**Motivation:**

Proteins are the basis of life. They play an important role in major biological processes, e.g., maintaining the structural integrity of the cell, transport and storage of small molecules, catalysis, regulation, signaling, and the immune system. There are 20 different amino acids that form proteins in nature. Protein structure is essential for the understanding of protein function. In order to recognize the protein functions of proteins at a molecular level, it is sometimes necessary to determine their 3D structure. Protein secondary structure prediction provides a significant first step toward tertiary structure prediction, as well as offering information about protein activity, relationships, and functions.

(<https://www.nature.com/articles/s41598-018-28084-8>)

Proteins translocating through nanopores are not an rare occurrence in the biological world. During biological processes like protein degradation, protein translocation across organelle membranes, and the delivery of bacterial toxins into host cells, proteins are subjected to external pulling forces and threaded through nanopore openings by ATP-dependent proteases, mitochondrial import machines, and toxin protective antigens. Often the inner cavities of the associated nanopores are too narrow to accommodate structured protein substrates, meaning proteins must unfold in the vicinity of the pore opening and translocate as a peptide chain.

(<https://pubs.rsc.org/en/content/articlelanding/2016/nr/c6nr00410e/unauth>)

Nanopores have been used to detect and analyze biological sample at single molecule level. In nanopore sensing, the interaction of the molecule with the nanopore, e.g. its translocation through the pore, alters one or more properties of the system that can be recorded by appropriate instruments

(<https://www.nature.com/articles/s41598-019-42867-7>)

Along with application to prediction of properties of DNA such as DNA sequencing, Nanopores sensing has also been employed to find the primary structure in protein. In this work we try to find a way to employ this technique to uncover the details regarding the secondary structure of the protein.

**Introduction to problem:**

**Aim-** To study the unfolding of a protein molecule translocating through a nanopore. Figure out a way to predict secondary structures present in the protein based on the translocation time distribution.

**Elaborated-** As established in the motivation section, proteins are an important part of biological processes and deeper knowledge of how these proteins behave in different biological processes can be of great use. To know more about these behaviors knowing the 3D structure of the protein is a key understanding and secondary structure helps us in realizing this goal. We study the simulation results of a protein molecule translocating through a nanopore, having to unfolding in the process, and giving a variety of translocation time depending upon the initial state (which decides which part of protein will interact with the nanopore when the reach contact). This translocation time distribution would contain the information of the protein structure as the translocation time was dependent on the proteins 3D structure. If we can extract this information by mapping the behavior to the secondary structures in the protein, we can possibly predict the secondary structures of any protein by translocating it through a nanopore.

**Methodology:**

Before attempting the protein translocation simulation, we decided to simulate a polymer chain translocation through a nanopore under the influence of an electric field induced force. We used MD simulations to simulate the environment containing the polymer chain containing 100 beads (each with diameter of 1 reduced unit) and a pore having the diameter of 2 reduced unit and depth of 20 reduced unit. The reduced units used here are LJ units with the following scales:

Mass scale - 130 g/mol = (130\*10^-3)/Na kg/molecule

length scale(sigma) = 3 angstroms

energy scale(epsilon) = KB \* T (T = 298.15K) = 4.11 \* 10 ^ (-21) J

time scale = root(m\*sigma^2/epsilon) = 2.174 \* 10^-12 s (Calculated from above scales)

An example movie of such a translocation with different pore size and different electric field is attached:



Each bead has a (negative) charge of 1 reduced unit and a electric field 0.00855 reduced units is applied inside the pore and another electric field 0.000855 reduced units is applied outside the pore.

The above simulation is run for a time of 1 ms and for a total of 500 distinct initial velocities and position of the beads in the polymer. The video below a simulation where each frame corresponds to an initial configuration of the polymer:



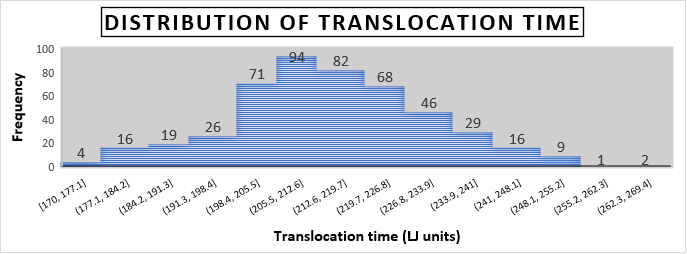
A python-based script was then used to calculate the translocation time for each of these translocations and a translocation time distribution could thus be found.

Our actual goal was though to simulate a protein translocation instead of a polymer translocation. For this, we wanted to take a coarse-grained MD simulation approach. We chose a horse heart cytochrome-c protein (PDB-1hrc) as our protein of interest as a starting point. And converted the structure into an coarse-grained model using martini model.

