UNIVERSITÉ LIBRE DE BRUXELLES

TRAN-F501

INTERNSHIP - 201819

Project: A stochastic simulation system for protein aggregation

Supervisor:
Tom LENAERTS
Author:
Prateeba RUGGOO

September 3, 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 work on the implementation of a simulation system required to study aggregation between proteins. This work is performed in collaboration with the Switch lab in the KU Leuven, who have an extensive expertise in studying aggregation and related diseases.

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

(Maybe add the administrative part, introduction to the team members and all...) During the first two weeks of the intership, I researched the subject and mostly did a of state of the around the subject to have a baseline knowledge around the subject. !!!!!!!!!!!!!! TO ADD MORE DETAILS !!!!!!!!!!!!!!!!!!

3 Third week (26th-31st August)

Now having a basic understanding around the subject, I could refine my reasearch and focus on the specific papers discussing the problem in order to have a deeper understanding of the theory before starting any kind of implementation. !!!!!!!!!!!!!! TO ADD MORE DETAILS !!!!!!!!!!!!!!!!!!

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 Step 1 : Mathematical Descriptions of Chemical Processes

A coupled system of chemical reactions of the form : $X_1 + X_2 \rightarrow X_3 + ...$ states that one molecule from substance X_1 reacts with one molecule of substance X_2 to give one molecule of substance X_3

4.1.1 Hypothesis :

The solution is well mixed $\rightarrow$ nonreactive collisions occur far more than reactive collisions $\rightarrow$ fast dynamics of motion can be neglected $\rightarrow$ Can use the number of each kind of molecule to represent the system.

Theorem 1. *The probability that a certain reaction μ will take place in the next instant of time dt is given by $a_\mu dt + o(dt)$, where a_μ is independent of dt.*

4.2 The stochastic framework

1. We have a set of reactions

$$A + B \rightarrow^{k_1} C \tag{1}$$
$$B + C \rightarrow^{k_2} D \tag{2}$$
$$D + E \rightarrow^{k_3} E + F \tag{3}$$
$$F \rightarrow^{k_4} D + G \tag{4}$$
$$E + G \rightarrow^{k_5} A \tag{5}$$

2. The propensities of the reactions are given by $k_1, k_2, ...k_5$
3. The probability that a given molecule A reacts with a given molecule B in a small time dt is $k_1 dt + o(dt)$.

4.2.1 Gillespie's stochastic framework of chemical kinetics

The principle task is to develop a method for simulation 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 2. *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.*

Definition 3. *Type of reactions*

$$* \rightarrow reactionproducts \tag{6}$$
$$S_j \rightarrow reactionproducts \tag{7}$$
$$S_j + S_k \rightarrow reactionproducts \tag{8}$$
$$2S_j \rightarrow reactionproducts \tag{9}$$
$$S_i + S_j + S_k \rightarrow reactionproducts \tag{10}$$
$$S_j + 2S_k \rightarrow reactionproducts \tag{11}$$
$$3S_j \rightarrow reactionproducts \tag{12}$$

Hypothesis 4. *Fundamental Hypothesis The reaction parameter c_μ can be defined as follows :*

Definition 5. *$c_\mu \delta t \equiv$ avarage probability that a particular combination of R_μ reactant molecules will react accordingly in the next time interval δt (first order)*

!!!!!!!!!!!!!!!!!!!! TO DO : FINISH THE DETAILS AND EXPLANATION OF EACH FORMULA !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

Definition 6. *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 nest instant of time is given by $a_\mu dt + o(dt)$*

Example 7. *Let S be the set of states i.e $S = (\#A, \#B, \#C, \#D, \#E, \#F, \#G)$, S will change to $S^{'} = (\#A - 1, \#B - 1, \#C + 1, \#D, \#E, \#F, \#G)$ if Reaction 1 is executed. The probability of this occurence is given by : $P(S^{'}, t + dt|S, t) = a_1 dt + o(dt)$*

