Report I

Q1

For the two entries listed in the first section, evaluate the structures based on their pLDDT and PAE scores, examine the TED domains, and provide a detailed discussion of the annotated domain's function. (max 4 lines for each entry)

In the first entry, the overall pLDDT score of the RPP7 protein is 82.88, indicating a high level of predicted secondary structure confidence, particularly for the global fold, however, there is a reduction in secondary structure confidence in several loop regions and some surface helices which should be treated with caution for further treatment. Based on the PAE map, four distinct domains can be identified, with the intra-chain interactions between domains 1–3, 2–3, and 3–4 showing low predicted error, indicating confident relative positioning. TED domain annotations failed to capture the biologically relevant domain architecture of the protein, resulting in misleading representations. This NLR-type disease-resistance protein enables *Arabidopsis thaliana* to recognize specific avirulent strains and trigger a strong immune response that blocks infection, with the orientation of its loops, liner regions and surface helices likely playing a role in effector recognition and pathogen interaction. Therefore, despite the acceptable global confidence indicated by the pLDDT and PAE metrics, the low-accuracy loop and helix regions warrant careful consideration for further structural and functional interpretation.

In the second entry, the overall pLDDT score of the STRIP1 protein is 83.81, indicating a high level of predicted confidence in the global secondary-structure fold, however, the pLDDT drops below 0.5 in the intra-loop region, suggesting potential intrinsic disorder (UniProt automatically annotates this region (333-423) as disordered based on sequence analysis), and falls below 0.7 for several residues at the C-terminal region. Based on the PAE plot, two distinct domains can be identified,with low error in their relative positioning, connected by a potentially disordered loop region, whereas the TED domain annotations failed to identify any structural domains. Because this protein regulates cell morphology and cytoskeletal organization as part of a scaffolding complex, its interactions with partner proteins are crucial. There is only one cryo-EM structure of this protein available in the PDB (ID: 7k36), in which it is complexed with other structural partners and in this structure, the intra-chain loop between residues 333 and 423 is not resolved, further supporting its intrinsic disorder, whereas the C-terminal region participates in interactions through extended linker elements.

Q2

For the second entry, assess the structures with respect to their pLDDT, PAE, and TED domains, and provide a detailed discussion on disorders. (max 4 lines)

In this entry, only three helices displayed high pLDDT scores, while almost 70% of the protein is assigned low-confidence regions (pLDDT < 50) in terms of the secondary structure, strongly indicating intrinsic disorder across the global fold (intrinsically disordered protein). Based on the PAE map, three distinct structural regions can be identified, however, domain assignment cannot be reliably performed because (i) the pLDDT scores for these regions are very low and (ii) the predicted error for their relative positioning exceeds 30 Å, similarly, the TED annotation for domain 1 is not informative for the same reasons. It can be inferred that the model positioned these structural regions in a largely random manner. Although the prediction highlights the intrinsically disordered nature of the protein, the generated structural model is not appropriate for downstream analyses.

Q3

To identify the interface, please use the Python script provided in the week 4 section. Once the interface has been identified, calculate the ipLDDT, iPAE, and ipTM values by averaging the scores corresponding to the interface residues (You can do averaging using a short Python script or simply in Excel).

From rank 1 to rank 5, all predicted complexes exhibited different binding modes between IL-37 and the receptor. Low pLDDT scores observed in both the receptor and the cytokine indicate low confidence in the predicted secondary structure of the complex, while the PAE maps reveal a single-domain cytokine (domain 1) and two distinct receptor units (domain 2, 3), showing high uncertainty in the relative positioning of the receptor domains (domain 2-3), as well as a high positional error between the cytokine and the receptor (domain 1-2 and domain 1-3), in other words, complex formation. According to the computed interface scores, none of the predicted models exceed the iPTM threshold of 0.5, suggesting that the predicted interfaces are likely incorrect. Similarly, all models exhibited low iPLDDT and high iPAE values, further supporting the low confidence in the predicted complex. It can be concluded that AF2 did not perform well in capturing the structure of this novel cytokine–receptor complex.

