Cytoplasm Simulation

David van der Spoel[1]

E-mail:

Abstract

Introduction

Biomolecules move and function in an environment densely packed with high concentrations of macromolecules. The presence of macromolecules leads to steric effect due to excluded volume effect and intermolecular attrative/repulsive forces due to distributed charges on the surface of macromolecules.

Structure and dynamics of biomolecules are well characterized *in vitro*, however how these features differ *in vivo* remains unclear. Investigating biomolecular properties *in vivo* is possible through development in the fields of nuclear magnetic resonance,[? ?] or fluoroscence spectroscopies.[? ? ?] An alternative is to use computational models and simulations techniques. Biomolecular simulations are often carried out under dilute conditions or simple models of macromolecular crowdings.[? ? ?] However, more attemps in modeling bacterial cytoplasm have been made recently.[? ?]

Here we report on a model of *Escherichia coli* cytoplasm at atomistic level. The challenges in modeling and simulation with Molecular Dynamics are pointed and discussed and solutions are provided.

This work is the first model of *E. coli* at atomistic resolution that spans cellular dynamics on a microsecond scale.

Material and Method

Cytoplasm model

In order to investigate the effects of crowding on cellular components in a more realistic way, we built a cytoplasm model based on *Escheria coli* as a model organism[? ? ? ?]

Because of restriction of computational resources, incorporating a full list of proteins, nucleic acids and metabolites in full atomistic model is prohibitive.

In order to build the RNA fraction of our cytoplasm model, we consider data showing that 2.9% of the dry weight of *E. coli* is composed by tRNAs, which correspond to 74% of the dry weight of non-ribosomal RNAs, while 55.0% of it is composed of proteins.[?] Thus, the tRNA weight should correspond to 5% of the total weight of all protein chains used in our model. Specifically, tRNA(Phe) was selected as a representative due to the availability of a recent crystallographic structure.[?] Copies were added to the simulation box to account for the correct protein/RNA weight ratio (Table X).

The metabolite fraction of our cytoplasm model was build based on data showing that the number of metabolite molecules in the cytoplasm of *E. coli* is about 42.86 times higher than the number of proteins. We considered the most abundant molecule as representative of each metabolite class, i.e. Glutamate for amino acids, ATP for nucleotides, FBP for central carbon intermediates and Glutathione redox cofactors. The copy number for each molecule was calculated as the ratio of their experimentally observed concentration in *E. coli*.

Components Preparation

The proteins and tRNA were downloaded from Protein Data Bank (PDB). We looked for the proteins structures that were either from or expressed in *E-coli*. In case of 1U22 (MetE)

and 2EIP (Ppa) we used a loop-closure modelng tool based on Random Coordinate Descent (RCD) method[?] to correct the information for missing residues. The four metabolites were parametrized using GAFF and Antechamber.

All-Atom Molecular Dynamics Simulation

General Simulation Setup: The proteins were simulated at 30% biomolecular mass fraction in a physiological salt concentration (0.15M NaCl). For all simulations, Amber99SB-ws force field was used[?] in combination with the TIP4P/2005 water model.[?] Electrostatic interactions were treated using the particle mesh Ewald algorithm.[?] All chemical bonds were constrained at their equilibrium length using the LINCS algorithm[?] allowing an integration time step of 2 fs. Temperature was controlled at 310 K using the v-rescale algorithm[?] and a coupling time of 0.5 ps. The pressure was controlled at 1 bar using the Parrinello-Rahman algorithm[?] with a time constant of 10 ps.

Table 1: Cytoplasm Components

Class	Name (PDB ID)	Number
8*Protein	TufA (1DG1[?])	6
	MetE (1U22[?])	7
	IcdA (1P8F[?])	2
	AhpC (1YEP[?])	1
	CspC (1MJC[?])	3
	Ppa (2EIP[?])	1
	GapA (1S7C[?])	1
	Eno (1E9I[?])	1
RNA	$tRNA^{Phe}$ (4YCO[?])	5
4*Metabolite	GLU	1436
	ATP	144
	FBP	225
	GSH	255
3*Inorganic Ion	K^+	4602
	Mg^{2+}	400
	Cl^-	1320
Solvent	Water	306221

Initially, each component was prepared with the same simulation setup as cytoplasm in

a box of water. After minimization and equilibration steps, each component together with a layer of water molecule with specific size were extracted. The size of the water shell was probed to make up the total number of water molecules the model needed to make a total biomolecular concentration of 30%. Possible ions in the water shell were not discarded and at the end counted for the total ionic strength of the cytoplasm model. The components were placed in a preliminary box of size $L = 30nm$ according to their abundancy. The resulting box were equilibrated with temperature and pressure coupling for total time of $200ns$, this step is called shrinking in this text. The size of the box shrank to $L = 22.90nm$ after $150ns$.

For error analysis, each simulations were repeated three times with independent starting velocities.

All simulations were performed with Gromacs 2018. Single simulations were started from crystal conformations. The cytoplasm simulations starting conformation were taken from the equilibrated conformation of each single simulation.

Single Simulation Setup: Each component was simulated with the same parameter as the cytoplasm for $200ns$.

Analysis

Before any analysis the periodic boundary condition (pbc) artifacts have been removed. We used GROMACS tools to do the analysis. For single component simulations, first the components were made whole and jump removed and then all the atom were put inside the compact box. The same treatment were applied to the cytoplasm simulations. Additionally, each component's trajectory were extracted and fitted by rotation and translation for later rotational correlation time analysis.

A Mean Square Displacement (MSD) analysis was used to calculate the translational diffusion coefficient.[?] The diffusion coefficients were extracted by a linear fit to MSD analysis by averaging blocks with a length of $10\,ns$. In principle diffusion coefficient needs to be corrected for finite size effects[?] but due to relatively large simulation boxes this correction

is negligible.

Protein protein interaction: Who is interacting with who?! (Residue)

Results

- Translational Diffusion
  - Translation diffusion as a function of size for both soup components and ingredients in the same plot
  - D/D0 for all components
- RMSD analysis
  - Comparison of RMSD in the soup and ingredients
  - RMSD for metabolites around macromolecules
- Protein-Protein interaction
- Metabolite-Protein interaction

Discussion