

Structural Biology: Assignment 1

University of Cambridge

Henrik Åhl

March 31, 2017

Preface

This is an assignment report in connection to the *Structural Biology* module in the Computational Biology course at the University of Cambridge, Lent term 2017. All related code is as of March 31, 2017 available through a Github repository by contacting hpa22@cam.ac.uk.

Part A

i Inspecting the unbound structure as assigned, we find a molecular complex consisting of principally two anti-parallel beta-sheets (one of three strands, one of 6) wrapping around to form a tubular shape. The beta-sheets are connected through intermediate carbon-dominated strands, holding the structure together. A very small helix is also inferred in the data, as can be seen as in blue in fig. 1 and fig. 2 in *NewCartoon* and *VDW* visualisation format respectively. Both figures are also aligned in order to show the capture cavity of the molecule, as is easiest seen in fig. 1. The backbone on one of the ends of the protein also appears to be open in such a way that it appears to possibly utilise this end to interact and attach to the larger Toll-like receptor complex. In contrast, the other end appears more stable, with more inferred closed circuits.

Over the course of the trajectory, the molecule also appears to be quite flexible with respect to its subdomains, with the cavity opening in particular seeming quite variable. When the water molecules are also accounted for, we can note that no such molecules appear to be present within the ligand cavity, which might be because of polar repulsion and an intractable free energy landscape. However, due to the insufficient resolution to analyse this, we are unable to draw any decisive conclusions.



Figure 1: Unbound protein at $t = 0$ in the *NewCartoon* visual theme. The capturing cavity is present foremost in this figure.

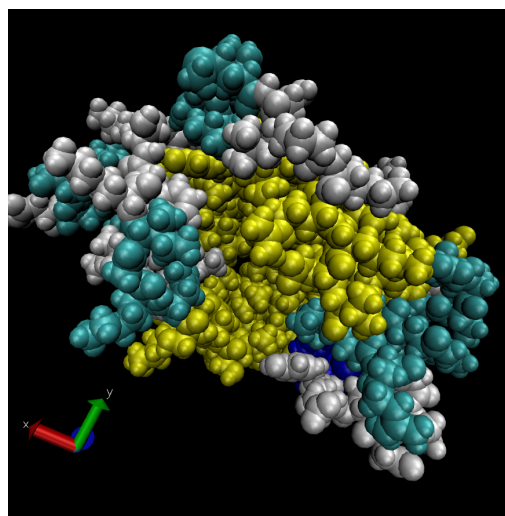


Figure 2: Same as above, but in the *VDW* theme, giving a more realistic estimate of the atoms' electrostatic surfaces.

ii, iii Simulating the protein from its initial state, it is clear that the binding cavity quickly closes off, making it seemingly inaccessible from the outside without prior interaction. The protein then resumes in this state over the course of the simulation, with mostly its tube-end backbone strands moving about significantly; especially the end containing the helix conformation appears visually particularly variable. Also the backbone on the backside of the protein as seen from when observing the cavity is highly variable.

Indeed, looking at the protein aligned to its initial state over the whole trajectory confirms the above mentioned observations. Figure 3 shows a comparison between the initial (blue) and final state (red) of the protein in the simulation trajectory. In particular, movement in the part connecting to the helix, as well as the capturing cavity are particularly noticeable.

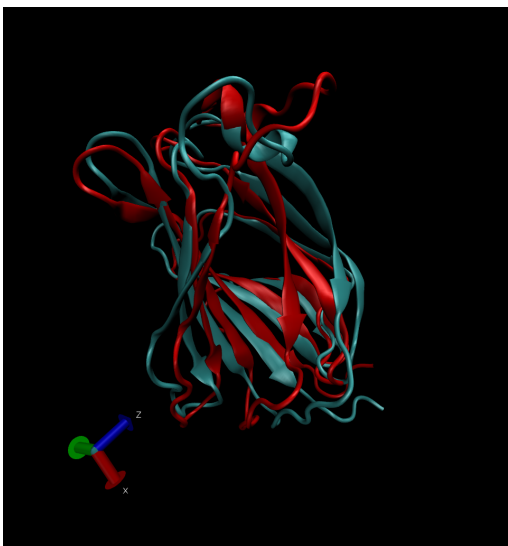


Figure 3: Comparison between the structure at the initial (blue) and final (red) conformation. Note in particular how the cavity has closed.

