrow} = 50.35 \quad N_{S\downarrow} = 49.65$$

$$|N_{S\uparrow} - N_{S\downarrow}| / (N_{S\uparrow} + N_{S\downarrow}) = 0.007$$

(b) How does the population difference change with increasing temperature?

It gets smaller

32. Identify all relevant energetic interactions and their functional forms found for the following amino acid constituents:

(a) Intermolecular interaction between CH_4 and CH_4 (model for alanine-alanine)

No net charge, no other permanent poles, and unfavorable electron pairing (same electronic configuration)

$$E_{total} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

(b) Intermolecular interaction between NH_3^+ and COO^- (model for zwitter ionic termini)

Net charge, no other permanent poles, and unfavorable electron pairing (same electronic configuration)

$$E_{total} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}}$$

Long Problems (25 points each)

33. *E. coli* DNA polymerase introduces about 1 incorrect base in 10^4 internucleotide linkages during replication “*in vitro*”. However, “*in vivo*” the proofreading function of *E. coli* DNA polymerase is much better, making 1 mistake in 10^8 nucleotide polymerization events.

(a) If an average *E. coli* gene is ~ 100 bases long, what is the probability of introducing an error per gene *in vitro*?

$$p(\text{in vitro}) = 10^{-2} \text{ mutations per gene}$$

(b) What is the probability of introducing an error per gene *in vivo*?

$$p(\text{in vivo}) = 10^{-6} \text{ mutations per gene}$$

(c) Given the large number of nucleotide polymerization events, N , in *E. Coli*, we will assume the number of errors, Err , is distributed according to a Gaussian distribution

$$P(Err) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(Err - \langle Err \rangle)^2 / 2\sigma^2};$$

where $\langle Err \rangle$ is the mean error and σ is the standard deviation. What is $\langle Err \rangle$ and σ *in vivo* for 10^8 gene polymerization events? What is $\langle Err \rangle$ and σ *in vitro* for 10^8 nucleotide polymerization events?

$$\langle Err \rangle_{\text{in vivo}} = p \times N = 10^{-6} \times 10^8 = 10^2$$

$$\langle Err \rangle_{\text{in vitro}} = p \times N = 10^{-2} \times 10^8 = 10^6$$

$$\sigma_{\text{in vivo}} = [p \times (1-p) \times N]^{1/2} = [10^{-6} \times (\sim 1) \times 10^8]^{1/2} = 10$$

$$\sigma_{\text{in vitro}} = [10^{-2} \times \sim 1 \times 10^8]^{1/2} = \sim 1000$$

(d) What is the probability that the number of errors will be 1 standard deviation larger than the mean *in vivo*?

$$P(110) = \frac{1}{10\sqrt{2\pi}} e^{-(110-100)^2 / 2(10)^2} = \frac{0.606}{25.1} = 0.024$$

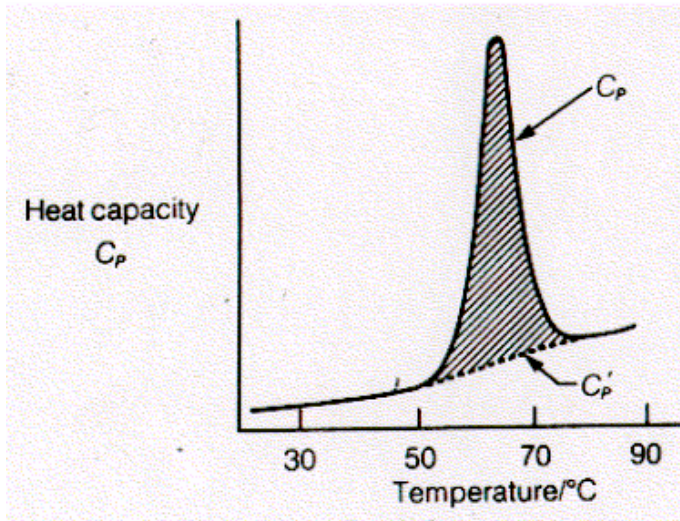
(e) What is the probability that the number of errors *in vivo* will be 5000 times smaller than the average number of errors *in vitro*?

