Final Project: Analysis of Hemagglutinin (HA), Neuraminidase (NA) or Your Favourite Protein (YFP)

This document should be read all the way through before you start working. It contains a collection of questions and things you can consider when doing a structure analysis. When you do your posters, you must choose to focus on one or two aspects of the modelled structure and describe this in more detail. Therefore, you do not need to answer all of the questions in detail – in fact, it is better to answer a few of them well than trying to cover them all. Some of the questions mentioned under NA may be equally relevant to HA or YFP and vice versa, so read all the questions. Also, you do not need to stick to the questions mentioned here – if you find some more interesting questions to address, you are welcome to do so! The more different the posters are, the more we will all learn about these proteins in the end. You will be working on this project until the end of the course.

Please keep in mind, that when you do your poster presentations at the end of the course, you will all be expected to participate actively in the oral presentation, and you can all be asked about all parts of the work.

Papers
A number of introductory papers on HA or NA have already been posted at CampusNet in the documents folder, but you will need to find others if you choose a different protein. When everyone in your group has read these papers, discuss them and try to explain the main points of the papers to each other. Spend some time on this and make sure that you all have a good understanding of the structure and function of the protein(s). You should then discuss in your groups which aspect of the structure you would like to focus on, and try to come up with a working title for your poster. It may change before you prepare your final poster, but it helps a lot to define your problem clearly before you start.

You may use all the information you can get hold of for your posters, and you may choose to use other papers than the ones I have given you. If you run into problems getting hold of a paper you need, ask me, and I will try to help you.

Getting some structure in your project
When you have decided on a protein, it is a good idea to define an overall goal for the project. The goal should be to produce a measurable result, i.e. the outcome of your project should be some prediction or suggestion, which can be tested experimentally. Not only will this make it easier to decide on the types of questions you need to answer in order to get there, it will also make your project much more interesting!

For HA and NA: To make it even more interesting, I suggest that you choose sides for your project either trying to improve or combat the virus. Some suggestions for core analyses in such projects are given below:

Virus viewpoint
This could for instance include an analysis of
• How influenza virus will/has escaped known drugs.
• How it has avoided detection by the immune system.
• How the virus could evolve in nature or be engineered by humans (!) to make it more infectious.

Doctor's viewpoint
This could for instance include an analysis of
• How known drugs could be modified to improve their efficiency or prevent virus escape.
• (Conserved) areas, which could be targeted by new drugs and/or the immune system.
• How the virus could be made less infectious (for vaccination purposes).
Hemagglutinin & neuraminidase types for project

You will be working on proteins from type A Influenza virus. As some of the HA and NA subtypes can already be found in the PDB, modelling these is no fun. I have therefore checked the PDB, to make sure that you focus your efforts on some of the subtypes, for which there are no structures (yet). The list of HA and NA sub-types for which you can currently find structures are listed below. Your projects should NOT deal with the modelling of these.

Hemagglutinin: H1, H2, H5, H7  
Neuraminidase: N1, N2, N6, N8, N9

If you want HA and NA subtypes known to be associated with disease in humans select H10 or N3. When choosing these from the Influenza Virus database (http://flu.lanl.gov), remember that not all variants have been observed in humans.

Also, make sure that you are not modelling structures already available in the PDB. If in doubt, please consult with me before starting your project.

Some questions to consider:

Neuraminidase enzyme-ligand interactions:
Does the template you have used have a ligand bound in the active site (substrate, product or inhibitor)? Try to identify which residues are interacting with the ligand. (You can for example use Ligplot: click on the relevant ligand from the PDBSUM entry for the structure. PDBSUM is linked from the PDB site). Compare the template with ligand bound to one, which does not have ligand bound. Often, water molecules found in the structure of the native enzyme suggest the most important interactions between the enzyme and its ligand. Is that the case here? Does the ligand seem to induce any major conformational changes in the whole enzyme structure? What about the local environment around the active site? Are there any differences between the two structures in the amino acids that you identified as interacting with the ligand?