iv Observing the various parts of the molecule, and their general movement produces figure fig. 4. We find that the majority of the variability is attributed to movement of the backbone, which is mainly goverend by movement of the beta sheets. Due to the helix being small in comparison to the rest of the molecule, its contribution to the overall variability is minor, which in particular can be seen when the summed RMSD of the beta-sheets and the helix is plotted. This put further insight into the above mentioned observation in that the helix itself is not very variable, but rather the backbone connected to it.

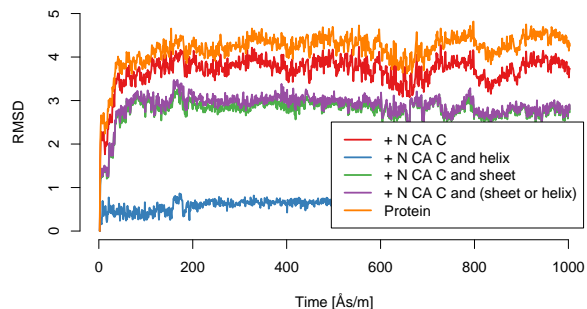


Figure 4: Spatial variations for select parts of the protein over the simulation. Notably, the protein as a whole changes significantly in its conformation, which is mainly driven by movement of the major structures, such as the beta-sheets and the overall backbone.

v The Ramachandran plot for time $t = 0$ in fig. 5 shows the ϕ and ψ angles for all the residues in the protein. Narrowing down the analysis, the beta sheet residues all have their data points located in the upper left corner, whereas the helix is located in the middle-left one. In other words, the important residues are all located in energetically allowed regions. Nevertheless, when accounting for the whole protein, some residues are indeed located outside of these, as the figure shows. At the end of the simulation, some clustering towards the coloured regions has occurred, although many residues are still part of an incorrect conformation; most of these being C_{α} atoms. Disregarding these faulty conformations, the C_{α} atoms are otherwise spread out among the three islands, and continue to be so as the simulation progresses.

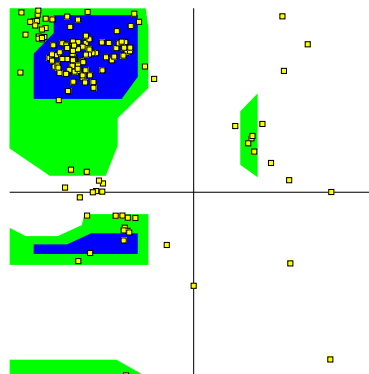


Figure 5: Ramachandran plot at $t = 0$, showing a clustering of the residues in the allowed energy conformations. However, some residues occur outside of these, and stay or get exchanged throughout the simulation.

