We are going to talk about the protein data bank. It holds our accumulated knowledge on macromolecular structure (proteins, DNA ,RNA,…). The knowledge is freely available for anyone to use. There are multiple proteins involved in disease solved with the drugs that inhibit them. To access the Protein Data Bank (PDB) go to https://www.rcsb.org/. Most of the structures have been determined by one of three techniques: X-ray Crystallography, Nuclear Magnetic Ressonance (NMR), or CryoElectronMicroscopy (CryoEM). These techniques are expensive and not viable on all systems, hence our structural databases are not that large.

The PDB contains many tutorials and statistics which you can browse at your own pace. These structures are often the starting point for drug design strategies Scientist look at the structures of proteins involved in disease, find how they interact with other molecules and find out about which interactions they might inhibit that stops the disease mechanism. Once they have identified a protein to target, they design molecules capable of interacting with the target protein to displace the disease causing interaction.

First we need to understand several things about the PDB. Based on genome sequencing of organisms (think 23andme or ancestry at a large scale!) we know of millions of protein sequences… is our structural knowledge on par with this sequence knowledge?

The PDB provides a unified format for all of these structures, you can download files as file.pdb. This task is an introduction to PDB format.

Tasks on the Computer:

1. From the rcsb.org website: How many protein structures are currently in the PDB? Draw a pie chart showing the number of protein structures and number of protein sequences (UniProtKB/TrEMBL contains 229580745 sequences).
2. The PDB provides a unified format for all of these structures, you can download files as file.pdb. Open the `tri-Alanine.pdb` file using a plain text editor like Visual Studio Code. Inspect the contents. Each line in this file that starts with ATOM corresponds to an atom of the system. Each line is comprised of several columns. Here is a breakdown of the columns:

```
1         2  3    4   5   6      7       8       9       10    11             12
ATOM      1  N    ALA A   1      -1.470  -3.246  -1.171  1.00  0.00           N
```

    1. Indicates whether the atom is a main atom (ATOM) or a heteroatom (HETATM)
    2. Atom number
    3. Atom name
    4. Residue name
    5. Chain ID
    6. Residue number
    7. X-coordinates
    8. Y-coordinates
    9. Z-coordinates
    10. Occupancy

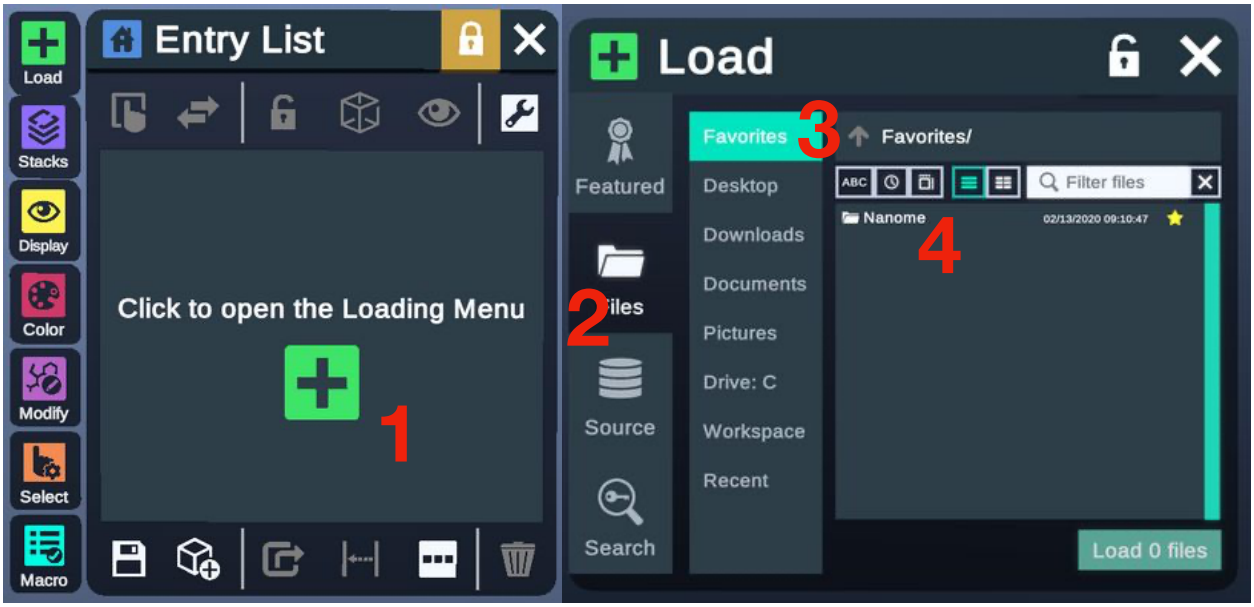
11. Temperature Factor
12. Element name according to periodic table.

You can read more about the PDB format here.

3. What is the coordinates of atom CB of the first residue?
4. Locate the atom with coordinates of X=-1.223,Y=1.500,Z=1.142. What residue and atom number does if correspond to?
5. What is your guess on the unit of the coordinates?
6. All of the residues are Alanines, but what is the difference between them?

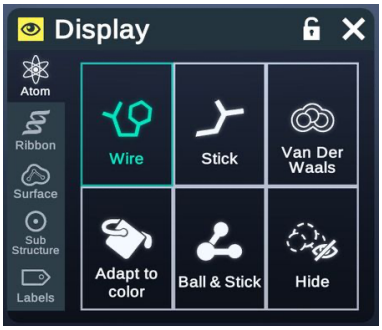
Tasks in VR:

1. Open Nanome and load the `tri-Alanine.pdb` file by using the load module from `favorites/Nanome/Task-1/`

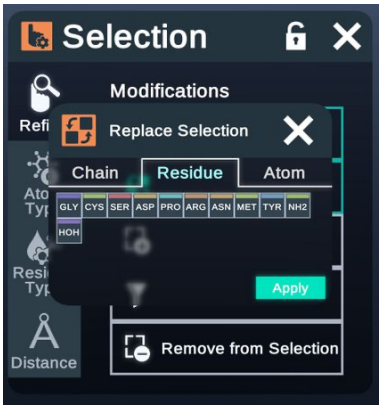


Inspect the structure.

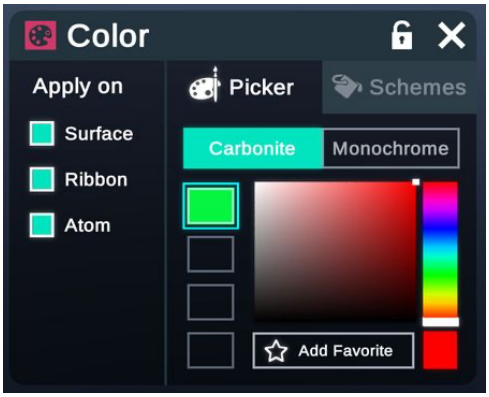
2. Use the Display panel to set the representation of the whole molecule to ball and stick.



3. Use the select panel to select the middle residue and label each atom by atom name.



4. Modify the color of the first residue by selecting it and using the color panel.



5. Take a picture of your molecule (with the labels and adjusted colors) using the camera function on your left wrist. Once you take the picture, you can email it to yourself.