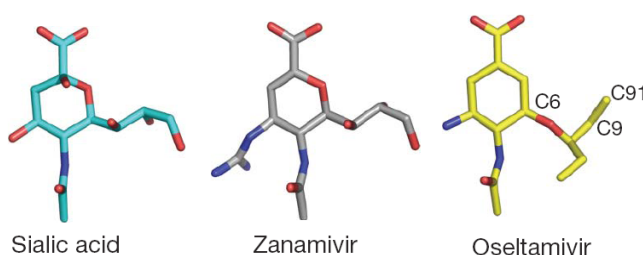


HOMOLOGY MODELING EXERCISE

We will investigate the structure of the influenza virus neuraminidase protein and look at how its function may be blocked by using the neuraminidase inhibitor oseltamivir, the active ingredient in the drug Tamiflu.

1. The substrate of influenza virus neuraminidase is sialic acid. The structures of sialic acid and the drugs oseltamivir (in Tamiflu) and zanamivir (in Relenza) are shown below (from P.J. Collins *et al.*, Nature **453**, 1258 (2008)).



Briefly describe the similarities and differences between enzyme substrate, sialic acid, and the enzyme inhibitors.

All three compounds have a six-membered ring with or without an oxygen-atom. They also have the same carboxyl group and amide substituent. Sialic acid and zanamivir have the same propyl-substituent with three hydroxyl groups while the substituent in oseltamivir in this position is both larger and more hydrophobic. The last ring-substituent is different in all three compounds.

2. The structure of an N2 neuraminidase with sialic acid bound in the active site can be found in the PDB entry 2BAT. A H5N1 neuraminidase with the active site blocked by oseltamivir can be found in 2HU4. Some naturally occurring mutated variants of viral neuraminidase have been found to give rise to Tamiflu-resistant strains of influenza. One such mutant is H5N1 His274Tyr. According to P.J. Collins *et al.* (Nature **453**, 1258 (2008)), this mutant binds oseltamivir with much lower affinity than the wild-type enzyme. Enzyme inhibition is reduced by a factor ~ 265 . The structure of oseltamivir bound to His274Tyr N1 neuraminidase can be found in the PDB-structure 3CL0.

Download the “PyMOL scene” or “PyMOL session” from this website <http://folk.uio.no/jonkl/pubstuff/NeuraminidasesExc4.pse>. Save the file on your Desktop and open it in PyMOL. This PyMOL session contains:

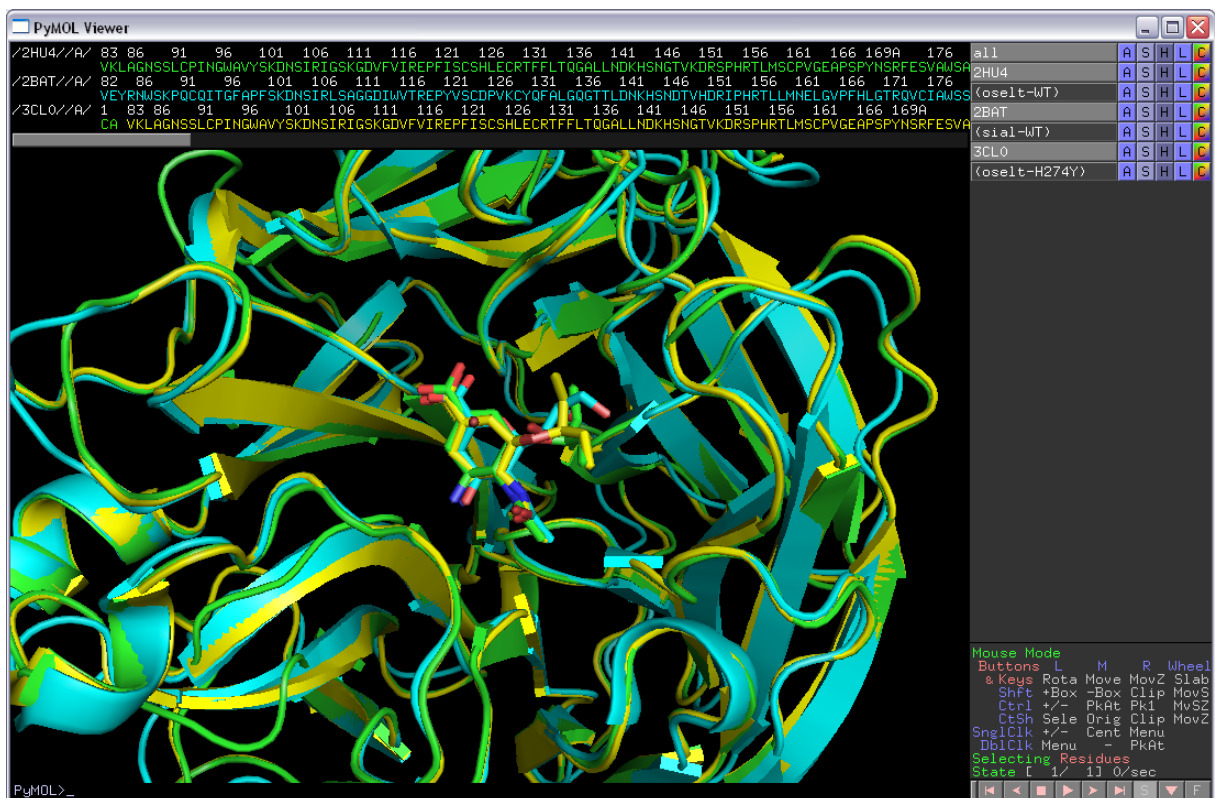
- 2HU4, chain A: A wild-type N1 neuraminidase bound to oseltamivir (green)
 - 2BAT, chain A: A wild-type N2 neuraminidase bound to sialic acid (blue)
 - 3CL0, chain A: A His274Tyr N1 mutant bound to oseltamivir (yellow)
3. The proteins are shown in cartoon rendering and the ligands as sticks. You could easily have generated this “scene” yourself (and saved it as a PyMOL session), but we do not have time for that now! Actually, if you have used PyMOL a bit before and have the time, please