Q4

For the fourth entry, assess the structures with respect to their pLDDT, PAE, and TED domains and consider the TmAlphaFold database to provide a discussion of the results with respect to membrane localization. Compare the structural differences between the AF2-predicted structure and the ground truth (PDB ID: 2LOS) by calculating the RMSD after structural alignment. Taking into account the pLDDT and PAE scores, as well as the ground truth structure, discuss the prediction accuracy of the AF2-generated model. Based on the structural accuracy, provide a discussion on the membrane localization. (max 6 lines)

In this entry, the overall pLDDT score of 62.62 reflects low secondary-structure confidence, and the PAE map further indicates the presence of four structurally distinct regions, consistent with four individual helices, while the TED domain annotation classifies nearly half of the protein as a single domain. For the membrane localization of this protein, CCTOP predicts three transmembrane segments; however, AF-TMDET embeds only two of segments into the membrane, leaving one predicted transmembrane segment in non-membrane region. When compared with the related entry PDB ID 2LOS, it becomes evident that AF2 mispredicted the orientation of the N-terminal transmembrane helix, and one of the TM segments is erroneously positioned parallel to the membrane plane in the AF2 model. In contrast, the 2LOS structure contains three parallel helices embedded in the membrane, while the terminal helix lies parallel to the membrane surface, most likely due to its aliphatic character. The RMSD value of 22.746 Å across all 112 pruned atom pairs further highlights the substantial global structural differences between the experimentally resolved structure and the AF2 prediction. Although AF2 is not aware of the membrane axis, TMAlphaFold further demonstrates in this case that the predicted structure is biologically/physically not plausible.

Q5

For the fifth entry, assess the structures with respect to their pLDDT, PAE, and TED domains, and survey the human myocilin "fragments" deposited in the PDB and compare them with the AlphaFold2-predicted model, highlighting any global structural differences (Please pay attention that the olfactomedin domain is the C-terminal fragment of the myocilin). Then check the UniProt database to determine the protein's function, and use this information to rationalize the structural differences observed between the computationally predicted model and the experimental structures.

In this entry, the PAE map indicates two distinct domains in myocilin: a coiled-coil domain (residues 74–184, also verified by UniProt sequence

analysis) and an olfactomedin domain (residues 244–503, adopting a -propeller fold), with very high local confidence (pLDDT) and high secondary structure accuracy, respectively. These domains are connected by linkers, loops, and additional disordered regions located at the N-terminal, C-terminal regions and between the domains, where the local confidence scores markedly decrease. The model also fails to accurately predict the relative positioning of the coiled-coil and olfactomedin domains, as reflected by the high PAE values in the inter-domain areas. Among the 24 experimental myocilin structures available in the PDB, only the olfactomedin domain is structurally resolved and corresponding region is present in the canonical sequence. According to the subcellular localization and function annotations in UniProt (with experimental evidence), the olfactomedin domain is secreted into the extracellular matrix following proteolytic cleavage at the linker region that connects it to the rest of the myocilin protein, which provides the underlying reason why structural determination has been limited to the olfactomedin domain. When the AF2-predicted structure is superimposed with all available myocilin structures in the PDB, a surface helix located at the interface between blades E and A becomes apparent. In all structures, this surface helix at the E–A blade interface is assigned to blade A, except in the structure with PDB ID 6PKF, where it is instead classified as part of blade E. This highlights the importance of this side helix as a structural support element contributing to the overall integrity of the olfactomedin domain.

