Bioinformatics Project Guidelines

1. **Pick 3 same proteins from 5 different organisms**
- List name of the protein and name of each organism
2. **Do a multiple sequence alignment**
- Show the results
- Explain the results in your own word
- **Explanation examples:**
- "… we can see that hemoglobin is quite similar in both chimpanzees and humans, but a bit different in rats, which is as expected because we are more similar to chimpanzees than rats."
- " … we can see that actin is not so much different among these organisms which was actually accepted because we know that actin is conserved among the species."
3. **Make phylogenetic tree**
- Show the results
- Explain the results in your own words
- **Explanation examples:**
- "… we can see that chimpanzees, humans, rats and cows are grouped together when compared with rose, which is expected since the rose is a plant while other organisms are mammals."
- "… we can see that humans and mice are grouped together, and chimpanzee is the outside group, however, we can also see that the support for the tree is 0.3, meaning that the results are not really reliable (only 30%)"
- "… we can see that the evolution rate between *A. thaliana* as an outgroup and other organisms is quite huge since the coefficient is 3."
4. **Pick one protein out of these and predict its 3D structure**
- The protein that you pick will ideally be "hypothetical" or uncharacterized one, whose function is not known (remember that you can find such a protein by first picking a protein that has already been researched a lot, and then finding, through BLAST, protein which is as similar as possible to that one, but it hasn't been researched so far
- Use swissprot/expasy or Phyre 2 (the one that we've used)
- Show the picture
- Explain the results
- Ideally, you will also try to refine the structure of the protein and compare the quality of protein's structure before and after the refinement
- **Explanation examples:**
- "… based on the picture, we can see that the protein is quite different from the regular protein whose structure is already predicted. But it is not surprising since the software is "only 50%" sure that this is the right conformation. Also, we can see that the templates which are used in the prediction of my protein are actually based on proteins quite smaller than mine, meaning that most probably only the part of my protein is shown"
5. **Validate your 3D structure**
- Validate using at least Ramachandran plot (more is welcome)
- E.g. Verify 3D, Dfire
- Explain the results
- **Explanation examples:**

- "… Our initially predicted protein has 78% of residues in allowed region. The results are not so good, but it is expected since there were no major templates which could have been used while predicting the structure itself. Moreover, we did refine our 3D model using available software, and we've increased the number of residues in allowed region from 78% to 86%, which is a substantial increase and probably gives us more reliable model to work with."

6. **Improve structure if needed**

- Using relevant software like 3Drefine, Whatif
- http://swift.cmbi.ru.nl/servers/html/index.html

7. **Find domains in your protein of choice**

- Show the results
- Explain
- Pfam, SMART
- **Explanation example:**
- "… we can see that my protein has SH2 domain which is not surprising since my protein is involved in signaling and binds to receptor thus being phosphorylated."
- "… we can see that my protein has DNA binding domain which just confirms the proposed function of my protein that it is actually a repressor, being able to silence some genes."
- "… we can see that my protein has no major recognizable domains, but it is normal since I've picked structural filamentous protein and we can see from our 3D predicted picture that it only has some beta sheets which do not form domain themselves, but they form the rigid body of our structural protein."

8. **Find localization of your protein of choice**

- You will, ideally, use more than one software and compare the results between them, thus making a general conclusion as to where your protein might actually be
- Psi-predictor, Suba3, LocTree3
- **Explanation example:**
- "… our protein localization prediction results show that our protein is located in mitochondria. This is actually expected since our protein is involved in ATP synthesis, according to the results we've got so far. There is, also a small chance of our protein being in cytoplasm, which is interesting and requires further research to be conducted in order to be able to fully interpret such a result. Also, our next step from this point on would be to confirm in the laboratory whether the protein can be present in cytoplasm or not. It might simply be due to the fact that the protein is produced there and then immediately transported to mitochondria."

9. **Predict the interactome of your protein**

- Using available software.
- **Explanation example:**
- "… our results show that our protein interacts with 6 other proteins which are also involved in the ATP synthesis which strongly supports our other results and our argument that this protein is, in fact, involved in ATP synthesis itself."
- "… our results show that our protein interacts with 5 other proteins. 2 of these are expected as they also participate in ATP synthesis, but 3 others show no correlation to mitochondria whatsoever. This might be due to two reasons: the data available on our protein is really low and the software cannot accurately predict the interactome, or, since the data is low, the protein has actually other functions as well, which are yet to be determined. Our research suggests that this other function is in cell signaling."

10. **Render the picture in PDB viewer (or any other relevant downloadable software)**

- Download PDB file from swissprot/expasy

- Open it in PDB viewer
- Render it in solid 3D (or even better with secondary structure motifs – alpha helices and beta sheets)
- Compare the result with the protein predicted online (note: you might have to rotate the protein in order to match it)
- Are 3D predictions the same?

Notes:

- This is a paper, so it should have the characteristics of the paper: Name of student, ID of student, title (pick the one you like), introduction (if necessary), results, conclusion (which are actually explanations for every results, but you can also conclude everything in the end as well, if you want), references
- Send everything in one file (word or PDF) – I will not accept if you send me separate pictures in the mail and write your explanations in the mail
- Also submit hardcopy
- In order to do this project and to write explanations well, you have to do a bit of research, meaning that you have to research about the function of your protein, or domains that you find in them
- You have to reference your information! Points will be deducted if you write that, for example, domain in your protein participates in the signaling but you don't write where did you take that information from
- You must understand the results: e.g. results of phylogenetic tree and what do numbers represent
- At least one paragraph is needed for every explanation, so in the end you must have at least one page of text (intro, conclusion and references not included!)
- Font: Calibri, Arial, Times New Roman (you get the idea)
- Font size: 11 or 12
- Spacing: maximum 1.5
- Margins: 1 inch
- References: IEEE, APA or MLA style (search online if you don't know); mark references in your text (e.g. "this domain binds to iron[4]", or "this domain binds to iron (Huges, et al, 2004), or "this domain binds to iron [4]")
- The better you do – more points you'll get. Plagiarism – 0 points (this only means that you have to reference where did you get your information, it doesn't mean that you cannot copy from other sources – <u>but your own explanation is still required</u>)
- **Quality over Quantity!**
- If you have any questions – ask

Good luck ☺