download the PDB files and get the 3 structures into the same session. That is, do it “properly”.

It is much easier to see similarities and differences if we align the structures in 3D space. Type “**align 2HU4, 2BAT**” (watch the screen as you tap “enter”!) to make a superimposition of these two structures. This is PyMOL’s variant of intermolecular alignment. One of the structures is translated and rotated in order to get the lowest possible RMSD with respect to the other. Which of the two structures were moved?

The first structure (in front of the comma) is moved onto the second.

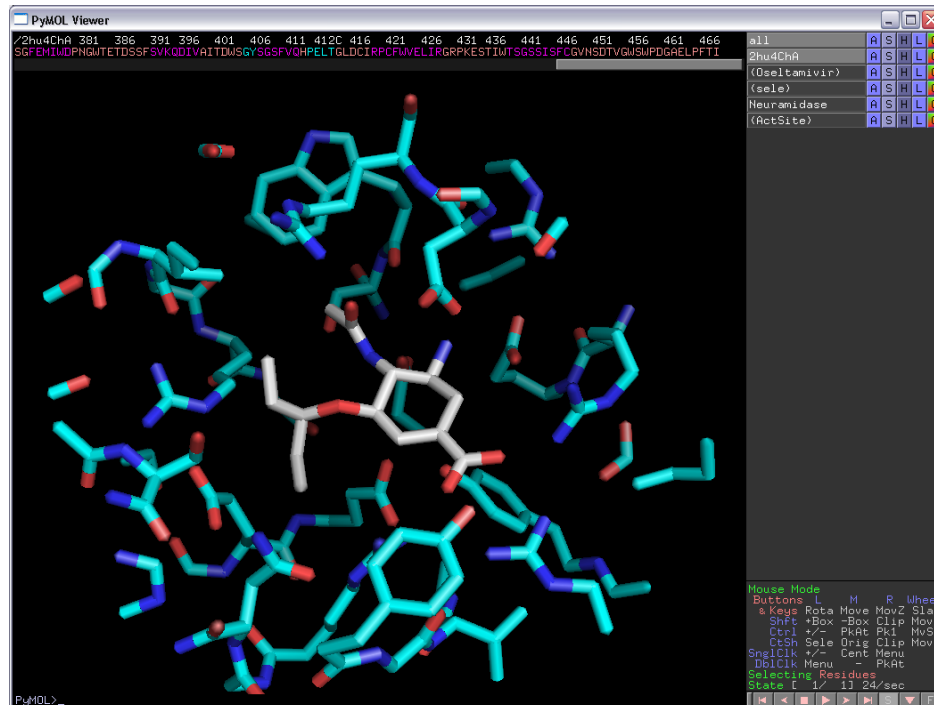
- Align also 3CL0 with the two others. Does sialic acid and oseltamivir bind in the same pocket on the enzyme surface? Would you describe the structures as identical? Similar? Dissimilar?



The structures are quite similar, but not completely identical. They obviously bind substrate/inhibitor in the same active site pocket. The two N1 neuraminidases are the most similar.

- We will now look at 2HU4 only. Turn off the two other objects in the mini-menu on the right-hand side of the viewer window. You now see only wild-type N1 neuraminidase bound to oseltamivir in the active-site pocket.
- “oselt-WT” is a selection containing only the oseltamivir molecule. Click on the selection to get that confirmed. Make a new selection containing only residues within 8 Å of oseltamivir, *i.e.* the “active-site region”. To do this type “select ActSite, oselt-WT around 8”.

