Solution Key to Physical CUMULATIVE EXAM Topic: "Transition State Theory applied to two-state protein folding kinetics: chevron plots, phi-value analysis, and rate limiting steps"

Vincent Voelz, Department of Chemistry, Temple University

June 17, 2019

Topic

The main reading assignment to prepare for the exam is Chapters 17, 18 and 19 of the book:

> Fersht, A. (1998). Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. WH Freeman.

You will be asked more specific questions/problems about:

- two-state folding kinetics (Chapter 18)
- chevron plots (Chapter 18, starting at p. 543)
- phi-value analysis (Chapter 18, starting at p. 559)
- m-values and β_T-values (Chapter 18, starting at p. 543)

Other good phi-value references can be found in the text and citations on: `https://en.wikipedia.org/wiki/Phi_value_analysis`

A very good reference on measuring and analyzing protein folding kinetic data, and chevron plots can be found here:

> Zarrine-Afsar, A., and Davidson, A. R. (2004). The analysis of protein folding kinetic data produced in protein engineering experiments. Methods, 34(1), 41–50. `http://doi.org/10.1016/j.ymeth.2004.03.013`

Questions

1. (5 pts) Why is folding "two-state"? Describe in words why proteins have (effectively) two favorable thermodynamic states, "unfolded" and "native". Express this in terms of the favorable free energy of folding $\Delta G = \Delta H - T\Delta S$, commenting on the role of enthalpy (H) and entropy (S) in each state.

While there are many conformational (micro)states available to polypeptide chain, only two (macro)states have appreciable population that can be detected experimentally at equilibrium . One state is the native state, which is essentially a single conformational state with very low conformational entropy, and the other is the unfolded state, which comprises a huge number non-native conformations and therefore has a large conformational entropy. Near the temperatures at which a protein normally operates, the folding reaction is favorable, with $\Delta G < 0$, from thermodynamic driving forces ascribed to both the protein and the solvent. The enthalpy of folding is favorable ($\Delta H < 0$), due mainly to the protein making backbone hydrogen bonds and favorable van der Waals interactions from tight packing. While the *conformational* entropy change of folding is highly disfavorable ($\Delta S_{\text{conf}} < 0$), the entropy change of the *solvent* is highly favorable ($\Delta S_{\text{solv}} > 0$), due to the hydrophobic effect–i.e. the burial of non-polar protein groups in the interior of a protein liberates bound waters, increasing their entropy. The free energy reward for desolvation is cooperative, in the sense that it is maximized only when all of the native structure of the protein has been formed; this is the main reason for why folding intermediates are not typically observed for two-state folders.

2. (5 pts) Relation between two-state rates and equilibrium. Consider the two-state kinetic model for folding:

$$U \underset{\mathrm{k_u}}{\overset{\mathrm{k_f}}{\rightleftarrows}} N$$

Here, U is the unfolded state, N is the native folded state, k_f is the folding rate, and k_u is the unfolding rate.

a. Write down first-order rate equations for $d[\mathrm{U}]/dt$ and $d[\mathrm{N}]/dt$, and use the steady-state criterion $d[\mathrm{U}]/dt = 0$ (or equivalently, $d[\mathrm{N}]/dt = 0$) to show that the equilibrium constant is given by $K = k_f/k_u$.

$$\frac{d}{dt}[\mathrm{U}] = -k_f[\mathrm{U}] + k_u[\mathrm{N}] = 0 \tag{1}$$

$$k_u[\mathrm{N}] = k_f[\mathrm{U}] \tag{2}$$

$$\frac{[\mathrm{N}]}{[\mathrm{U}]} = \frac{k_f}{k_u} = K \tag{3}$$

3. (10 pts) Observed kinetics in two-state folding. Consider the time-dependent relaxation to the equilibrium state. Let the unfolded, concentration at equilibrium be $[U_0]$.

a. Show that the total concentration can be written as $[\text{total}] = [U_0](1 + K)$.

