## Doc from supervisors before start of work:

* Create bounding boxes for proteins - python library (residue, pdb/gro, offset -> x, y, z, box, boxy, boxz)
* Visualization of proteins - python library
  + stick and point cloud visualization
  + surface visualization via MSMS tool
  + visualization of protein as 3d object via PyVista
  + visualization of chemical properties via PDB2PQR and APBS
  + visualization of secondary structures via DSSP (optional)
* Geometric measurements of distances - python library
  + pdb, residue numbers -> distance, centre-of-mass, total charge
* Active-site-characteristics (given residue numbers)
  + static (pdb): charge, surface area, centre-of-mass,
  + dynamic (trajectory): gyration, oscillation (principal frequencies oscilations), magnitude of oscillation

Thesis day 1 – 9/22/2021

# Meeting

<https://www.rcsb.org/> - protein database. Get PDB files here

pymol – existing tool

1. probably existing – loading pdb files (3d point cloud)

2. visualize with pyvista (size of atoms)

What surface algo does pymol use

Sphinx – for documentation

# Planning

Do research (on what though??)

Find pdb importing tool

Visualize point cloud

9/24/2021

# Notes:

While mmCIF is now the standard, there is still legacy support for PDB [files.](https://towardsdatascience.com/visualizing-and-analyzing-proteins-in-python-bd99521ccd)

Biopython “tutorial”:

<https://towardsdatascience.com/visualizing-and-analyzing-proteins-in-python-bd99521ccd>

# Log:

Managed to visualize point cloud and spline between atoms (as per the order they are given)

# To do:

Visualize residues (COM of cords) and make a spline with the residues in correct order (stick and ball)

9/27/2021

# Notes:

First looked at potential existing tools for [visualization](https://www.researchgate.net/post/What_are_the_best_python_libraries_or_apis_for_3D_protein_structure_visualization).

Spent the whole entire morning (up until 13:15), trying to download the open-source [version](https://github.com/schrodinger/pymol-open-source/blob/master/INSTALL) of pymol. First directly, then on the VM, then again on windows using a [tutorial](https://tubiana.me/how-to-install-and-compile-pymol-windows-linux-mac/). Without success.

Started the stick and ball. On that end made todo list below.

Currently atomic radii are “None” by default. So I edited the StructureBuilder, PDBParser constructor to add radius.

# Task at hand:

## Stick and ball:

1. Color atoms
2. Make that atoms are varying sizes based on actual atom
3. Figure out atom pairs that are bounded
4. Draw sticks

9/28

# Questions/hurdles

1. How to rebuild the Biopython library on anaconda?
2. What does first item (residue) in slack document mean?
3. How exactly will they use this? What exactly will it be used for?

# Notes

Managed to color the atoms. But can’t choose the color and is sub optimal. No scaling

9/29

# Notes:

No meeting. Worked most of the day. Managed to figure out how to scale the spheres (atoms) and also got the correct coloring. Then I made spheres for the residues (amino acids), also made them transparent. Then figured out the approximate bonds and plotted them.

In other words: stick and ball model completely done

9/30/2021

# Notes:

<http://mgltools.scripps.edu/packages/pmv> PMV: python molecule viewer

started researching surface visualization: <https://ssbio.readthedocs.io/en/latest/_modules/ssbio/protein/structure/properties/msms.html#get_msms_df_on_file>

10/01

# To do:

Figure out how to plot the surface: (MSMS, DDS)

1. Find method
2. Implement method

10/04

# Notes:

Up until now: ok, so rn I’m in the process of downloading the open source Pymol AND trying to figure out the surface visualization with MSMS

I should try to download pymol, as that might already have almost everything done. Not sure what my thesis will be about then, but I have to get that done. If that doesn’t have surface vis and basically everything the supervisors are looking for then I can start on the MSMS

Spent the whole entire day trying to download pymol (and get MSMS to work for a tiny bit)

10/05

# Notes:

Spent another 45 min downloading Pymol. This time with SUCCESS!

Spent rest of day trying to figure out how to access pymol from python file and understand msms and ssbio

# Notes for meeting tomorrow:

* Why not use pymol
* Not sure msms works
* Is my stick and point model good as it is?
* What exactly do you need this tool for? How will you use it?

10/06

# Meeting 2:

Pdb to openbabel – visualize bonds fromthere

Install binary for msms manually

Offset to bounding box

Pdb to Mol2 has bond info

10/07

# To do:

* Remove solvent atoms (done)
* Correct the bonds
  + Get bond information
    - Convert to different file format OR
    - Use library to extract bond info
* Visualize surface
  + Get MSMS to work OR
    - Make sure correct binary
    - Change ssbio files to work with biopython
  + Find different visualization library
* Refracture code

10/08

# Notes:

Continue to do from yesterday, in particular bonds.

Change of plans. My laptop’s DNS protocol is not working properly, can’t download openbable -> refractored code. Now back to trying to solve DNS issue

10/10

# Notes:

Worked on surface plotting. Rough surface achieved

10/11/2021

# Notes:

Figured out how to smooth surface (indexing from 1, not 0 in xyz file…).

Working to see how I can convert pdb to xyz on windows.

Decided against converting bash/awk script to python -> created python script to call bash script for linux only

10/12

# Notes:

Continue work on solely linux supported script.

Added scripts to main file. Modified them so that they realize the OS they are running on and call the script accordingly.

I might have to develop a pdb to xyzr python script either way, as I think solvent is included in current xyzr files…

10/13/2021

# Planning:

Ok, so I have a program that, on linux, can create the surface AND stick and ball visualization plots from a PDB. It only runs on Linux. The bonds are still not at all perfect. And surface is also 1 color. Apart from that I should look at runtime object creation.

# To do:

* Make it windows compatible – create pdb\_to\_xyzr(n) python script
* Runtime object creation
* Figure out bonds
* Color surface

# Notes:

Converted bash-awk script to python!

Not completely… Doesn’t work on special cases -> have to look into awk more

Special named atoms (so most of them) don’t work. Found source of problem, have not started on solution yet

10/14

# Notes:

Above mentioned script done.

Wrote a few more scripts to convert raw (wiki) atomic radius data to dictionaries.

Tested whole program for hemoglobin. Takes ages (30 min)

10/20

# Meeting:

* Surface
* Refractored code
* Scripts

# Questions

* What to do with bonds (Avogadro)?
* Do I need a paper?

# Meeting notes

Openbabel mol2 - DONE

Msms non binary? – couldn’t find open source code

Triangulation surface instead of msms convex hull (scipy)

Temp files in temp folder (don’t pollute)

Color given residues and atoms differently

Colorful (hydrophobic hydrophilic) visualization

Colorful by charge

10/22/2021

# Notes:

* Figure out bonds
  + From mol2 file
  + Using pandas df

Have: list of atom indexes for two endpoints of bonds

Need: list of 3D coordinates of these indexes

To do: filter out x,y,z coordinates by atom index from main df

DONE

Visualization of Van der Waals radius for atoms possible. Refractored code a bit more. Started working on residue info.

<https://medium.com/quick-code/understanding-self-in-python-a3704319e5f0>

PDB FILE INFO:

<https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/tutorials/pdbintro.html>

10/27

# Notes:

Working on surface (msms binary should be worked around)

# Meeting:

Create plan of work for semester