Hide everything and then show both oseltamivir and ActSite as “sticks”. Color them “by element” but use different colors for the C-atoms making it easy to see both the oseltamivir and the protein residues. Hide everything else. You should now be able to see all the residues of the neuramidase packing around oseltamivir (as seen below).



7. For the (ActSite) selection, do “L” → “residues” to get these residues labeled.
8. You can let PyMOL make an attempt on localizing H-bonds between oseltamivir and the protein by doing for the (oselt-WT) selection: “Actions” → “find” → “polar contacts” → “to other atoms in object”. As you see, there are too many H-bonds, but at least you get some idea.
9. List two acidic residues forming H-bonds to oseltamivir. Which basic residue is donating an H-bond to the amide group of oseltamivir? Two other basic residues contact the carboxyl group of oseltamivir. Which are they? One of them is even involved in strong, so-called “bidentate” H-bonding. Which one? List some other residues involved in van der Waals interactions packing with oseltamivir.

Two acidic residues containing H-bonds to oseltamivir are Glu119 and Asp151. Arg152 is binding the amide group. The two basic residues donating H-bonds to the carboxylate group are Arg292 and Arg371. The latter is forming bidentate H-bonds.

Some other residues forming the pocket where oseltamivir is bound are: Arg118, Arg224, Glu276, Glu277, and Tyr347.

10. Turn on all three objects again, 2HU4, 2BAT, and 3CL0. Find residue 274 in the three structures. Show them as sticks. What are they? What can you say about the properties of these residues?

The 274 residue is His in 2HU4 and 2BAT and Tyr in the His274Tyr mutant in 3CL0. Tyr has a larger and more bulky side chain than His.

11. Color the three objects differently, but “**by element**”. Now you see the (aligned) active site residues of the three enzymes.
12. Above you have identified several residues involved in forming H-bonds with oseltamivir: Glu119, Asp151, Arg152, Arg292 and Arg371. Are these residues conserved in all three neuraminidases? What can you say about the conformations/rotamers of the side chains for these residues? Why do you think these residues are conserved? What would happen if for example the mutation Glu119Trp was introduced? Would the enzyme be inhibited by oseltamivir? Would it have any activity on sialic acid? Can you find any active site residues that are not conserved between the three neuraminidases?

The conservation/rotamer is:

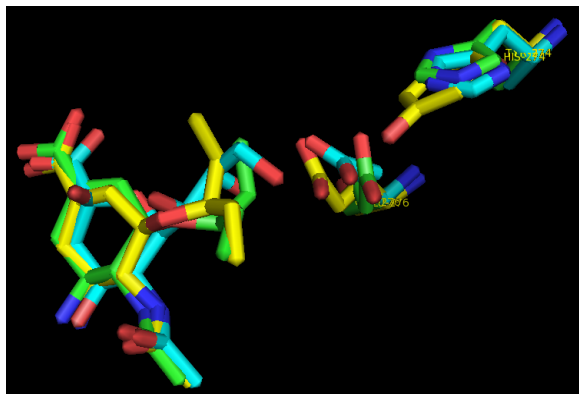
- **Glu119: Conserved in all three/identical rotamer**
- **Asp151: Conserved in all three/nearly identical rotamer**
- **Arg152: Conserved in all three/different rotamer in 2BAT**
- **Arg292: Conserved in all three/identical rotamer**
- **Arg371: Conserved in all three/identical rotamer**

These residues are necessary for the function of the enzyme. Most likely mutations would lead to reduced activity on sialic acid. Since the substrate/inhibitors have very similar structure (Question 1 above) it is not surprising to find that these residues directly contacting substrate/inhibitors have mostly the same conformations in all three structures.

The mutation Glu119Trp would put a very bulky residue in the active site. Neither oseltamivir nor the substrate would be able to bind. This mutation would certainly not occur in a viable virus.

Residue Tyr347 in N1 (2HU4 and 3CL0) is Gln347 in N2 (2BAT).

13. Locate residues 274 and Glu276. Take a closer look at the conformation of Glu276 in the 3 structures. Are you able to explain why the His274Tyr mutant binds oseltamivir less efficiently than the wild-type enzyme? Can you explain why His274Tyr is “allowed”, *i.e.* the enzyme has wild-type activity on its substrate? Why is this mutation particularly good for the virus and bad for the doctor trying to treat a patient with Tamiflu?