If the initial concentrations of the unfolded and native states are $[U_0]$ and $[N_0]$, respectively, then according to the relation from 2a, the total concentration can be written:

$$[\text{total}] = [U_0] + [N_0] = [U_0] + [U_0]K = [U_0](1 + K)$$

b. Next, substitute [N] = [total] - [U] into the first-order rate equation for $d([U] - [U_0])/dt$ and show that the time-dependence of [U]-$[U_0]$ is given by an exponential decay with the observed rate of $k_{obs} = k_f + k_u$. *Hint*: Note that $d([U] - [U_0])/dt$ is the same as $d[U]/dt$, because $[U_0]$ is a constant.

$$\frac{d}{dt}([U] - [U_0]) = -k_f[U] + k_u[N] \tag{4}$$
$$= -k_f[U] + k_u\big([\text{total}] - [U]\big) \tag{5}$$
$$= -k_f[U] + k_u\big([U_0](1 + K) - [U]\big) \tag{6}$$
$$= -k_f[U] + k_u\big([U_0](1 + \frac{k_f}{k_u}) - [U]\big) \tag{7}$$
$$= -k_f[U] + k_u[U_0] + k_f[U_0] - k_u[U] \tag{8}$$
$$\frac{d}{dt}([U] - [U_0]) = -(k_f + k_u)([U] - [U_0]) \tag{9}$$
$$\tag{10}$$

The time-dependent solution to this equation is $([U]-[U_0])(t) = ([U]-[U_0])(0)\exp(-k_{\text{obs}}t)$, where the observed relaxation rate is $k_{\text{obs}} = k_f + k_u$.

c. Name and describe at least two experimental methods for measuring protein folding kinetics.

There are many possible answers to this question. In general, many proteins fold on the millisecond (or faster) time scale, so it is essential that time-resolved experiments use methods to induce fast mixing from denaturant to buffer solutions (or vice versa) on this time scale. This can be achieved in so-called 'stopped-flow' mixers, or in continuous-flow mixers. An example of the latter is a microfluidic mixer, in which two solutions are mixed in a turbulently mixing region; the time after mixing can be calibrated to the spatial displacement after this region, assuming a fixed output flow velocity.

Any spectroscopic signal that changes between the folded and unfolded state can be used to monitor the change in the folded population. Fluorescence or absorbance is a common observable, with typical fluorophores coming from either aromatic resides or covalently attached fluorescent dyes. A very common signal is the UV fluorescence of tryptophan

(Trp), which absorbs around 280 nm, and is sensitive to solvent exposure. Site-directed mutagenesis is often employed to engineer a single Trp residue into proteins for this sole purpose. IR spectroscopy can be used to monitor folding through shifts in the amide bond absorptions (namely C=O amide stretches) due to hydrogen bonding. This technique is often coupled with fast refolding from a laser-induced temperature jump.

Hydrogen-deuterium exchange (HDX) is an NMR method that can measure folding/unfolding rates using the observed rate of exchange between H and D on backbone amides, depends both on the intrinsic exchange rates in the folded and unfolded states, and the rate at which unfolded and folded states interconvert.

4. (10 pts) The chevron plot. One way to determine protein folding and unfolding rates is to measure observed folding kinetics in varying concentrations of denaturant ([den]). Common denaturants for proteins include guanidinium hydrochloride (GuHCl) or urea. These results of these experiments are typically presented in a so-called *chevron plot*.

a. Draw a typical chevron plot of $\ln(k_{\text{obs}})$ vs [den]. On the plot, show what the m_f and m_u values correspond to.

Your drawing should look something like the figure below. The m_f and m_u values are the slopes of the left and right asymptotes, respectively.

