Diffusion coefficient:

Diffusion coefficient, also called Diffusivity, is an important parameter indicative of the mobility of a system(e.g. protein) in an environment(e.g. water). Calculating the diffusion constant of proteins, antibodies, and enzymes is crucial for gaining a deeper understanding of the dynamic properties and interactions of biomolecules in various contexts, ranging from fundamental biology to applied biotechnology and medicine. Calculating the diffusion using molecular dynamics (MD) simulations is important for several reasons:

- **Understanding Molecular Mobility:** The diffusion constant provides insight into how fast and efficiently a molecule moves through a solution. This is critical in understanding the dynamic behavior of biomolecules in different environments, such as within cells or in extracellular fluids.
- **Biological Functionality:** The mobility of proteins, antibodies, and enzymes is closely related to their biological functions. For example, enzymes need to diffuse through cellular environments to find their substrates, and the rate of diffusion can influence reaction rates. Similarly, the diffusion of antibodies affects their ability to recognize and bind to antigens.
- **Drug Design and Delivery:** Understanding the diffusion of biomolecules can inform drug design, particularly in predicting how a drug will interact with its target or how it will be distributed within the body. For instance, the diffusion rate can affect how quickly an antibody reaches its target or how a protein therapeutic behaves in the bloodstream.
- **Impact on Aggregation and Stability:** The diffusion constant plays a role in understanding protein aggregation, a process relevant in diseases like Alzheimer's. Understanding how proteins diffuse and potentially aggregate can aid in developing strategies to prevent or mitigate these processes.

Several methods can be used to calculate the diffusion constant of proteins, enzymes, or other biomolecules in molecular dynamics (MD) simulations. The most common approaches include the Mean Square Displacement (MSD) method and the Velocity Autocorrelation Function (VACF) method which we are going to explain in this summer school. We will learn about MSD method in this class and will learn about VACF method late this week:

1. Mean Square Displacement (MSD) Method:

The MSD method involves tracking the positions of the atoms or the center of mass of a molecule over time. The diffusion constant D is obtained from the linear relationship between the MSD and time, according to the Einstein relation. For a 3D system, the diffusion constant D is given by Einstein Relation:

$$D = \frac{1}{6} \lim_{t \to \infty} \frac{d}{dt} \langle \Delta r^2(t) \rangle$$

Where $< \Delta r^2(t) >$ is the mean square displacement(MSD) at time t.

The MSD method is straightforward and widely used in MD simulations to calculate diffusion constants for both small molecules and larger biomolecules like proteins.

2. Velocity Autocorrelation Function (VACF) Method:

The VACF method calculates the diffusion constant by analyzing the decay of the velocity autocorrelation function over time. The diffusion constant D can be obtained by integrating the VACF. The diffusion constant is given by Green-Kubo Relation:

$$D = \frac{1}{3} \int_0^\infty \langle \mathbf{v}(0) \cdot \mathbf{v}(t) \rangle dt$$

Where $< v(0) \cdot v(t) >$ is the velocity autocorrelation function at time t.

The VACF method is useful when the MSD approach is less effective, particularly in systems with short simulation times or when high-frequency dynamics are of interest.

Mean squared displacement

The mean square displacement is a measure of the deviation of the position of a particle with respect to a reference position over time.

$MSD(t) = < \Delta r^2(t) > = \langle |r(t) - r(0)|^2 \rangle$

r(t) is the position of the particle at time t.

r(0) is the initial position of the particle.

$\langle \cdot \rangle$ denotes the average over all particles or over multiple time intervals.

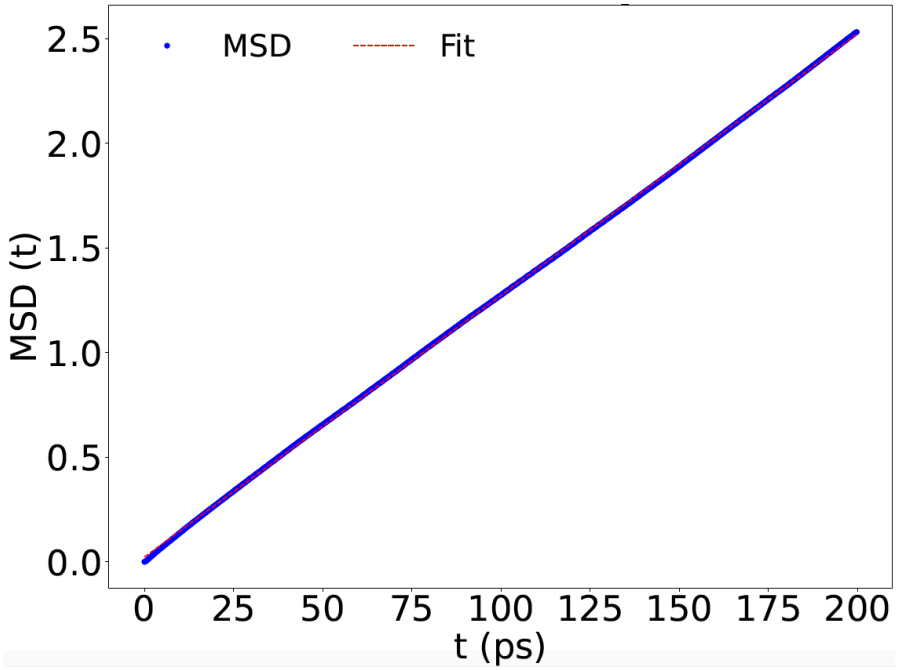
The formula calculates the average of the squared distances of the particles from their initial positions after time t.

Diffusion Constant (D):

The diffusion constant D quantifies how fast a substance spreads out or diffuses in a medium. It's directly related to the mean square displacement. For a system in three dimensions, the diffusion constant can be derived from the MSD using Einstein Relation:

$$D = \frac{1}{6} \lim_{t \to \infty} \frac{d}{dt} \langle \Delta r^2(t) \rangle$$

Where $< \Delta r^2(t) >$ is the mean square displacement(MSD) at time t.



Instructions for Calculating Mean Square Displacement (MSD)

Overview

One approach to calculate the Diffusion coefficient (D) is using Mean Square Displacement (MSD). MSD can be calculated by using the position of the center of mass (COM) of the protein. The coordinate.tcl script calculates the coordinates of the COM of the protein for all frames in a DCD trajectory file.

Command Details

- **center selection [weight weight]**: Returns the geometric center of atoms in the selection using the given weight.
- **com**: Computes the center of mass of the selection based on the idealized coordinates.

Running the Script

To calculate the coordinates of the COM for each frame, you can use the run_coordinate.pbs file to run the coordinate.tcl script. This script requires the following input files:

1. **DCD file**: ../2500Frames/1ubqNVT.dcd (relative path to the DCD file)
2. **PSF file**: ../molPdbPsf/1ubq.psf (relative path to the PSF file)

Making the Script Executable

To make the shell script executable, follow these steps:

1. Open your terminal.

Type the following command and press Enter:

```
chmod +x run_coordinate.pbs
```

Running the Script

After making the script executable, you can run it with the following command:

1. Open your terminal.

Type the following command and press Enter:

```
./run_coordinate.pbs
```

Output

The output file containing the COM coordinates will be available in the coordinateProtein folder.

Further Analysis

To calculate the MSD/Diffusion, additional analysis is required. Typically, Python is used for this purpose. The coordinate file generated from the coordinate.tcl script will serve as the input for the Python code that calculates the MSD/Diffusion.