Hyun-Joon Yang

yanghyun@usc.edu

BISC 481

HW 5

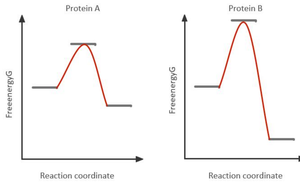
In [1]:
```
from IPython.display import Image
from IPython.core.display import HTML
HTML("""
<style>
.output_png {
    display: table-cell;
    text-align: center;
    vertical-align: middle;
}
.output_jpeg {
    display: table-cell;
    text-align: center;
    vertical-align: middle;
}
.center {
    text-align: center;
    margin: auto;
}
</style>
""")
```

Out[1]:

1) Two proteins A and B fold into their native structure at room temperature. The diagrams below show the free energy diagram of the hypothetical folding process. (10 pts)

In [2]:
```
Image('img1.jpg')
```

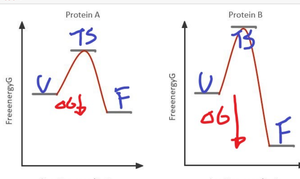
Out[2]:



A) In the diagrams below label the unfolded state [U], the folded state [F], the transition state [TS] and the free energy difference $\Delta G_{U\rightarrow F}$ between the unfolded [U] and folded state [F].

In [3]:
```
Image('1-A.jpg')
```

Out[3]:



B) Which of the two proteins A or B is expected to fold faster into its structure?

Protein A since the height of the energy barrier (ΔG^{TS}) is lower

C) One of the two proteins folds with a folding rate of $k_f = 0.013\ sec^{-1}$ and the other at a rate of $k_f = 1.3\ sec^{-1}$. Which of the two folding rates belongs to the fastest folding protein? What would the time constant τ be for the two proteins?

Folding rate of 1.3 sec^{-1} belongs to the faster protein (protein A)

$$\tau = \frac{1}{k_f}$$

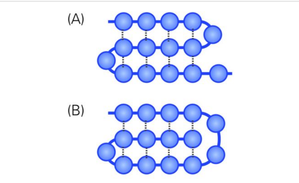
Protein A: $\tau = \frac{1}{1.3} = 0.77\ sec$

Protein B: $\tau = \frac{1}{0.013} = 77\ sec$

2) Shown are the hypothetical folds of two proteins (A) and (B). A residue is indicated by a sphere and a contact between two residues by a dashed line. (15 pts)

In [4]:
```
Image('img2.jpg')
```

Out[4]:



A) What is the contact order of a protein? Calculate the contact order for protein A and B

$$CO = \frac{1}{N_{res}C}\sum^{C}\Delta S_{i,j}$$

A: $\frac{1}{12*4}[(8+6+4+2)+(8+6+4+2)] = \frac{1}{2}$

B: $\frac{1}{12*4}[(11+11+11+11)+(8+6+4+2)] = \frac{8}{15}$

B) Explain which of the two proteins is expected to fold faster.

Protein A since lower contact order is related to faster the folding

3) How is the melting temperature of a protein defined? (10 pts)

Melting temperature of a protein is the temperature at which number of folded proteins equals the number of unfolded proteins

4) How can hydrophobic interactions between residues provide favorable entropy for protein folding? (10 pts)

The hydrophobic effect restricts rotations of water molecules reducing entropy, which in turn allow ΔG to become negative

5) Lets define a simplistic protein model, in which each residue can assume 6 distinct conformations in unfolded state, and just one unique conformation in the folded state. Show your calculations and equations. (15 pts)

A) How many different conformations a protein of 100 amino acids can assume in unfolded state?

$$6^{100}$$

B) What is the entropy contribution to the Free Energy of folding ΔG for this protein? (The gas constant R= 8.314 J/(K*mol)

$$\Delta S_{U\rightarrow F} = -N_{res}R\ln 6$$

$$-100(8.314)\ln 6 = -1489.67 J$$

6) What is the biological function of a chaperone protein? List different types of chaperons (10 pts)

Function of chaperone proteins is to assist folding of macromolecules or in assembly of large macromolecular complexes.

Types of chaperons:

- small heat-shock proteins
- Hsp70
- chaperonins
- Hsp90
- Hsp100

7) List the key computational steps present in all comparative (homology) modeling algorithms. (10 pts)

- template search
- template alignment
- model building
- model evaluation

8) What 3 methods are employed by DeepMind's AlphaFold protein folding algorithm (see the AlphaFold paper attached)? Which of them yielded the best results in CASP13 benchmark? (10 pts)

- Memory-augmented simulated annealing with neural fragment generation with GDT-net potential
- Memory-augmented simulated annealing with neural fragment generation with distance potential
- Repeated gradient descent of distance potential

Gradient descent of distance potential yielded best results

9) What was the key approach used in all 3 AlphaFold methods to derive residue distance predictions? (10 pts)

Distance predictions from multiple sequence alignment

10) Analyze the results of CASP13 modeling assessment. (25 pts)

We will focus on the target T0990 and predictions made by AphaFold.
Go to this page with CASP13 results https://predictioncenter.org/casp13/results.cgi?tr_type=all&offset=T0979 and find target T0990 and its three domains D1, D2 and D3:

- Click on the T0990 structure (protein on the black background) which will download its .pdb file to your computer.
- Click on T0990 (text), which will bring you to the result page. Download the best model .pdb file (A7D group= AlphaFold).
- Download the best models for all three T0990 domains.

Load the T0990 structure and all four models to Pymol (you have installed it for Part 1 of the course) and superimpose the structure and the models the align command, e.g. `align T0990////CA, T0990TS043_1////CA`.

A) What are RMSD values between the T0990 structure and the models?

- T0990: 35.269
- T0990 D1: 1.096
- T0990 D2: 29.657
- T0990 D3: 4.346

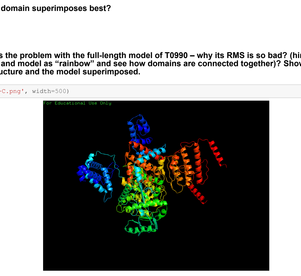
B) Which domain superimposes best?

D1

C) What is the problem with the full-length model of T0990 – why its RMS is so bad? (hint: color the structure and model as "rainbow" and see how domains are connected together)? Show the picture of the structure and the model superimposed.

In [5]:
```
Image('10-C.png', width=500)
```

Out[5]:



It seems the full-length model was "assembled" wrong - the domains are in incorrect positions.