(following <http://cgmartini.nl/index.php/tutorials-general-introduction-gmx5/proteins-gmx5>)

**Results:**

After imposing a debatable assumption of restricting the leading bead to only move along the z direction, the following translocation time distribution was found from the 500 simulations of translocations of polymer chain through nanopore:



The following are the horse heart cytochrome-c proteins all atom and coarse grained pdbs respectively:

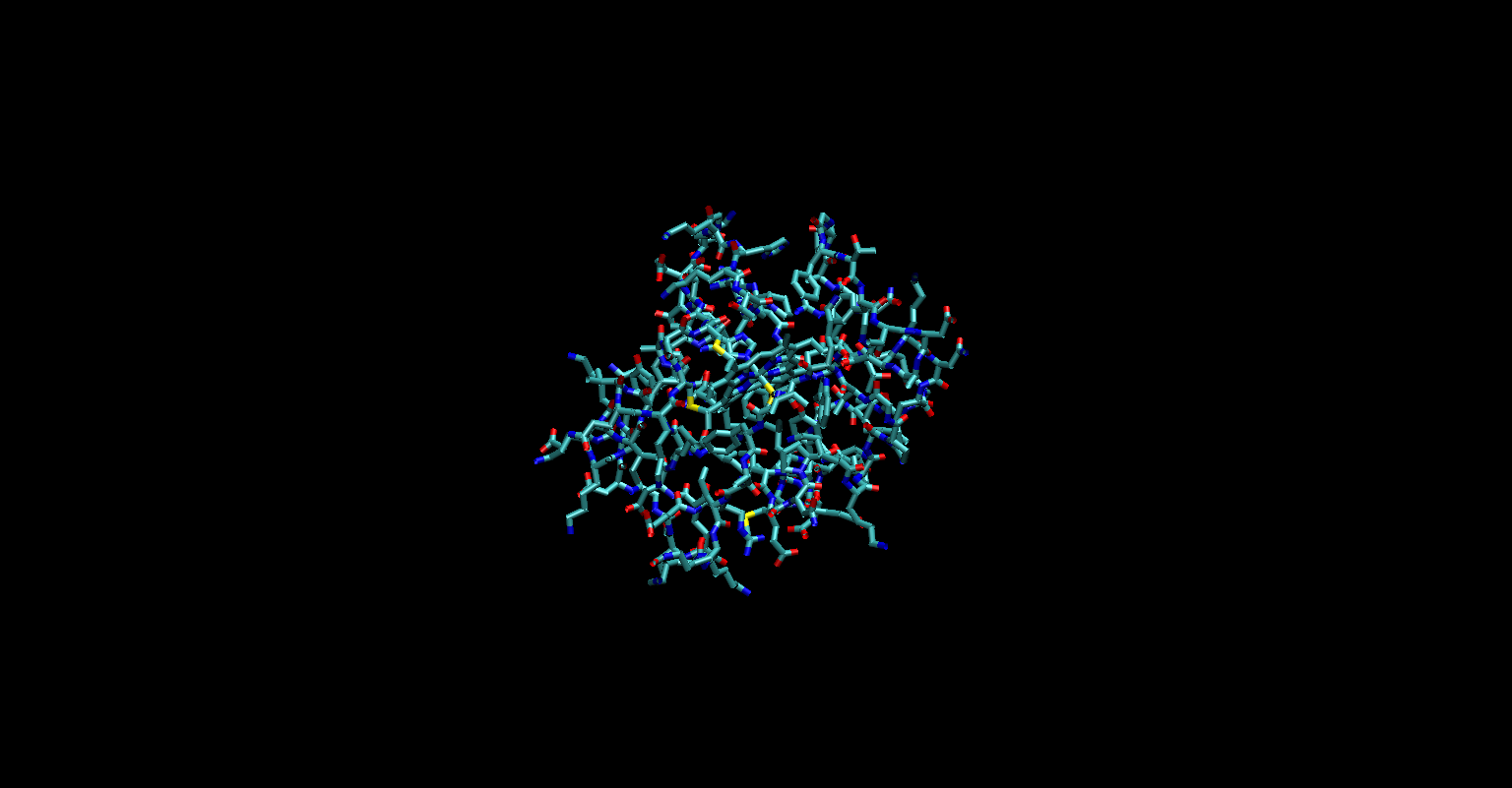


Fig: All atom PDB of the horse heart cytochrome-c protein

Fireworks in the sky

Description automatically generated with medium confidence

Fig: Coarse-grained PDB of the horse heart cytochrome-c protein

A picture containing silhouette, night sky

Description automatically generated

Fig: Comparison of All-atom and Coarse-grained PDB of the horse heart cytochrome-c protein (Blue and red lines represent all atom PDB and yellow and pink beads represent coarse-grained PDB)

**Future Plan:**

1. Run a simulation with this coarse-grained model instead of the polymer chain
2. Running the above simulation 500 or so times with different initial conditions, producing a translocation time distribution
3. Analyzing the translocation time distribution to figure out correlations between the distribution and secondary structures present in the protein (possibly repeating this on other proteins to confirm the found correlation).