

UNIVERSITÉ LIBRE DE BRUXELLES



TRAN-F501

INTERNSHIP - 201819

Project: A stochastic simulation system for protein aggregation

Supervisor:

Tom LENAERTS

Author:

Prateeba RUGGOO

November 9, 2019

Contents

1	Introduction	2
1.1	Mathematical modeling	2
2	Gillespie's stochastic framework of chemical kinetics	2
2.1	Gillespie's Direct Method	3
2.1.1	Gillespie's Direct Method formulas	3
2.1.2	Gillespie's Direct Method : Pseudo code	3
2.1.3	Gillespie's Direct Method : Example	3
2.2	Gillespie's Next Reaction Method	4
2.2.1	Gillespie's Next Reaction Method : Pseudo code	5
2.2.2	Gillespie's Next Reaction Method : Example	5
3	Molecular mechanisms of protein aggregation	7
3.1	Obtaining qualitative constraints : Half times	7
3.1.1	Log(half time) versus Log(monomer concentration) plot	7
3.2	From Half-times to models	9
4	Results	9
4.1	Amylofit's parameters estimation	9
4.2	My program's parameter estimation	9
4.3	Amylofit's estimated parameters in Gillespie stochastic framework	9
4.4	My estimated parameters in Gillespie stochastic framework	9
5	Conclusion	9

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 and is performed in collaboration with the Switch lab in the KU Leuven, who has an extensive expertise in studying aggregation and related diseases.

1.1 Mathematical modeling

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 $\{c_\mu\}$. There are two fundamental approaches to the mathematical modelling of chemical reactions.

1. Deterministic models which are based on differential equations.
2. Stochastic simulations where the fundametal principle is that molecular reactions are essentially random processes i.e it is impossible to say with complete certainty the time at which the next reaction will occur. This approach uses basic Newtonian physics and thermodynamics to arrive at a form defined as the propensity function that gives the probability a_μ of reaction μ occuring in time interval $(t, t + \delta t)$.

2 Gillespie's stochastic framework of chemical kinetics

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

Definition 2. *The reaction parameter c_μ is defined so that $c_\mu \delta t$ gives the probability that a randomly chosen molecule of chemical species S reacts during the time interval $[t, t + \delta t]$ where t is time and δt is an infinitesimally small time step.*

Definition 3. *The probability that exactly one reaction μ occurs during the infinitesimal time interval $[t, t + \delta t]$ is equal to $S(t)k_\mu \delta t$ where $S(t)$ is the number of chemical species S at time t .*

Definition 4. *State of the system is defined by the number of molecules of each species and changes discretely whenever one of the reactions is executed.*

2.1 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 ?

2.1.1 Gillespie’s Direct Method formulas

1. Probability density $P(\mu, \tau)$ that the next reaction is μ and that it occurs at time τ is given by :

$$P(\mu, \tau)d\tau = a_\mu \exp(-\tau \sum_j a_j)d\tau.$$

2. Probability that the next reaction is reaction μ is given by :

$$Pr(Reaction = \mu) = a_\mu / \sum_j a_j.$$

3. Probability distribution for times

$$P(\tau)d\tau = (\sum_j a_j) \exp(-\tau \sum_j a_j)d\tau.$$

2.1.2 Gillespie’s Direct Method : Pseudo code

Algorithm 1 Gillespie’s Direct Method

Input: N chemically active species S_i , $\{X_i^{(0)}\}$ initial values of each species 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 2.
4. Choose τ according to an exponential with parameter $\sum_j a_j$ as in 3.
5. Update the number of molecules to reflect execution of reaction μ . Set $t \leftarrow t + \tau$.
6. Go to step 2.

end while

2.1.3 Gillespie’s Direct Method : Example

The source code of the Direct method simulation model can be found at [Direct-method-github](#). The reaction set is the example used in Gillespie’s paper [Gil76].



where the reaction parameters $c_i \in (i = 1, 2, \dots, 6)$ is equal to $[1, 1, 2, 1, 2, 1]$ respectively and the initial number of each species is $= 10$.