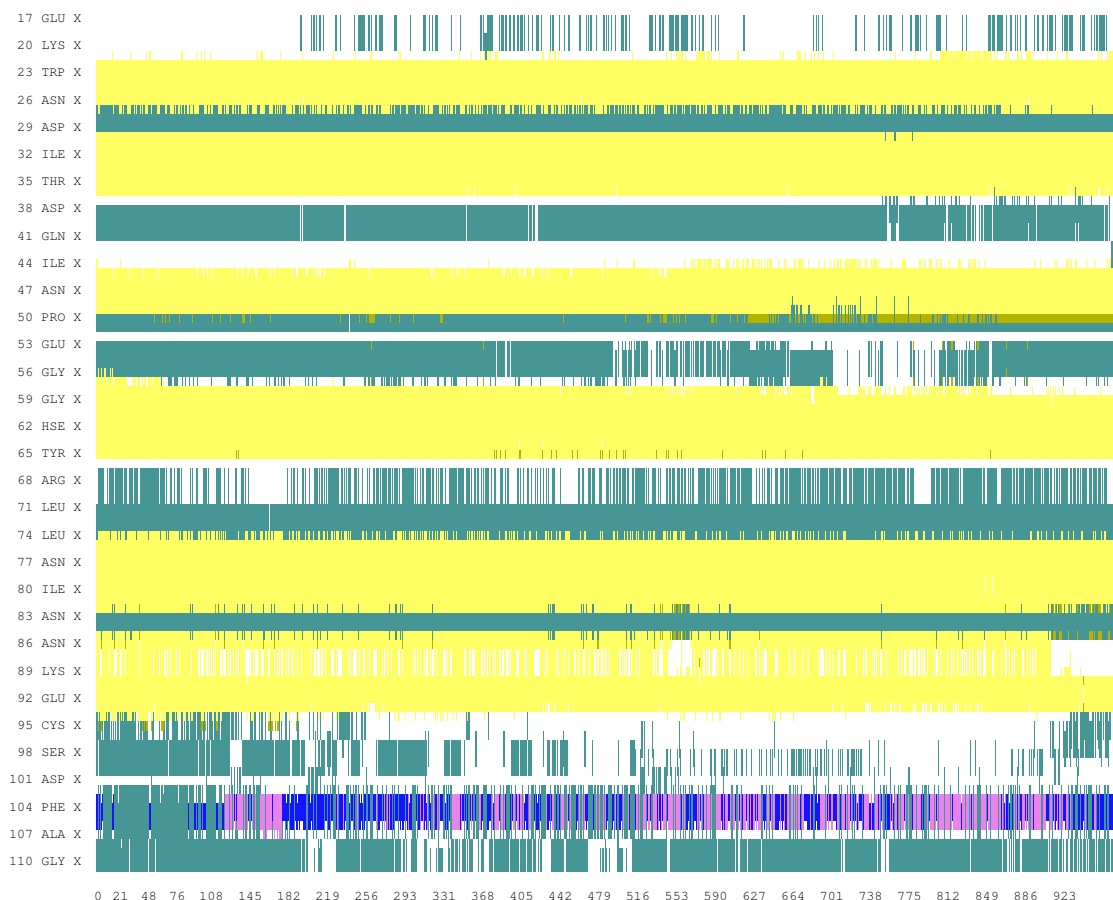


Figure 6: Structural inference on the residues in the system, with colour signifying inferred secondary structure. Most beta-sheets remain very stable throughout, with variations mostly in the edgemost parts of the conformation. Some regions are more variable than others, e.g. the ALA107 residue, which likely is such because of its unshielded position on the protein.

vi, vii In the Timeline view, we see how the domains change over time for the different residues. As is to be expected from having observed the full trajectory, the beta sheets are fairly stable, with the varying parts being the edgemost ones. Similarly, the parts involved in forming the helix domain are likewise very variable. We also see the formation of new, pink domains. In general, most domains are stable overall, but most residues also alternate into some other domain at some point. Figure 7 shows a midway alteration in domains under the trajectory. Note in particular the occurrence of a new, pink loop-like domain up top.



Figure 7: Snapshot during simulation of change in inferred domains. Note in particular the occurrence of a new, pink, loop-like domain near the top of the complex.

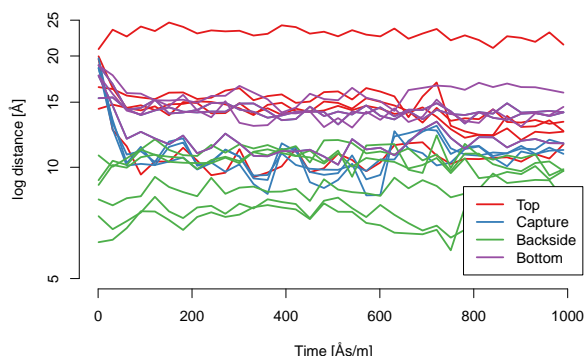


Figure 8: Analysis in change in distance of four select parts of the protein. Here 'top' denotes the end of the protein which has an inferred 'open' backbone, i.e. with unconnected backbone loops. It can clearly be seen that the capture gap is the one structural part of the protein which changes significantly in the initial stages of the simulation.

viii Analysis of the variability in distance between certain molecules is visible in fig. 8, where the logarithmic distance for the distances between select atoms is given. Note in particular how the one part of the protein which changes in variability is the capturing cavity, because of the enclosing of the capture. In addition, we see how the topmost part, which we identify as the end of the tube with an open backbone, is constantly very variable, as we also see visually.

ix Figure 9 shows the charged molecules overlaid on the inferred structure of the protein. Noticeably, most charged compounds are located far away from the backbone, as well as from each other. Similarly, the polar amino acids are all located in the outmost parts of the protein (fig. 10). The more flexible parts of the protein also tend to consist of smaller amino acids (not shown).