The residue 274 in the “wild-type” neuraminidases (green/blue) is His. In the His274Tyr mutant (yellow) it has been replaced by the more bulky residue Tyr. This pushes the Glu276 residue ~2Å out into the active site pocket. This leaves less room for the bulky substituent on oseltamivir which is then pushed out of the pocket (compare yellow/green). Apparently this leads to less efficient oseltamivir binding and inhibition of the enzyme. The substituent on sialic acid is less bulky and the His274Tyr mutant does not affect binding of the substrate. Interestingly, zanamivir, with the same, less bulky substituent inhibits also the His274Tyr mutant. This is thus a viral strain that is Tamiflu, but not Relenza, resistant.

Now let us try some modeling!

14. You have a cousin working for Médecins Sans Frontières near Goroka in Papua New Guinea. You get this e-mail from her:

Dear cousin,
we have some serious problems here with an outbreak of an influenza-like disease. We have high mortality rates and it appears to be highly contagious. We have some indications that this might be an oseltamivir-resistant strain of influenza. I remember you told me about that structural biology course and perhaps you can help me with information on the structure of the viral neuraminidase? But be quick, we are certainly in a hurry here :-)

Through another contact we managed to get the neuraminidase gene from this virus sequenced:

```
>Possible neuraminidase [Putative influenza A virus (Goroka)]  
MNPNQKILCTTATAIVIGSIAVLIGIANLGLNIGLHLKPICNCSHSQPEATNASQTIINNYNETNITQI  
SNTNIQMEERASRGFNNLTGKGLCTINSWHIYGKDNAVRIGENSVDLVTREPYVSCDPDECRFYALSQGT  
IRGKHSNGTIHDRSQYRALISWPLSSPPTVYNSRVECIGWSSTSCHDGKSRMSICISGPNNNASAVVWYN  
RRPVAEINTWARNILRTQESECVCHNGVCPVVYTDGSATGPFDYRIYYFKEGKILRWESLTGTAKWIEEC  
SCYGERTGITCTCRDNWQGSNRPVIQIDPVAMTHTSQYICSPVLTNDNPRPNDPNVGKNDPYPGNNNGV  
KGFSYLDGVNTWLGRITISTASRSGYEMLKVPNALTDERSKPIQGQTIVLNTDWSGYSGSFMDYWAEGDCY  
RACFYVDLIRGRPKEDKVWWTNSIVSMCSSTEFIGQWNWPDGAKIEYFL
```

Can you have a look at this and give me feedback? Can structural biology tell us anything about this strain? Might it be resistant to Tamiflu?

Please help us,
Your cousin

What should you do? In this case, would you recommend trying to crystallize the protein and investigate the structure by X-ray crystallography? Why not?

Determining the structure by experimental methods will take at least a month, perhaps much longer. There is no time for this! In many cases you can get *the answer to your biological question* much quicker and less expensively employing bioinformatics modeling.

You decide that you will try homology modeling for the Goroka neuraminidase. What are the 6 steps involved?

Go to the NCBI website (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and do “protein blast”. Use the Goroka-sequence as a query sequence, use blastp and search the pdb-database for possible templates. Make sure you use the pdb-database, and nothing else! Do you find any templates? Search for the string “2HTY” on the results page. Below the sequence alignment for this hit click on “23 more sequence titles” to find 2HU4. Is 2HU4, containing oseltamivir, a possible template? What is the sequence identity between the template (2HU4) and the target (Goroka sequence)? Does homology modeling appear to be a possibility? Will you be able to use 2HU4 for modeling the full-length protein?

The six steps in homology modeling are:

1. **Template selection**
2. **Alignment correction**
3. **Backbone modeling**
4. **Loop and side chain modeling**
5. **Structure refinement**
6. **Structure testing and validation**

There are many structures (more than 50) with sequence identity to the Goroka-sequence above 30%. All these can be used as templates. The sequence identity between target and template 2HU4 is 49%, but only spans residues 92 – 468 of the target. Only this segment can be modeled.

2HU4 is most likely not the best template to model overall structure as good as possible since there are other candidates with more than 80% sequence identity with the target. However, we are interested in drug resistance and a template binding a drug is therefore ideal. In addition, we already have some experience with 2HU4 (from above). In order to save time, we use 2HU4.

15. You decide to do homology modeling for the Goroka target sequence and go to the SWISS-MODEL website (<http://swissmodel.expasy.org>). Click on “Start Modelling” and create a new account by clicking on “Create Account”. Follow the instructions you get and login at the SWISS-MODEL workspace.
16. The sequence alignment you got from blastp is the following, in clustal and fasta format, respectively:

CLUSTAL

Target/1-377	LCTINSWHIYGKDNAVRIGENSDVLVTREPYVSCDPDECRFYALSQGTTRGKHSNGTIH
Templ/1-375	LCPINGWAVYSKDNSIRIGSKGDVVFVIREPFISCSHLECRTFFLTQGALLNDKHSNGTVK
Target/1-377	DRSQYRALISWPLSSPPTVYNSRVEICIGWSSTSCHDGKSRMSICISGPNNNASAVVWYNR
Templ/1-375	DRSPHRTLMSCPVGEAPSPYNSRFESVAWSASACHDGTSLWTIGISGPDNGAVAVLKYN
Target/1-377	RPVAEINTWARNILRTQESECVCNNGVCPVVYTDGSATGPFYRIYYFKEGKILRWESLT
Templ/1-375	IITDTIKSWRNNILRTQESECACVNGSCFTVMTDGPSNGQASYKIFKMEKGKVVKSVELD
Target/1-377	GTAKWIEECSCYGERTGITCTCRDNWQGSNRPVIIQIDPVAMTHTSQYICSPVLTDNPRPN
Templ/1-375	APNYHYEECSYPNAGEITCVCRDNWHGSNRPWVSFNQ-NLEYQIGYICSGVFGDNPRPN
Target/1-377	DPNVGKCNDPYPGNNNGVKGFSYLDGVNTWLGRITISTASRSGYEMLKVPNALTDDRSPK
Templ/1-375	D-GTGSCG-PVSSNGAYGVKGFSEFKYNGVWIGRTKSTNSRSGFEMIWDPNGTETDSSF
Target/1-377	IQQQTIVLNTDWSGYSGSFMDY--WAEGDCYRACFYVDLIRGRPKEDKVWWTNSIVSMC
Templ/1-375	SVKQDIVAITDWSGYSGSFVQHPGLTGLDCIRPCFWVELIRGRPKESTI-WTSGSSISFC


```
Target/1-377    SSTEFIGQWNWPDGAKIEY
Templ/1-375    GVNSDTVGSWPDGAELPF
```

```
>Target
LCTINSWHIYGKDNAVRIGENSVDLVLTREP YVSCDPDECRFYALSQGT TIRGKHSNGTIH
DRSQYRALISWPLSSPPTVYNSRVECIGWSSTSCHDGKSRMSICISGPNNNASAVVWYNR
RPVÆINTWARNILRTQSECEVCCHNGVCPVVYTDGSATGPF DYRIYYFKEGKILRWESLT
GTAKWIEECSCYGERTGITCTCRDNWQGSNRPVIQIDPVAMTHTSQYICSPVLTDNPRPN
DPNVGKCNDPYPGNNNGVKGFSYLDGVNTWLGRTISTASRSGYEMLKVPNALTD DRSKP
IQGQTIVLNTDWSGYSGSFM DY--WÆEGDCYRACFYVDLIRGRPKEDKVWWT SINSIVSMC
SSTEFIGQWNWPDGAKIEY
>Templ
LCPINGWAVYSKDNSIRIGSKGDV FVIREPFISCSHLECR TFFLTQGALLNDKHSNGTVK
DRSPHRTLMSCPVGEAPSPYNSRFESVAWSASACHDGT SWLTIGISGPDNGAVAVLK YNG
IITDTIKSWRNNILRTQSEECACVNGSCFTVMTDGP SNGQASYKIFKMEKGKVVKSVELD
APNYHYEECSYPNAGEITCVC RDNWHGNSRPWVSFNQ-NLEYQIGYICSGVFGDNPRPN
D-GTGSCG-PVSSNGAYGVKGFSFKYGNVWIGRTKSTNSRSGFEMIWDPNGW TETDSSF
SVKQDIVAITDWSGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKESTI-WTSGSSISFC
GVNSDTVGSWPDGAELPF
```

How many indels (insertions/deletions) are there? Is getting the correct sequence alignment important for homology modeling? How could you improve the alignment?

There are 5 indels. Getting the correct sequence alignment is the most important step in homology modeling. One would get a better alignment by generating an MSA from target, template and a large number of homologs (50-100?) and then use the alignment of target and sequence from this MSA. One should also make sure indels are found in or near loops in the template, if possible.

17. Instead of using “Automatic mode” which is the simplest (but might give errors) we will try “Alignment Mode“. We need a good alignment to start this job. However, due to lack of time, we will try to use the alignment we got from blastp above. Open the template structure in PyMOL (for example the file <http://folk.uio.no/jonkl/pubstuff/2hu4ChA.pdb>). Show as “cartoon” or “ribbon”. Locate the indels. Are they in loops/coils or in helices/sheets? Do we need to correct the alignment manually by moving indels out of helices/sheets?

All indels are in loops and not within sheets/helices. It is not necessary to manually move any indels.