Example 5. *Generating a sample trajectories of a chemical process in the stochastic framework : Direct Method*

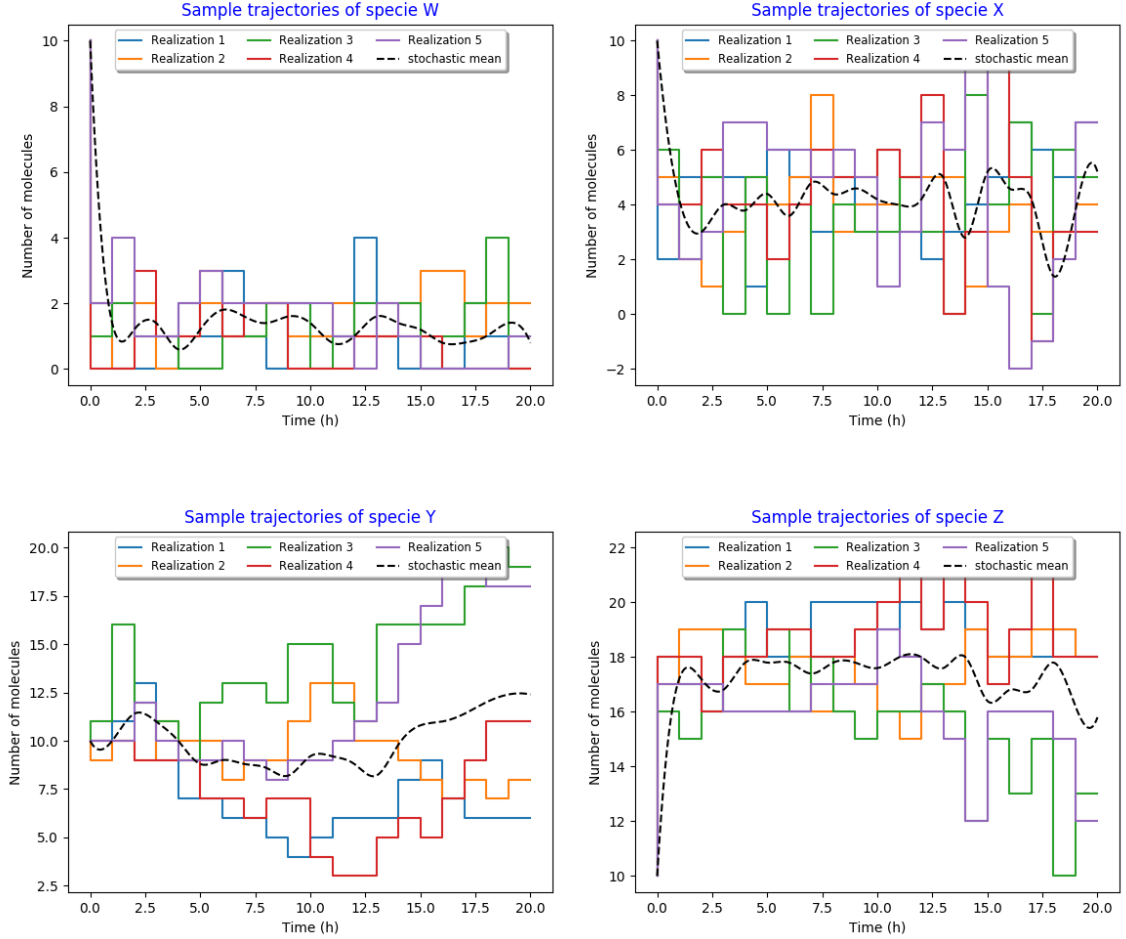


Figure 1: The x-axis denotes the duration time of the simulation and the y-axis denotes the number of molecules of each species S_i at each time step.

