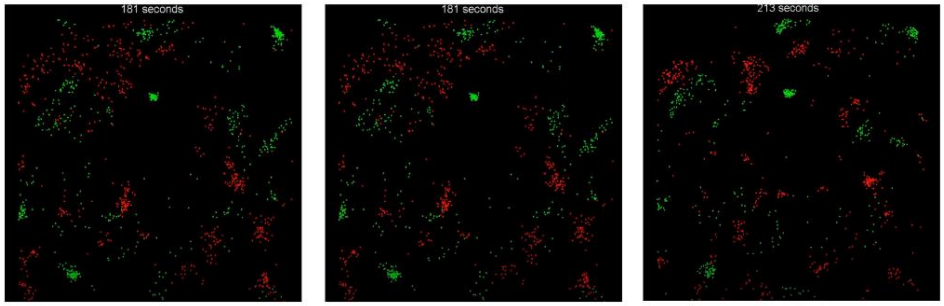
The Virtual Microscope

Project By: Ariel Abramovitz, Shahar Zuntz, Avishai Wizel, Noy Porat.

Code availability: https://github.com/arielabramovitz/Bio-Hackathon-Project



Introduction:

In the realm of immunology, understanding the interactions between molecules is crucial for comprehending the intricate mechanisms of immune response and cellular signaling.

Understanding the intricate interactions (especially over time) between the molecules: CD45, LCK, and TCR is vital for unraveling the complexities of T cell signaling and immune response regulation. They can have profound implications for immune system function and can contribute to various immunological disorders.

Due to the limitations imposed by the size of single molecules in relation to the wavelengths of light, visualizing and tracking them using even advanced light microscopy techniques, such as confocal microscopy, is extremely challenging. Additionally, time-resolved imaging becomes sluggish, impeding the observation of complete sequential motions. Considering these constraints, our primary objective is to improve the temporal resolution or frame rate of synthetic or real microscopy time-series comprising individual molecules. To achieve this, we plan to leverage prior statistical and physical knowledge. By employing these methods, we anticipate enhancing the resolution and smoothing the transitions between frames.

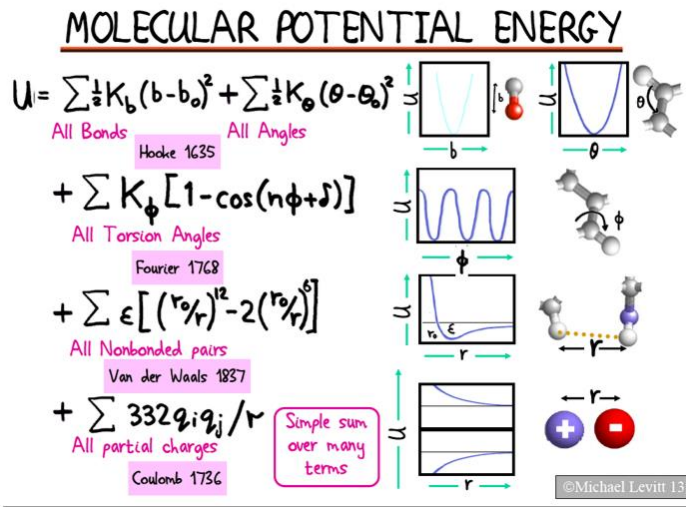
Our project is a software toolbox consists of two key components. The first component utilizes advanced algorithms to enhance the resolution of input image sequences, resulting in higher-resolution movies that closely resemble reality. This enhancement enables a more detailed and accurate representation of molecular motion captured in microscopy time-series. The second component focuses on generating synthetic movies that realistically simulate molecular motion. Leveraging sophisticated algorithms and techniques, this feature produces synthetic movies that closely mimic real-world scenarios.

Background:

Molecular dynamics is a fundamental approach that investigates the dynamic behavior of molecules by leveraging the principles of physics. Unlike the static representations of molecules, we have encountered thus far in the course, molecules exhibit intricate motion in space, responding to applied forces, interacting with other molecules (such as water molecules), and engaging in mutual interactions (Träuble, H., 1971).