18. Back at the SWISS-MODEL workspace now paste in the alignment of target and template in fasta format that you find above. Wait for the possible templates to be loaded. Choose structure 2HU4, chain A, the structure you have looked at earlier. Wait for the job to finish... Or continue further down while you wait

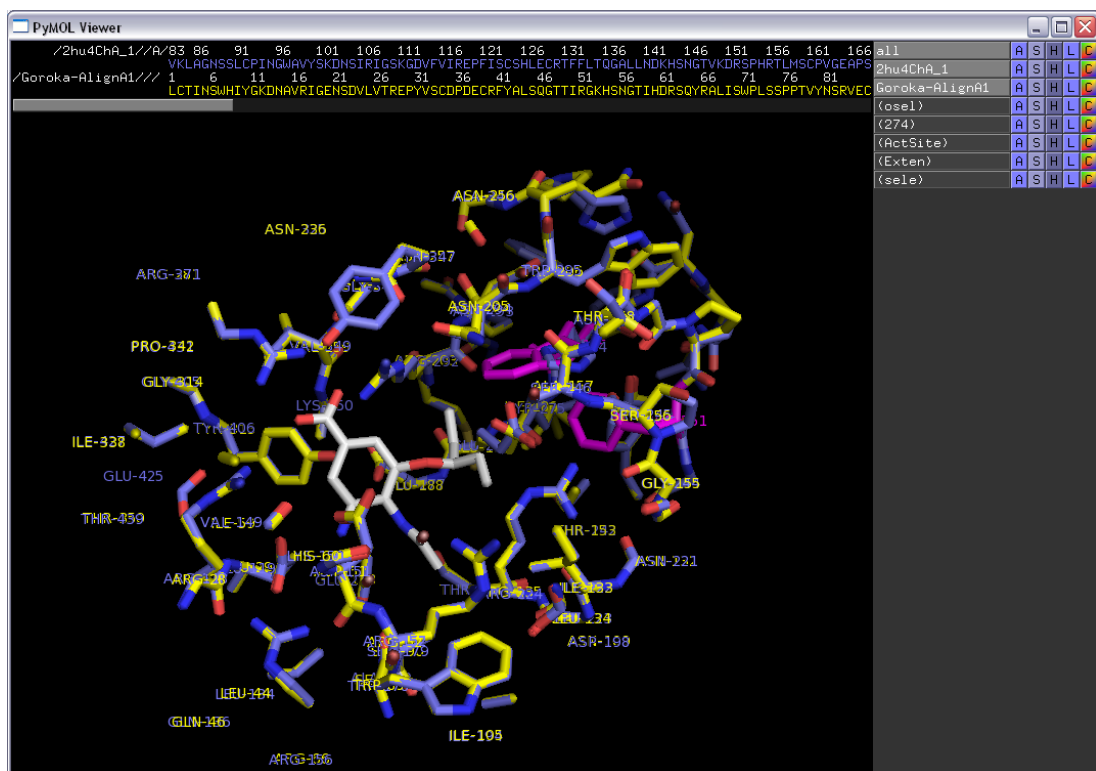
19. Did the “Alignment Mode” job finish? Look at the “Modeling Log”. How many loops were modeled? Do they appear to be ok?

The “Alignment Mode” job finished successfully. 5 loops were modeled, one for each indel. They all appear to be fine.

20. Download and save the model PDB-file to your Desktop. Now you might check the quality of the model by submitting it to some of the “Structure Assessment” tools under “[Tools]” at the SWISS-MODEL workspace. Do this *if you have time*. If not, we will just

trust the model for now. Anyway, this model is not very good since we have used a quick and dirty alignment. But it is good enough for our task: find out if it is likely that the new virus strain can be treated with Tamiflu

21. Open the model and 2HU4, chain A, in the same PyMOL window. First show “**as cartoon**”. Can you locate any of the loops that have been modeled at the indels?
22. Make a selection containing only G39 of 2HU4, chain A. Rename it “**oselt**”. Make a selection containing only residue 274 of the same chain. Rename it “**274**”. Color 2HU4 blue and the model yellow. Hide everything. Now type “**select Actsite, oselt around 8**” and “**select Exten, 274 around 8**”. Show both these selections as sticks and color “**by element**”, the first choice on the list. Do the same with “274”. Show also G39 as sticks and in a third type of “**by element**” coloring.



