

Physics 177: Midterm Exam Solutions

Jeffrey Moffitt

March 9, 2008

■ Problem 1: Polyelectrolytes

■ Part a

In class, we derived the condition $\xi \leq \frac{1}{|Z_i Z_p|}$ for ions to be in thermal equilibrium around the charged polymer. Recall that $\xi = \frac{e^2}{\epsilon k_B T b}$ (where $b = \frac{l}{N}$ is the distance between charged groups on the polymer), Z_i is the valency of the polymer, and Z_p is valency of each of the charges on the polymer. Z_p can be absorbed into ξ by simply defining a value of b that includes this parameter, i.e. if $Z_p = 2$, then we can simply imagine a polymer with twice as many charges per unit length and set $Z_p = 1$. Thus, the value of Z_p is arbitrary and can be set to 1.

Thus, for a divalent salt, $Z_i = 2$ and $\xi \leq 1/2$.

■ Part b

If $\xi = \frac{e^2}{\epsilon k_B T b} \leq \frac{1}{2}$, then the distance between charges after condensation must be $b = 2 \frac{e^2}{\epsilon k_B T}$. Plugging in numbers

$$b = 2 \frac{(1.6 \times 10^{-19} \text{ C})^2}{4\pi(78.5)(8.85 \times 10^{-12} \frac{\text{C}^2}{\text{Jm}})(1.38 \times 10^{-23} \text{ J/K})(298 \text{ K})} = 1.43 \times 10^{-9} \text{ m} = 1.43 \text{ nm}$$

Thus, the final charge density after condensation is one negative charge per 1.43 nm, or 0.70 e/nm. Contrast this to the original value for DNA, 2 charges per base pair, i.e. 1 charge per 1.7 Å

■ Part c

Recall that the Poisson-Boltzmann equation is

$$\nabla^2 \psi(r) = -\frac{4\pi\rho(r)}{\epsilon}$$

$$\rho(r) = Z_+ e n_+(r) + Z_- e n_-(r)$$

where $\rho(r)$ is the charge density at a distance r , $Z_{+/-}$ is the valency of the positive and negative ions, e is the charge of an electron, and $n_{+/-}(r)$ are the number densities of the positive and negative ions.

The Boltzmann equation tells us that the number density of the positive ion at a distance r is

$$n_+(r) = c_+ e^{-\frac{E_+(r)}{k_B T}}$$

where c_+ is the original concentration of the positive ion and $E_+(r) = Z_+ e \psi(r)$. Similarly for n_- . Thus,

$$\rho(r) = Z_+ e c_+ e^{-\frac{Z_+ e \psi(r)}{k_B T}} + Z_- e c_- e^{-\frac{Z_- e \psi(r)}{k_B T}}$$

The linear-Poisson-Boltzmann equation assumes that $e \psi \ll 1$, allowing us to expand the exponentials, yielding

$$\rho(r) \approx Z_+ e c_+ \left(1 - \frac{Z_+ e \psi(r)}{k_B T}\right) + Z_- e c_- \left(1 - \frac{Z_- e \psi(r)}{k_B T}\right)$$

Since charges are balanced in the original salt, the concentration of each ion $c_{+/-}$ must satisfy the equation, $Z_+ c_+ = -Z_- c_-$. Thus,

$$\rho(r) \approx \frac{Z_+ e^2 c_+}{k_B T} (-Z_+ + Z_-) \psi(r)$$

Thus, the linearized-Poisson-Boltzmann equation is

$$\nabla^2 \psi(r) = -\frac{4\pi}{\epsilon} \frac{Z_+ e^2 c_+}{k_B T} (-Z_+ + Z_-) \psi(r)$$

This in the general form

$$\nabla^2 \psi(r) = \kappa^2 \psi(r)$$

Since the left-hand-side of this equation has a second derivative with respect to space, the term κ^2 can be used to scale the distance, i.e. $x = \kappa r$. Thus, κ is the inverse of the Debye length. Comparing this expression to the above expression it is clear that

$$\kappa^2 = \left(\frac{4\pi e^2 Z_+ (Z_+ + |Z_-|) c_+}{\epsilon k_B T} \right)$$

For a monovalent salt like NaCl, $Z_+ = 1$, $Z_- = -1$, $c_+ = n_s$ and κ^2 reduces to the expression derived in class.