2.2 Gillespie's Next Reaction Method

The goal of the next reaction method is to do away with the above processes and update only data that has been modified. The main idea is to generate a putative time τ_i for each reaction i that will 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. To make the sampling of reactions more efficient, an indexed priority queue is used to store the reaction times. On the other hand, to make the recomputation of propensities more efficient, a dependency graph is used. This dependency graph tells which reaction propensities to update after a particular reaction has fired.

2.2.1 Gillespie's Next Reaction Method : Pseudo code

2

Algorithm 2 Gillespie's Next Reaction Method

Input: N chemically active species S_i , $\{X_i^{(0)}\}$ initial values of each species 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

2.2.2 Gillespie's Next Reaction Method : Example

The source code Next Reaction Method can be found at [Next-Reaction-Method-github](https://github.com/Next-Reaction-Method).

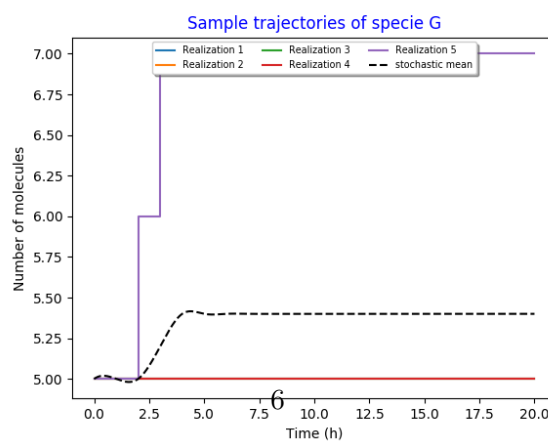
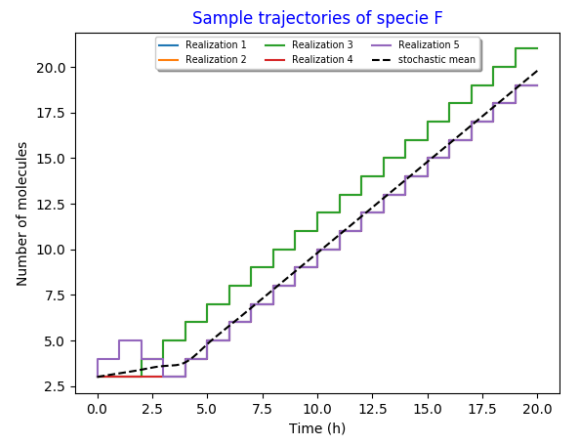
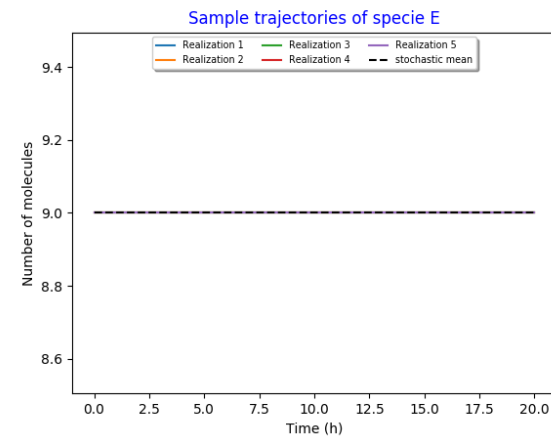
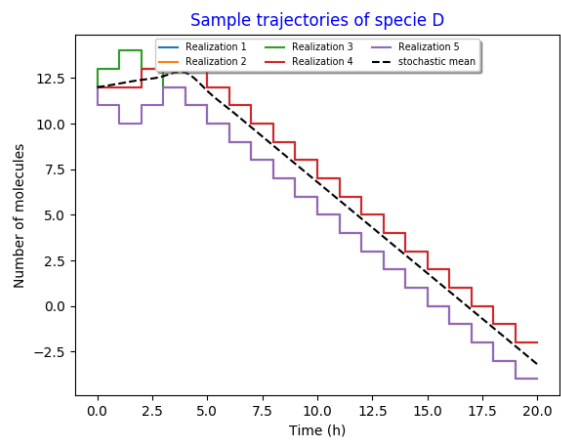
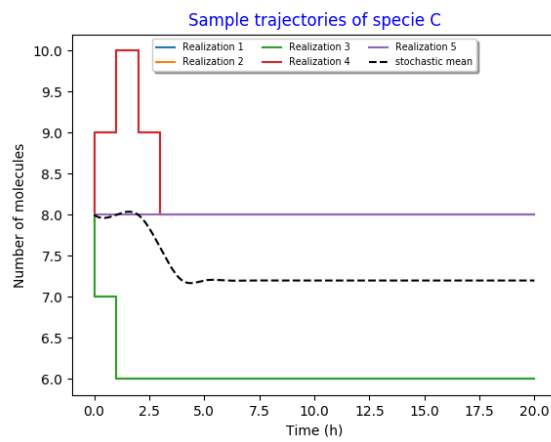
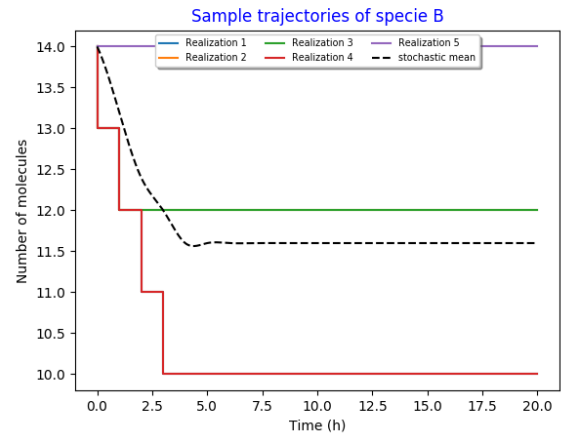
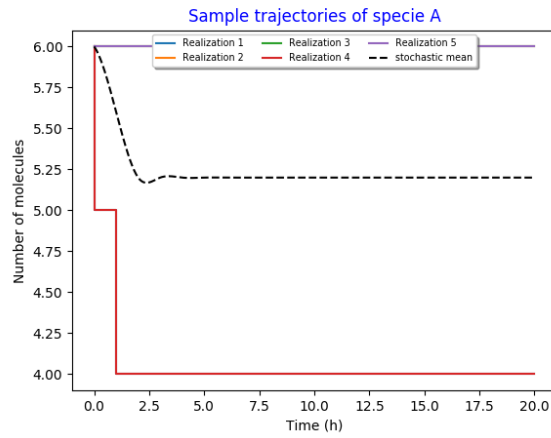
The reaction set is the example used in the following paper [GB00].