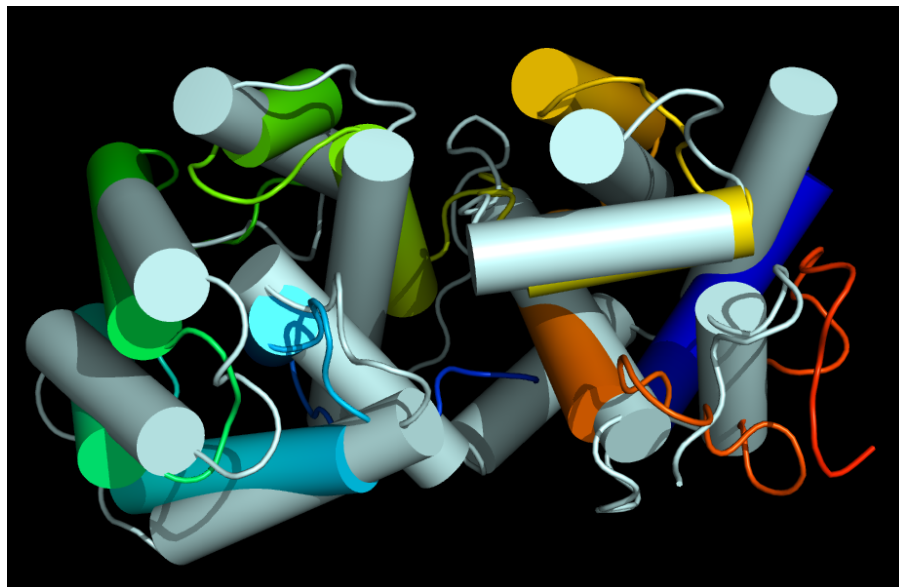
23. Take a close look at the residues interacting with oseltamivir in the pocket of 2HU4 and in the model of the Goroka neuraminidase. Can you see any differences between the two structures that might explain the putative oseltamivir resistance? What do you suggest to tell your cousin?

All residues contacting oseltamivir appear to be conserved and have the same rotamer in the target and in the template. The exception is Glu276 which is rotated (pushed?) slightly towards the bulky group in oseltamivir due to the two bulky groups “behind it” (pink above) caused by mutations in the Goroka-sequence. Like the N1 His274Tyr mutant, it appears likely that the Goroka-strain is oseltamivir resistant.

Note: The Goroka-sequence is fake and generated by Jon just for this exercise. He stuffed several bulky residues “behind” Glu276 in the structure to get this residue slightly pushed into the active site. There are more bulky residues there than what could realistically fit in. In order to get a better model one would possibly have to run some molecular mechanics geometry optimization steps, but this is beyond what we

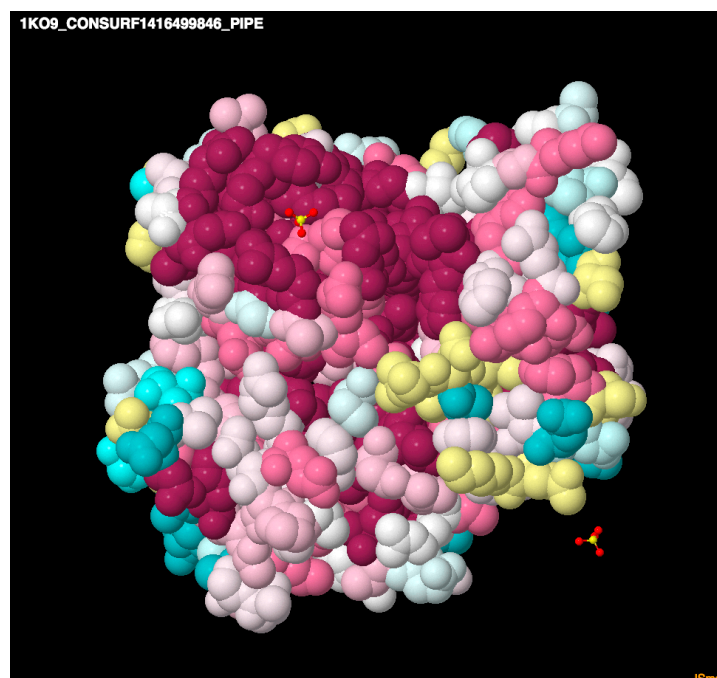
have time for in this course. This would give you a “more optimal” model in terms of overall structure, but not necessarily a better model for answering the question we are asking: is the strain likely to be oseltamivir resistant?

24. Make a new, clean PyMOL session. Get the PDB files for the PDB entries 2ABK and 1XQO. Read about these structures at the PDB website. What are these proteins doing? Open these two files in PyMOL. Try to align the 2 structures with the align command (intermolecular alignment, first doing a sequence alignment). Inspect the alignment in PyMOL. Are the two structures well aligned? Now instead try to do “cealign”: <http://www.pymolwiki.org/index.php/Cealign> Does it give a better alignment? Show both proteins as cartoon with alpha helices as cylinders (“Setting” → “Cartoon” → “Cylindrical helices”). Color 2ABK in rainbow and 1XQO in cyan. Carefully investigate where you have similarities and where you have differences in the two structures. Add 1EBM to the session and align that structure to the other two. Try “set grid_mode, 1”. What do you get?