At its core, molecular dynamics encompasses a range of methodologies that employ physical laws to drive the movement of molecules. In this context, a dynamic system refers to a system comprised of interacting components that evolve over time, guided by a set of rules dictating their changes and interactions. These components reside in a configuration space, with their positions and coordinates serving as the focal points of interaction, leading to temporal modifications in the system's configuration. By considering the forces acting on the molecules, we can probe the system's state at a given time interval.

Interactions among molecules are described in terms of potential energy, which is mathematically approximated using a specific functional form that approximates the forces present in nature. This modeling primarily incorporates elements such as spring forces (Hooke's law), torsional effects (Fourier), and other relevant factors. Furthermore, to capture the full picture, it is necessary to account for atom-to-atom communication, whereby atoms attract each other when they approach (repel each other with a power of 12 and attract each other with a power of 6). The goal is to achieve an energy minimum for atoms, considering the forces acting upon them and preventing them from approaching too closely to avoid substantial energy increase.



Merely analyzing the components and their interactions does not provide insight into the system's temporal evolution or how these interactions manifest in movement. To address this, a dynamic model must be established. Newton's laws elucidate the dynamics of the universe by relating force and mass to the calculation of time-varying acceleration. Energy, defined as the integral of force over a path, enables the derivation of forces from the energy function in space. By determining the mass of each molecule, the rate of change of each particle's velocity can be computed, giving rise to molecular dynamics.

Despite the utility of Newton's laws, direct calculations based on them are computationally intensive. Moreover, it is desirable to incorporate the presence of water molecules surrounding the target molecule, as these mimic the displacements experienced in a real environment. Introducing water molecules adds an element of randomness, creating the process of diffusion. To optimize computational efficiency, an analytical analysis is employed to examine the statistical properties of these random collisions. This enables the formulation of novel laws of dynamics that capture the effects of water molecules, even in their absence. Notably, the random movement resulting from these collisions can be effectively modeled using a normal distribution. This leads to two types of dynamics: Langevin dynamics and Brownian dynamics.

In Brownian Dynamics simulations, the focus is primarily on the collective behavior of particles rather than tracking the detailed motion of individual atoms. The method considers the forces acting on the particles, including intermolecular interactions and external forces, and incorporates random thermal fluctuations based on the principles of Brownian motion (Huber et.al, 2019).

This approach ensures that the particle's movement is influenced by both the applied forces and the random movements simulating collisions with water molecules. It provides insights into the diffusive behavior, equilibrium properties, and dynamic processes occurring in systems where the influence of the surrounding solvent or environment is crucial.

Three key molecules that play significant roles in immune cell activation and signaling that we are going to investigate their dynamics and interactions are CD45, LCK, and TCR:

CD45 is a cell surface protein found on immune cells, particularly T cells, B cells, and natural killer cells. It plays a crucial role in regulating immune responses by modulating the activation and signaling of these cells. CD45 acts as a protein tyrosine phosphatase, involved in controlling the balance between immune activation and inhibition.