(12)

where the reaction parameters $k_\mu \in (i = 1, 2, \dots, 5)$ is equal to [1,2,2,2,1] respectively and the initial number of each species is = 10.

Example 6. *Generating a sample trajectories of a chemical process in the stochastic framework : Next Reaction Method*



3 Molecular mechanisms of protein aggregation

Until now, all what was done is using random reaction parameters for each reaction in the stochastic simulation. However in order to study the protein aggregation, understanding the kinetics of the aggregation is fundamental. In the following paper [MKA⁺16], a protocol is described on how to make full use of kinetic descriptions derived from a model of the underlying microscopic reactions that make up the aggregation network. The fitted parameters are therefore meaningful and correspond to physical properties of the system, such as nucleus sizes, rate constants of individual reactions and saturation concentrations.

The key idea is to analyze a large data set of multiple kinetic traces at different reagent concentrations simultaneously, with a single rate law and yield strong mechanistic constraints that will be helpful in choosing the appropriate models.

3.1 Obtaining qualitative constraints : Half times

Given a dataset, several fundamental models can be considered for the fitting process to yield the rate constants. In order to choose the best model, adding some constraints based on the concentration dependence of the aggregation reaction is a solution. A helpful qualitative constraint is half times.

Definition 7. *Half-time is defined as the time at which the signal has reached half of its final plateau value.*

Normalised data and the monomer concentration of each repeat is used for this process. By plotting Log(half time) versus Log(monomer concentration) plot, the slope known as the scaling exponent gives insight as to which about the dominant process of fibre multiplication, which in turn can be used to decide on possible models for the fitting. By considering how the half times scale with monomer concentration and how this scaling depends on the monomer concentration, one can obtain constraints on possible mechanisms.

Very generally, a negative curvature in the double logarithmic plots (i.e., when the slope becomes steeper at higher monomer concentrations, and therefore the process more monomer dependent) is indicative of competition between several processes in parallel. A positive curvature, in contrast (i.e., a flattening of the curve at higher monomer concentrations, and thus a decrease in monomer dependence), suggests the presence of a saturation effect in a serial process or, in rare cases, at monomer concentrations close to solubility, it can be due to a change in nucleus size⁹. The behavior of half-times with varying monomer concentration is, therefore, a good first guide to narrowing down the number of possible models, because it limits the number of acceptable reaction networks by determining the reaction order of the dominant process and probing for competition or saturation effects. The model for fitting needs to be chosen to reflect these findings.

3.1.1 Log(half time) versus Log(monomer concentration) plot

Example 8. *To set title*

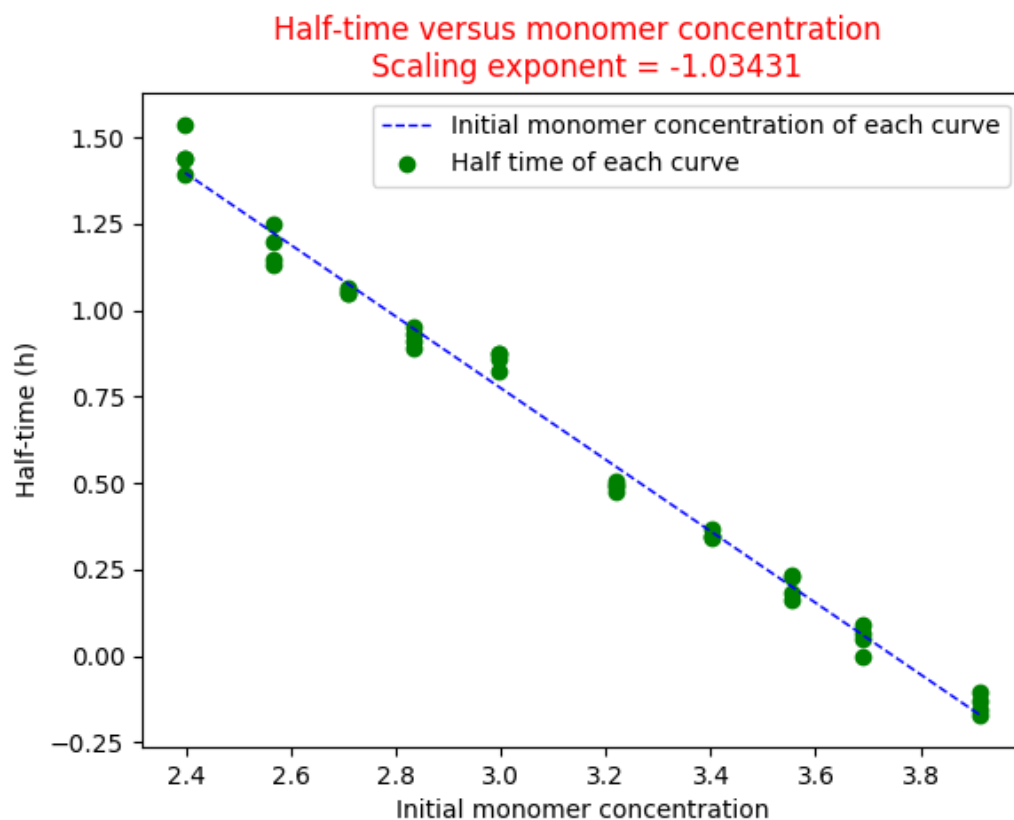


Figure 3: The x-axis denotes the time used and the y-axis denotes the number of molecules : Sample 2

The source code to generate the log off plot of a given dataset can be found at [Next-Reaction-Method-github](#).

