Evolutionary and Structural Analysis of Pathogen Proteins.

Final year UG project 2025-26

Learning Agreements

Learning Agreement

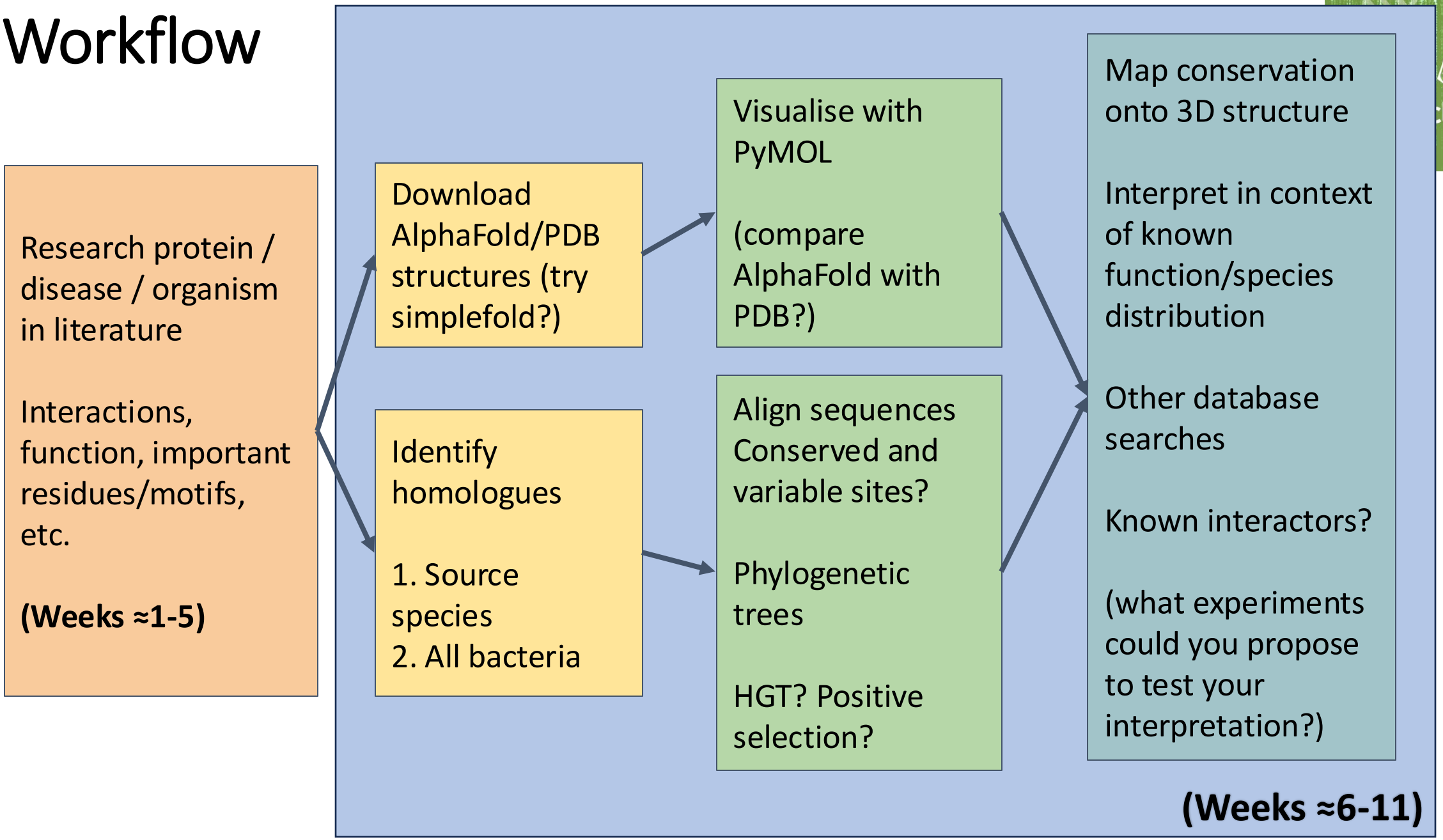
I have (almost) all of your signed learning agreements

