A SIMPLE MODEL TO STUDY A COMPLEX CYTOPLASM

{Picture of main idea}

{Logos}

Work by: Mubarrat Mahin Mursalin

Under the supervision of: Stefan Klumpp

Title

-A simple model to study the complex cytoplasm

Preface

-this work was done for the partial completion of

Introduction

-The main goal of this paper was to better understand the nature of the inside of a cell.
-We will define the scope of this work in the study of science. Specifically we show that coarse grained models, built from first principles, can still describe the motions of the cytoplasm constituents.
-More exactly we try to describe this system in a way that is useful to biologist, chemists and physicists, and gives a physically correct description to the nature of things inside of cells.
-This is becoming more relevant as advances in simulation, systems biology, and cellular biophysical and biochemical methods become more advanced.
-Specifically, we can look at what we know about the insides of cells from 3 perspectives, those of experiments, simulations, and theory, and the expectations of scientists in the fields of biology physics and chemistry.

-The works of Charles Darwin started to define life and the study of living things under the scope of evolution by natural selection.
-Later on DNA was found to be the the code for living systems (that we know of), and the central dogma of molecular biology has since been the paradigm for thinking in biology on the small scale. The central dogma states that the genome of an organism is defined by its DNAs, then in cell this DNA is read and used as RNAs. RNAs then make proteins.
-Proteins effect what we see phenotypically, that is to say, protein content controlled by gene expression,leads to the different types of cells we see.
-Biology is concerned with the dynamics of living things, which based on their genetic and the external environment.

-Physics is concerned with the motions of objects in time.
-The most famous works of Isaac Newton were that of how an apple falls, and his laws governing the motion of objects.
-The works of Albert Einstein that are commonly known are that of the Energy Mass equivalency, E=MC^2. However, his lesser known work regarding the photoelectric effect and Brownian motion are what won him the Nobel prize.
-There are many directions in the study of physics, and often they are not related to each other because of the difference in scales.
-Biophysics is the study of the motions of living objects in time. It is founded on statistical physics.

Theory

Ab-Initio Life

-Starting from the atomic level we can build to the cellular level
-Cells are the basic unit of life and we want to study their complex behavior
-We pick a simple cell analogous to a bacteria, and we model it’s cytoplasm which constitutes a large part of the cell

The simple model

-We use a Lennard-Jones sphere with a screened charge to represent any component of the cytoplasm
-

Why a glass

Phase separation

It’s about the complexity

Methods

Making Soup; Building a cytoplasm

- We started from the work of Feig et al as a complete annotated database of the contents of a piece of cytoplasm. This database contained many constituents with concentrations. It was supplemented with data from PaxDB and UniProt DB which are proteomics and genomics databases respectively. This was done to account for proteins in the Feig et al data which were in PDB format. Our final database has 452 constituents with some species having a minimum concentration of 1um. From this database we can generalize our average cytoplasm to be in a box of (100nm)^3 with 94.81% filled with water molecules.

- From the model cytoplasm we also created a test system. Our aim was to create an analogous system to the full cytoplasm, which means keeping physical qualities such as mass and radii abundance, while simplifying the system. The main quantities of the cytoplasm constituents were charge and radius so the test system was modeled using 9 molecule types corresponding to 3 radii small medium and large and 3 types of charges positive negative and neutral. The abundances of the constituents were made to represent volume fractions in the real cytoplasm.

The Simulation Engine

-Data from our database was given as an input to a simulation engine built on the HOOMD-Blue library on Python. HOOMD-Blue was selected for its ease of use, good documentation, and for its ability to run molecular dynamics simulations of GPUs.

-The simulation box was constructed according to occupied volume fraction. The total volume of the box was set so that the molecules from database would occupy a specified volume fraction. To create the system, a scaled up box was filled up with molecules in the database. These molecules were placed randomly with the PACMOL software implemented in python via the Mbuild library.

-Multiple steps were taken to ensure that while each instance of the simulation was different with the same constituents, each instance was also thoroughly equilibrated before any quantities of interest were recorded. Specifically, the enlarged box filled with molecules was shrunk to the correct size. After compression, the system was allowed to equilibrate with potentials that were both cut off at a small radius, and modified to raise the potential energy minimum. This was rationalized as erroneous results would only result if there was significant overlap of the spheres representing the molecules, and this would not happen from the cutoff nor the boosted potential. The pressure and potential energy was monitored during the equilibration to verify that the system had stabilized.

-Molecular dynamics simulations of this equilibrated system was run as the main computer experiment in this study. The specific data that was given to the simulation was a simulation box filled with molecules, an n-n pairwise interaction matrix with n as the number of species which scales the LJ epsilon parameter, physical parameters of temperature viscosity and ionic strength, and simulation constraints such as the timestep the total time and how often to write data to the trajectory. Very specifically, at each timestep the simulation box was initialized. For each particle the force was calculated as generated from pairwise LJ and Yukawa interactions with every other particle in the cutoff radius. During said timestep, the force and a random noise term drove the particles, while a viscous dampening slowed them as according to brownian dynamics. The usual simulation was done with in a box of size (100nm)^3, with a timestep of 0.1 femtoseconds, at 30C and 150mM ionic strength and a viscosity

of 1 (water). HOOMD-Blue wrote the trajectory at a given rate as a binary gsd file. To minimize file sizes, the simulation was broken into 2 time regimes, slow and fast. 2 trajectories were created from 1 simulation, with 2 write rates, with the slow writer ending after a certain number of steps. The data was stitched together from both trajectories during analysis.

-To monitor the accuracy of the simulation, the kinetic, potential and total energies along with the kinetic temperature and pressure were monitored. This was done by extracting these quantities along with others at each trajectory step. Stability was determined when the total energy, potential energy and pressure stabilized during equilibration.

-To monitor the mean squared displacement of the particles from the trajectory, the Freud library was used. The library is able to calculate the MSD for each particle in the trajectory over a sliding time window, which decreases computational cost. For a certain species the MSD was averaged over all particles of said species. Other quantities derived from the MSD v time such as log MSD v log time was also calculated from the averaged MSDs.

Analysis

Results

Validating the model

Varieties of the cytoplasm

Exploring the Interactome

Conclusions

Crowding does not lead to anomalous diffusion