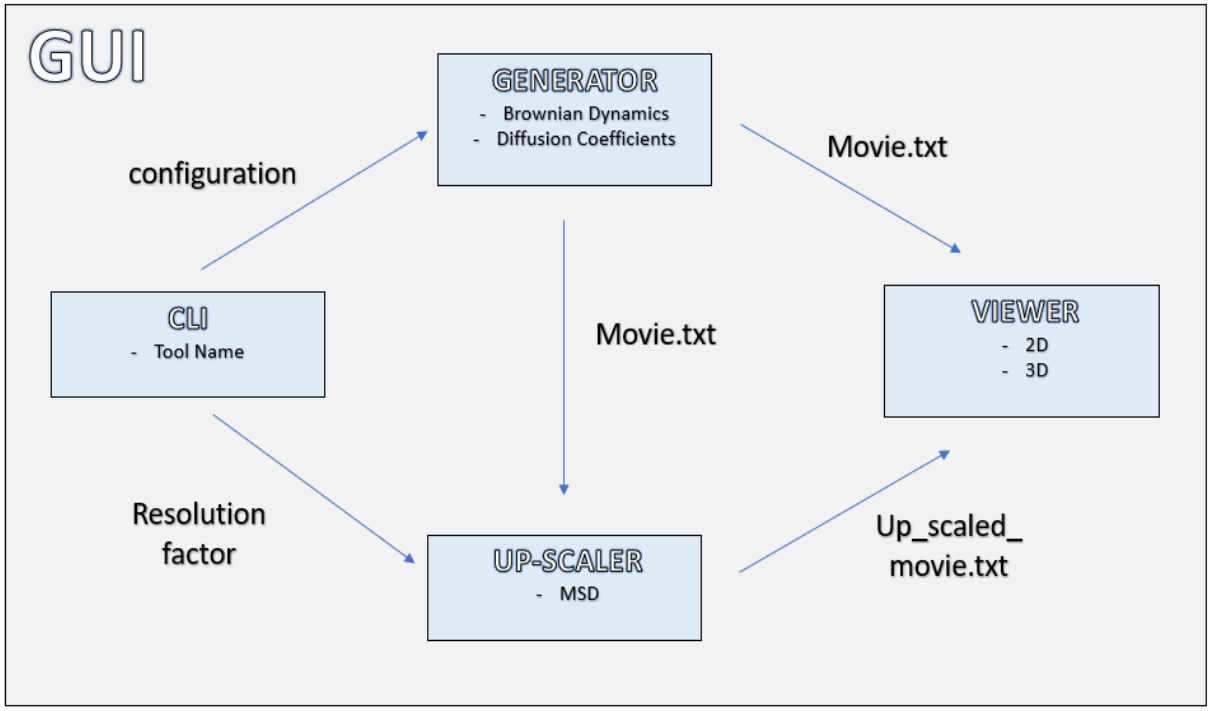
LCK, also known as lymphocyte-specific protein tyrosine kinase, is an enzyme found in T cells. It plays a key role in initiating T cell receptor (TCR) signaling, which is essential for T cell activation and the immune response. LCK helps phosphorylate signaling molecules, allowing downstream cellular processes to occur.

TCR, short for T cell receptor, is a protein found on the surface of T cells. It recognizes and binds to specific antigens presented by antigen-presenting cells, such as infected cells or cancer cells. This binding triggers a signaling cascade that leads to T cell activation and the initiation of immune responses against the recognized antigen.

These molecules are integral components of the immune system, working in concert to orchestrate a cascade of events leading to the initiation of an immune response. (Filipp et.al 2012)

Model and Methods:

Model Overview:



LIBRARIES:

- Numpy
- Matplotlib
- Scipy
- React- Frontend

GENERATOR:

Generation of synthetic frames of molecules in a 2D space. The frames should simulate realistic molecular motion and interactions, and we will be using them as our data set for the up-scaling challenge.

The generator gets a configuration- diffusion parameters for each type (3) of molecules, and 3 interaction parameters for every pair of molecules:

D_rest- the minimal distance between two molecules, for them to be considered as "interacting".

R- the maximal distance between two molecules, for them to be considered as "interacting".

K- the force factor between two interacting molecules, i.e., how strong is the force between them. It is also referenced as a spring factor.

The generator also receives a resolution parameter- it determines the number of frames to generate.

Using the Brownian Dynamics algorithm we calculate the next configurations in time, therefore creating a "movie" divided into frames- a text file that describes the x-y coordinates of molecules in time.

$$X_{n+1} = X_n + \frac{\Delta t}{k_B T} D f(X_n) + \sqrt{6D\Delta t} \cdot R$$

X_n particle configuration at time step n

$f(X_n)$ sum of all forces acting on particle

R random diffusion vector of magnitude ~ $Norm(0,1)$

D diffusion coefficient

Δt time step

UP_SCALER:

Functional tool to increase the temporal resolution of a given movie. It is done by getting a low-resolution movie from the generator. The up scaler learns the diffusion coefficients and adds new frames between every two frames, using Brownian motion and the MSD algorithm. The interaction parameters is currently assumed as known due to time limits of the project, although it is possible to add a function to learn them in the future. In addition, the identity of each molecule in space is also known, but again, a future improvement can be adding a tracking algorithm.

The MSD algorithm analyzes the motion and diffusive behavior of particles in a molecular dynamic's simulation or experimental data. It quantifies the average displacement of particles over time and provides insights into their mobility, diffusion coefficients, and underlying dynamics.