■ Problem 2: Kinked Worm-Like Chain

■ Part a

Recall that for a worm-like chain

$$\langle R^2 \rangle = \int_0^L \int_0^L \langle \vec{r}(s) \vec{r}(s') \rangle ds ds'$$

Since there is permanent bend or kink at a distance l from one end, we can divide this double integral into three regions

- 1: Both $\vec{r}(s)$ and $\vec{r}(s')$ are before the bend
- 2: Both $\vec{r}(s)$ and $\vec{r}(s')$ are after the bend
- 3: $\vec{r}(s)$ is before the bend and $\vec{r}(s')$ is after it (and vice versa)

Mathematically this can be expressed by breaking our integral up into three separate integrals over different bounds.

$$\int_0^L \int_0^L \langle \vec{t}(s) \vec{t}(s') \rangle ds ds' = \int_0^l \int_0^l \langle \vec{t}(s) \vec{t}(s') \rangle ds ds' + \int_l^L \int_l^L \langle \vec{t}(s) \vec{t}(s') \rangle ds ds' + 2 \int_0^l \int_l^L \langle \vec{t}(s) \vec{t}(s') \rangle ds ds'$$

where each corresponds to each of the cases above and the factor of two in the last term comes from the fact that there are two possible choices for which vector is in which region.

The first and second integrals were done in class and in homework 2:

$$\int_0^l \int_0^l \langle \vec{t}(s) \vec{t}(s') \rangle ds ds' = 2 P l \left(1 - \frac{P}{l} (1 - e^{-l/P}) \right)$$

$$\int_l^L \int_l^L \langle \vec{t}(s) \vec{t}(s') \rangle ds ds' = 2 P (L - l) \left(1 - \frac{P}{(L-l)} (1 - e^{-(L-l)/P}) \right)$$

To evaluate the last integral we need to define two reference vectors, $\vec{t}_-(l)$ and $\vec{t}_+(l)$, where these are the tangent vectors on either side of the bend. Thus, $\langle \vec{t}_-(l) \vec{t}_+(l) \rangle = \cos(\pi - \theta) = -\cos(\theta)$. Note that there is a negative sign here because the vectors both point towards increasing s ; thus, the angle between them is actually the complement to θ , $\pi - \theta$. The dot product in each region of the last integral can now be referenced to these two vectors

$$\begin{aligned} 2 \int_0^l \int_l^L \langle \vec{t}(s) \vec{t}(s') \rangle ds ds' &= 2 \int_0^l \int_l^L \langle \vec{t}(s) \vec{t}_-(l) \rangle \langle \vec{t}_-(l) \vec{t}_+(l) \rangle \langle \vec{t}_+(l) \vec{t}(s') \rangle ds ds' \\ &= -2 \cos(\theta) \int_0^l \int_l^L e^{-(l-s)/P} e^{-(L-l-s)/P} ds ds' \\ &= -2 P^2 \cos(\theta) (1 - e^{-l/P}) (1 - e^{-(L-l)/P}) \end{aligned}$$

Thus, the full expression is

$$\langle R^2 \rangle = 2 P l \left(1 - \frac{P}{l} (1 - e^{-l/P}) \right) + 2 P (L - l) \left(1 - \frac{P}{(L-l)} (1 - e^{-(L-l)/P}) \right) - 2 P^2 \cos(\theta) (1 - e^{-l/P}) (1 - e^{-(L-l)/P})$$

The sign on the additional term makes physical sense: when $\theta > \pi/2$ then $\cos(\theta) < 0$ and $\langle R^2 \rangle$ goes up, just as we would expect for a kink with $\theta > \pi/2$.

■ Part b

If $l \gg P$ and $(L - l) \gg P$, then all exponentials in the above expression ~ 1 , and the mean square end-to-end distance reduces to

$$\begin{aligned} \langle R^2 \rangle &= 2 P l + 2 P (L - l) - 2 P^2 \cos(\theta) \\ &= 2 P L - 2 P^2 \cos(\theta) \end{aligned}$$

Moreover, if $l \gg P$ and $(L - l) \gg P$, then $P L \gg P^2$, so

$$\langle R^2 \rangle \approx 2 P L$$

This is the expression for a normal worm-like chain, as expected since in the limit that the polymer is very long, local defects such as the bend at position l should not affect the chain statistics.