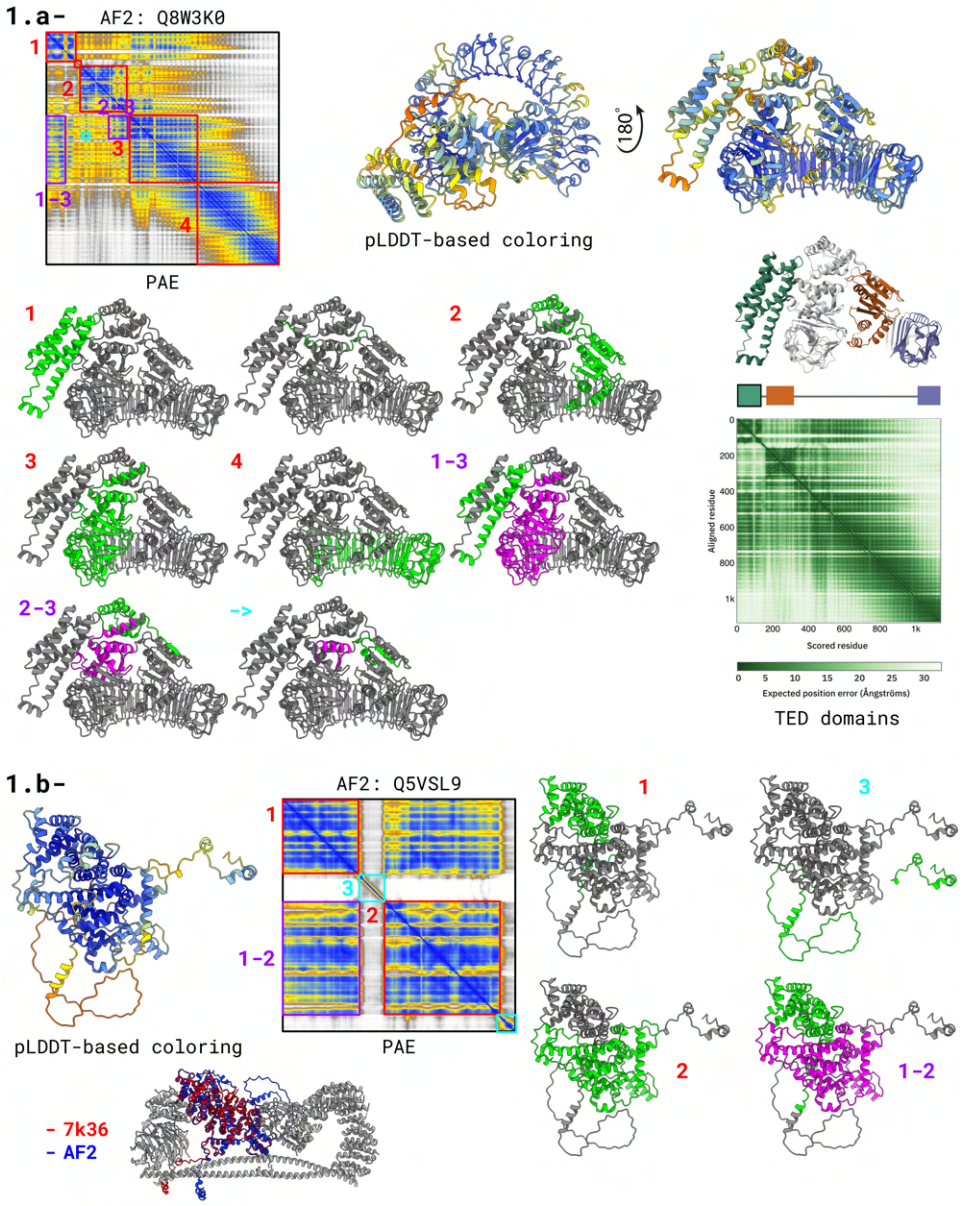
