

# COMPUTATIONAL PREDICTIONS FOR THE NUCLEATION MASS AND LAG TIMES INVOLVED IN A $\beta$ 42 PEPTIDE AGGREGATION

Preetam Ghosh<sup>1</sup>, Bhaswati Datta<sup>2</sup> and Vijayaraghavan Rangachari<sup>3</sup>

<sup>1</sup>Department of Computer Science, Virginia Commonwealth University, Richmond, VA, U.S.A.

<sup>2</sup>School of Computing, <sup>3</sup>Department of Chemistry & Biochemistry, University of Southern Mississippi Hattiesburg, MS, U.S.A.

**Keywords:** Alzheimer's Disease, Protein self-assembly, Ordinary differential equation, Mass-kinetics.

**Abstract:** The aggregates of amyloid- $\beta$  (A $\beta$ ) peptide are the primary neurotoxic species in the brains of Alzheimer's patients. We study the molecular-level dynamics of this process employing chemical kinetic simulations by dissecting the aggregation pathway into *pre-nucleation*, *post-nucleation* and *protofibril elongation* stages. Here, we discuss how our earlier identified rate constants for protofibril elongation were incorporated into a simplified simulation of the complete aggregation process to understand the lag-times in the sigmoidal fibril growth curves of fibril formation. We also present some initial findings on the rate constants and possible hypotheses on the nucleation mass involved in the pre-nucleation stage.

## 1 INTRODUCTION

In Alzheimer's disease (AD), the aggregates of a protein called, A $\beta$  are strongly believed to be the cause for neuronal death and cognitive decline (Selkoe, 2003). A $\beta$  aggregates to form large fibrillar deposits that follows a sigmoidal growth pattern involving a 'lag-phase' prior to fibril growth. The lag-phase is generated due to an initial rate-limiting step of nucleation (Jarrett, 1993); (Harper, 1997). However, the precise mechanism of nucleation and size of the nucleus are not known. Accurate *in vitro* analyses of the process is difficult as the intermediate oligomers are difficult to isolate and characterize. However, one intermediate from the post-nucleation phase, called protofibrils were identified (Walsh, 1997) that show propensities to both *elongate* as well as *laterally associate* to grow into mature fibrils. However, many previous works on A $\beta$  aggregation kinetics have not incorporated the pre-nucleation events that constitute a critical step of the aggregation process, more likely due to the difficulty in doing so for stochastic processes.

It is important to identify the nucleation mass and the kinetic rate constants involved in all the different phases of A $\beta$  aggregation: pre-nucleation, post-nucleation and fibril elongation. Various

models on A $\beta$  aggregation reviewed in (Morris, 2009), use curve fitting without considering the pre-nucleation events. Recently, (Lee, 2007) reported a molecular-level model of insulin aggregation that forms the basis for the model presented here. Earlier, we modelled the protofibril elongation and lateral association stages to report the kinetic rate constants involved (Ghosh, 2010). Here, our contributions are summarized as follows: (i) use the rate constants from protofibril elongation into the biophysically similar post-nucleation phase; (ii) create a model to estimate the lag-times and nucleation mass of A $\beta$ 42; (iii) report *in vitro* A $\beta$ 42 aggregation experiments that motivate our nucleation mass estimates; (iv) discuss the problems in directly comparing the simulated lag-times to those from experiments.

## 2 A $\beta$ PROCESS SIMULATION

### 2.1 *In Vitro* Results on A $\beta$ 42 Aggregation

We monitored A $\beta$ 42 aggregation in five different concentrations, 10, 25, 50, 75 and 100  $\mu$ M by thioflavin-T (ThT) fluorescence (lag-times shown in Fig 1). Since A $\beta$  aggregation is nucleation-

dependent, increase in concentration decreases the lag-time besides increasing the rate of aggregation. Hence, we observed the least lag-time for 100  $\mu\text{M}$  followed by 75, 50, 25 and 10  $\mu\text{M}$  concentrations respectively. In addition, there was an *inverse linear correlation between the logarithm of A $\beta$ 42 concentration and the corresponding lag-time as shown in Fig 1*. This observation was later used to accurately characterize the A $\beta$ 42 nucleation mass based on the simulated lag-times.