To calculate the diffusion coefficients, the Up scaler uses Mean Square Displacement (MSD) algorithm:

$$\text{MSD} = \langle \Delta r^2(\tau) \rangle = \langle |r(t+\tau) - r(t)|^2 \rangle_t$$

where r(t) and r(t + τ) are the positions of a single particle at two time points τ seconds apart. The angular brackets in the MSD indicate a time average on all such position pairs in a trajectory.

According to the MSD, for each particle and for time interval τ, we calculate the displacements between consecutive time steps (t and t+τ) by subtracting the initial

position from the final position (Fig 1, A). Then, the average of the displacements is squared to focus on the magnitude of the displacements (Fig 1, B). Finally, they are averaged over all particles from the same type to get the MSD for a specific particle type for time interval τ, so we get MSD for every τ, for every molecule type.

To visualize the result, we can plot the MSD for each time interval τ (Fig 1, C). A linear plot indicates normal diffusion, so the diffusion coefficient can be calculated by the equation:

$$\langle \Delta r^2(\tau) \rangle = 2dD\tau$$

where D is the diffusion coefficient and d are the degree of freedom (Gal et al,2013).

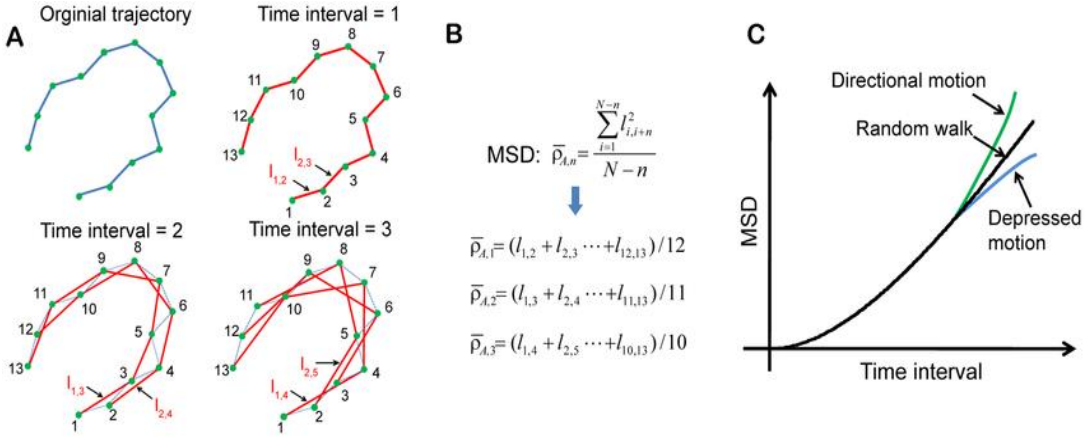
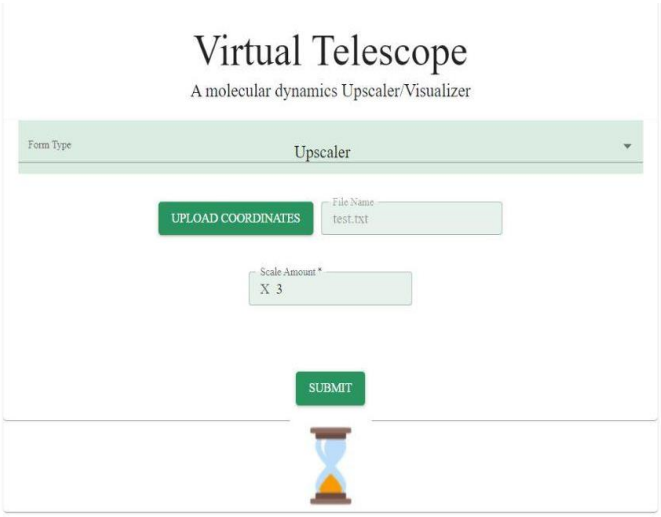


