Supporting information

Engineering more stable proteins

Romas Kazlauskas

Contents

Derivation of equation relating spectral changes and equilibrium constant, eq. 5.

Example of linear extrapolation of protein unfolding data in urea.

Example of using Rosetta Design to predict a stabilizing substitution.

Derivation of equation relating spectral changes and equilibrium constant, eq. 5.

Deriving the relationship between spectral changes and equilibrium constant is straightforward. The protein is either in the native state or in the denatured state, so the sum of the two fractions is one: $F_N + F_D = 1$. The observed fluorescence at each urea concentration, Y_{obs}, is the sum of the contributions from the native and denatured states:

$$Y_{obs} = Y_N \cdot F_N + Y_D \cdot F_D \qquad (S1)$$

Substituting $F_N = 1 - F_D$ yields:

$$F_D = (Y_{obs} - Y_N)/(Y_D - Y_N) \qquad (S2)$$

Similarly substituting $F_D = 1 - F_N$ yields:

$$F_N = (Y_D - Y_{obs})/(Y_D - Y_N) \qquad (S3)$$

Dividing these two equations yields eq S4, which is the same as eq 8 in the main text.

$$K_{unfold} = F_D/F_N = (Y_{obs} - Y_N)/(Y_D - Y_{obs}) \qquad (S4)$$

Example of linear extrapolation of protein unfolding data in urea.

Protein A was dissolved in solutions of different concentrations of urea and allowed to reach equilibrium. The fluorescence spectra of these solutions showed an increase in fluorescence at 2-5 M urea, Fig. S1. Using this fluorescence data, calculate the Gibbs energy of unfolding in pure water for protein A.

[urea]	Y_{obs}
0	72
1	70
2	81
3	95
4	140
5	150
6	155
7	152
8	156
9	156
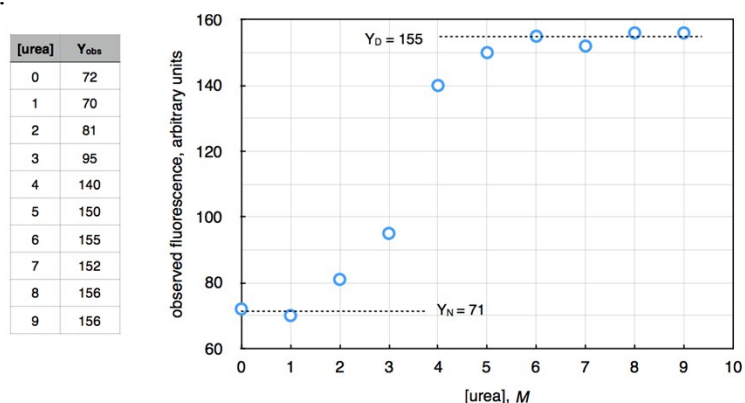
Fig. S1. Simplified experimental data from a urea-unfolding experiment to measure the Gibbs energy of unfolding of a protein. The intensity of fluorescence can either increase or decrease upon protein unfolding; in this example, it increases. Analysis and replot of this data are in Fig 9 below.

Solution

1. Estimate the fluorescence of the native and denatured states.

> The fluorescence of the native state is the flat part of the curve before unfolding begins. The average of the values at 0 and 1 M urea yields $Y_N = 71$. The fluorescence of the denatured state is the flat part of the curve after unfolding is complete. The average of the values at 6-9 M urea yields $Y_D = 155$.
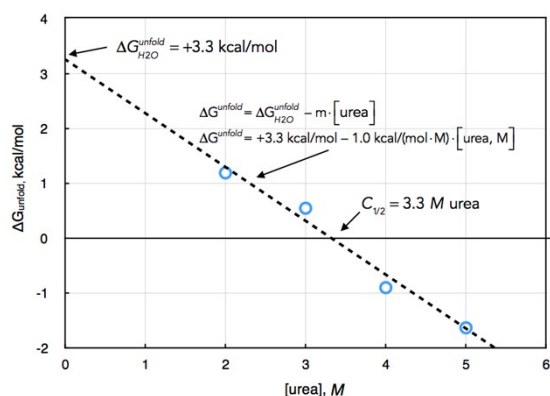
2. Estimate the equilibrium constant between the native and denatured forms at each urea concentration in the unfolding region.

> For this example: $K_{unfold} = (Y_{obs} - 71)/(155 - Y_{obs})$, so one can calculate K_{unfold} at different urea concentrations by substituting the experimental values of Y_{obs}. This procedure yields the values in the table below.

[urea]	K_{unfold}	ΔG_{unfold} (kcal/mol)
2	0.14	1.2
3	0.40	0.54
4	4.6	−0.90
5	16	−1.6

3. Convert the equilibrium constant to Gibbs energy according to eq 9 in main text using R = 1.987 cal/mol °K and T = 298 °K, which yields the data in the table above.

4. Plot the Gibbs energy of unfolding as a function of urea concentration (eq 10 in main text), fit the data to a straight line and extrapolate to pure water ([urea] = 0). The slope of the line in the linear extrapolation plot is called the m-value.

