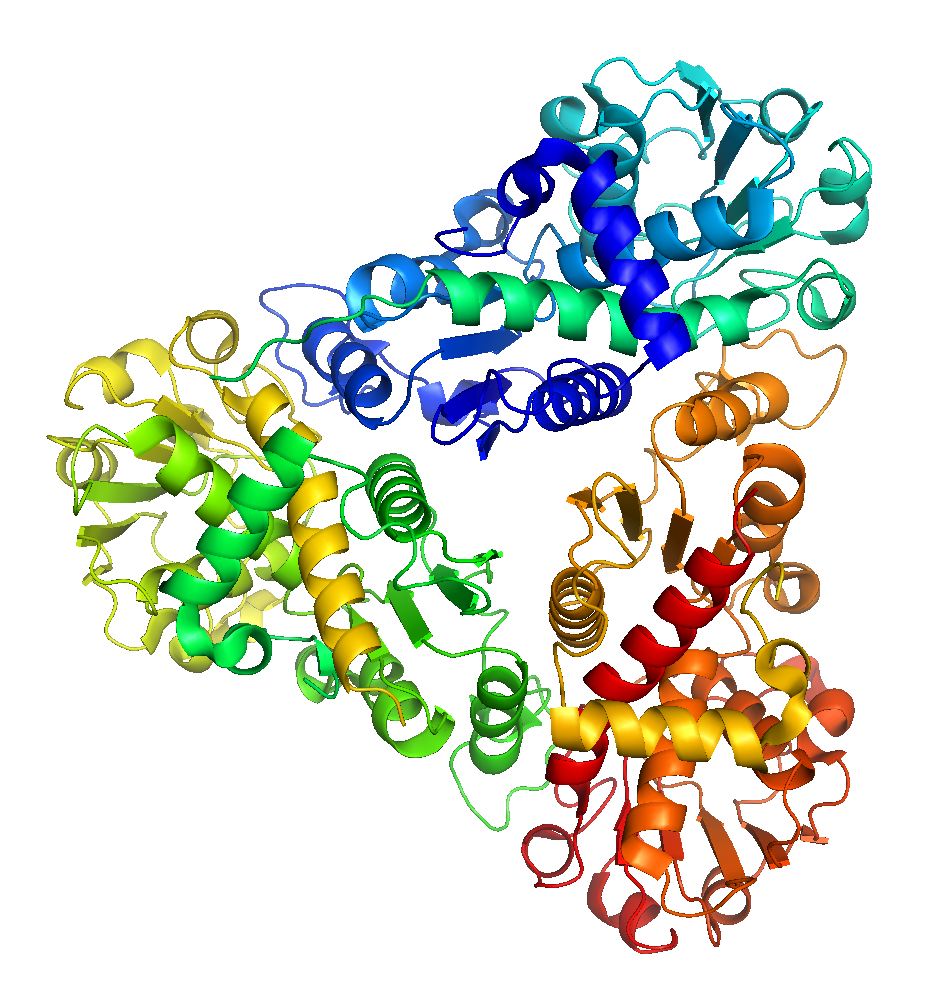
**Introduction**

Proteins are fundamental entities for all living beings. Proteins aid in almost every function that is carried out inside the body, mostly in the cells. They are large and complex molecules. Proteins are made up hundreds or thousands of amino acids which are connected to each other making them long chains. Different amino acids sequences form different proteins. Every protein has a unique structure. The amino acid sequence determines the structure and the function of any given protein.

Proteins are naturally not available in the form that we often see in visualizations, every protein structure is determined by using various methods like x-ray crystallography, Cryo EM or NMR spectroscopy. The amino acid sequence obtained by these methods is stored in a special text file called PDB – protein data bank and has a file extension .pdb. A PDB file usually contains the metadata and the structural information of the protein. Every protein has a unique PDB ID with which it is referenced.

Proteins are visualized using specialized visualization tools that allow the user to check every amino acid in the sequence. Every visualizing tool requires a PDB file of the protein. The popular protein visualization tools available today are PyMol, Jmol/JSmol, RasMol, etc. Using these tools and their supporting libraries we can tweak the visualizations to support our view.



**Figure 1:** PyMol illustration of knotted protein 1JS1.

Knots and slipknots in protein chains make the protein folding process complicated and should generally be avoided. Biochemists have discovered many different types of knots in proteins with varied number of crossings in them. Reaching the native fold or the state with the most minimum energy in these types of proteins is very difficult and hence we need computational support to analyze the protein structures and determine if they are knotted or not.

The motivation behind developing this algorithm was to simplify the protein folding process in general by making sure that any given amino acid sequence does not have a knot or slipknot in it and also simplifying them rapidly and effectively. Biochemists around the world can use this algorithm to check if their protein models have knots and slipknots in them and can avoid knots and slipknots in their models without spending too much time to check for them manually. In most cases visualizing knots and slipknots is a tedious task, in this report we also present an integrated script to support knot and slipknot visualizations.

1. We provide a supporting script for PyMol to highlight knots and slipknots in the analyzed chain.
2. We have established a server to analyze the protein chains and give a clear visualization of the knot or slipknot using JSmol library.

The rest of this thesis is organizes as follows: Section 2 introduces to all the required background knowledge to help understand knotfind and slipknotfind. Section 3 talks about the related work that has been done by other institutions to work with knots and computing the knot types. Section 4 starts by formally introducing knotfind and explaining it in depth, section 4 then progresses to slipknotfind which is an extension of knotfind and describes slipknots and how to detect them and we conclude section 4 by explaining varies types of visualizations that we came up with to highlight knots and slipknots in protein chains. Finally, in section 6 we conclude our discussion and propose possible future ideas to extend this project.