Interestingly, the protein consists of to a large extent of hydrophobic amino acids (figs. 11 and 12). This suggests that incoming lipids will bind with their tails inside the cavity, as the hydrophilic head would be comparatively inclined to be aligned outwards. The hydrophobic amino acids are also fairly evenly spread out across the cavity, with only minor gaps.

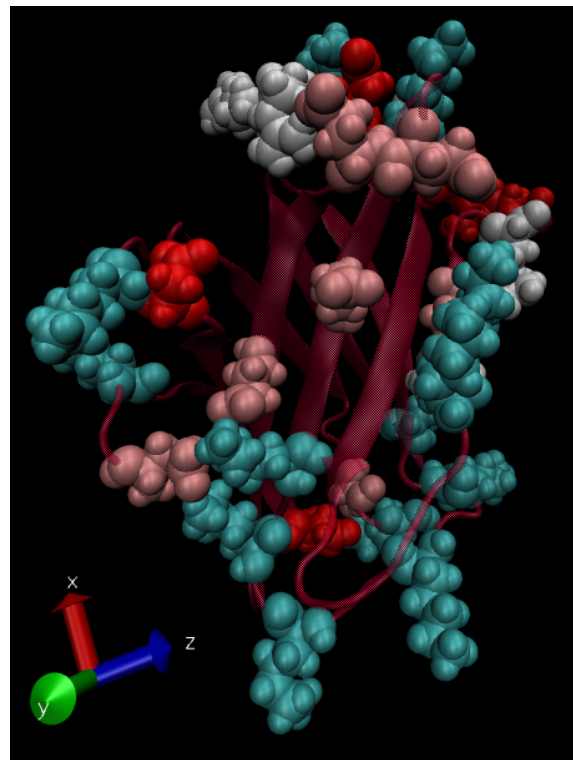


Figure 9: Charged residues overlaid on the *NewCartoon* template of the molecule. Most of these are situated on the outsides of the protein, and not close to the cavity.

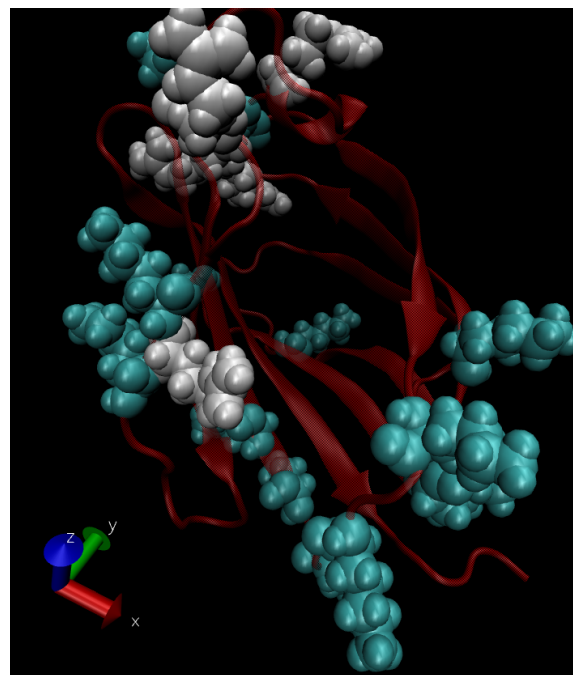


Figure 10: Likewise, the polar amino acids are like the encapsulating group located primarily on the outermost parts.

