UNIVERSITÉ LIBRE DE BRUXELLES

TRAN-F501

INTERNSHIP - 201819

Project: A stochastic simulation system for protein aggregation

Supervisor:
Tom LENAERTS
Author:
Prateeba RUGGOO

September 23, 2019

Contents

1 Introduction

Diseases like Alzheimer and Parkinson disease are the result of proteins aggregating into large fractal structures that hinder the cell function or even destroy them. Understanding how aggregates are formed and change over time is important to understand when they become harmful and how maybe treatments affect aggregate formation.

The goal of this Internship is to implement a simulation system to study aggregation between proteins. This work is performed in collaboration with the Switch lab in the KU Leuven, who has an extensive expertise in studying aggregation and related diseases.

2 First 2 weeks (12th-23rd August)

During the first week of the Internship, I spent the first few days getting used to the work environment, did my computer and office set up, got to know my team mates and the routines (lunch time, group lunches etc) and some paperwork so that I was given access to IB2 on a daily basis. Once I settled down, I researched the subject and mostly did a state of the art to have a baseline knowledge.

3 Third week (26th-31st August)

Having acquired a basic understanding, during the third week, I refined my reasearch and focused on the specific papers discussing the problem in order to have a deeper understanding of the theory before starting any kind of implementation. The papers I focussed on are the following [Gil76], [MKA+16], [GB00], [JSMI12].

4 Fourth week (2nd-6th September)

The goal of this week was to implement a simple prototype (in python for now) of an exact numerical simulation method to simulate trajectories of discrete, stochastic systems.

4.1 Gillespie's stochastic framework of chemical kinetics

The principle task is to develop a method for simulating the time evolution of the N quantities $\{X_i\}$, knowing only their initial values $\{X_i^{(0)}\}$, the form of the M reactions $\{R_\mu\}$ and the values of the reaction parameters $\{c_\mu\}$.

Definition 1. *Problem definition: We are given a volume V containing molecules of N chemically active species $S_i (i = 1, \ldots, N)$. Let $X_i \equiv$ current number of molecules of chemical species $S_i \in V, (i = 1, 2, \ldots, N)$ and let $R_\mu (\mu = 1, \ldots, M)$ be the chemical reactions in which the species S_i can participate where each reaction R_μ is characterized by a numerical*

reaction parameter c_μ and let R be the different type of reactions :

$$* \rightarrow reactionproducts \tag{1}$$
$$S_j \rightarrow reactionproducts \tag{2}$$
$$S_j + S_k \rightarrow reactionproducts \tag{3}$$
$$2S_j \rightarrow reactionproducts \tag{4}$$
$$S_i + S_j + S_k \rightarrow reactionproducts \tag{5}$$
$$S_j + 2S_k \rightarrow reactionproducts \tag{6}$$
$$3S_j \rightarrow reactionproducts \tag{7}$$

, the goal is to simulate the trajectories of the N chemically active species S_i. and predict which reaction will occur at each time step according to the correct probability distribution.

Hypothesis 2. *The reaction parameter c_μ can be defined as follows : $c_\mu \delta t \equiv$ average probability that a particular combination of R_μ reactant molecules will react accordingly in the next time interval δt (first order).*

Definition 3. *State of the system is defined by the number of molecules of each species and changes discretely whenever one of the reactions is executed. The probability that a certain reaction μ will take place in the next instant of time is given by $a_\mu dt + o(dt)$.*

4.2 Gillespie's Direct Method

Given the problem defined above, the Gillespie's Direct Method answers two questions :

1. Which reaction occurs next ?
2. When does it occur ?

4.2.1 Gillespie's Direct Method formulas

1. Probability density $P(\mu, \tau)$ that the next reaction is μ and it occurs at time $\tau \rightarrow P(\mu, \tau)d\tau = a_\mu exp(-\tau \sum_j a_j)d\tau$.
2. Probability for reactions $\rightarrow Pr(Reaction = \mu) = a_\mu / \sum_j a_j$.
3. Probability distribution for times $\rightarrow P(\tau)d\tau = (\sum_j a_j)exp(-\tau \sum_j a_j)d\tau$.

4.3 Gillespie's Direct Method : Pseudo code

1

4.4 Gillespie's Direct Method : Example

Example 4. *Generating a sample trajectory of a chemical process in the stochastic framework.*

Algorithm 1 Gillespie's Direct Method

Input: N chemically active species $S_i, \{X_i^{(0)}\}$ initial values of each specie S_i, the set R of chemical reactions and the reaction parameter c_μ for each reaction.