Now **you** upload the countersigned copy on MyPlace **as a PDF file**

[upload link]

The Project

Workflow


Research protein / disease / organism in literature
Interactions, function, important residues/motifs, etc.
(Weeks ≈1-5)
Download AlphaFold/PDB structures (try simplefold?)
Identify homologues
1. Source species
2. All bacteria
Visualise with PyMOL
(compare AlphaFold with PDB?)
Align sequences Conserved and variable sites?
Phylogenetic trees
HGT? Positive selection?
Map conservation onto 3D structure
Interpret in context of known function/species distribution
Other database searches
Known interactors?
(what experiments could you propose to test your interpretation?)
(Weeks ≈6-11)

Candidate proteins – start points

Any changes needed?

Organism	Host	Gene/Protein	PHI accession	Student
Escherichia coli	*Homo sapiens*	*espY*	PHI:8647	LB
Shigella flexneri	*Homo sapiens*	*ipaJ*	PHI:9253	LT
Candida albicans	*Mus musculus*	*sap6*	PHI:10193	IM
Pseudomonas aeruginosa	*Homo sapiens*	*tplE*	PHI:6646	AE
Vibrio vulnificus	*Mus musculus*	*vvhA*	PHI:6877	JT

http://www.phi-base.org/

Literature Review

How is yours going?

- Could/did you find…
  - Supporting literature for function/virulence/mechanism/biochemical activity?
  - Sequence homologues?
  - Structural predictions/solved structures?
  - Any other useful data
    - Protein-protein interactions?
    - Experimental data (e.g. GEO data for knockout experiments)?

- **What did you find out?**

Topics that could be covered

The host-pathogen system [What is it? Importance and disease context? What molecules mediate the conflict?]

Your protein (family) [Function? What does it interact with? Structures? Sequences available? Organism diversity?]

Sequence-structure-function relationships [e.g. how does sequence conservation relate to structure and/or function]

Databases [e.g. Phi-Base, UniProt, RCSB/PDB, EBI's AlphaFold database]

Bioinformatics methods/tools [e.g. AlphaFold, phylogenetic methods, dN/dS (positive selection)]

