Soup Simulation

David van der Spoel[1]

November 13, 2017

Abstract

Introduction

The observation of dynamics and kinetics details of whole cell all at once is an extremely challanging job. Partly due to lack of comprehensive model behind even a simplest single cell organism, such as *Escheria coli*. And partly due to the limitations of *in vivo* experiments. Because of these barriers, the focus is usually narrowed down to a sub-system of the whole organism, often *in vitro* where the system of interest is isolated and studied out of its physiologal context, but also more recently the system of interest are tagged with ... techniques and studied *in vivo*, thanks to developments in single cell studies [**?**] This experimental challenges makes the whole cell study a suitable subject for theoretical/computational investigation.

Simulations can provide the ultimate detail concerning individual particle motions as a function of time. If the underlying model is accurate and the simulation time is long enough for reaching equilibrium.

- write about whole cell simulation,
- ecoli as a model system
- molecular dynamics simulation of crowding
- questions to answer

  The key question is whether and how the in vivo behavior of biological macromolecules differs from their well-characterized in vitro properties.

The model shows that some proteins have a much less rigid structure in cells than they do in isolation, whilst others are able to work together more closely to carry out certain tasks (study protein structure)

The model predicted that small molecules such as food, water and drugs would move more slowly through cells as they become stuck or trapped by larger molecules.

- our work

  Driven by data from high-throughput experiments, we built a comprehensive cytoplasmic model based on Escheria coli

Method

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

Cytoplasm model: Eight different non-ribosomal proteins which are the most abundant and that account for 50% of the the number of proteins found in the cytoplasm of *Escherichia coli* K-12 were selected as the basis to build our cytoplasm model []. The oligomerization state observed in their crystallographic structures was kept, water molecules, ions and ligands were removed from the crystallographic structure. The number of chains added to the simulation box is proportional to their experimentally observed abundance (Table X).

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 by 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*.

General Simulation setup:

- Force Field
- Water model
- Gromacs
- Temperature and pressure
- (details of soup box)
- Minimization
- Equilibration
- Production
- Box building
- Droplets
- Shrinking

General Simulation Setup: The proteins were simulated at 30% biomolecular mass fraction in a physiological salt concentration (0.15M NaCl). For all simulations the Amber99sb-ildn force field was used [**?**, **?**] in combination with the TIP4P/2005 water model [**?**]. Electrostatic interactions were treated using the particle mesh Ewald algorithm [**?**] and XXXX. 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.

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

Table 1: Proteins PDB ID

Class	Name (PDB ID)	Number
8*Protein	TufA (1DG1 [**?**])	12
	MetE (1U22 [**?**])	7
	IcdA (1P8F [**?**])	4
	AhpC (1YEP [**?**])	10
	CspC (1MJCc [**?**])	3
	Ppa (2EIP [**?**])	6
	GapA (1S7C [**?**])	4
	Eno (1E9I [**?**])	2
RNA	$tRNA^{Phe}$ (4YCO [**?**])	5
4*Metabolite	Glu	1436
	ATP	144
	FBP	225
	GSH	255
3*Inorganic ion	K^+	x
	Mg_8^{2+}	y
	Cl^-	z
Solvent	Water	w

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, this step is called shrinking in this text.

Single Simulation Setup:

Analysis: Before any analysis the pbc artifacts have been removed. We used GROMACS tools to do the 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.

We followed the Lipari-Szabo approach to calculate the rotational correlation time, using a second order Legendre polynomial $P_2(\mathbf{r}_{NH})$ to be able to compare the results to experimental relaxation data [**?**]. First the average order parameter $<S^2>$ over all $\mathbf{r}_{NH}$ vectors for each protein was determined from trajectories after the biomolecular rotation- and translation-movements were removed. Then the rotational correlation function $C(t)$ of the original trajectories was fitted to

$$C(t) \;=\; <S^2> \exp(-t/\tau_M) + (1- <S^2>) \exp(-t/\tau_T) \qquad (1)$$

where τ_M is the rotational correlation time (tumbling) and $\tau_T = (\tau_M^{-1} + \tau_e^{-1})^{-1}$, where τ_e is the internal bond movement.

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

Results

Discussion