■ Problem 3: DNA Binding Proteins

■ Part a

The dissociation constant K_D is related to the free energy of dissociation via

$$\Delta G = -RT \ln K_D$$

where R is the gas constant, i.e. $k_B N_A$. For the linear piece of DNA the ΔG_l of binding is

$$\Delta G_l = (-1.38 \times 10^{-23} \text{ J/K}) (6.022 \times 10^{23}) (298 \text{ K}) \ln(1.78 \times 10^{-11}) = 61.3 \text{ kJ/mol} = 14.7 \text{ kcal/mol}$$

and for the circular DNA

$$\Delta G_c = (-1.38 \times 10^{-23} \text{ J/K}) (6.022 \times 10^{23}) (298 \text{ K}) \ln(3.68 \times 10^{-12}) = 65.2 \text{ kJ/mol} = 15.6 \text{ kcal/mol}$$

$\Delta G_c > \Delta G_l$ because we are considering the dissociation reaction. The ΔG for binding is the negative of these two values. Thus, binding to the circular DNA is more spontaneous than binding to the linear DNA.

■ Part b

The difference between the linear DNA and the circular DNA is that circular DNA has been pre-bent. Bending DNA requires energy, so if the DNA binding protein bends the DNA as part of its binding reaction, then prebinding the DNA will change the dissociation constant.

■ Part c

If the DNA binding protein bends the DNA, the free energy for binding can be written as

$$\Delta G_{\text{bind}} = \Delta G_{\text{inter}} - \Delta G_{\text{bend}}$$

where ΔG_{inter} is the free energy associated with binding to the DNA, this includes all the interaction terms and the entropic terms associated with the protein, and ΔG_{bend} is the free energy associated with bending the DNA from its initial state to its final state.

Moreover, $\Delta G_{\text{bind}} = -\Delta G_{\text{dissociation}}$ from above.

Thus, the difference in the ΔG values from above, $\Delta\Delta G$ is

$$\Delta\Delta G = \Delta G_c - \Delta G_l = (\Delta G_{\text{inter}} - \Delta G_{\text{bend},c}) - (\Delta G_{\text{inter}} - \Delta G_{\text{bend},l}) = \Delta G_{\text{bend},l} - \Delta G_{\text{bend},c}$$

Now, $\Delta G_{\text{bend},l}$ is the free energy necessary to bend the DNA from the linear form to its final conformation, whereas $\Delta G_{\text{bend},c}$ is the free energy necessary to bend the DNA from its bent state in the mini-circle to its final conformation. Thus, the difference in these two values is equal to the energy necessary to bend the DNA from its linear form to its bent conformation in the DNA circle.

A 200 bp DNA circle has a radius of

$$r = (200 \text{ bp}) (0.34 \text{ nm/bp}) / 2\pi = 10.8 \text{ nm}$$

Recall that the energy per unit length for a DNA with radius of curvature r is $E/L = \frac{1}{2} A \frac{1}{r^2}$, where r is the radius of curvature and A is the bending rigidity of DNA.

Since the protein only interacts with 20-bp of DNA, it is only these 20-bp of DNA that are distorted in the final bound form, and it is only the energy associated with binding these 20-bp that is important. Thus, the energy to bend this 20-bp region into a curve with radius of curvature, r is

$$E = \frac{1}{2} A l \frac{1}{r^2}$$

This free energy is the free energy difference calculated above, thus solving for the bending rigidity

$$A = 2 \Delta \Delta G \frac{r^2}{l} = 2 (65.2 \text{ kJ/mol} - 61.3 \text{ kJ/mol}) (1000 \text{ J/kJ}) \left(\frac{1 \text{ mol}}{6.022 \times 10^{23} \text{ molecules}} \right) \frac{(10.8 \text{ nm})^2}{(20 \text{ bp})(.34 \text{ nm/bp})} = 2.22 \times 10^{-19} \text{ J nm} = 32.0 \text{ kcal nm/mol}$$

