Name: Aryan Shirbhate
Roll: BS22B009

Assignment (3)

Q1] a) Search experimental structures on RCSB PDB

Go to uniprot → search P83916 → CBX1 → Scroll to sequences → Copy Fasta format seq → Search on RCSB PDB → Paste Fasta seq and set cutoff sequence identity to 100 %. → click Search.

↓

- In the results we got 8 structures all corresponding to Chromobox protein homolog 1 (CBX1) which matches P83916

Results: (1) 3Q6S
(2) 5T1G
(3) 2FMM
(4) 1DZ1
(5) 1AP0
(6) 1GUW
(7) 3F2U
(8) 6D07

Amino acid count:→ The full length of sequence for P83916 (CBX1) from Uniprot contains 185 amino acids

Does the experimental structure cover full length?

No the experimentally determined structures do not represent the full length protein (185 AA). All 8 structures retrived from the RCSB database, correspond to individual domains or fragments of protein.

- 3Q6S → Covers chromoshadow domain (approx 117 - 185 residues)
- 1GUW, 1AP0, 1DZ1 → Represent Chromodomain or its complex with other molecules (21-71 residues)

* 2FMM, 3F2U, 5T1G: Also map to parts of chromodomain or its complex with other molecules.

* 6D07: Represents a short peptide fragment interacting with chromodomain.

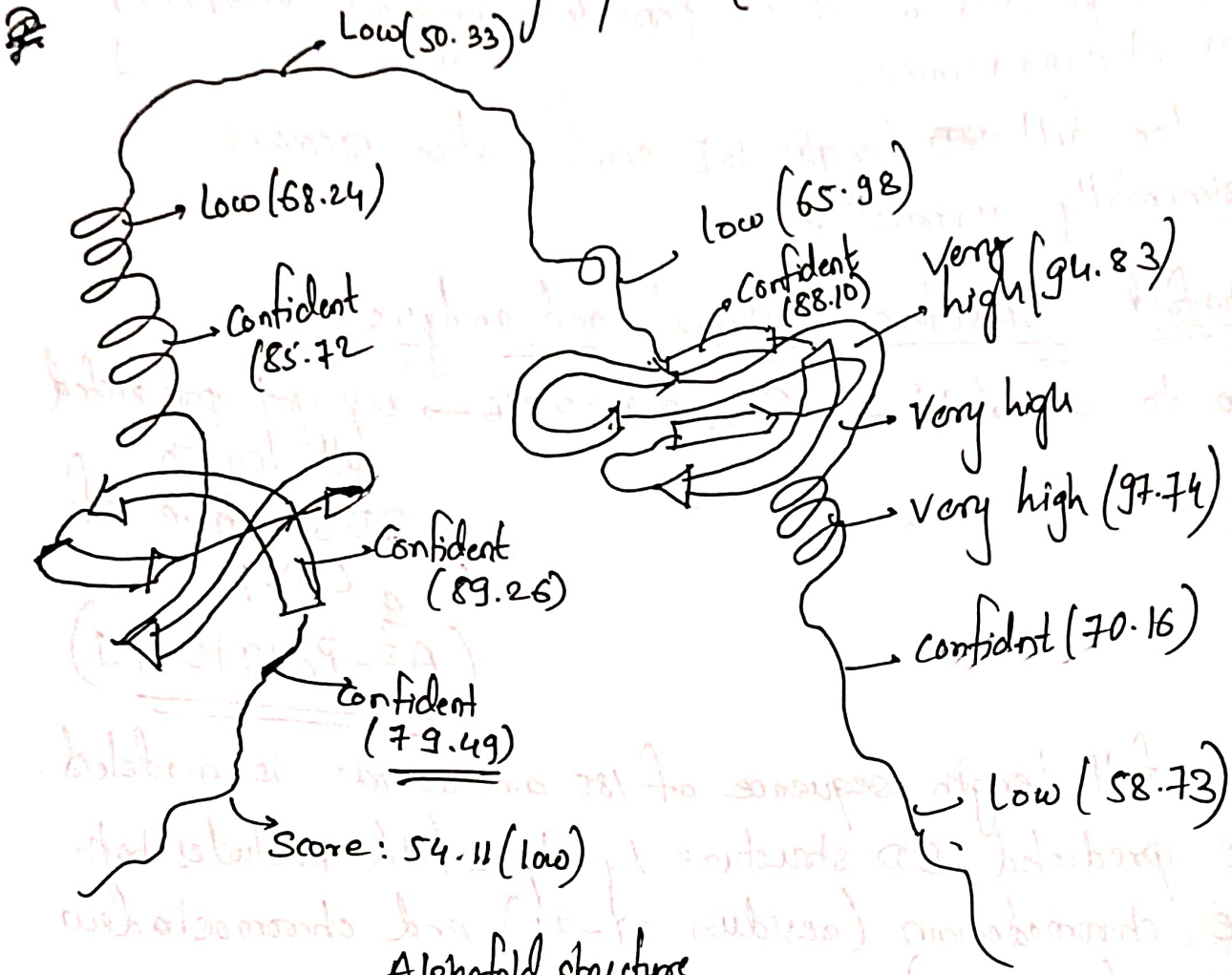
Thus the full length 3D conformation remains experimentally unresolved.