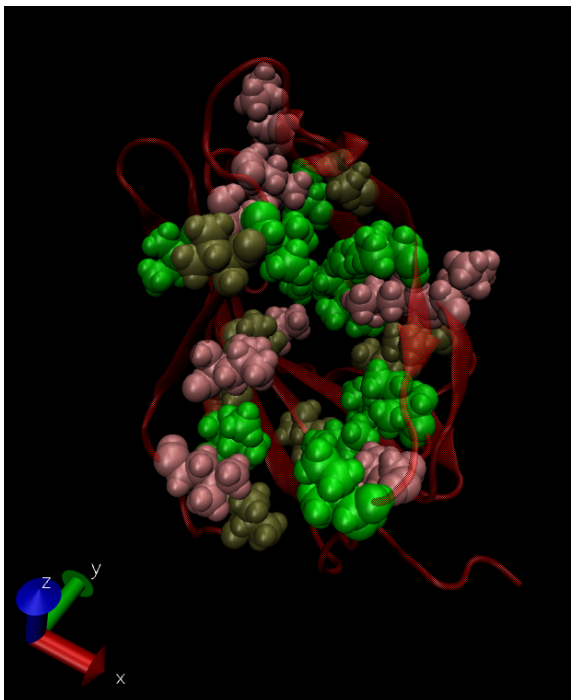


Figure 11: Hydrophobic and aliphatic amino acids.

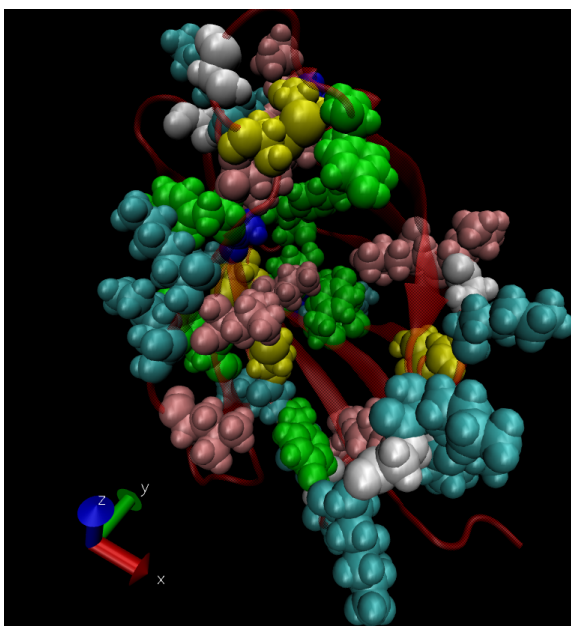


Figure 12: All hydrophobic amino acids. These are fairly evenly spread out across the protein and also present within the cavity.

x In the beginning of the simulation, a few water molecules are present inside the cavity, whereas later on no such molecules are enclosed (see earlier remarks). Throughout the timecourse, the wa-

ter molecules appear to interact with the structure mainly through its hydrogen, and appears to otherwise be evenly spread out across the whole protein.

xi The molecule turns around one of its axes, initially being positioned with its functional ring being positioned inside of the cavity. At the end of the simulation, however, the residue has been pushed out of the cavity and effectively turned ca 180° such that the aromatic ring is on the outside of the complex.

Part B

i-iv We first note that with the ligand bound, there is no inferred alpha helix domain anymore. Compared to the unbound protein, the Ramachandran plot for the bound counterpart tells of a similar distribution of residues with the same angles.

Identifying changes in distances for the protein's vital parts (fig. 13), i.e. the front- and backend of the cavity, we see that there are no significant changes in distances over the whole timecourse, likely because the structure is already close to its functional energy minimum. Also, just as we can expect, the functional parts, i.e. the parts involved in the capture process are the most variable, whereas the backside of the protein is fairly stable throughout. In addition, the beta sheets are more stable than the outermost parts of the cavity.

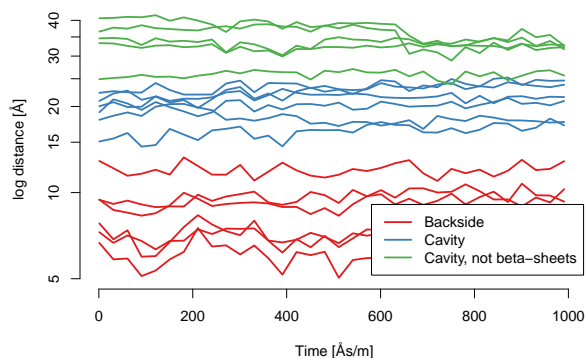


Figure 13: Analysis of the distance between atoms on either end of the labelled structures. As expected, the backside of the protein is the most stable, whereas the beta-sheets are affiliated with larger movements.