$$Err = 1,000,000 / 5000 = 200$$

$$P(200) = \frac{1}{10\sqrt{2\pi}} e^{-(200-100)^2 / 2(10)^2} = \frac{1.9 \times 10^{-21}}{25.1} = 7.7 \times 10^{-24}$$

34. The variation in the constant pressure heat capacity, C_p , of a protein as a function of temperature is known as the melting curve, as it measures the relative populations of the folded and unfolded proteins in water. We would like to use C_p measurements made along this curve to determine a single equation that gives the standard Gibbs free energy for unfolding, $\Delta G_{unfold}^0(T)$, at any temperature.

(a) Draw a typical heat capacity curve for a protein in water from 25C to 90C. Name 3 intramolecular energy interactions that break and therefore new energy levels that become populated as the protein goes from the folded to the unfolded state.



hydrogen bonds, van der Waals interactions (dispersion), salt bridges

(b) Express $\Delta G_{unfold}^0(T)$ in terms of $\Delta H_{unfold}^0(T)$ and $\Delta S_{unfold}^0(T)$

$$\Delta G_{unfold}^0(T) = \Delta H_{unfold}^0(T) - T\Delta S_{unfold}^0(T)$$

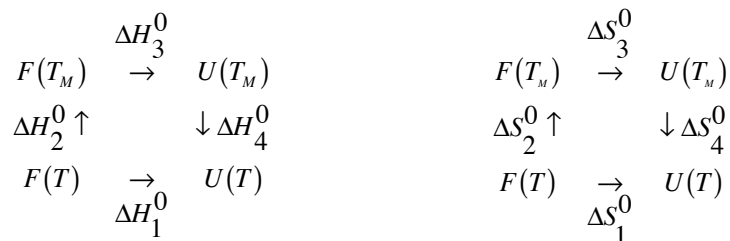
(c) Express $\Delta H_{unfold}^0(T)$ in terms of C_p^{unfold} ; express $\Delta H_{fold}^0(T)$ in terms of C_p^{fold}

$$\Delta H_{unfold}^0(T) = \int_{T_1}^{T_2} C_p^{unfold} dT \qquad \Delta H_{fold}^0(T) = \int_{T_1}^{T_2} C_p^{fold} dT$$

(d) Express $\Delta S_{unfold}^0(T)$ in terms of C_p^{unfold} ; express $\Delta S_{fold}^0(T)$ in terms of C_p^{fold}

$$\Delta S_{unfold}^0(T) = \int_{T_1}^{T_2} \frac{C_p^{unfold}}{T} dT \qquad \Delta S_{fold}^0(T) = \int_{T_1}^{T_2} \frac{C_p^{fold}}{T} dT$$

(continued) You are given the following thermodynamic cycles, where F corresponds to folded and U to unfolded



(e) Which leg of the following thermodynamic cycle allows us to determine $\Delta H_{unfold}^0(T)$ and $\Delta S_{unfold}^0(T)$ and how would I determine it from the other three legs of the cycle?



$$\Delta H_{unfold}^{\circ}(T) = \Delta H_2^{\circ}(T) + \Delta H_3^{\circ}(T) + \Delta H_4^{\circ}(T) \qquad \Delta S_{unfold}^{\circ}(T) = \Delta S_2^{\circ}(T) + \Delta S_3^{\circ}(T) + \Delta S_4^{\circ}(T)$$

(f) Which leg of the thermodynamic cycle measures $\Delta H_{unfold}^0(T_M)$ and $\Delta S_{unfold}^0(T_M)$? How would you get these quantities from experiment and what equations would you use?

$$F(T) \xrightarrow{\Delta H_3^0} U(T) \qquad F(T) \xrightarrow{\Delta S_3^0} U(T)$$

$$\Delta H_3^{\circ}(T_m) = \int_{T=T_M-\Delta}^{T=T_M+\Delta} C_P^{unfold} dT \qquad \text{area under peak of melting curve}$$

$$\Delta G_3^{\circ}(T_m) = \Delta H_3^{\circ}(T_m) - T \Delta S_3^{\circ}(T_m) = 0 \qquad \rightarrow \qquad \Delta S_3^{\circ}(T_m) = \frac{\Delta H_3^{\circ}(T_m)}{T}$$