$\Delta G^{unfold}_{H2O} = +3.3$ kcal/mol

$\Delta G^{unfold} = \Delta G^{unfold}_{H2O} - m \cdot \left[urea \right]$

$\Delta G^{unfold} = +3.3$ kcal/mol $- 1.0$ kcal/(mol·M) $\cdot \left[urea, M \right]$

$C_{1/2} = 3.3\ M$ urea

Fig S2. Linear extrapolation plot of the data from Fig S1 above yields $\Delta G_{unfold} = 3.3 - 1.0[urea]$ as the best fit.

The extrapolation yields a y-intercept of +3.3 kcal/mol, which is the Gibbs energy of unfolding of this protein in pure water. This protein is less stable than typical proteins, but even for this one, the equilibrium amount of denatured protein in pure water is only 0.38%, which would be difficult to measure directly.

Example of using Rosetta Design to predict a stabilizing substitution

Predict a single amino acid substitution to stabilize the esterase SABP2 (structure pdb code = 1y7i). Previous modeling suggested that a replacement at position 60 may stabilize the protein. Use the RosettaDesign server http://rosettadesign.med.unc.edu to predict the substitution.[1] The server requires registration, but it is free to use.

Computational approach: Test all 19 possible replacements at position 60 to identify any that have lower energy while repacking the side chains at nearby residues to adjust for the amino acid substitutions. To do these calculations, we need to provide Rosetta with 1) the pdb file of SABP2 and 2) instructions for the desired calculation. An example set of instructions is below:

NATRO

start

47 A NATAA
56 A NATAA
57 A NATAA
58 A NATAA
59 A NATAA
60 A ALLAA
61 A NATAA
62 A NATAA
63 A NATAA
64 A NATAA

To send these instructions to Rosetta Design, copy the text above and save them as a text file with the name SABP2_LEU60.res or something similar. The file extension must be .res. Submit the pdb file and the res file to the Rosetta design server.

Hint: The results do not indicate the instructions that you used for the calculations. Once you submit the job and get a job number, rename your res file by adding the job number to later match it to the results (e.g., 35161_SABP2_LEU60.res)

Explanation of the instructions:

The first line "NATRO" is the default for the whole protein. NATRO means natural rotamer and tells Rosetta Design to keep the existing amino acid and the existing side chain rotamer. That is, don't change anything.

The next line, "start" indicates the start of specific instructions for particular amino acids.

The following lines are all formatted as [amino acid number] [chain letter] [instructions]. In this case, several amino acids from chain A are being redesigned.

At positions 47, 56-59, 61-64, NATAA means native amino acid and tells Rosetta design to keep the existing amino acid, but allow the side chain to move to a new rotamer. This flexibility allows the surrounding region to adjust as the amino acid at position 60 changes.

At position 60, ALLAA means all amino acids and tells Rosetta Design to try all amino acids and all rotamers of these amino acids.

Results

The results consist of a log file and a pdb file. Open the pdb file with a text editor.

Scroll to residue 60 of chain A; it has been redesigned to Arg.

Scroll to nearly the end of the file to the row starting with the word 'pose'. The last number in this row is the total energy = 41.7 (the units are Rosetta Energy Units)

As a control, run another calculation that only readjusts the positions of the side chains. The res file is the same, except line 60 should read: 60 A NATAA. In this case the total energy is higher, 41.8, confirming that the Arg substitution is an improvement. Experiments confirmed that the Leu60Arg substitution improved heat stability at 60 °C by about a factor of two.[2]

More advanced calculations

1. Try the calculation several times

Change the "Number of independent trajectories" from 1 to 5, which will run the same calculation five times. Each calculation may give slightly different results. You will receive five pdb files.

2. Expand the range of possible side chain rotamers

The conformation of the amino acid side chains is defined by angles chi 1, chi 2, chi 3, and chi 4. For phenylalanine, two angles, chi 1 and chi 2, define the conformation; for lysine, all four angles are needed to define the conformation. The command EX 1 EX 2 tells Rosetta Design to include additional rotamers for chi 1 and chi 2 at ±1 standard deviation from the mean chi angle for each rotamer for buried residues. The command USE_INPUT_SC means include the native rotamer, since it may be an unusual one. For 60, there is no USE_INPUT_SC because the initial amino acid is being replaced. The EX 1 and EX 2 tells Rosetta Design to try extra rotamers. In this case, the prediction is the same - Arg - but the energy values will be lower.

NATRO
start

47 A NATAA USE_INPUT_SC EX 1 EX 2
56 A NATAA USE_INPUT_SC EX 1 EX 2
57 A NATAA USE_INPUT_SC EX 1 EX 2
58 A NATAA USE_INPUT_SC EX 1 EX 2
59 A NATAA USE_INPUT_SC EX 1 EX 2
60 A ALLAA EX 1 EX 2
61 A NATAA USE_INPUT_SC EX 1 EX 2
62 A NATAA USE_INPUT_SC EX 1 EX 2
63 A NATAA USE_INPUT_SC EX 1 EX 2
64 A NATAA USE_INPUT_SC EX 1 EX 2

Other suggestions

It may be easier to view the results pdb file if you open the file as a spreadsheet with Excel or similar program. (The data is in tabular form.) The energy scores section has energy values broken down by a number of categories across different columns and values for each amino acid going down in rows.

References

1. Y. Liu and B. Kuhlman, *Nucl. Acids Res.*, 2006, **34**, W235−W238.

2. B. J. Jones, H. Y. Lim, J. Huang and R. J. Kazlauskas, *Biochemistry*, 2017, **56**, 6521–6532.