Doing the same for the ligand itself, measuring distances between atoms on the heads and tails respectively, it is easily seen that the tail parts are significantly more volatile than the heads, showing signs of

reorganisation within the cavity throughout the simulation. In contrast, distances between atoms on the heads stay surprisingly stable, which ought to be because the stabilising bonds they are connected to each other with, and which the five tails encapsulated in the cavity lack.

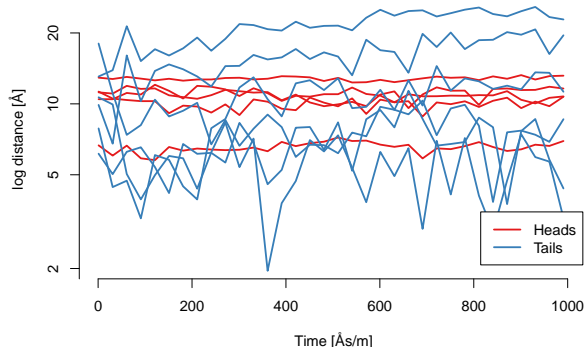


Figure 14: Distances between select atoms on a few structurally interesting parts of the protein. The headgroups are clearly much more stable than the tails, possibly due to electrostatic interactions with aminoacids involved in locking in the

The only positive amino acid binding to a phosphate group which we are able to identify is a sole arginine, which binds to phosphate group one of the ligand, as seen in fig. 15. However, also an aspartate and a glutamic acid, which are both negative, bind to the two phosphate groups through intermediate Na ions, which therefore indeed bridge the interaction. We also observe (not shown) that basic residues form contacts primarily with neutral ones and aliphatic ones.

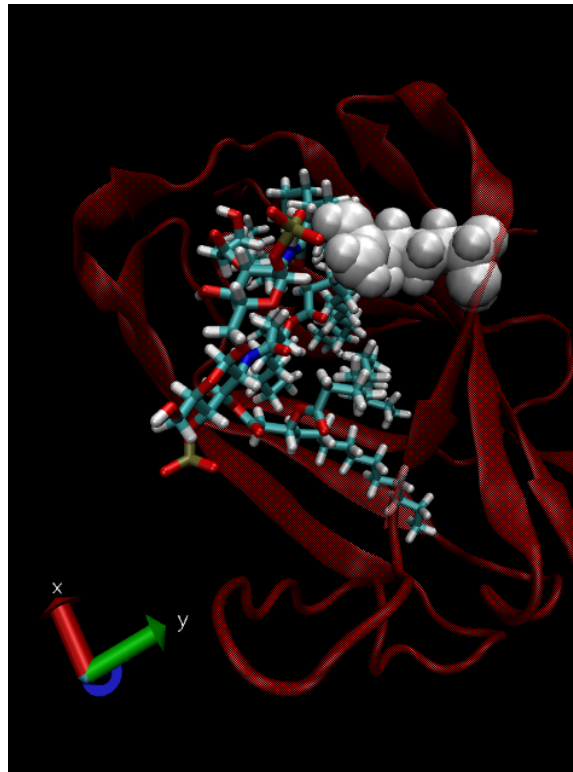


Figure 15: Positive amino acids within 5 Ångström of the ligand. Snapshot taken at the very end of the simulation. Only a single ARG is present. However, when also including negative amino acids, which are intuitively less likely to interact with the negatively charged headgroups, we find that these bind to the phosphate groups via intermediate Na ions (not shown).

Interestingly, in order to bind the ligand, the bound complex appear to be more flexible with respect to the non-sheet surfaces, indicating that the ligand intercedes with the ordinary path of least resistance, which ends up being what effectively entangles it in the cavity. Withing the cavity, the hydrophobic residues in the protein creates a hospitable environment, and does along with locking of the polar headgroups keep ligand in place.