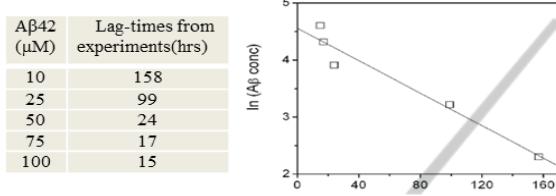


Figure 1: Lag-times from *in vitro* experiments.

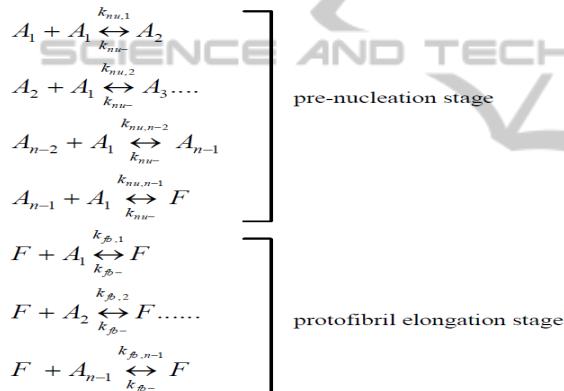


Figure 2: Reactions towards fibril formation.

## 2.2 Modified Model on A $\beta$ Aggregation

Here, we adapt the insulin aggregation model in (Lee, 2007) for the A $\beta$ 42 system. We characterize the pathway using biochemical reactions, compute the reaction fluxes and formulate the differential equations for each oligomer concentration as a function of time. Solving the set of homogeneous ODEs allow us to study the temporal dynamics of each oligomer. Fig 2 shows the modified set of reactions considered in our simulation.

Here,  $A_i$ 's denote  $i$ -mers,  $n$  is the nucleation mass and  $F$  is a fibril. The following assumptions were made: *a*) monomer adds to  $i$ -mers until fibril formation; *b*) nucleation involves monomer addition as well as a structural change in the oligomer  $A_n$  (this conformational change is implicit); *c*) post-nucleation events are faster, as the forward rate

constants for post-nucleation are much higher than those in  $k_{nu,i} = \frac{1}{2} k_{nu,1}(1 + \frac{1}{\sqrt[3]{i}})$ ;  $k_{fb,i} = k_{fb,1} \frac{1}{\sqrt[3]{i}}$

nucleation (i.e.,  $k_{nu,n+i} > k_{nu,i}$ ) (a  $\sim 10^8$  fold difference

$$J_{nu,i} = k_{nu,i} A_i A_i - k_{nu,i} A_{i+1}; \forall i = 1, \dots, n-1$$

$$J_{fb,i} = k_{fb,i} A_i F - k_{fb,i} F; \forall i = 1, \dots, n-1$$

$$\frac{dA_1}{dt} = -2 J_{nu,1} - \sum_{i=2}^{n-1} J_{nu,i} - J_{fb,1}$$

$$\frac{dA_i}{dt} = J_{nu,i-1} - J_{nu,i} - J_{fb,i}; \forall i = 2, \dots, n-1$$

$$\frac{dF}{dt} = k_{nu,n-1} A_1 A_{n-1} - k_{nu,n-1} F;$$

was reported in (Lee, 2007)); d) the reverse reaction rate constants are assumed to be independent of size  $i$ , and abbreviated as  $k_{nu,-}$  and  $k_{fb,-}$ . e) since agitation drastically shortens the lag-times,  $k_{nu,i}$  and  $k_{fb,i}$  are assumed to be diffusion-limited; using the Stokes-Einstein equation, the diffusivity is proportional to the inverted cubic root of  $i$ , resulting in:

Hence, the reaction fluxes and differential equations can be derived as follows:

## 2.3 Integrating Prototubril Elongation

A complete simulation of the A $\beta$  system requires an estimate of the following six parameters:  $k_{nu,1}$ ,  $k_{fb,1}$ ,  $k_{nu,-}$ ,  $k_{fb,-}$ ,  $n$  and  $b$ , where,  $b$  is the constant that maps ThT fluorescence to concentration estimates. It is impossible to try out different values for each of these variables to properly match the experimental plots due to the huge solution space. Hence we dissected the sigmoidal fibril-growth curve in (Ghosh, 2010) into: (i) pre-nucleation stage (ii) post-nucleation stage and (iii) prototubril elongation stage. The pre- and post-nucleation stages are well-approximated by the set of reactions shown in Fig 2. However, prototubril elongation stage needs to combine reactions from both post-nucleation and lateral association. This requires the estimation of two more rate constants: the forward and backward rate constants for the lateral association stage denoted by  $k_{la}$  and  $k_{la,-}$  respectively. In our previous report (Ghosh et al., 2010), we estimated the post-nucleation rate constants ( $k_{fb,1}$ ,  $k_{fb,-}$ ,  $k_{la}$  and  $k_{la,-}$ ) separately and verified them with *in vitro* experiments as follows:  $k_{fb,1}=9.0 \times 10^3 (\text{h}^{-1}\text{mM}^{-1})$ ,  $k_{fb,-}=4.5 \times 10^2 (\text{h}^{-1})$ ,  $k_{la}=9.0 \times 10^{-1} (\text{h}^{-1}\text{mM}^{-1})$ ,  $k_{la,-}=6.0 \times 10^{-3} (\text{h}^{-1})$ . We next directly substitute the fibril elongation rate constants into our modified model to predict the lag-times.

### 3 RESULTS AND ANALYSIS

Our model makes all possible oligomers in the *pre-nucleation stage* mathematically tractable due to the abstraction that any post-nucleation stage aggregate (starting from the nucleation mass itself, i.e.,  $A_n$ ) is treated as a fibril (i.e.,  $F$ ). However, this model does not consider the *length of the fibrils* as variables and hence cannot match the plateaus of the ThT fluorescence curves generated by experiments. This is because the fibrils of differing length will have different contributions on ThT intensity which cannot be directly captured using this model. Indeed, in Fig 4, we have mapped the concentration of  $F$  to ThT intensity for different initial A $\beta$  concentrations, and each curve saturates at the same peak. This problem was circumvented by assuming different mapping constants in (Lee, 2007) to separate the peaks for different A $\beta$  initial concentrations, which is not a biophysically correct assumption (as discussed in (Ghosh, 2010)). Thus, this model cannot implement an entirely accurate simulation of the pathway. In this paper, however, our main goal is to study the lag times in the pathway, and hence predict a working range for the nucleation mass. As our model in Fig 2 can accurately study the pre-nucleation stage oligomers, we will henceforth use it to study only the lag times in the aggregation pathway generated for different values of the nucleation mass ( $n$ ).

Table 1: Lag-times (in hrs) from our simulation for various estimates of nucleation mass.

Nucleation Mass	10 (μM)	25 (μM)	50 (μM)	75 (μM)	100 (μM)
n=7	39	9	5.3	4	3
n=8	26	10	6.2	4.9	4
n=9	43	10.9	7	5.5	4.5
n=10	36.5	12	8	6.2	5
n=11	23	12	8.7	7	5.5
n=12	12.2	8	5.2	4.2	3.9
n=13	38	8	6	5	4.1
n=14	21	7.8	6.8	5.1	4.9
n=15	55	5	5	5	5
n=16	37	6	6	6	6
n=17	176	10	10	10	10
n=18	38	9	7	7	7
n=19	35	10	8	8	8
n=20	30	11	10	10	10
n=21	26	14	12	10	10

#### 3.1 Lag Time Predictions

In Table 1, we show the simulated lag times for different nucleation mass and initial A $\beta$  concentration. In order to find the pre-nucleation rate constants along with the nucleation mass, we use the following scheme: estimate the rate constants that give the *maximum lag times* for each value of the nucleation mass. Note that, changing the rate

constants further to achieve higher lag-times render the system of differential equations unstable.

Interestingly, the simulation shows 4 distinctly different regimes of lag times corresponding to 4 different pairs of rate constants in pre-nucleation (highlighted using different colors in Table 1). At the same time, this also characterizes four different regimes of nucleation masses associated with A $\beta$  aggregation summarized as follows: *Regime 1*: n=7,8,9,10,11; *Regime 2*: n=12,13,14; *Regime 3*: n=15,16,17; *Regime 4*: n=18,19,20,21.