Output: Sample trajectory of a chemical process in the stochastic framework.

while !(simulation time exceeded) **do**

1. Initialization: Set initial number of molecules in the system, set $t \leftarrow 0$.
2. Calculate the propensity function, $a_i \forall i$.
3. Choose μ according to the distribution in eq 5.
4. Choose τ according to an exponential with parameter $\sum_j a_j$ (as in eq 6).
5. Update the number of molecules to reflect execution of reaction μ. Set $t \leftarrow t + \tau$.
6. Go to step 2.

end while

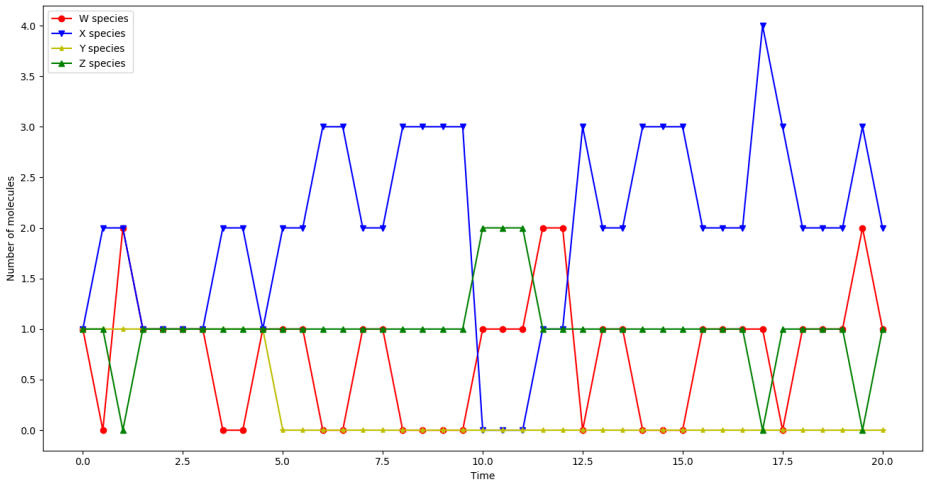

Figure 1: The x-axis denotes the duration time of the simulation and the y-axis denotes the number of molecules of each species S_i

4.5 Gillespie's Direct Method : Implementation

The source code of the simple simulation model can be found at To do -> put github link. The reaction set is the example used in Gillespie's paper [Gil76].

$$X \rightarrow Y \tag{8}$$

$$Y \rightarrow X \tag{9}$$

$$2X \rightarrow Z \tag{10}$$

$$Z \rightarrow 2x \tag{11}$$

$$W + X \rightarrow 2X \tag{12}$$

$$2X \rightarrow W + X \tag{13}$$

and the reaction parameter c_μ is chosen randomly.

5 Fourth week (9th-20th September)

5.1 Gillespie's Next Reaction Method

The goal of those two weeks was to implement the Gillespie Next reaction method in order to be able to consider bigger systems. The main idea behind the Next Reaction method to generate a putative time τ_i for each reaction i to occur and choose the reaction μ whose putative time τ_μ is least. In order to minimize the computation time, the Next reaction method is implemented using specific data structures.

Definition 5. *Let Reactants(ρ) and Products(ρ) be the sets of reactants and products of a reaction ρ.*

Definition 6. *Let a_μ be the value computed by the propensity function and DependsOn(a_μ) be the set of substances that affects the value a_μ*

Definition 7. *Let Affects(μ) be the set of substances that are affected if reaction μ occurs.*

Definition 8. *Given a set of reactions R, the graph $G = G(V, E)$ is a directed graph where V is the set of Reactions and there exists an edge from v_i to v_j if and only if Affects(v_i) $\cap$ DependsOn(a_{v_j}) $\neq \emptyset$. G is then referred to as the dependency graph of the set of reactions R and is useful to know which $a_i s$ to change when a given reaction occurs.*

Definition 9. *Indexed prioroty Queues. To do -> definition*

Algorithm 2 Gillespie's Next Reaction Method

Input: N chemically active species S_i, $\{X_i^{(0)}\}$ initial values of each specie S_i, the set R of chemical reactions and the reaction parameter c_μ for each reaction.
Output: Sample trajectory of a chemical process in the stochastic framework.
while !(simulation time exceeded) **do**
1. Initialization :
   1.1 Set initial number of molecules, set $t \leftarrow 0$, generate a dependency graph G.
   1.2. Calculate the propensity function, $a_i \forall i$.
   1.3. For each i, generate a putative time, τ_i, according to an exponential distribution with parameter a_i.
   1.4. Store the τ_i values in an indexed priority queue P.
2. Let μ be the reaction whose putative time, τ_μ, is least.
3. Let τ be τ_μ.
4. Update the number of molecules to reflect execution of reaction μ. Set $t \leftarrow \tau$.
5. For each edge(μ, α) in the dependency graph G,
   5.1 update a_α.
   5.2 if $\alpha \neq \mu$, set $\tau_{alpha} \leftarrow (a_{\alpha,old}/a_{\alpha,new})(\tau_\alpha - t) + t$.
   5.3 if $\alpha = \mu$, generate a random number, ρ, according to an exponential distribution with parameter a_μ and set $\tau_\alpha \leftarrow \tau + t$.
   5.4 Update the old τ_α value in P.
6. Go to step 2.