In the non-bound states, the protein visually appears more rigid with respect to the overall structure, whereas it in the bound case has to adapt to the incoming ligand. In other words, the trajectory in the bound case is more varied, due to the induced environmental change. However, at the point where the ligand has stabilised somewhat, the other structure appears more variable; especially the helix moves away a bit at some of the later timepoints.

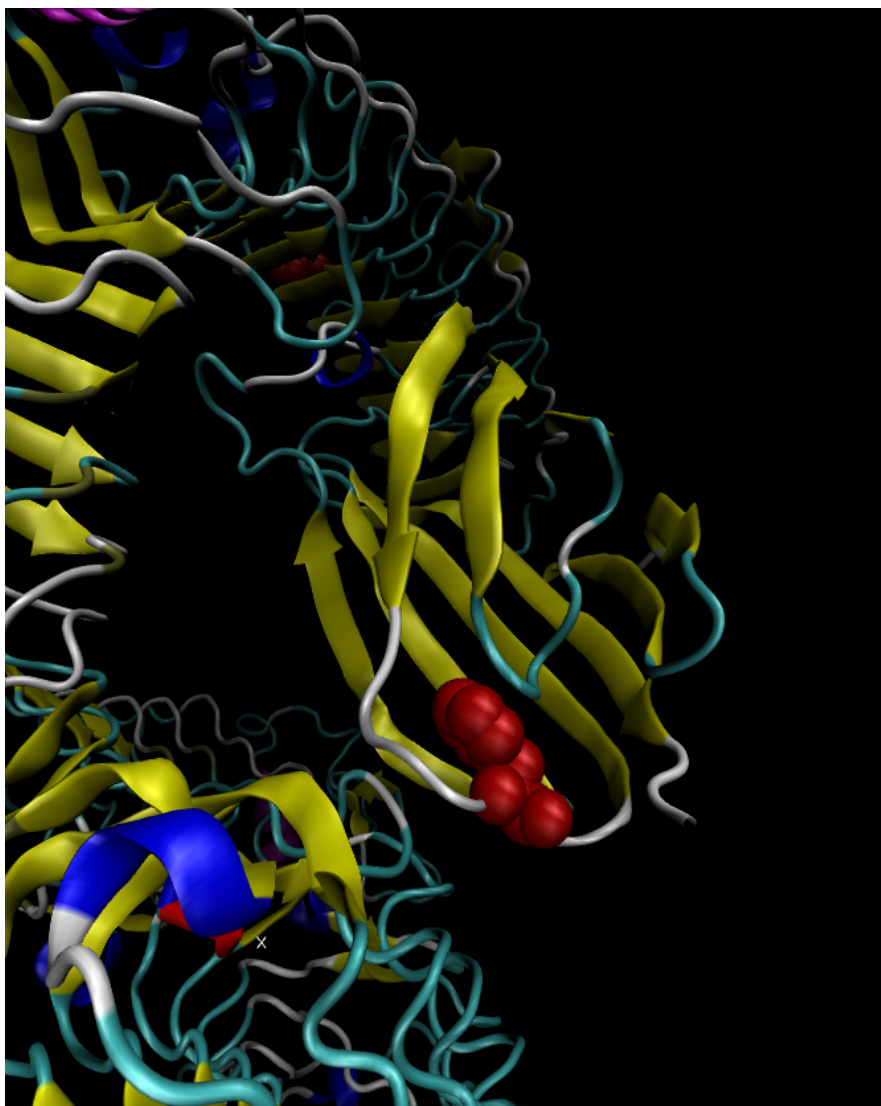


Figure 16: TLR4 associated with the MD2 protein, with residue Phe126 shown in red, aligned inwards to the protein.

v In fig. 16 the protein associated with the rest of the complex can be seen, with Phe126 coloured in red. Reminiscing the earlier observations of Phe126 turning during activation, it appears possible that the conformational change accompanied with the residue inhibits the protein when set free, but that this effect is hampered when the ligand is bound. However, if the ligand is smaller, e.g. due to fewer tails or similar, the change will still be able to occur to some extent, thus inhibiting TLR4. This would explain the observed decay in pronunciation for ligands with fewer tails.