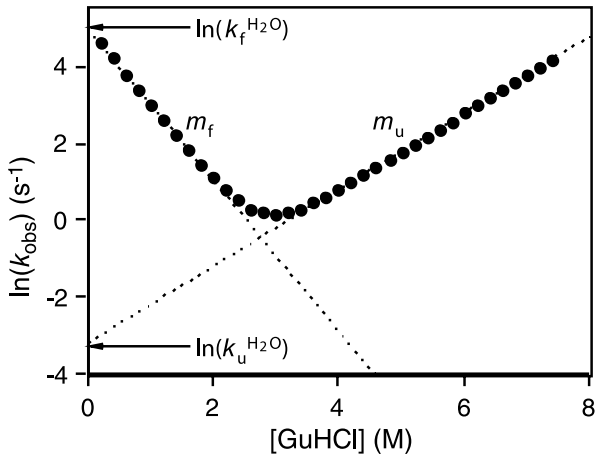

Figure 1: A typical chevron plot. Graphic reproduced from Zarrine-Afsar, A., and Davidson, A. R. (2004).

b. What is special about the concentration [den] at the minimum values of k_{obs}?

The concentration [den] where k_{obs} is at a minimum is the denaturant concentration where exactly half of the population at equilibrium is folded, and half is unfolded, i.e. the denaturation midpoint. At this point, $k_f = k_u$.

c. How can one use the chevron plot to calculate the folding and unfolding rates in the absence of denaturant? Mark these values on the chevron plot.

Assuming linear dependence of the folding rate with denaturant concentration, one can linearly extrapolate the folding and unfolding arms of the chevron plot to zero denaturant, where the extrapolated values of $k_f^{\text{H}_2\text{O}}$ and $k_u^{\text{H}_2\text{O}}$ are where the interpolated lines cross the y-axis.

d. How can one use the chevron plot to calculate the folding stability in the absence of denaturant?

From the extrapolated values of $k_f^{\text{H}_2\text{O}}$ and $k_u^{\text{H}_2\text{O}}$, the free energy of folding in the absence of denaturant can be calculated as $\Delta G^{\text{H}_2\text{O}} = -RT \ln(k_f^{\text{H}_2\text{O}}/k_u^{\text{H}_2\text{O}})$.

5. (ungraded) Chevron plot m-values. Consider m_f and m_u in the chevron plot, which correspond to the denaturant dependence of the folding and unfolding rates separately:

$$\ln k_f = \ln k_f^{\text{H}_2\text{O}} - m_f[\text{den}] \tag{11}$$

$$\ln k_u = \ln k_u^{\text{H}_2\text{O}} + m_u[\text{den}] \tag{12}$$

Use the definition of ΔG in terms of K to show that $\Delta G = \Delta G^{\text{H}_2\text{O}} - m[\text{den}]$, where $m = m_f + m_u$.

Unfortunately, there's a typo in this question, so I left this problem ungraded.

The question should read: $m = -RT(m_f + m_u)$. (This is shown in the Zarrine-Afsar et al. paper, but not explained well in Fersht's book. While m_f and m_u are proportionality constants in units of M^{-1}, the thermodynamic proportionality is in terms of molar free energy per molar concentration of denaturant, kJ mol^{-1} M^{-1}). The derivation is straightforward substitution:

$$\Delta G = -RT \ln \frac{k_f}{k_u} \tag{13}$$

$$= -RT\big(\ln k_f + \ln k_u\big) \tag{14}$$

$$= -RT\big(\ln k_f^{\text{H}_2\text{O}} - m_f[\text{den}] - \ln k_u^{\text{H}_2\text{O}} - m_u[\text{den}]\big) \tag{15}$$

$$= -RT \ln \frac{k_f^{\text{H}_2\text{O}}}{k_u^{\text{H}_2\text{O}}} + RT(m_f + m_u)[\text{den}] \tag{16}$$

$$= \Delta G^{\text{H}_2\text{O}} - m[\text{den}] \tag{17}$$

6. (10 pts) Structural information about transition states from m-values. In our simple model of two-state folding kinetics, folding and unfolding rates obey Arrhenius kinetics, where the folding rate is determined by an activation barrier $G_\ddagger - G_U$ and the unfolding rate is determined by an activation barrier $G_\ddagger - G_N$, as shown in the free energy diagram below:

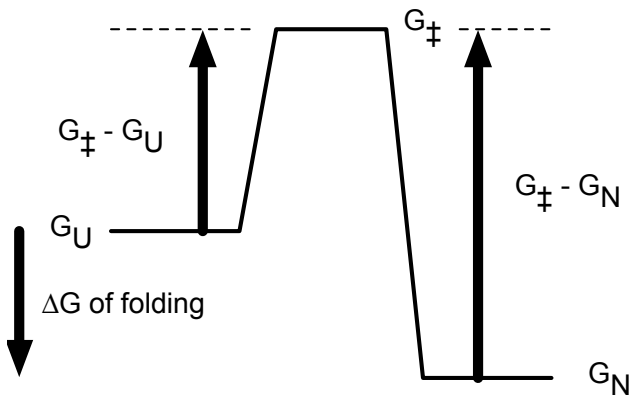

Figure 2: Folding free energy diagram

According to the Arrhenius rate law:

$$k_f \sim k_0 \exp(-(G_\ddagger - G_U)/RT) \tag{18}$$
$$k_u \sim k_0 \exp(-(G_\ddagger - G_N)/RT) \tag{19}$$

where G_U, G_N, and $G_\ddagger$ are the free energies of the unfolded state, folded state, and transition state, respectively; and k_0 is a pre-factor constant (related to intramolecular diffusion).

a. Take the logarithm of the above equations, and compare with problem 5, to find the denaturant-dependence of the folding barriers.

Taking the logarithm of k_f, we get

$$\ln k_f = \ln k_0 - (G_\ddagger - G_U)/RT = \ln k_f^{\mathrm{H_2O}} - m_f[\mathrm{den}],$$

and solving for the folding barrier $(G_\ddagger - G_U)$, we obtain:

$$(G_\ddagger - G_U) = -RT\big(\ln \frac{k_f^{\mathrm{H_2O}}}{k_0}\big) + m_f[\mathrm{den}],$$

This means that the activation barrier for folding *increases* as the concentration of denaturant increases.

Using a similar procedure for $(G_\ddagger - G_N)$, we obtain:

$$(G_\ddagger - G_N) = -RT\big(\ln \frac{k_u^{\mathrm{H_2O}}}{k_0}\big) - m_u[\mathrm{den}],$$

This means that the activation barrier for unfolding *decreases* as the concentration of denaturant increases.

b. It is known that chemical denaturants have more significant denaturing effects (i.e. a greater m value) for protein states that have greater average solvent exposure. Suppose we perform a series of kinetics experiments and find that m_f is four times as large as m_u. What does this say about the average solvent exposure of the transition state? Is it more like the unfolded state (very solvent-exposed), or is it more like the folded state (compact, with low solvent exposure)? Use the idea of the Tanford β values to explain.

If $m_f \approx 4m_u$, that means that the transition state has a solvent exposure more like that of the *native* state than the unfolded state. The Tanford β_T value measures the extent to which the solvent exposure of the transition state resembles that of the folded state:

$$\beta_T = \frac{m_f}{m_f + m_u}$$

In this case, $\beta_T \approx 4m_u/(4m_u + m_u) = 4/5$, so the transition state would be about 4/5 of way to native state in terms of solvent exposure. Below is a picture illustrating this.

7. (15 pts) Structural information about transition states from phi-values. a. Explain what a phi-value is, and how they can be measured using single-point mutation and folding kinetics studies.

The Φ-value is a quantity measured for (usually) a single-point mutation at a specific residue. It is defined as the ratio of the apparent free energy change $\Delta\Delta G_\ddagger$ in the unfolded-to-transition state barrier, to the change in the folding free energy $\Delta\Delta G$:

$$\Phi = \frac{\Delta\Delta G_\ddagger}{\Delta\Delta G}$$

where $\Delta\Delta G_\ddagger = -RT\ln(k_f^{\mathrm{mut}}/k_f^{\mathrm{wt}})$ is inferred from the change in the folding rates from wild type (wt) to mutant (mut).

b. What key assumption is made about the unfolded state when calculating phi-values?

Besides the larger assumption of two-state folding, a key assumption is that mutational effects for the unfolded state are *negligible*, so that all changes in $\Delta\Delta G$ result from changes in native-state stability. This can be seen above in Figure 4.