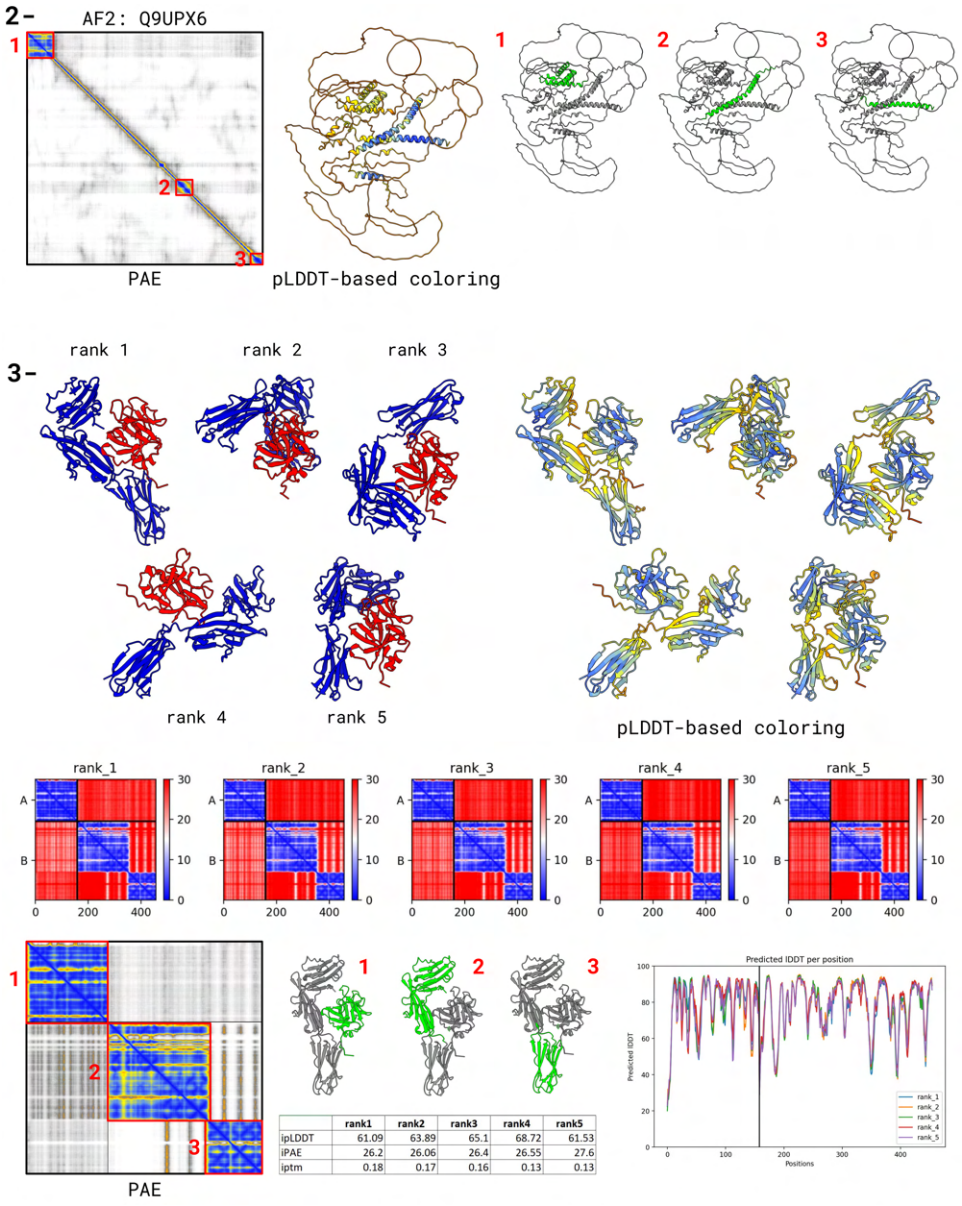