(g) What are the expressions for ΔH_n^0 and ΔS_n^0 for the other two legs of the cycle?

$$\Delta H_2^{\circ}(T) = \int_T^{T_M} C_P^{fold} dT = C_P^{fold} (T_M - T) \qquad \Delta H_4^{\circ}(T) = \int_{T_M}^T C_P^{unfold} dT = C_P^{unfold} (T - T_M)$$

$$\Delta S_2^{\circ}(T) = \int_T^{T_M} \frac{C_P^{fold}}{T} dT = C_P^{fold} \ln \frac{T_M}{T} \qquad \Delta S_4^{\circ}(T) = \int_{T_M}^T \frac{C_P^{unfold}}{T} dT = C_P^{unfold} \ln \frac{T}{T_M}$$

(h) Write the expressions for $\Delta G_{unfold}^0(T)$ and show your work

$$\Delta H_{unfold}^{\circ}(T) = C_P^{fold}(T_M - T) + \Delta H_3^{\circ}(T_M) + C_P^{unfold}(T - T_M) \quad \Delta S_{unfold}^{\circ}(T) = C_P^{fold} \ln \frac{T_M}{T} + \Delta S_3^{\circ}(T_M) + C_P^{unfold} \ln \frac{T}{T_M}$$

$$\Delta G_{unfold}^{\circ}(T) = C_P^{fold}(T_M - T) + \Delta H_3^{\circ}(T_M) + C_P^{unfold}(T - T_M) - T \left(C_P^{fold} \ln \frac{T_M}{T} + \Delta S_3^{\circ}(T_M) + C_P^{unfold} \ln \frac{T}{T_M} \right)$$

$$\Delta G_{unfold}^{\circ}(T) = \Delta H_{unfold}^{\circ}(T_M) + \Delta C_P(T - T_M) - T \left(\Delta S_3^{\circ}(T_M) + \Delta C_P \ln \frac{T}{T_M} \right)$$

34. For the folding of a 100 residue protein, containing 6 glycines and 4 prolines, at 300K:

- (i) When unfolded, each residue except glycines and prolines can take 3 conformations of equal energy; glycine can take 4 conformations; proline can take 2 conformations.
- (ii) The conformation of the folded state has every residue in 1 possible conformation except glycine which has 2 conformations.
- (iii) A His-Asp ion pair, whose interaction energy is -350kJ/mol in vacuum, is found in the interior of a folded protein with dielectric constant $\epsilon_p=4$. The ion pair remains intact when unfolded in water whose dielectric constant is $\epsilon_w=80$.
- (iv) Assume water molecules have 6 possible configurations when surrounded by other waters, protein backbone or polar sidechains, otherwise they have only 2 possible configurations.
- (v) Every hydrophobic sidechain in unfolded state interacts with 2 water molecules, and all hydrophobic sidechains are buried in folded state.
- (v) Assume water-protein interaction energy is zero.

(a) Calculate the Helmholtz free energy for the protein-protein interactions only ($\Delta A_{\text{protein}}$). Show your work. Does $\Delta A_{\text{protein}}$ favor folding?

$$\Delta U_{\text{protein}} = \frac{-350}{4} - \left(\frac{-350}{80} \right) = -83.125 \text{ kJ/mol}$$

$$-T\Delta S_{\text{protein}} = -2.5 \text{ kJ/mol} (\ln 1^{94} 2^6 - \ln 3^{90} 4^6 2^4) = -2.5 \text{ kJ/mol} (4.159 - 109.97) = 264.528 \text{ kJ/mol}$$

$$\Delta A_{\text{protein}} = \Delta U_{\text{protein}} - T\Delta S_{\text{protein}} = -83.125 \text{ kJ/mol} + 264.528 \text{ kJ/mol} = 181.403 \text{ kJ/mol} \quad \text{no does not favor folding}$$