Another assumption is that native-like structure is acquired along a single reaction pathway, in an all-or-nothing way. This assumption is discussed below.

c. How does a phi-value tell you information about the structure of a protein folding transition state?

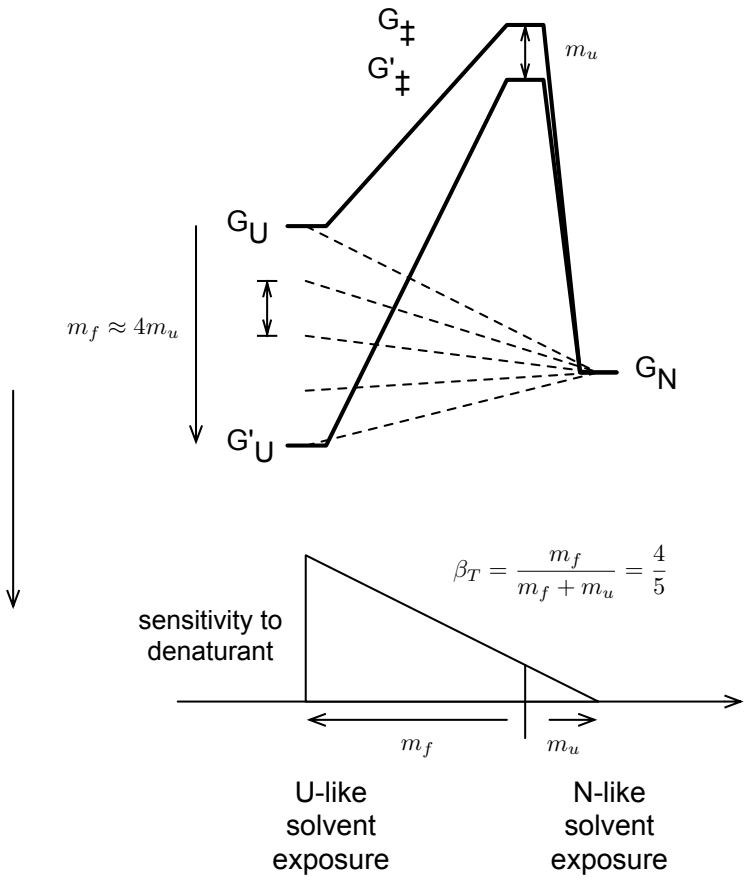

Figure 3: an illustration of the Tanford β_T value.

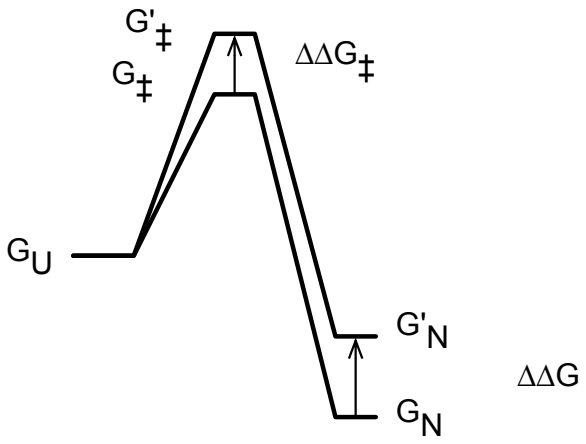

Figure 4: an illustration of the quantities used in computing Φ values.

The interpretation of a Φ-value comes thinking about the order in which native structure is acquired along the folding reaction coordinate. The key question is whether native-like structure is formed *before*, or *after* the transition state (TS). A mutation at a given residue will disrupt the stability of native state through the interactions made in the native conformation. If this structure forms only *after* the TS is reached, then we would expect mutations to have negligible impact on the TS, making $\Delta\Delta G_{\ddagger} = 0$, resulting in a Φ-value of 0. If, on the other hand, the residue participates in native-like structure formation *before* the TS is reached, we would expect $\Delta\Delta G_{\ddagger} = \Delta\Delta G$, resulting in a Φ-value of 1.