To calculate the persistence length of DNA, recall that $P = \frac{A}{k_B T}$, thus the persistence length, P , is

$$P = \frac{A}{k_B T} = \frac{2.22 \times 10^{-19} \text{ J nm}}{(1.38 \times 10^{-23} \text{ J/K})(298 \text{ K})} = 54.0 \text{ nm}$$

■ Problem 4: Entropy and Enthalpy of an Alpha Helix

■ Part a

The melting point of a protein is defined as the temperature at which the 50% of the protein is unfolded. This implies that the equilibrium constant for the $F \leftrightarrow U$ reaction is 1 (where F stands for the folded state and U for the unfolded state). This in turn implies that the free energy of the system is zero.

Thus, since $\Delta G = \Delta H - T \Delta S$, this implies that

$$\Delta H = T \Delta S$$

Thus,

$$\Delta S = \Delta H / T = \frac{(170 \text{ kJ/mol})}{(323 \text{ K})} = 526 \text{ J/mol K}$$

■ Part b

If each amino acid has 3 additional degrees of freedom each with 2 possible values when the protein is unfolded, then the multiplicity of the unfolded state, Ω , is $2^3 = 8$, per degree of freedom, and the total multiplicity is $\Omega = 8^N$, where N is the number of amino acids.

From Boltzmann's equation, $\Delta S = k_B \ln(\Omega) = k_B \ln(8^N) = k_B N \ln(8)$.

Thus, solving for N

$$N = \frac{\Delta S}{k_B \ln 8} = 526 \frac{J}{\text{mol}} \frac{\text{mol}}{6.022 \times 10^{23}} \frac{1}{1.38 \times 10^{-23}} \frac{1}{\text{Log}[8]} = 30.4$$

Thus, there must be ~ 30 amino acids in the alpha helix.

■ Part c

Recall that the only chemical groups that hydrogen bond in an alpha helix (ignoring side chains) are the amino group and the carbonyl group of the peptide backbone. Moreover, the amino group hydrogen bonds with the carbonyl group 4 amino acids up the chain, i.e. residue i bonds with residue $i+4$. Thus, if there are N amino acids in an alpha helix, there are $N - 4$ hydrogen bonds in the backbone. The minus four comes from the last four amino groups that don't have a carbonyl group with which they can H-bond.

Thus, the enthalpy per hydrogen bond is

$$\frac{\Delta H}{N-3} = \frac{(170 \frac{\text{kJ}}{\text{mol}})}{(30-4)} = 6.5 \frac{\text{kJ}}{\text{mol}} \text{ per amino acid} = 1.08 \times 10^{-20} J = 2.6 k_B T$$

■ Problem 5: Protein Binding to a Charged Surface: I

■ Part a

To derive the electrostatic potential, recall that the Poisson-Boltzmann equation is

$$\nabla^2 \psi = -4 \pi \frac{\rho}{\epsilon}$$

with charge density, ρ ,

$$\rho = -n_s e (e^y - e^{-y})$$

where $y = e \psi / k_B T$.

Since the electrostatic energy is always less than the thermal energy, $y < 1$, and we can expand the exponential terms in ρ

$$\rho \approx -2 n_s \frac{e^2}{k_B T} \psi$$

Thus, the Poisson-Boltzmann equation becomes

$$\nabla^2 \psi = 8 \pi n_s \frac{e^2}{k_B T \epsilon} \psi = \kappa^2 \psi$$

where κ is the inverse of the Debye length.

Because of the symmetry of the surface, the only direction in which the potential will change is along the direction perpendicular to the surface. Let's call this the x direction, in which case this becomes a first order differential equation

$$\frac{d^2}{dx^2} \psi = \kappa^2 \psi$$

The solution to this differential equation is

$$\psi(x) = a e^{-\kappa x} + b e^{\kappa x}$$

where a and b are arbitrary parameters that must be determined from boundary conditions.

Our first boundary condition is that the potential must smoothly decay to zero as $x \rightarrow \infty$; thus, $b = 0$.