(b) Calculate the free energy for water, non-hydrophobic interactions ($\Delta A_{\text{non-hphobe, H}_2\text{O}}$). Show your work. Does $\Delta A_{\text{non-hphobe, H}_2\text{O}}$ favor folding over nonfolding?

$$\Delta U_{\text{non-hphobe, H}_2\text{O}} = 0$$

$$-T\Delta S_{\text{non-hphobe, H}_2\text{O}} = -2.5 \text{ kJ/mol} (\ln 6^{\# \text{nonhphobe-water}} - \ln 6^{\# \text{nonhphobe-water}}) = 0$$

$$\Delta A_{\text{non-hphobe, H}_2\text{O}} = A_{\text{fold}} - A_{\text{unfold}} = \Delta U_{\text{non-hphobe, H}_2\text{O}} - T\Delta S_{\text{non-hphobe, H}_2\text{O}} = 0 \quad \text{no does not favor folding}$$

(c) Calculate the free energy for water, hydrophobic interactions (ΔA will be per hydrophobic residue). Show your work. What is minimum number of hydrophobic sidechains required to ensure that folding in water is spontaneous at room temperature (300 K)?

$$\Delta U_{\text{hphobe, H}_2\text{O}} = 0$$

$$-T\Delta S_{\text{hphobe, H}_2\text{O}} = -2.5 \text{ kJ/mol} (\ln 6^2 - \ln 2^2) = -5.493 \text{ kJ/mol/\#hphobe}$$

$$n_{\text{hphobe}} \Delta A_{\text{hphobe, H}_2\text{O}} = A_{\text{fold}} - A_{\text{unfold}} = \Delta U_{\text{hphobe, H}_2\text{O}} - T\Delta S_{\text{hphobe, H}_2\text{O}} = -5.493 n_{\text{hphobe}} \text{ kJ/mol}$$

$$\Delta A_{\text{protein}} + \Delta A_{\text{non-hphobe, H}_2\text{O}} + n_{\text{hphobe}} \Delta A_{\text{hphobe, H}_2\text{O}} = 0 \quad 181.403 \text{ kJ/mol} = 5.493 n_{\text{hphobe}} \text{ kJ/mol} \quad n_{\text{hphobe}} = 33$$

(Extra Credit). To find the probability, p_v , that both minimizes the Gibbs free energy and maximizes the entropy, we start with the following equation

$$L = -k_b \sum_v p_v \ln p_v + \alpha \sum_v p_v E_v + \lambda \sum_v p_v V_v + \gamma \left(\sum_v p_v - 1 \right)$$

(a) What is the first term on the right-hand side of the equation?

The probabilistic version of entropy

(b) What is the meaning of the second term on the right-hand side of the equation?

The system exchanges energy with the bath (energy fluctuates)

(c) What is the meaning of the third term on the right-hand side of the equation?

Volume fluctuations

(d) What is the meaning of the fourth term on the right-hand side of the equation?

The probability summed over all microstates adds up to 1

(e) What is the meaning of L (the Lagrangian)?

It is the maximization of the entropy (2nd Law) under the constraints that energy and volume fluctuate

(f) How do I determine the probability, p_v ?

$$\frac{dL}{dp_v} = 0; \quad \frac{dL}{dp_v} = -k_b (1 + \ln p_v) + \alpha E_v + \lambda V_v + \gamma = 0 \quad p_v = \exp \left(\frac{\alpha E_v + \lambda V_v + \gamma - k_b}{k_b} \right)$$

(g) Can you solve for the α Lagrange multiplier?

$$S = -k_b \sum_v p_v \ln p_v \quad -k_b \ln p_v = -(\alpha E_v + \lambda V_v + \gamma - k_b)$$

$$S = -\sum_v p_v (\alpha E_v + \lambda V_v + \gamma - k_b) \quad S = -\alpha \langle E \rangle - \lambda \langle V \rangle + (\gamma - k_b) \sum_v p_v$$

$$\left(\frac{\partial S}{\partial \langle E \rangle} \right)_{N, \langle V \rangle} = \frac{1}{T} = -\alpha$$