(b) Alphafold Structure retrival and analysis

Go to alphafold → Enter P83916 → We got predicted full length 3D structure of CBX1 (AF-P83916-F1)

- The full length sequence of 185 aminoacids is modeled.
- The predicted 3D structure by alphafold includes both the chromodomain (residues 21-71) and chromoshadow domain (117-185).
- The structure is color coded by confidence using pLDDT score (per residue confidence score).
  - Very high Confidence: (pLDDT > 90): Deep blue (mainly structured regions as α helices).
  - High ⇒ (70-90): Lighter blue regions still fairly reliable
  - Low to very low confidence: (< 70) Yellow to orange segments, indicates flexible / unstructured regions, especially in N terminal & residues (1-20) some loops.
- Alphafold model showed entire protein with both domains included.
- Alphafold gives modeled (but with low confidence) while Experimental (RSCB) gives not resolved (often omitted in crystallization).

- Experimental (PCSB) only includes partial domain while Alpha fold has full length protein (1-185 residues).



Alphafold structure

(AF-P83916-F1)

Reliability of Alphafold Prediction:

- pLDDT > 90 - Very high confidence (trustworthy structure)
- pLDDT 70-90: Good Confidence.
- pLDDT < 70: Lower confidence, often disordered or flexible regions.

- In this model:- Most core domains (chromodomain and chromoshadow domain) are predicted ~~with~~ with high to very high confidence supporting structural accuracy.
- Some loop regions and termini are ~~accurately~~ modeled with low or very low confidence suggesting intrinsic disorder or flexibility.

(2) Reading Exercise

(3) The Gibbs free energy of protein folding provides insights into stability of a protein at a given temperature. Negative ΔG indicates spontaneous folding, while +ve ΔG value suggests unfolding is favoured.

ΔG is calculated by this formula:

$$\Delta G = -RT \ln\left(\frac{1-F_f}{F_f}\right)$$

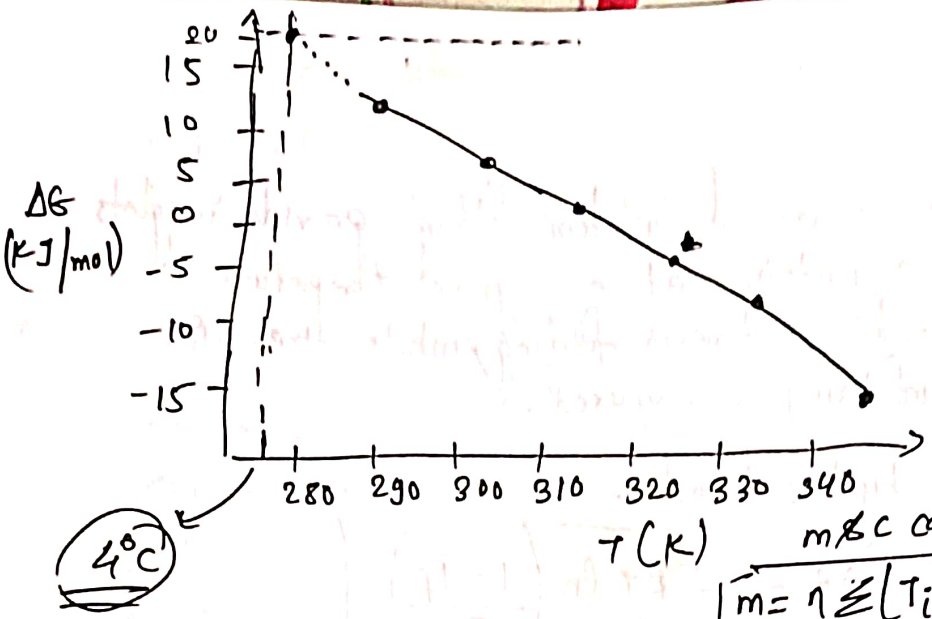
Where ΔG = Gibbs free energy change in J/mol
R = gas constant = 8.314 J/mol-k
T = Temperature in Kelvin
F_f = fraction folded (between 0 and 1)