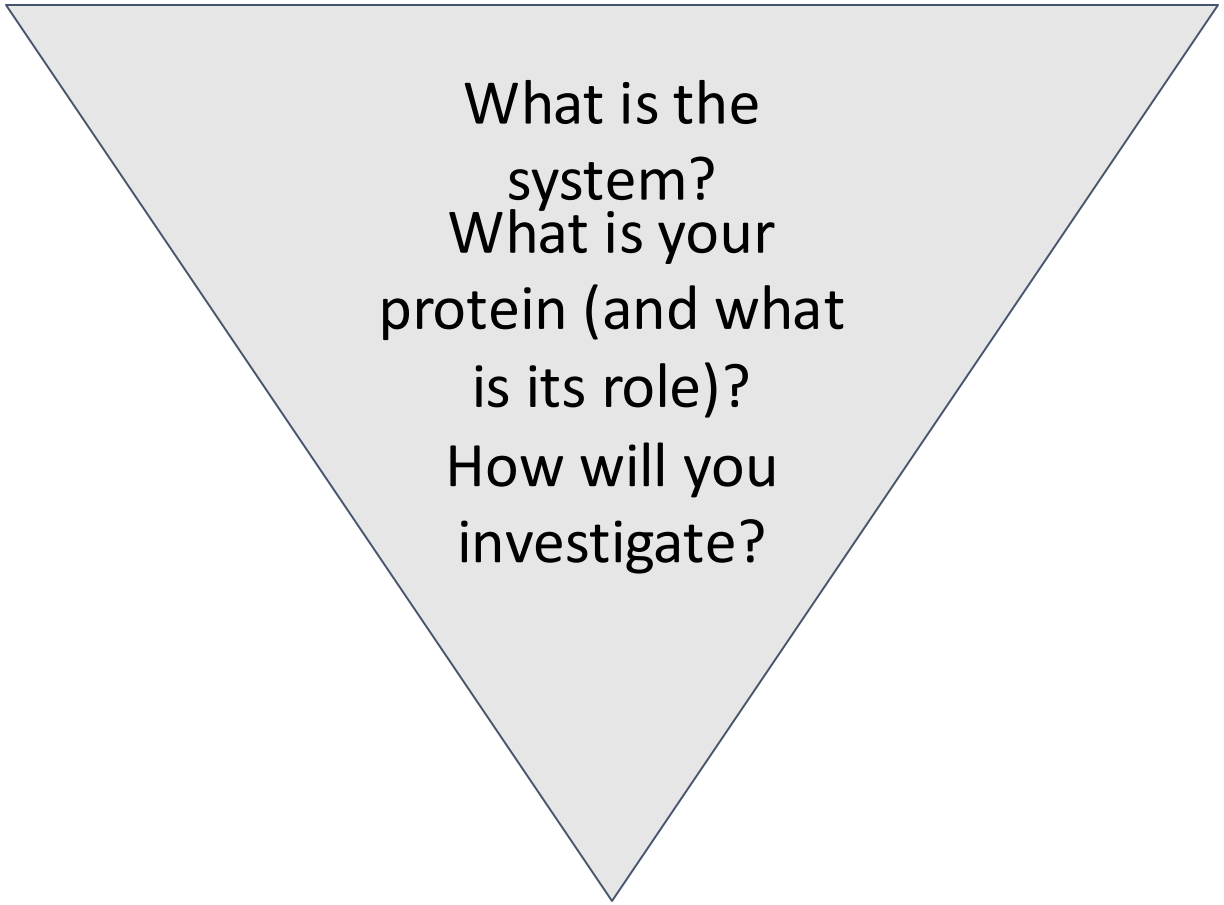
What I'm looking for in an introduction

- **Either** a *narrative account of your protein and what's known about it* **or** a *fully-described critical analysis*
  - This project suits a narrative account best, in my opinion
- Introduce the context of the protein
  - What is the pathogen? What is the disease caused? What is the host?
  - What disease symptoms/signs is the protein thought to be responsible for?
  - What specific role or pathway, if any, is known for the protein?
- Tell me about the protein
  - Is it big? Is it small? Is it secreted, a surface protein, or something else?
  - Is it an enzyme? Is it a binding protein? Does it do something else?
  - Are any specific parts of the protein known to have a contribution to function?
  - Are relatives of this protein known in other organisms and if so, which? What is known about those relatives?
- What are ***you*** going to do?