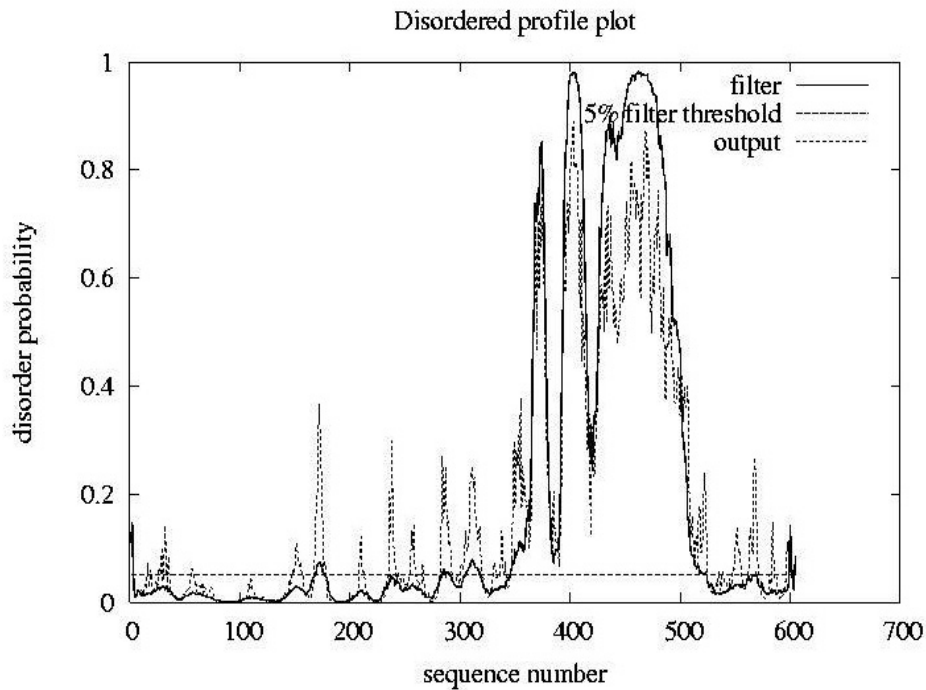


25. Find 2HU4, 2BAT, and 3CL0 in the PDB and have a look at the corresponding pdb-files in PyMOL. There are several chains in all these PDB entries. Can you say anything about the quaternary structure of neuraminidase from these structures? Do the pdb-files contain the structures of full-length proteins? Use chain A from the three pdb-files and make a session such as the one you started with in this exercise.
26. Take a look at the various model quality estimation tools at the SWISS-MODEL Workspace. Is the model ok, or not. What seems to be the problem? Upload the 2HU4 chain A PDB file and run it through the same tools. Are the results any better?
27. Try to improve the alignment of target and template by using the information from homologous proteins. Use the improved alignment to generate a new model for the Goroka neuraminidase. Do you get a better model?

28. Browse CATH (<http://www.cathdb.info>) and SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop>) to look for your favorite protein. Or look for Nth, MutY, OGG1, MBD4, and homologs.
29. Search with the 1XQO, chain A, structure in the PDB. This is 3D structure searching. Use Dali: http://ekhidna.biocenter.helsinki.fi/dali_server
30. In the MSA exercise you made an alignment of human NTHL1 and MBD4 (and several more homologs). Use blastp to search for PDB structures that are similar or identical to these two proteins. Use Dali pairwise comparison to align the two best structures and get a sequence alignment from this 3D alignment. Compare the sequence based and structure based alignment and note similarities/differences. Which one should be better? For the structural alignment, use Dali: http://ekhidna.biocenter.helsinki.fi/dali_lite
31. Do a blastp search with your favorite protein as query in the PDB sequence database. Do you find any templates that can be used for homology modeling? You might also try fold recognition, for example GenTHREADER (<http://bioinf.cs.ucl.ac.uk/psipred>) or Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2>).
32. Here, <http://folk.uio.no/jonkl/StuffForMBV-INFx410/96seqs.fasta>, you find a lot of OGG1 homologs. Open the MSA in JalView and remove all sequences which there appears to be something wrong with (missing exons, obviously wrong start, etc.). Also use “Edit” → “Remove Redundancy...” to get rid of most of the sequences that are very similar (use 95% redundancy as cut-off threshold). Keep at least 25-35 sequences and realign with T-Coffee. Use <http://consurf2.tau.ac.il> to map the MSA onto the OGG1 structure with PDB identifier 1KO9. Study the results carefully. Assume you did not know where on the surface of OGG1 the active site was. How could you get an idea from the ConSurf results?



33. A structural disorder prediction (DISOPRED2, <http://bioinf.cs.ucl.ac.uk/disopred>) for a protein is given below. Is it possible to make a good structural model for the full-length protein?



No. There is most likely not a rigid, regular 3D structure between residues 350 and 500 (approximately). This part is predicted to be floppy, flexible and have structural disorder.