Data (°C)	Fraction folded (F_f)
20 = 293.15 K	0.995
30 = 303.15 K	0.962
40 = 313.15 K	0.756
50 = 323.15 K	0.289
60 = 333.15 K	0.039
70 = 343.15 K	0.004

$$\Delta G = -8.314 \times T(K) \times \ln\left(\frac{1-F_f}{F_f}\right)$$

~~Data~~

Temp (K)	ΔG (kJ/mol)
293.15	12.901
303.15	8.144
313.15	2.944
323.15	−2.4189
333.15	−8.876
343.15	−15.741



m & c calculated by :⇒

$$m = \frac{n\Sigma(T_i \Delta G_i) - \Sigma T_i \Sigma \Delta G_i}{n\Sigma T_i^2 - (\Sigma T_i)^2}$$

$$c = \frac{\Sigma \Delta G_i - m \Sigma T_i}{n}$$

At 4°C (277.15 K)

$\Delta G = mT + C$

$\Delta G = -0.5704(T) + 180.96$

Slope $= R^2 = 0.995$

$\Delta G = -0.5704 \times 277.15 + 180.96$

$\Delta G = +22.88$ KJ/mol at 4°C

Ans.

(4) a) No ~~mutation~~ mutations did not change the overall confirmation of BPTI, the abstract clearly says "this buried 30-51 disulphide adopt to a confirmation very similar to that of native state of wild type ~~BPTI~~ BPTI, although they were severly destabalized relative to wild type."

- This means when thermodynamic stability decreases the global 3D structure remains largely intact suggesting minimal conformational change.

(b) from the unfolding curves (B figure) we assess stability using the mid point of denaturation (C_m) - the higher the C_m, more stable the protein.

- Most Stabalizing Mutant (Highest Cm): C30S | C51S (filled circles) shows the rightmost shift on the denaturation plot highest stability. The hydrophobic interaction between these likely ensures structural stability. The next most stable is C30V | C51A.

- The C30G | C51M mutant is most destabalizing. Glycine owning to its small size is likely to interact with methonine. Mets large size could destabalize protein by increasing solvent accessibility.

c] Non polar side chains tend to be more stabalizing than polar side chains but this stabalization is not necessarily additive. The C30S | C51A and C30A | C51S mutants show similar level of stability, both are only slightly more stable than C30S | C51S. This modest increase in stability can be attributed to the non polar nature of alanine's side chain. However the C30A | C51A mutant is significantly more stable than the others, suggesting that the combined effect of nonpolar side chains is not simply additive. The enhanced stability likely arises from favourable hydrophobic interactions.

d] The C30S | C51S mutant is sterically the most similar to the wild type protein because of the structural resemblance between cysteine and serine. Despite this the mutant is considerably less stable than the control C30A | C51A. This is because serine is polar and hydrophillic which allows it to interact with the surrounding solvent thereby disrupting the proteins internal structure and reducing its stability. Overall stabalization provided by non-polar residues is complex & cannot be fully explained by simple concepts

like additivity or the hydrophobic effect alone. However hydrophobic interactions clearly contribute significantly to the stability observed in control mutant.