3.2 From Half-times to models

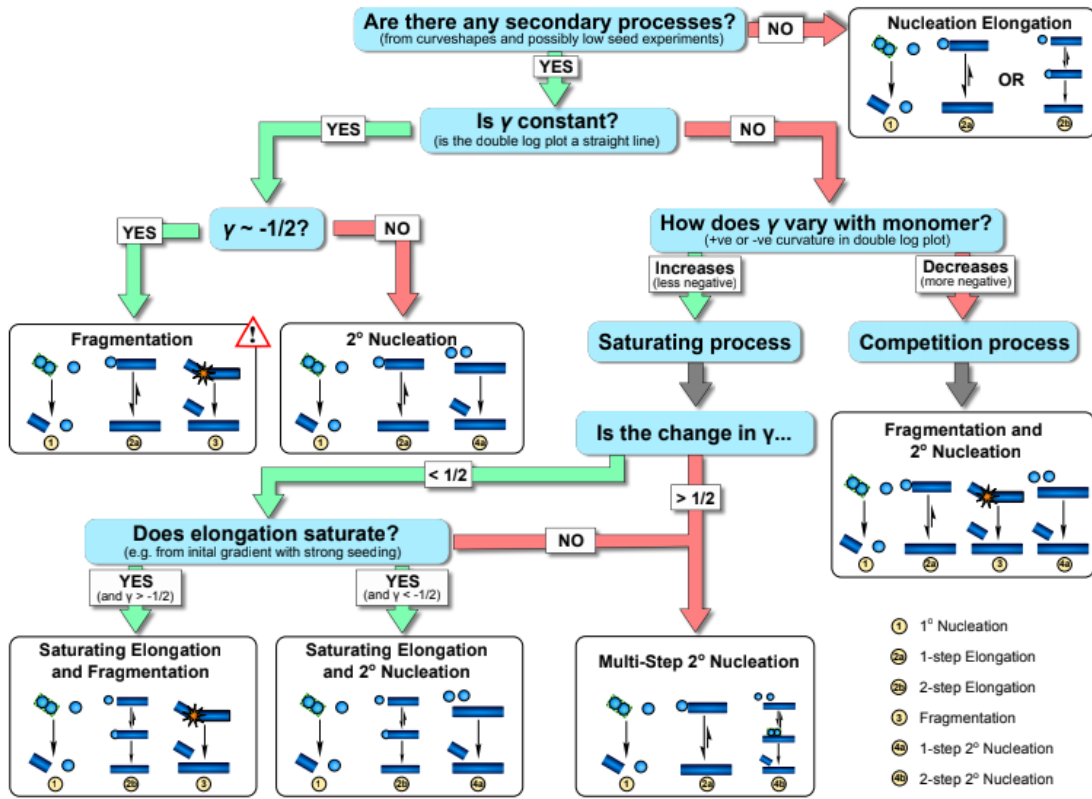


Figure 4: The curvature of the double logarithmic plots and the value of their slopes gives insights into which aggregation mechanisms are dominant. The flowchart illustrates the decision process to arrive at a likely model, based on the half time plots.

4 Results

4.1 Amylofit's parameters estimation

4.2 My program's parameter estimation

4.3 Amylofit's estimated parameters in Gillespie stochastic framework

4.4 My estimated parameters in Gillespie stochastic framework

5 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.
- [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.