Solvent-transfer models to study the energetics of the hydrophobic effect

To understand the energetics to bury an amino acid—whether during protein folding or at a protein-protein interface—we need to *model* the process in a way that is simpler than a complete protein structure. One common, powerful approach is a "solvent transfer" model. The idea is simple: measure the energy to move small molecules from water into various solvents, and then use that information to try to come up with emprical rules that reproduce the transfer energy. The solvent is a *model* for the protein interior that may capture some molecular features well, and other aspects poorly. We can use the rules we come up with for solvent transfer to predict the energetics of moving an amino acid into the protein interior.

The data

In this exercise, we are going to explore the transfer energy for amino acids going from water into four different solvents: 1-octanol, N-methylacetamide, ethanol, and hexanes. The data we are going to use comes from transfer of blocked peptides of the form:

$$Ala - X - Ala$$

where X is the amino acid of interest. The transfer free energy of $Ala-Gly-Ala$ is the reference free energy ($0\ kJ \cdot mol^{-1}$); all transfer energies are relative to this value. The measured values come from Damodara & Song (1986) *JBC* (https://www.ncbi.nlm.nih.gov/pubmed/3711086) and Wimley, Creamer & White (1996) *Biochemistry* (https://www.ncbi.nlm.nih.gov/pubmed/8611495).

We will correlate these transfer free energies to the solvent-accessible surface areas (SASA) of $Ala - X - Ala$ peptides calculated in PyMOL. As with the transfer free energies, the SASA of the $Ala - Gly - Ala$ peptide has been subtracted from the SASA value of other peptides.

These data are given in the "transfer-free-energy.xlsx" file that came with this lab.

Approach

We are going to try to find a quantitative relationship between the surface area and transfer free energy into different solvents. We will ask what we can and cannot rerporduce, and hopefully gain some molecular insights into the solvent-transfer/protein-folding process along the way.

We are going to use R^2 as a metric in this analysis. Quick review: R^2 measures the fraction of the variation in some obsevable that is explained by a computational model. An $R^2 = 1.0$ means the computational model explains everything; $R^2 = 0$ means the computationala model explains nothing. Having

an $R^2 = 0.6$ would mean that a computational model captures 60% of the variation in the observable; meaning 40% of our variation is explained by something that is not captured in our computational model. Such comparisons are powerful. If our computational model only describes the hydrophobic effect, whatever variation is left over is likely due to something besides the hydrophobic effect—a pointer towards what to improve in the computational model.

Since our goal is to figure out what correlates and why between transfer energy and accessible surface area, some of what we will do below will involve deleting points to improve the fit. This might feel like cheating. The point, however, is to understand *why* some amino acids work well and others do not.

Prompts

1. Is the surface area of carbon, nitrogen, or oxygen a better predictor of the octanol transfer free energy? Can you explain why this might be?

2. Graph the octanol transfer free energy for the neutral amino acids against the surface area you identified above. (Put the surface area on the x axis).

   (a) Are there points that systematically fall off the line?

   (b) What do they have in common?

   (c) What molecular explanation might explain this?

3. Remove the weirdo data points you identified in #2.

   (a) How many $kJ \cdot mol^{-1}$ of free energy do you gain for transferring a square angstrom of carbon to octanol?

   (b) What is the R^2 for the linear fit?

   (c) What does this mean molecularly?

4. Again, using the non-weirdo points from #2:

   (a) How well do the results correlate between octanol, NMA, EtOH, and hexane?

   (b) Make a table for the four solvents that has two columns: molecular features of that solvent that model the protein interior well; molecular features of that solvent that model the protein interior poorly.

   (c) Given how much these four solvent models differ from one another, how different might you expect transfer to the protein interior is from any one of these models?

5. These surface areas were calculated using the structures in the "pdb" directory, using a single conformation for each sidechain.

   (a) Choose two (non-weirdo) amino acids of your choice. Load up the structures from the "pdb" directory in PyMOL. Use the mutagenesis wizard to put the sidechain in different conformations. Measure the solvent-accessible surface area of carbon for each conformation.

i. Load the pdb file
ii. set dot_solvent, 1
iii. get_area name C*
iv. Introduce mutation using the Wizard and hit "Apply".
v. Repeat iii. and iv. for all rotatmers.

(b) How big is the spread of SASA relative to the trend $\Delta G_{transfer}$ vs. SASA?

6. The mathematical model below was carefully parameterized to reproduce the energetics of protein folding associated with the hydrophobic effect. ΔASA is calculated in Å^2, energies are given in $J \cdot mol^{-1}$, and temperatures are in K.

$$\Delta C_{p,nonpolar} = 1.883 \times \Delta ASA_{nonpolar}$$

$$\Delta H_{nonpolar} = -35.313 \times \Delta ASA_{nonpolar} + \Delta C_{p,nonpolar}\,(T - 333.15)$$

$$\Delta S_{nonpolar,Tref} = \Delta C_{p,nonpolar} \times ln\left(\frac{T}{385.15}\right)$$

(a) Use this mathematical model to estimate $\Delta G_{transfer}$ for each amino acid.

(b) How well does this corrleate with the octanol scale? Does this "protein interior" model match to the extent you predicted in 4(c)?

7. Combining what you observed in #4, #5, and #6: can you estimate the confidence you would have in an energy calculated using a surface area calculation? (You may not have the stats skills to do this rigorously. That's fine. The goal is to wrestle with it intellectually).