The rate constants for each of these regimes are shown in Table 2. Note that the forward rate constants were fixed for each nucleation mass, while the backward rate constant were varied to achieve the highest lag times as reported in Table 1. The problem here is that each of the nucleation masses does allow us to find a pair of rate constants for the pre-nucleation stage. It is however, not possible to match the simulated lag times to that observed experimentally (as reported in Fig 1). We will discuss this problem in the next section.

Table 2: Rate constants for prenucleation stage for various estimates of nucleation mass.

Nucleation Mass	$k_{\text{fnn},1}$ (h $^{-1}$ mM $^{-1}$ )	$k_{\text{fnn}}$ (h $^{-1}$ )
n=7	$1.38 \times 10^0$	$9.5 \times 10^{-1}$
n=8	$1.38 \times 10^0$	$5.6 \times 10^{-1}$
n=9	$1.38 \times 10^0$	$3.7 \times 10^{-1}$
n=10	$1.38 \times 10^0$	$2.3 \times 10^{-1}$
n=11	$1.38 \times 10^0$	$1.9 \times 10^{-1}$
n=12	$4.6 \times 10^0$	$2.9 \times 10^{-1}$
n=13	$4.6 \times 10^0$	$2.8 \times 10^{-1}$
n=14	$4.6 \times 10^0$	$7.0 \times 10^{-2}$
n=15	$1.38 \times 10^1$	$1.4 \times 10^{-2}$
n=16	$1.38 \times 10^1$	$1.0 \times 10^{-2}$
n=17	$1.38 \times 10^1$	$1.0 \times 10^{-3}$
n=18	$2.1 \times 10^1$	$1.1 \times 10^{-2}$
n=19	$2.4 \times 10^1$	$1.1 \times 10^{-2}$
n=20	$2.6 \times 10^1$	$1.5 \times 10^{-2}$
n=21	$2.8 \times 10^1$	$1.2 \times 10^{-2}$

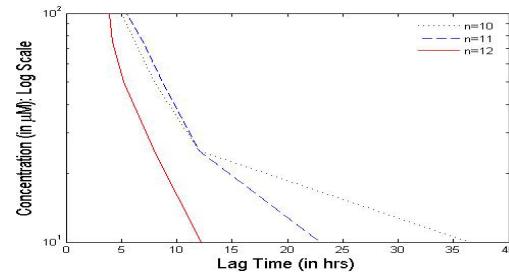


Figure 3: Close to linear semi-log plots.

However, as seen in Fig 1 (and also from other experiments in our lab consistently), the semi-log plot of the lag-times against initial concentration of A $\beta$  is linear. So we used this property to figure out what values of nucleation mass are most feasible for the A $\beta$ 42 pathway. Note that n=10, 11 (in Regime 1) and n=12 (in Regime 2) are *close to linear* and

hence may serve as good approximations of the nucleation mass (Fig 3). It was also observed in course of the simulations that an initial concentration of 10  $\mu\text{M}$  made the simulation erratic for a wide range of rate constants (because of increased dynamism and stochasticity in the system with lower molecular count of the species rendering the ODEs unstable). So, we generated these semi-log plots for the different regimes (data not shown) by removing the data points for 10  $\mu\text{M}$ . Indeed, these curves show a more stable relationship between the lag times and the initial concentrations, and we find close to linear behavior for  $n=10,11$  (in Regime 1),  $n=12,13$  (in Regime 2) and  $n=15,16,17$  (in Regime 3).

The next question is whether 10,11,...,17 is the right range for the nucleation mass, or can we further reduce it? Fig 4 shows the concentration curve for  $F$  against time and different initial concentrations. One requirement for the rate constants reported above is that these curves *must saturate to the same peak* as expected mathematically. So we considered this to be another constraint that reduced the range of feasible nucleation masses to  $n=10, 11...,14$ . Note that  $n=15, 16, 17$  did not allow the concentration curves to saturate (data not shown), and hence were ruled out as possible candidates for the nucleation mass.

