BT3020 Structural Biology Problem Set 3

Use all available resources and peer group learning.
Deadline to submit handwritten answer sheets is 2nd May 2025
The google form link to upload is https://forms.gle/aFveimMrwa12pi296

1) Predicting the 3D structure of proteins is one of the fundamental grand challenges in biology. AlphaFold, the state-of-the-art AI system developed by Google DeepMind, is able to computationally predict protein structures with unprecedented accuracy and speed. Working in partnership with EMBL European Bioinformatics Institute (EMBL-EBI), over 200 million protein structure predictions by AlphaFold are freely and openly available to the global scientific community.
a) For a protein with the Uniprot sequence ID (P83916), search the RCSB website to identify whether the experimentally determined structure of the protein is available. Submit this protein sequence as a query in (https://www.rcsb.org/search/advanced/sequence) and search (Use a 100% identity cutoff). Examine the structure. How many amino acids constitute the protein sequence? Does the structure correspond to that of the full-length protein?
b) Next go to the AlphaFold database (https://alphafold.ebi.ac.uk/) to retrieve the structure of the full-length protein as predicted by AI methods. Use the Uniprot ID to search.
Examine the experimental and predicted 3D structures. Comment on the differences between the AlphaFold prediction compared to the experimental structure? Comment on the reliability of the predicted structure? (Use ChatGpt or equivalent tools to find out how to evaluate the reliability of AlphaFold predictions).

2) **Reading exercise:** A paper describing an undergraduate level experiment for calculating free energy of folding of a model protein called lysozyme is given for your reference. I strongly suggest you read this to get a feel for how the experiment is done.
Background: The stabilization free energy of a folded protein can be estimated by measuring shifts in the equilibrium from folded to unfolded states in response to physical conditions like change in temperature or chemical denaturants like urea or Guanidine hydrochloride. One approach is to estimate the fraction of folded protein present in the sample using a measure of functional activity in response to different denaturants. Usually a two-state sigmoidal process going from folded to unfolded state is assumed with increase in temperature/chemical denaturant. The initial and final states are 100 % folded and 100% unfolded, respectively and the intermediate values of fraction of folded states are normalized with respect to these.

3) Data from such experiments described above can be used to calculate the ΔG value for the folding process of a protein using the formula $\Delta G = -RT \ln((1-F_f)/F_f)$, where F_f is the fraction folded. One such data set for a protein is given below. Calculate ΔG of folding at 4°C using a plot of ΔG as a function of temperature.

Temp ˚C	Fraction folded (F_f)
20	0.995
30	0.962
40	0.756
50	0.289
60	0.039
70	0.004

4) The abstract of a journal paper describing an actual experiment on the protein Bovine Pancreatic Trypsin Inhibitor (BPTI) is given below. First read up on BPTI from your textbook/wiki etc. Wild type is the original natural sequence and mutations are substitutions of an amino acid with another, for example,

C30G/C51A means that Cys30 and Cys51 are mutated to Gly and Ala, respectively.
The following three pieces of information are given below. The abstract/summary of the experiment, the tertiary structure of BPTI and the unfolding profile data.

Experiment Abstract: We have used a combination of spectroscopic techniques to assess the thermodynamic consequences of paired amino acid substitutions at positions 30 and 51 in bovine pancreatic trypsin inhibitor (BPTI). Correctly folded, wild type BPTI contains a disulfide at the 30−51 positions, with the non-backbone atoms of this cystine being relatively solvent inaccessible. Mutants missing this buried 30−51 disulfide adopt a conformation very similar to that of the native state of wild type BPTI, although they are severely destabilized relative to the wild type molecule. We have conducted a systematic effort to find the energetically most favorable substitution for this buried 30−51 disulfide in BPTI. To this end, we have studied and characterized the chemically induced unfolding of BPTI using guanidine hydrochloride for a family of mutants in which the amino acid residue(s) at positions 30 and 51 have been systematically altered. Specifically, we studied the unfolding of the following 8 sets of paired-double substitution mutants: C30A/C51A, C30V/C51A, C30G/C51A, C30S/C51A, C30T/C51A, C30A/C51S, C30S/C51S, and C30G/C51M.

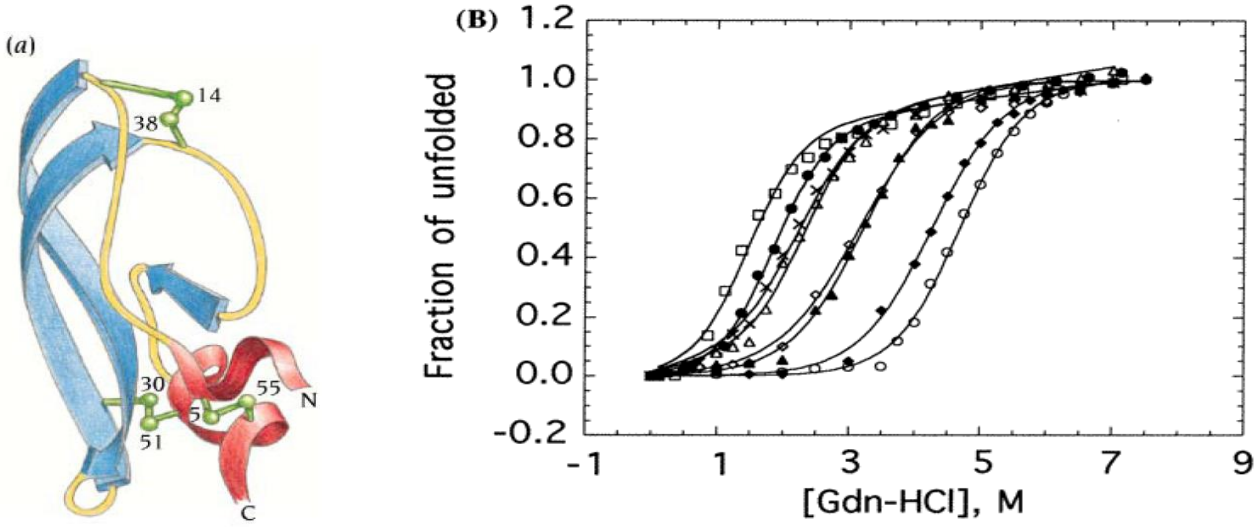

Figure legend B: Guanidine hydrochloride-induced (Gdn-HCl) equilibrium unfolding of 30−51 disulfide mutants of BPTI: C30A/C51A (open circles); C30V/C51A (filled diamonds); C30G/C51A (open triangles); C30T/C51A (open diamonds); C30S/C51A (filled triangles); C30A/C51S (×); C30S/C51S (filled circles); C30G/C51M (open squares). Unfolding was monitored by CD spectroscopy at 222 nm, in 10 mM potassium phosphate (pH 7.0) and 100 mM NaCl, at 25 °C. The plots show the normalized data (fraction of unfolded, 0% to 100% unfolded). The solid lines are the best fit of the experimental data to a two-state process.

Using these three pieces of data above, try to answer the questions below. You could download the structure from the PDB to have a better understanding of the structural features of the protein. Use the C30A/C51A double mutant as a relative control protein to carry out comparisons and attempt answers.

a) Did the mutations change the overall conformation of BPTI?
b) Which set of side chains are the most destabilizing/stabilizing?
c) Are non-polar side chains more destabilizing or stabilizing than polar side chains? Is this effect additive? Propose an explanation for your observation.
d) Identify the mutant with the most sterically (by size) similar side chain to the wild-type protein. Compare the relative stability of this mutant to the control and provide a possible explanation for the observed behavior.