end while

Example 10. *Generating a sample trajectory of a chemical process in the stochastic framework.*

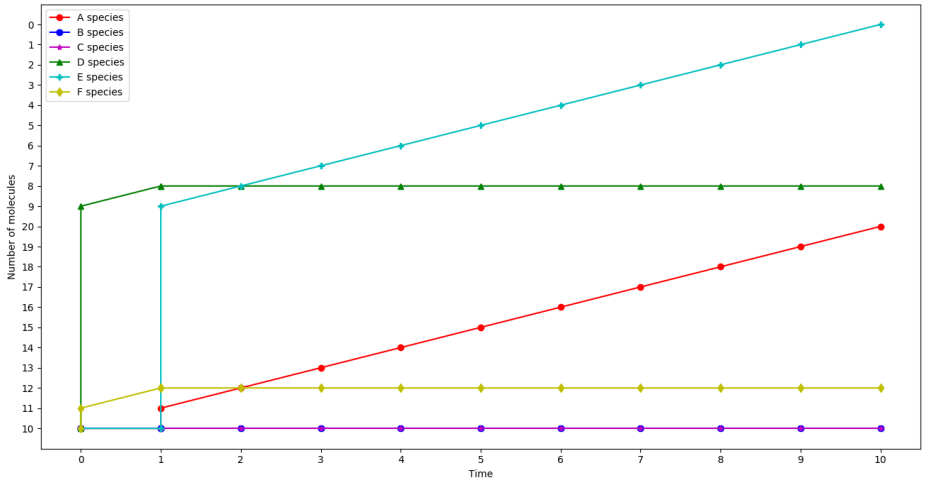

Figure 2: The x-axis denotes the time used and the y-axis denotes the number of molecules

Example 11. *Generating a single sample trajectory of a chemical process in the stochastic framework.*

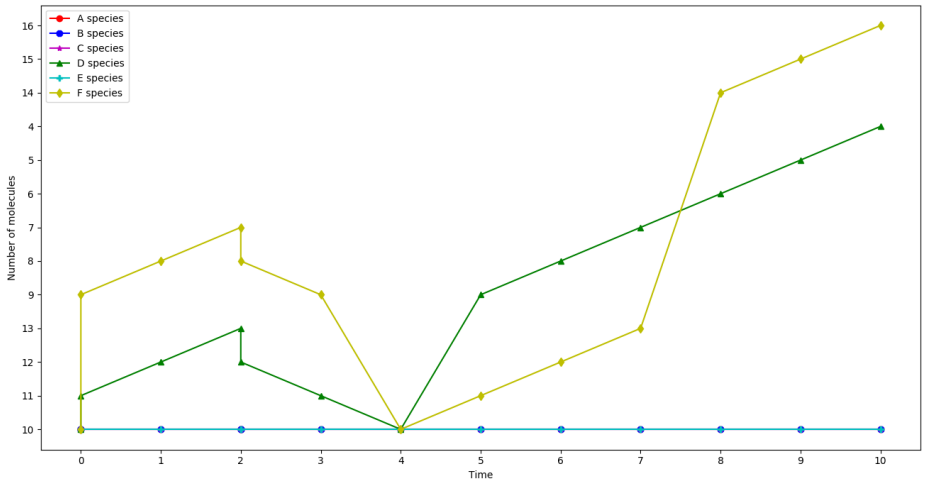

Figure 3: The x-axis denotes the time used and the y-axis denotes the number of molecules

6 Conclusion

References

[GB00] Michael A. Gibson and Jehoshua Bruck. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889, March 2000.

[Gil76] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, December 1976.

[JSMI12] Sigurður Ægir Jónsson, Iskra Staneva, Sandipan Mohanty, and Anders Irbäck. Monte Carlo Studies of Protein Aggregation. *Physics Procedia*, 34:49–54, 2012.

[MKA+16] Georg Meisl, Julius B Kirkegaard, Paolo Arosio, Thomas C T Michaels, Michele Vendruscolo, Christopher M Dobson, Sara Linse, and Tuomas P J Knowles. Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nature Protocols*, 11(2):252–272, February 2016.