Figure 1

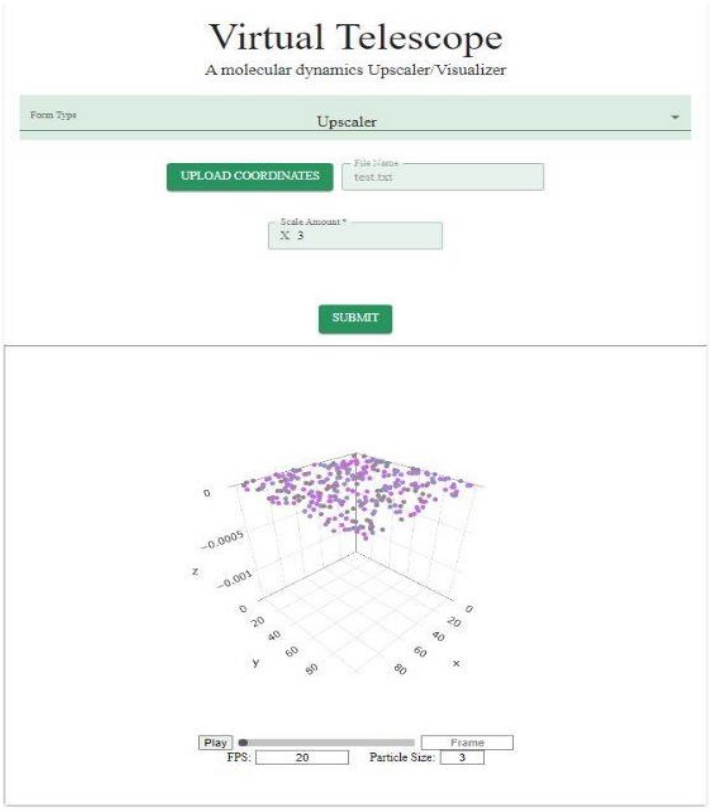
After calculating the MSD, we generate the frames by simulating Brownian dynamics. This simulation uses the diffusion coefficients that we learned, the given interaction coefficients and a direction vector which defines the force. The vector is updated every few iterations to direct the molecules to their "destination", which is the position determined by the frames the Generator outputted. It is calculated by the difference vector between the next "Generator frame" and the last frame generated by the Up-Scaler.

GUI & VIEWER:

Opening screen with the ability to choose the tool to be used, and upload files:



After submitting a file, entering the parameters needed for the tool, and the relevant calculations are finished- the viewer window opens and presents the movie results in 2D or 3D:



Results and Conclusions:

In this project, our goal was to develop a software toolbox capable of enhancing the resolution of molecular movement movies by adding frames in-between the given frames, thereby creating smoother and more realistic representations of molecular dynamics. We have made significant progress in achieving this objective and obtained several noteworthy results.

Firstly, we successfully applied the Mean Square Displacement (MSD) algorithm to analyze the diffusion coefficients of the molecules in the movies. The MSD algorithm allowed us to track and calculate the squared displacements between consecutive frames, providing valuable insights into the mobility and diffusive behavior of the molecules. We were able to approximate the diffusion coefficients with a high level of precision, which is crucial for accurately modeling molecular motion.

To do so, we generate using the Generator 100 frames with 100 molecules for each type, each frame is 0.001 seconds, and the D for TCR is 1, for CDK45 is 2, and LCK is 3, resulting in the following MSD plot (Fig 2). As the figure shown, the bigger D the molecular have, the larger MSD (⟨Δr^2 (τ)⟩) for each dT (τ). After using the equation above and mean all the dT, our approximation was: 'D_TCR'= 1.01, 'D_CD45'= 1.975, 'D_LCK'= 2.76. A remarkably close result the “true” values.