To determine A , we use Gauss' Law to determine the electric field right next to the surface. Recall that Gauss' Law states that

$$\oint E dA = \frac{4\pi}{\epsilon} Q_{\text{enc}}$$

where Q_{enc} is the charge enclosed within the Gaussian surface. We choose a box of surface area A and infinitesimal height, which because the surface charge density is σ , will enclose a total charge of σA . Note, we cannot use a box of finite height because we would then enclose some of the free ions which would change the value of Q_{enc} .

Thus,

$$E(x=0) A = 4\pi \frac{\sigma}{\epsilon} A \Rightarrow E(x=0) = 4\pi \sigma / \epsilon$$

Since $E = -\nabla \psi = \kappa A e^{-\kappa x}$, we can determine A by evaluating this expression at $x = 0$; thus,

$$\kappa a = \sigma / \epsilon \Rightarrow a = \frac{4\pi \sigma}{\epsilon \kappa}$$

Thus, the potential is

$$\psi(x) = \frac{4\pi \sigma}{\epsilon \kappa} e^{-\kappa x}$$

Alternative derivation: Many people did not linearize the Poisson-Boltzmann equation initially but instead solved the full equation. The solution of the full Poisson-Boltzmann equation can be found in Duane, page 343. It is

$$\psi(x) = \frac{2k_B T}{e} \ln\left(\frac{1 + \gamma e^{-\kappa x}}{1 - \gamma e^{-\kappa x}}\right)$$

where γ is determined from boundary conditions as above. Taking the derivative of this ψ yields

$$E(x=0) = -\nabla \psi |_{x=0} = \frac{4k_B T \kappa}{e} \frac{\gamma}{1-\gamma^2}$$

From Gauss' law as above, $E(0) = 4\pi \sigma / \epsilon$

Thus, γ satisfies the following relation

$$\frac{\gamma}{1-\gamma^2} = \frac{4\pi \sigma e}{4k_B T \epsilon \kappa} = \frac{1}{2} \frac{1}{A} \frac{1}{\kappa}$$

where

$$A = \frac{k_B T \epsilon}{2\pi\sigma e}$$

The solution to this equation is

$$\gamma = -\kappa A + \sqrt{1 + \kappa^2 A^2}$$

where we have picked the positive solution.

■ Part b

First we calculate the Debye length in 50 mM NaCl. Recall that

$$\kappa = \left(\frac{8\pi e^2}{k_B T \epsilon} c \right)^{1/2}$$

In this case,

$$e = 1.6 \times 10^{-19} \text{ C}$$

$$k_B = 1.38 \times 10^{-23} \text{ J/K}$$

$$T = 273 + 25 = 298 \text{ K}$$

$$\begin{aligned} \epsilon &= 4\pi \epsilon_r \epsilon_0 = 4\pi (78.5) (8.85 \times 10^{-12} \text{ C}^2/\text{J m}) \\ &= 8.73 \times 10^{-9} \text{ C}^2/\text{J m} \end{aligned}$$

$$\begin{aligned} n_s &= .05 \text{ M} = (0.05 \text{ mol/L}) (6.022 \times 10^{23} \text{ molecules/mol}) (1000 \text{ L/m}^3) \\ &= 3.011 \times 10^{25} \text{ molecules/m}^3 \end{aligned}$$

Thus,

$$\kappa = \left(\frac{8\pi (1.6 \times 10^{-19} \text{ C})^2}{(1.38 \times 10^{-23} \text{ J/K}) (298 \text{ K}) 4\pi (78.5) (8.85 \times 10^{-12} \text{ C}^2/\text{J m})} \right)^{1/2} (3.011 \times 10^{25} \text{ molecules/m}^3)^{1/2} = 7.346 \times 10^8 \text{ m}^{-1}$$

Thus, the Debye length, κ^{-1} , is

$$\kappa^{-1} = 1.36 \text{ nm}$$

Now if the protein is binding to at a distance r_0 from the charged surface, it has an electrostatic energy that corresponds to $q\psi(x=r_0)$, where q is the charge of the protein.