In a perfectly two-state folding mechanism, with sequential acquisition of native structure along a well-defined reaction coordinate, one would expect that alanine-scanning mutagenesis studies (i.e. measuring folding kinetics for a series of single-point alanine mutants) would reveal a set of residues with Φ-values of 1, representing the set of residues that is folded in the TS. The remaining residues should have Φ-values of 0. In practice, this is almost never is the case. Fractional Φ-values are instead measured, indicating partial structuring, or suggesting the existence multiple pathways. Non-canonical Φ-values that are less than zero, or greater then one are also often measured. One reason for this is no doubt statistical uncertainty, because the measured values of the $\Delta\Delta G$ are often so close to zero that the calculated ratio $\Delta\Delta G_{\ddagger}/\Delta\Delta G$ is highly uncertain. Another reason may be unfolded-state effects, which is not accounted for in Fersht's model.

d. A protein mutation F46A destabilizes the folded state and has a phi-value of 1. Draw a possible scenario for wild-type and mutation chevron plots. (Draw and label these on the same set of axes.)

A Φ-value of 1 means that the folding rate would be slowed, but the unfolding should remain the same.

e. A protein mutation D55A is stabilizes the folded state and has a phi-value of 0.

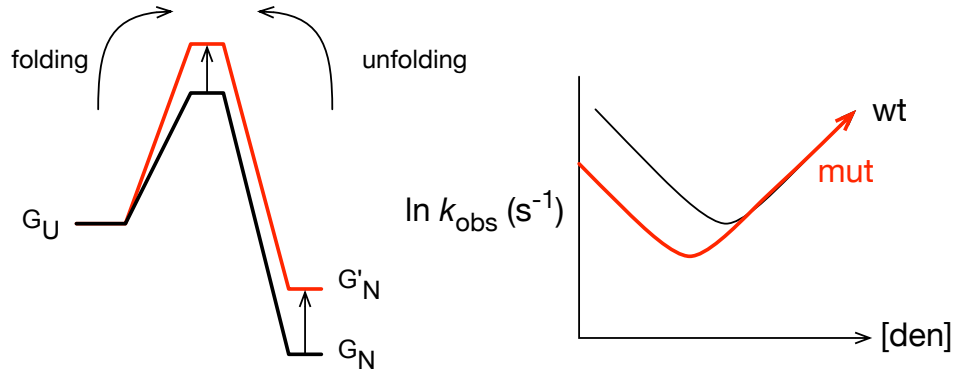

Figure 5: A protein mutation F46A destabilizes the folded state and has a Φ-value of 1. Note: this assumes there are no changes in the m-values.

Draw a possible scenario for wild-type and mutation chevron plots. (Draw and label these on the same set of axes.)

A Φ-value of 0 means that the folding rate will be unchanged, but the unfolding rate should be slower (because the native state is stabilized).

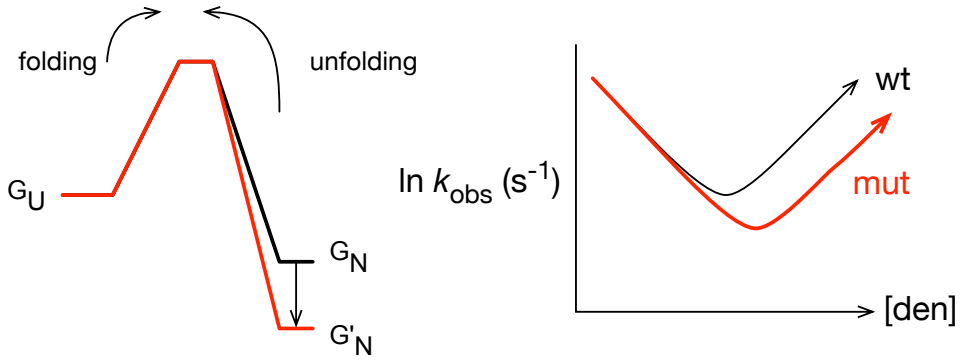

Figure 6: A protein mutation D55A stabilizes the folded state and has a Φ-value of 0. Note: this assumes there are no changes in the m-values.