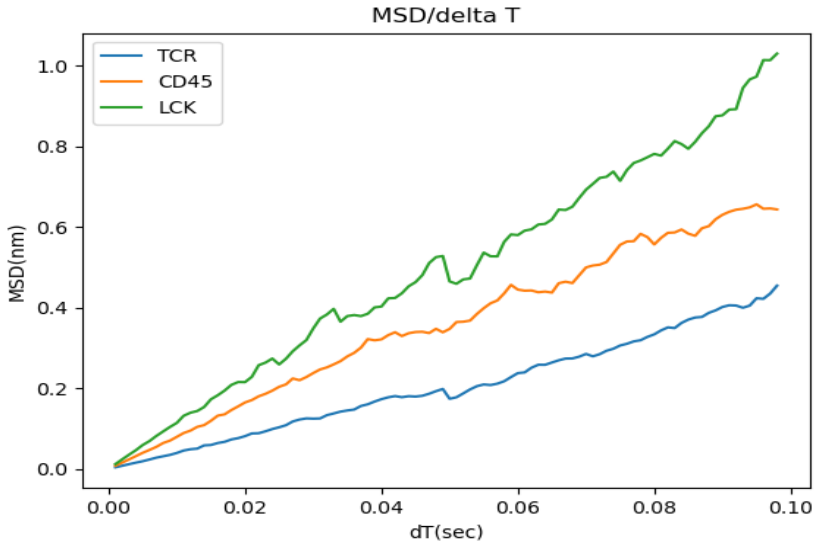


Figure 2

Furthermore, we employed a difference vector approach to model the Brownian movement of single molecules and the interactions between pairs of molecules. We observed that pairs of molecules in a range of distance between each other are defined as "interacting", resulting in moving together next to each other. Moreover, by considering the forces involved and simulating the random collisions with water molecules, we were able to capture the stochastic nature of Brownian Dynamics.

This enabled us to better understand the underlying forces influencing molecular motion and their impact on the overall behavior of the system.

However, we encountered an undesired result in our approach. When adding frames in-between the given frames from the Generator, we observed a noticeable "jump" between the last added frame of the Up-Scaler and the subsequent Generator frame. We attribute this jump to the stochastic factor inherent in Brownian Dynamics. Due to the unpredictable nature of molecular motion, sometimes the molecule does not reach its destination within the given timeframe, resulting in a discontinuity between frames.

Because of this problem, it was hard to evaluate our model. Nevertheless, did tried to evaluate our model by creating 1000 frames by the generator, deleting 9 frames between every 10th frame, and up scale the new movie, adding the same number of frames, resulting a movie of the same length. Then, we compared the original frames to the those outputted by the up scaler: we calculated the average Euclidean distance of every molecule between the original frame to the up-scaler frame. To reduce Volatility of the graph, we used Moving average with window size of 5. The result shows a ~1.2 Euclidean distance between the movies (Fig 3, blue line)

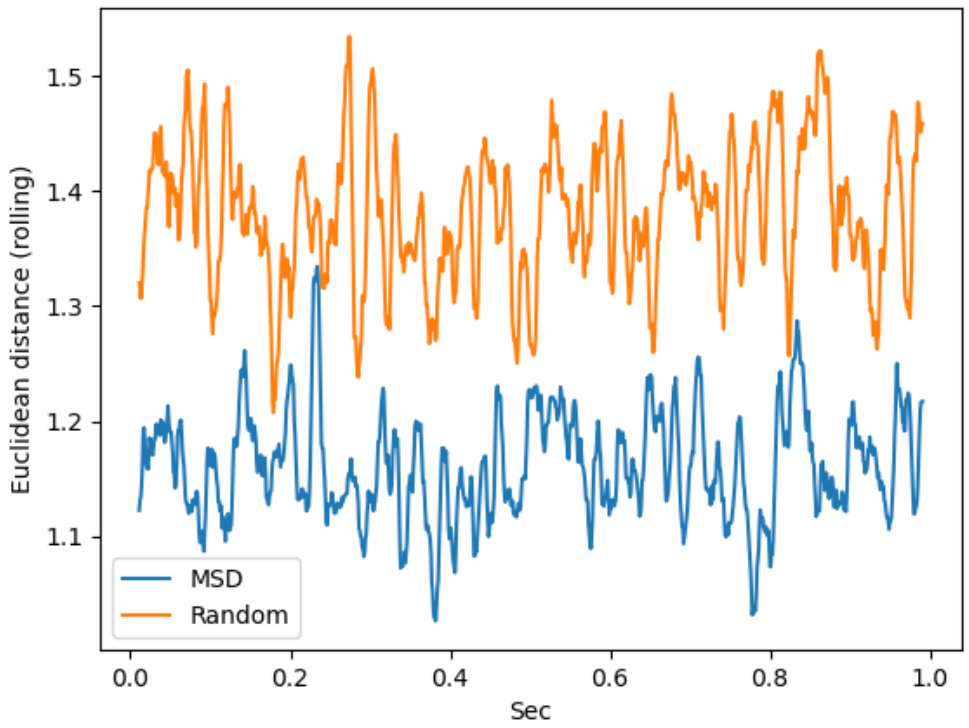


Figure 3

But because of the randomness issue, it is hard to assess this result. However, we did try to compare this model with a model that do not learn the diffusion coefficients

by the MSD algorithm, but rather with random choice of an integer between 1 to five. We did notice an improvement, causing the random model to have ~1.4 Euclidean distance (orange line).