1.a-
AF2: Q8W3K0
1
2
2-3
3
1-3
4
PAE
pLDDT-based coloring
180°
1
2
3
4
1-3
2-3
->
Aligned residue
Scored residue
Expected position error (Ångströms)
TED domains
1.b-
AF2: Q5VSL9
1
3
2
1-2
pLDDT-based coloring
PAE
1
3
2
1-2
- 7k36
- AF2

Fig 1



	rank1	rank2	rank3	rank4	rank5
ipLDDT	61.09	63.89	65.1	68.72	61.53
iPAE	26.2	26.06	26.4	26.55	27.6
iptm	0.18	0.17	0.16	0.13	0.13

Fig 2

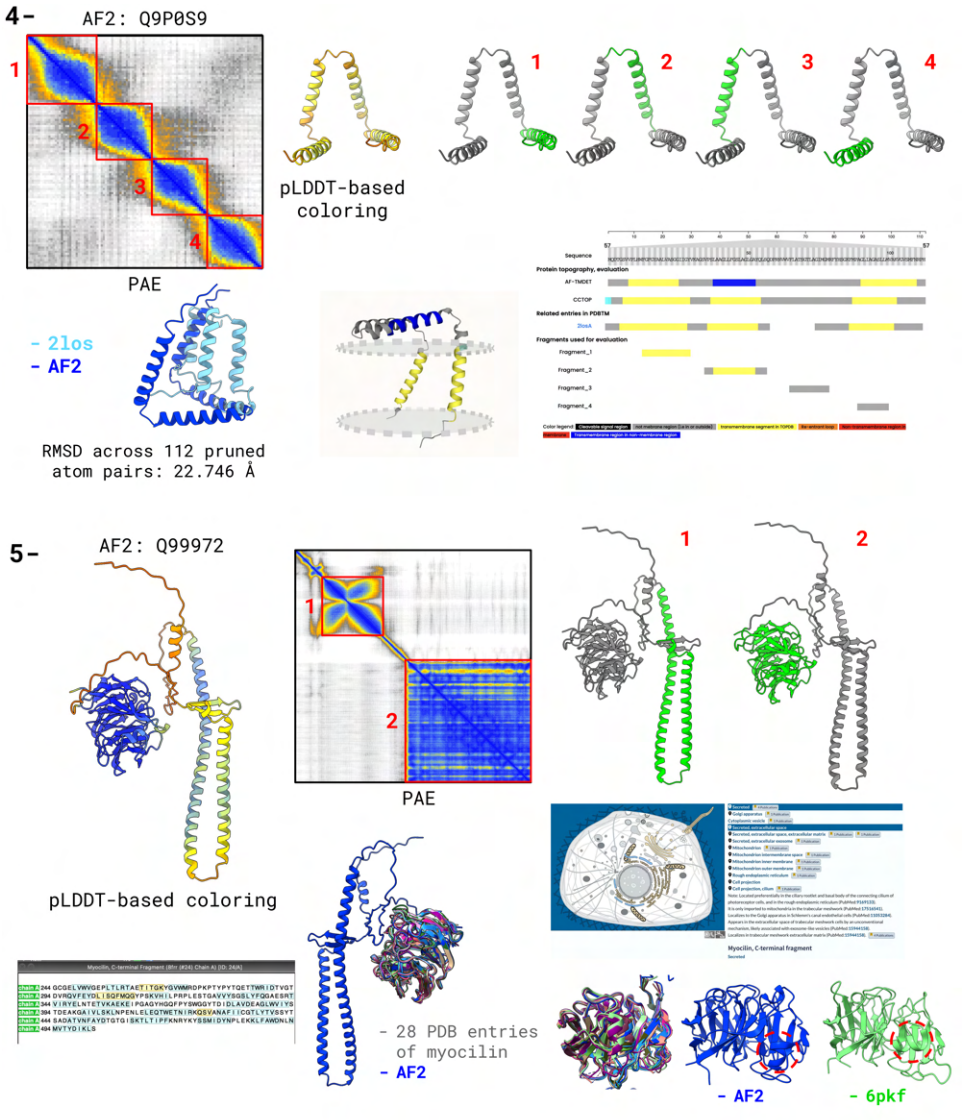

4-
AF2: Q9P0S9
1
2
3
4
PAE
pLDDT-based coloring
1
2
3
4
- 2los
- AF2
RMSD across 112 pruned atom pairs: 22.746 Å
5-
AF2: Q99972
pLDDT-based coloring
1
2
PAE
1
2
- 28 PDB entries of myocilin
- AF2
- AF2
- 6pkf

Fig 3