Thus, its energy E (not to be confused with the electric field above) is

$$E = q \frac{4\pi\sigma}{\epsilon\kappa} e^{-\kappa r_0}$$

If the surface charge density is $\sigma = \frac{-e}{2 \text{ nm}^2}$, then we can calculate the energy of binding at a distance r_0 .

$$E = - \frac{4\pi (1.6 \times 10^{-19} \text{ C})^2}{(4\pi) (78.5) (8.85 \times 10^{-12} \frac{\text{C}^2}{\text{J m}}) (7.346 \times 10^8 \frac{1}{\text{m}}) (2 \text{ nm}^2) (10^{-18} \frac{\text{m}^2}{\text{nm}^2})} \text{Exp}\left[\frac{-0.1 \text{ nm}}{1.36 \text{ nm}}\right] = -2.33 \times 10^{-20} \text{ J} = -14 \text{ kJ/mol}$$

From the dissociation constant, we can now calculate the free energy change upon binding using the fact that

$$\Delta G = RT \log(K_D) = 8.31 \frac{\text{J}}{\text{K mol}} (298 \text{ K}) \text{Log}(1 \times 10^{-2}) = -1.89 \times 10^{-20} \text{ J} = -11.4 \text{ kJ/mol}$$

Note the sign is changed on the ΔG relation since the dissociation constant is the equilibrium constant for the process of dissociating whereas the energy we just calculated is for binding.

Thus, using the fact that $\Delta G = \Delta H - T \Delta S$, we can calculate ΔS

$$\Delta S = - \frac{\Delta G - \Delta H}{T} = - \frac{(-11.4 \text{ kJ/mol} + 14 \text{ kJ/mol})}{298 \text{ K}} = -8.7 \frac{\text{J}}{\text{K mol}}$$

Alternative: If we solved the more complicated equation above, we would find that the potential is

$$\psi(x) = \frac{2k_B T}{e} \ln\left(\frac{1 + \gamma e^{-\kappa x}}{1 - \gamma e^{-\kappa x}}\right)$$

First evaluating $A \kappa$

$$A \kappa = \frac{k_B T \epsilon}{2 \pi \sigma e} \kappa = \frac{(2 \text{ nm}^2)(10^{-18} \frac{\text{m}^2}{\text{nm}^2})(1.38 \times 10^{-23} \text{ J/K})(298 \text{ K})(4\pi)(78.5)(8.85 \times 10^{-12} \frac{\text{C}^2}{\text{J m}})}{2 \pi (1.6 \times 10^{-19} \text{ C})^2} (7.346 \times 10^8 \frac{1}{\text{m}}) = 0.328$$

This implies that

$$\gamma = -\kappa A + \sqrt{1 + \kappa^2 A^2} = -0.328 + \sqrt{1 + (0.328)^2} = 0.724$$

Thus, we can evaluate the potential at $x_0 = 1$

$$\psi(x = x_0) = \frac{2k_B T}{e} \ln\left(\frac{1 + \gamma e^{-\kappa x}}{1 - \gamma e^{-\kappa x}}\right) = \frac{2(1.38 \times 10^{-23} \text{ J/K})(298 \text{ K})}{1.6 \times 10^{-19} \text{ C}} \ln\left(\frac{1 + 0.724 \text{ Exp}\left[\frac{-0.1 \text{ nm}}{1.36 \text{ nm}}\right]}{1 - 0.724 \text{ Exp}\left[\frac{-0.1 \text{ nm}}{1.36 \text{ nm}}\right]}\right) = 0.0838 \frac{\text{J}}{\text{C}}$$

Thus, the energy (not electric field) is

$$E = -e\psi(x = 1 \text{ \AA}) = 0.0838 \frac{\text{J}}{\text{C}} (-1.6 \times 10^{-19} \text{ C}) = -1.34 \times 10^{-20} \text{ J} = -8.1 \text{ kJ/mol}$$

Thus, the entropy change in this case is

$$\Delta S = \frac{\Delta H - \Delta G}{T} = \frac{-8.1 \text{ kJ/mol} + 11.4 \text{ kJ/mol}}{298 \text{ K}} = 11 \text{ J/K mol}$$

Note that this answer is different from that calculated above using the linearized Poisson-Boltzmann equation. The reason is that for this surface charge density the energy is not always less than $k_B T$. In fact the value calculated above is already larger than $k_B T$. This implies that this is not a valid approximation; though, it definitely simplifies the calculation greatly! (That was the intent at least!)