Section structure – think about this when reading papers

“Inverted pyramid” - start general, become more specific

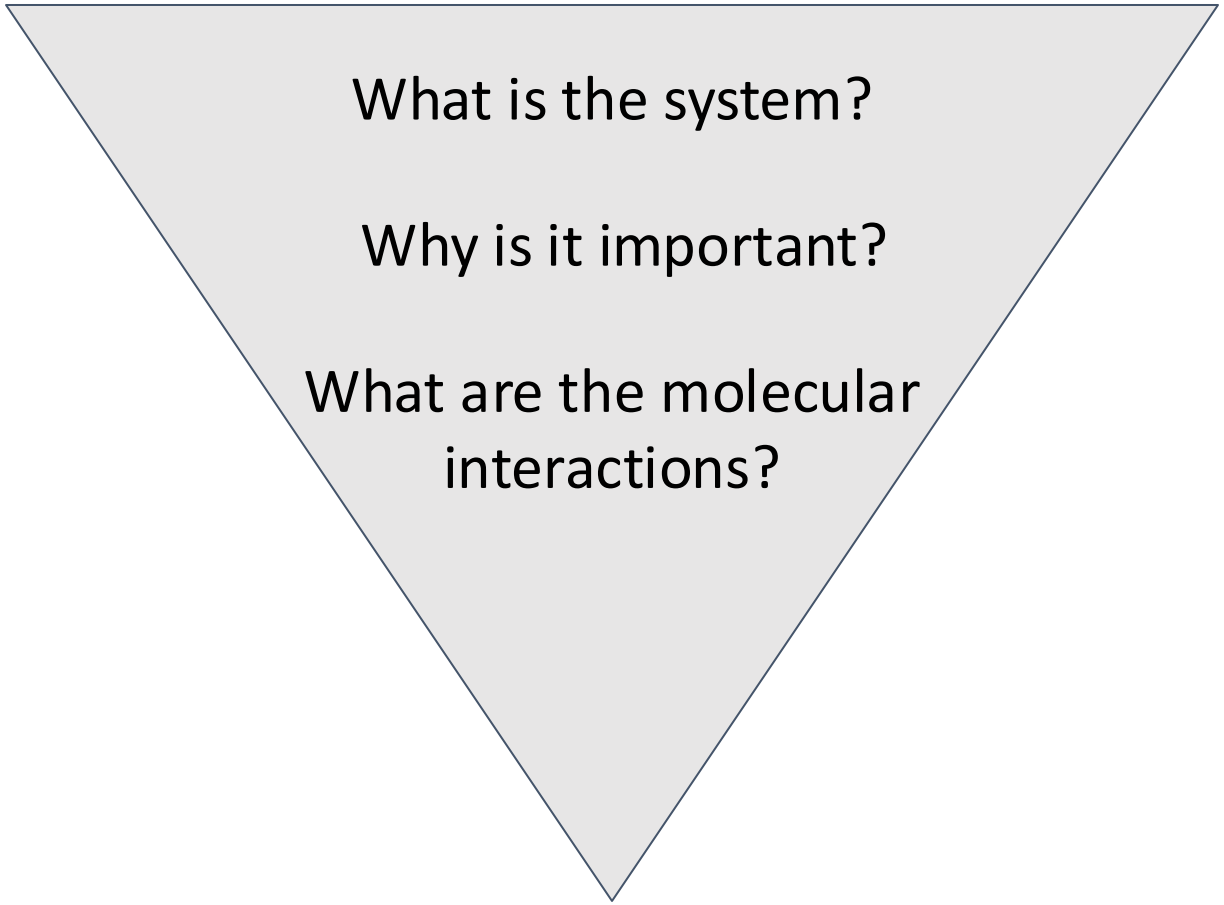
Overall



Section structure – where does this paper fit in?

“Inverted pyramid” - start general, become more specific

Host-pathogen system



Where the marks come from

	0-20	21-39	40-49	50-59	60-69	70+	Mark contribution
Introduction - content	Does not present material relevant to the issues. Insecure grasp of concepts. Copying* from other sources (even if source is cited). Failure to cite all sources.	Unsatisfactory depth of knowledge. Substantive errors. OR over-reliance on other work (e.g. close paraphrasing*, - perhaps with lack of understanding of the material).	Frequent use of secondary or out-of-date material. Possibly some substantive errors	Some use of secondary sources or out-of-date material. Some errors, indicating insecure grasp of issues. But does contain pertinent information appropriately discussed.	Appropriate and up-to- date original material. Few, if any, minor errors and no substantive errors.	Evidence of critical thought and analysis of material.	
Introduction - purpose	Aims and relevance to field omitted.	Poor placement with respect to field. No or incorrect hypotheses.	Poor placement and/or weak description of hypotheses.	Presents the major theories but weak or unclear hypotheses.	Study placed in context with clear hypotheses	Identifies and explains the theoretical importance with clear hypotheses	20%

What do you want to talk about?

Next Week's Group Meetings

Monday 13th October, 1:30pm, HW324

Thursday 16th October, 10:30am, HW324

Topics to Discuss at Next Meeting

- How the literature search is going
- How are you managing your time?
- Share advice
  - How to find useful papers
  - What databases are helpful
  - What software tools might be useful