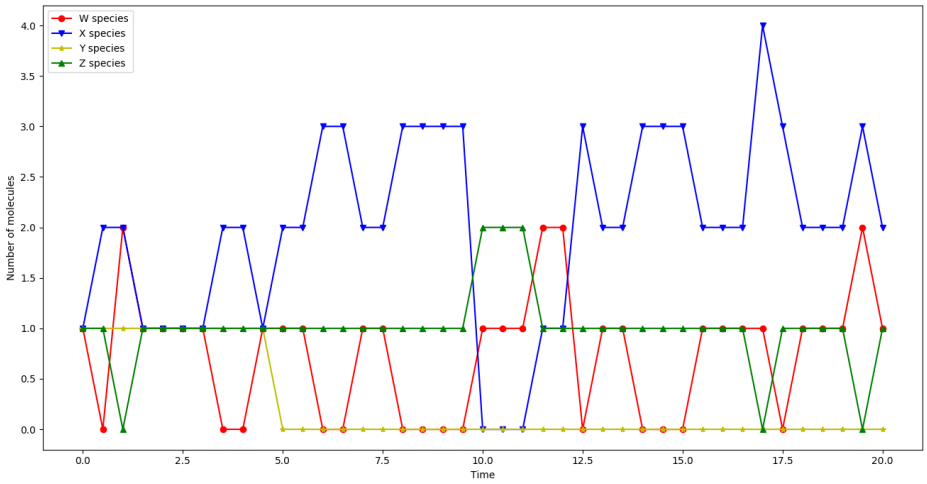

Figure 1: The x-axis denotes the dataset used and the y-axis denotes the computation time

5 Gillespie's algorithm

Algorithm 1 Gillespie's algorithm

Input:

Output:

while !(number of reactants = 0 || simulation time exceeded) **do**

1. Initialization: Initialize the number of molecules in the system, reaction constants, and random number generators.
2. Monte Carlo step: Generate random numbers to determine the next reaction to occur as well as the time interval. The probability of a given reaction to be chosen is proportional to the number of substrate molecules, the time interval is exponentially distributed with mean $1/R_{TOT}$
3. Update: Increase the time by the randomly generated time in Step 2. Update the molecule count based on the reaction that occurred.

end while

6 Initial values

1. $n_A = n_B = 10$

2. $n_{AB} = 0$
3. $t = 0$
4. $k_D = 2$
5. $k_B = 1$

7 Gillespie's algorithm : Simple example

7.1 Reaction Rates

1. If we consider two types of mocules : A and B, two types of reaction can happen.
   (a) A and B bind together to form AB dimer.
   (b) AB dimers dissociates into an A and a B molecule.
2. let k_D be the reaction rate for a AB dimer formation.
3. let k_B be the reaction rate for a AB dimer deformation.
4. So at a time t, the total reaction rate $= k_D \cdot$ number of type A molecules $\cdot$number of type B molecules$+k_B \cdot$ number of AB dimers $\Rightarrow R_{TOT} = k_D n_A n_B + k_B n_{AB}$.

7.2 Time evolution

Two steps to perform to advance forward in time.

1. Calculate the time to the next reaction.
   (a) Determining reaction concern predicting how much time we need to wait before a reaction happens. Since this time is unknown, it is often appropriate to think of it as a random variable having an exponential distribution.
   (b) Here the next reaction time is a random number drawn from exponential distribution function with mean $1/R_{TOT}$.
   (c) Thus we move from time t to δt
2. Determine the next reaction among the possible reactions.
   (a) $P(A + B \to AB) = k_D n_A n_B / R_{TOT}$
   (b) $P(AB \to A + B) = 1 - (k_D n_A n_B / R_{TOT})$

   So with these two probabilities we either :
   (a) Form a dimer by reducing n_A and n_B by one and increase n_{AB} by one.
   (b) Or dissociate an AB dimer by adding one to n_A and n_B and minus one to n_{AB}

The Gillespie algorithm repeats these two steps as many times as needed to simulate the system for as many reactions as we want.

8 Conclusion

References