An additional note: a factor of 4π is floating throughout the calculation. I have been careful to use the same definition for the Poisson-Boltzmann equation and Gauss' Law.

■ Part c

If one calculated that the entropy decreases then the answer should be that this is expected since the degrees of freedom of the protein are now restricted. If one calculated that the entropy increased, then this must of been due to the release of ions or water molecules condensed on the surface.

■ Problem 6: Protein Binding to a Charged Surface: II

■ Part a

Recall that the acid dissociation constant, K_a , is related to the $\text{p}K_a$ via

$$K_a = 10^{-\text{p}K_a}$$

and that the fraction ionized at a given pH is

$$f_{A^-} = \frac{K_a}{K_a + [H^+]}$$

Thus, the three K_a values and the fraction ionized are

N-terminus:

$$K_a = 10^{-8}$$

$$f_{A^-} = \frac{10^{-8}}{10^{-8} + 10^{-7}} = \frac{1}{11} = 0.091$$

C-terminus:

$$K_a = 10^{-3}$$

$$f_{A^-} = \frac{10^{-3}}{10^{-3} + 10^{-7}} = \frac{10,000}{10,001} \sim 0.9999 \sim 1$$

and lysine residue:

$$K_a = 10^{-10.5} = 3.2 \times 10^{-11}$$

$$f_{A^-} = \frac{3.2 \times 10^{-11}}{3.2 \times 10^{-11} + 10^{-7}} = 3.2 \times 10^{-4}$$

Some may argue that the term "ionized" corresponds to the charged state. This is a reasonable definition, so no credit was lost if you calculated the fraction for the charged states.

■ Part b

To calculate the average charge as a function of hydrogen ion concentration, recall that if we have a species that can ionize, $AH \leftrightarrow A + H^+$ then the fraction (or probability) of the species that will be ionized,

$$p_{A^-} = \frac{K_a}{K_a + [H^+]}$$

while the probability of not being ionized is

$$p_{AH} = \frac{[H^+]}{K_a + [H^+]}$$

Note that these probabilities add to 1 as expected.

The protein considered here has three ionizable groups, the N and C termini and the lysine residue, each of which can be ionized or not. For simplicity, let's label the K_a associated with each of these groups as K_N , K_C , and K_K .

To calculate the average charge of the protein we need to first consider all of the possible ionization states of the protein and the charge of those states, and the probability of observing having that state. Again, for simplicity let's define the symbols NH+ and N as an extra hydrogen bound to the n-terminus or not, KH+ and K as an extra hydrogen bound to the lysine or not, and finally the symbols C- and CH as the hydrogen missing from the c-terminus or not. Now, we can easily classify and label all of the possible combinations of ionization state.

The maximum charge of the protein is +2, which only occurs in the state N+K+CH. The probability of this state is

$$p_{++0} = \frac{[H^+]}{K_N + [H^+]} \frac{[H^+]}{K_K + [H^+]} \frac{[H^+]}{K_C + [H^+]}$$

where we have used the fact that since the ionization of each group does not affect the ionization probability of the other groups this probability is simply the product of all three individual probabilities.

Now, to consider the possible states that give a charge of +1: This can occur in three ways: NH+KH+C-; NH+; K; CH; N; KH+; CH. The probabilities of these individual states are

$$p_{++-} = \frac{[H^+]}{K_N + [H^+]} \frac{[H^+]}{K_K + [H^+]} \frac{K_C}{K_C + [H^+]}$$

$$p_{+00} = \frac{[H^+]}{K_N + [H^+]} \frac{K_K}{K_K + [H^+]} \frac{[H^+]}{K_C + [H^+]}$$

$$p_{0+0} = \frac{K_N}{K_N + [H^+]} \frac{[H^+]}{K_K + [H^+]} \frac{[H^+]}{K_C + [H^+]}$$

There are three states that can give a net charge of zero: NH+ K C-; NKH+C-; NKCH. The probabilities of these states are

$$p_{+0-} = \frac{[H^+]}{K_N + [H^+]} \frac{K_K}{K_K + [H^+]} \frac{K_C}{K_C + [H^+]}$$

$$p_{0+-} = \frac{K_N}{K_N + [H^+]} \frac{[H^+]}{K_K + [H^+]} \frac{K_C}{K_C + [H^+]}$$

$$p_{000} = \frac{K_N}{K_N + [H^+]} \frac{K_K}{K_K + [H^+]} \frac{[H^+]}{K_C + [H^+]}$$

Finally, there is only one state that can give a net charge of -1: NKC-. Its probability is

$$p_{00-} = \frac{K_N}{K_N + [H^+]} \frac{K_K}{K_K + [H^+]} \frac{K_C}{K_C + [H^+]}$$

The sum of all of these probabilities equals 1 as expected.