Hemagglutinin receptor specificity:
What does α2-3 and α2-6 linkages mean? (Show with a drawing, for example). What makes swine interesting in this connection? Which mutations could be important for receptor specificity or affinity? If you inspect a sequence alignment and compare your query sequence with that of the template structure – are there any mutations near the receptor-binding site? If so, could they somehow influence the affinity for the ligand? Are the mutations distributed evenly along the whole sequence in a multiple sequence alignment or do they tend to cluster in certain regions? If so, where are the more conserved regions in the structure and where are the variable regions? How does this relate to the secondary structure? How about the receptor-binding site? Does this make sense to you?

Recognition of HA or NA by the immune system:
Influenza A is known to be able to escape the immune system by introducing mutations in HA and NA. You can investigate an epitope on NA or HA. It is a good idea to use a template, which is in complex with an antibody. Identify the residues on NA/HA, which have an atom within 4 Å of an atom in the antibody (see how to do this in PyMOL in “PyMOL essentials”). These residues collectively make up a so-called conformational or discontinuous epitope. Investigate the sequence alignment between your query sequence and the template. Are the mutations distributed evenly along the whole sequence alignment or do they cluster in certain regions? Are there more or fewer mutations in the epitope compared to the rest of the structure? Where are the more conserved regions in the structure and where are the variable regions? How does this relate to secondary structure? Can you explain this?

YFP Projects

If you decide to work on Your Favourite Protein, you must first make sure that your project is realistic within the time frame of this course. In other words: Talk to me first!

The specific questions that you will address will most likely vary from the ones suggested above and the project might take a very different direction. However, the overall structure of the project and the way in which it is presented will be the same for all projects (see below).
**Poster contents**

I would like you to use the IMRAD (Introduction, Methods, Results and Discussion) format. It is important to keep in mind that the poster format puts a hard limit on the amount of text you can put into your posters, so you will really need to boil it down to the most important points. You will present your posters through oral presentations, which will give you a chance to elaborate somewhat on your work, but also here it is important to keep focused on the most important elements.

Some elements will be included in all posters:

1. Title - make it descriptive so the reader knows immediately which angle you have chosen to take and don’t just call it “The structure of hemagglutinin”.
2. Group name, logo and names of the group members.
3. An introduction to the protein: Write a bit about its biological context, a general description of the structure, and also introduce the special feature you have decided to focus on.
4. Methods: describe the tools you have used to create and validate your model and any other tools that may have been important for your work.
5. A special section of the poster (not too long) is to be dedicated to a description of those parts of your query sequence for which there is no template:
   a. What structural information can you get for a protein (sequence), which does not have a (high homology) template in the PDB?
   b. What type of (biological) questions can be addressed with this kind of information?
6. Results and discussion.
7. References.

**Poster format**

**Size:**
Posters will be printed on A0 paper, dimensions 1189mm x 841mm. It is up to you if you want it in portrait or landscape orientation. If making your presentation in PowerPoint, just look under Files/Page setup to adjust the size of your poster.

**Fonts:**
Keep font sizes for the main text in a size comparable to at least 24 point Arial. Figure texts should not be smaller than a 20 point Arial. A good rule of thumb is that your poster should be readable when printed on A4 paper. In general, sans serif fonts (like Arial and Helvetica) are more readable on a poster than serif fonts (like Times and Garamond).

**Colours:**
Avoid dark background colours and excessive colouring in general as this makes the posters very expensive to print (and it also doesn’t improve readability). Background colour for structure images produced in PyMOL should be white. **Hint:** For background colours as well as PyMOL figures (the protein part) you generally need to use a lighter shade of the colours you initially think of as suitable.

**Text-to-figure ratio:**
I suggest that you reserve approximately one third of your poster for figures. This in turn means that you will have space for no more than 1000 words of main text, corresponding to two pages of text in 12 point Times New Roman. Remember that project details, which don’t fit on the poster, can be used for your presentation.