In conclusion, our software toolbox has shown promising results in enhancing the resolution of molecular movement movies. Through the application of the MSD algorithm, we successfully approximated diffusion coefficients and gained insights into the molecular mobility. Additionally, our modeling of Brownian Dynamics and the forces involved provided valuable information on the interactions between molecules. While the observed "jump" between frames remains an area for improvement, we believe our findings contribute to the understanding and simulation of molecular dynamics.

Discussion:

This project aimed to address the challenging task of obtaining high-resolution movies of molecular movement, considering the limitations of current microscopy techniques. By combining knowledge from physics, mathematics, and biology with programming tools and learning methods, we strived to improve the resolution and fidelity of molecular dynamics representations.

Given the relatively short duration of our project, there are several areas for future exploration and improvement. One potential avenue is to leverage more advanced learning methods, including potentially neural networks, to predict additional coefficients such as interaction coefficients. By utilizing distance distributions between adjacent molecules, we can gain deeper insights into the complex interactions occurring within molecular systems.

Moreover, our current representation of molecules as 2D matrices limits our understanding of their true behavior in a three-dimensional space. Expanding our work to incorporate image processing techniques would enable us to capture the intricate movements of molecules in a more realistic manner. Introducing an additional axis to account for the membrane's movements within the cellular context would further enhance the accuracy and applicability of our methodology. By integrating real data from microscopy experiments, we can generate authentic molecular movies that closely resemble their biological counterparts.

Furthermore, adding a tracking algorithm for the molecules will able us to apply this tool on a real microscope observation, instead of a given identity to every position.

Finally, a significant aspiration for future research is the development of more precise methods to describe Brownian Dynamics. By identifying additional patterns

in molecular movement, we can increase our understanding and generate more accurate results. We believe that some of the limitations we have shown in the last chapter could be mitigated with further optimization of the algorithm, and perhaps even with exploring alternative techniques to further enhance the resolution and fidelity of the movies.

In summary, our project represents a promising step towards overcoming the challenges associated with obtaining high-resolution movies of molecular movement. With future endeavors focusing on advanced learning methods, three-dimensional representation, and the refinement of Brownian Dynamics modeling, we anticipate significant improvements in our ability to simulate and understand molecular dynamics. Ultimately, we envision our work playing a pivotal role in unraveling the intricate mechanisms of immune responses and cellular signaling, leading to further discoveries in the realm of immunology.

References:

1. Filipp, D., Ballek, O., & Manning, J. (2012). Lck, Membrane Microdomains, and TCR Triggering Machinery: Defining the New Rules of Engagement. *Frontiers in Immunology*, *3*. https://doi.org/10.3389/fimmu.2012.00155
2. Träuble, H. (1971). The movement of molecules across lipid membranes: A molecular theory. *The Journal of Membrane Biology*, *4*(1), 193–208. https://doi.org/10.1007/bf02431971
3. Huber, G. L., & McCammon, J. A. (2019). Brownian Dynamics Simulations of Biological Molecules. *Trends in Chemistry*, *1*(8), 727–738. https://doi.org/10.1016/j.trechm.2019.07.008
4. Chen, J. X., & Kim, A. H. (2004). Brownian Dynamics, Molecular Dynamics, and Monte Carlo modeling of colloidal systems. *Advances in Colloid and Interface Science*, *112*(1–3), 159–173. https://doi.org/10.1016/j.cis.2004.10.001

5. Bullerjahn, J. T., Von Bülow, S., & Hummer, G. (2020). Optimal estimates of self-diffusion coefficients from molecular dynamics simulations. *Journal of Chemical Physics*, *153*(2). https://doi.org/10.1063/5.0008312
6. Allen, B. L., & Mayo, S. L. (2009). An efficient algorithm for multistate protein design based on FASTER. *Journal of Computational Chemistry*, NA. https://doi.org/10.1002/jcc.21375
7. Gal, N., Lechtman-Goldstein, D. & Weihs, D. Particle tracking in living cells: a review of the mean square displacement method and beyond. Rheol Acta 52, 425–443 (2013). https://doi.org/10.1007/s00397-013-0694-6