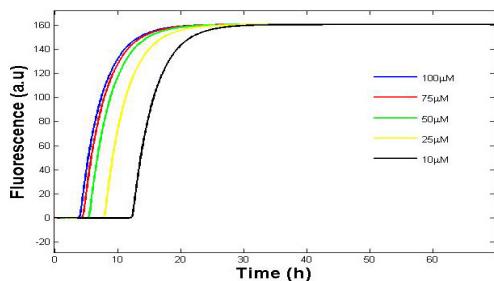


Figure 4: Simulated fluorescence change curves for different initial concentrations with  $n=12$ .

### 3.2 Can we Compare Simulated and Experimentally Observed Lag Times?

The experimental ThT fluorescence plots show the *cumulative effect* of all oligomers of a certain size (and beyond). The results shown above plot the concentration of  $F$  which model the cumulative effect of all the nucleated oligomers in the pathway. However, it is assumed that *all nucleated oligomers* show up on the ThT curves (this is generally not the case from actual experiments). Hence, the lag times estimated from our model are lower than that seen experimentally. Also, it is not yet known what size

of oligomers actually show up ThT positive and hence the experimental estimates are at best the maximum limits of the lag times for each initial  $A\beta$  concentration. To get around this problem, we varied the rate constants to estimate the maximum possible lag times for each value of the nucleation mass. This is still an approximation of the actual system and needs further study. Ideally, we need to know what sizes of oligomers are considered ThT positive such that the experimental curves can be meaningfully compared to the simulated plots. The present paper, however, gives us a feasible range of nucleation masses to work with in order to build a complete simulation of the on-pathway. The rate constants estimated in this exercise can serve as a guidance for the complete simulation where we will need a more detailed model (with separate parameters for each post-nucleation oligomer) to properly model their effects on the system.

## 4 CONCLUSIONS

In this paper, we have studied the lag times in the sigmoidal  $A\beta$  fibril formation pathway. We also reported that the nucleation mass can potentially be in the range 10,11,..., 14 mers. In order to reduce the complexity of the entire fibril formation pathway, we used the rate constants that we have earlier estimated for the post-nucleation stage into a modified model that can approximately characterize the complete pathway. These estimates will serve as the basis for implementing a complete and accurate simulation of the pathway wherein we have approximately estimated all the 6 variables involved. Such a simulation will pave the path to study the complete system dynamics of  $A\beta$  aggregation leading to a better understanding of AD in general.

## ACKNOWLEDGEMENTS

This work was supported by NSF-1158608.

## REFERENCES

- Selkoe, D. J., Schenk, D. (2003). Alzheimer's disease: molecular understanding predicts amyloid-based therapeutics, *Annu Rev Pharmacol Toxicol* 43, 545-584.
- Harper, J. D., Lansbury, P. T., Jr. (1997). Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of

- the time-dependent solubility of amyloid proteins, *Annu Rev Biochem* 66, 385-407.
- Jarrett, J. T., Berger, E. P., Lansbury, P. T., Jr. (1993). The C-terminus of the beta protein is critical in amyloidogenesis, *Ann N Y Acad Sci* 695, 144-148.
- Walsh, D. M., Lomakin, A., Benedek, G. B., Condron, M. M., Teplow, D. B. (1997). Amyloid beta-protein fibrillogenesis. Detection of a protofibrillar intermediate, *J Biol Chem* 272, 22364-22372.
- Morris, A. M., Watzky, M. A., Finke, R. G. (2009). Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature, *Biochim Biophys Acta* 1794, 375-397.
- Ghosh, P., Kumar, A., Datta, B., Rangachari, V (2010). Dynamics of protofibril elongation and association involved in A $\beta$ 42 peptide aggregation in Alzheimer's disease, *BMC Bioinformatics* 2010, 11(Suppl 6):S24, pp. 1-19.
- Lee, C. C., Nayak, A., Sethuraman, A., Belfort, G., McRae, G. J. (2007). A three-stage kinetic model of amyloid fibrillation, *Biophys J* 92, 3448-3458.