The final average charge on the protein is the sum of all of the probabilities of seeing each charge state: Thus $\langle q \rangle$ is

$$\langle q \rangle = +2e p_{++0} + 1e(p_{++-} + p_{+00} + p_{0+0}) + -e p_{00-}$$

In terms of the hydrogen ion concentration and the K_a values this is

$$\langle q \rangle = \frac{e}{(K_N + [H^+])(K_K + [H^+])(K_C + [H^+])} \left(2 [H^+]^3 + [H^+]^2 (K_N + K_K + K_C) - K_N K_K K_C \right)$$

To calculate the average charge of the protein at pH = 7.0, recall that this implies that $[H^+] = 10^{-7}$. Plugging this in to the above expression with the K_a values derived above, yields

$$\langle q \rangle = \frac{e}{(10^{-8} + 10^{-7})(3.2 \times 10^{-11} + 10^{-7})(10^{-3} + 10^{-7})} \left(2 (10^{-7})^3 + (10^{-7})^2 (10^{-8} + 10^{-3} + 3.2 \times 10^{-11}) - (10^{-8})(10^{-3})(3.2 \times 10^{-11}) \right) = 0.91 e$$

As a side note, the above expression simplifies nicely to

$$\langle q \rangle = e \left(\frac{[H^+]}{K_N + [H^+]} + \frac{[H^+]}{K_K + [H^+]} - \frac{K_C}{K_C + [H^+]} \right)$$

The sum of the fraction of each group that has a charge times the charge on that group.

■ Part c

Since the change in enthalpy is the charge of the protein times the electrostatic potential, this implies that

$$\Delta H = \langle q \rangle \psi(r = r_0)$$

■ Part d

The change in enthalpy will be zero when the net charge of the protein is zero. Inspection of the expression above yields the requirement that

$$2 [H^+]^3 + [H^+]^2 (K_N + K_K + K_C) - K_N K_K K_C = 0$$

This expression can be solved analytically or numerically to yield a the concentration at which the average charge is zero (Recall that this is known as the isoelectric point of the protein or pI)

$$\text{pH} = \text{pI} = 9.25$$

$$\text{or } [H^+] = 5.62 \times 10^{-11} \text{ M}$$

We also learned a quick trick for calculating this in class. The pI is the concentration of $[H^+]$ at which the fraction of +1 charged species equals the fraction of -1 charged species.

Thus,

$$p_{++-} + p_{+00} + p_{0+0} = p_{00-}$$

$$\left(\frac{[H^+]}{K_N + [H^+]} \frac{[H^+]}{K_K + [H^+]} \frac{K_C}{K_C + [H^+]} \right) + \left(\frac{[H^+]}{K_N + [H^+]} \frac{K_K}{K_K + [H^+]} \frac{[H^+]}{K_C + [H^+]} \right) + \left(\frac{K_N}{K_N + [H^+]} \frac{[H^+]}{K_K + [H^+]} \frac{[H^+]}{K_C + [H^+]} \right) = \frac{K_N}{K_N + [H^+]} \frac{K_K}{K_K + [H^+]} \frac{K_C}{K_C + [H^+]}$$

Notice the denominator is common in all expressions, and thus cancels, leaving

$$[H^+]^2 (K_C + K_K + K_N) = K_N K_K K_C$$

Notice that $K_C \gg K_K, K_N$, thus $K_C + K_K + K_N \sim K_C$. This implies that

$$[H^+]^2 \approx K_N K_K$$

which implies that

$$\text{pH} = \frac{\text{p}K_N + \text{p}K_K}{2} = \frac{8 + 10.5}{2} = 9.25$$