**Bioinformatics worksheet** – Prepare a word document that shows each of the following (screenshots are your friend).

Using the amino acid sequence of *Erwinia chrysanthemi* Pectate Lyase C determined from pdbID 1AIR.

1. Using the BLAST tool, verify that humans do not have a homologous protein. Make your life easy and type *homo sapiens* into the organism box. What is the highest score?
2. Do any eukaryotes have Pectate Lyase homologues? You can quickly determine this by using the BLAST tool and restricting the search to eukaryotes in the organism box.
   1. What organisms are the top five hits from?
   2. Using Google (or any other tool that you’d like to use), determine what type of organism these are? (e.g. plants, mammals, etc.)
3. Scroll down to the entry from Verticillium alfalfae VaMs.102 (this should be the ~3rd entry). Follow the gene link under ‘Related Information’. This takes you to the NCBI gene page for this protein. How many exons does this gene have? This can easily be found in the Genomic Context area.
4. Determine the gene sequence. The easiest way to get this is to click on the FASTA or GenBank link in the Genomic Context area. How many base pairs are in the gene? Paste the sequence into a Word document in FASTA format.
5. Determine the mRNA sequence of this gene. You can get to this sequence two ways. If you’re on the GenBank page, follow the link in the mRNA section of the ‘Features’. Alternatively, head back to the gene page and follow the Nucleotide link under ‘related information. How long is the mRNA sequence? Paste the sequence into the document from Problem 4 in FASTA format.
6. Translate the DNA gene sequence and compare it to the sequence you find from a translated mRNA. You’ll want to use the translator tool in ExPASy for the DNA – it might be useful for the mRNA as well for comparison purposes. How many stops are there in the translation from the DNA sequence?
7. Now do a sequence alignment using the Clustal W2 or Omega server (accessed from ExPASY). What is the sequence of the intron? After putting in the sequence in correct FAST format, make sure to select DNA in STEP 1 and it may be useful to format the output as CLUSTAL (with or without numbers).
8. Why is it important to work with mRNA sequence, NOT gene sequences if you’re a biochemist trying to study a protein?
9. Sketch a 2D Gel electrophoresis product if E. chrysanthemi Pectate Lyase and V. alfalfa Pectate Lyase are the only two proteins in a sample applied to the 2D gel. Hint: you need to determine the pI and MW of each.
10. Prepare a homology model using pdbID 1AIR as the template. Note that you’ll need to prepare a sequence alignment FASTA file as you did in your homework. Remember to save it as a .txt file. (see below for a refresher). Summarize the most noticeable differences. As a biochemist, what information can you gain from building this homology model?

If you have time:

1. Determine the intron sequences for the gene (hypothetical protein) just upstream from the pectate lyase (so VDBG\_08402).
2. Determine the function for the hypothetical protein (VDBG\_08402). You can find this information in the GenBank entry. Tryptophanyl-tRNA synthetase
3. Prepare a homology model of this protein using the best match by BLAST. Make sure to select the Protein Data Bank from the Database dropdown box.

Homology Model Directions from PS5:

* From the ExPASy page, select genomics then sequence alignment. As you can see, there are a lot of options – I like to use the ClustalW2 server, but you can try any of them.
* On the ClustalW2 site, input the sequences seen below.
* On the step 3 box, click on *more options* and select Pearson/FASTA as the output format.
* Submit this project and wait for the alignment to complete. This should be very quick.
* In this example, the sequences align very nicely, so there are not any gaps in the core region. Do note, however, that there are some dashes at the beginning and end of the unknown protein – this means that there are not any amino acids that align with the crystal structure at this point.
* Download the alignment file and paste it into a word document. Save this document as at .txt file (this is really important – Chimera won’t recognize the .doc or .docx format as an alignment file)

Open this file in Chimera (make sure to select aligned FASTA as the format).

* Select Structure 🡪 Modeller tools.
* The target will be the unknown protein and the template will be 3N73\_A. Make sure these are selected.
* Now select ok. If you are asked to input a code for Modeller, use MODELIRANJE
* Wait for the sequence to complete.