THEORY AND SIMULATION OF AMYLOID AGGREGATION PROCESS: SEQUENCE EFFECTS AND DEFECTS

by

ELAHEH GHANATI

B.S., Ferdowsi University of Mashhad, 2007

M.S., Shahid Beheshti University of Tehran, 2010

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Physics

College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2016

Approved by:

Major Professor

Jeremy Schmit

Abstract

In this work, we present a model for the kinetics of amyloid fibril aggregation. In the model we mapped the process of Hydrogen bond (H-bond) formation and breakage to a random-walk. we captured the effect of side chains using position dependent H-bonds free energies which allows us to calculated the residence time for different binding alignments with the fibril. The residence time can be compared to the diffusion-limited attachment rate to give net aggregation stability. This stability increases exponentially with increasing number of bonds or binding energy in homopolymer chains, however for chains with patterned sequences, the residence time shows strong effects of the binding alignment. Using the residence time for uniform structures combined with estimate of the diffusion rate, we modeled and simulated the kinetics of amyloid aggregation. Results of the simulations gives the bond energies and concentrations required for the onset of growth of aggregates.

Acknowledgements

Dedication

This work is dedicated to my family and especially my mother, Narges, my father, Mousa and my dear sister Azam. My family proved that physical distance never has meaning. they understand my passion and support me with all of their heart. They never blame my mistakes and kindly support me in happiness and pains, in every moment. Family is the most valuable thing in life. Thanks to all and I love you.

Contents

List of Figures

Chapter 1

Introduction

Amyloid fibrils are insoluble, degradation resistant assemblies formed from the aggregation of soluble proteins. These misfolded proteins deposit in organs and tissues, disrupt normal organ function and causes different diseases including many neurological disorders or systemic diseases. Alzheimer's disease, atrial amyloidosis, spongiform encephalopathies (e.g., Mad cow disease), primary systemic amyloidosis, senile systemic amyloidosis, haemodialysis-related amyloidosis, hereditary nonneuropathic systemic amyloidosis, type II diabetes, injection-localized amyloidosis, secondary systemic amyloidosis, hereditary cerebral amyloid angiopathy, finnish hereditary systemic amyloidosis, familial amyloid polyneuropathy I and ll, aging pituitary, prolactinomas, and British familial dementia are examples of these diseases[2–4, 6].

Because of the effect of protein aggregation in causing many diseases, understanding the mechanism in which proteins and peptides act in the dense *in-vivo* environment has received much attention. It is a question as to how, even in environments with high concentrations act in the way that is less favorable for aggregation [4]. The structure of amyloid fibrils and their stability has been studied and the improvement of experimental tools is suggesting better models for the mechanism of formation. These studies started in the mid-19th century, when Schleiden and Vichow gave the term "amyloid" for description of the iodine stained deposits in investigation of a liver at autopsy [6]. Although at first investigation they were thought to be carbohydrate, later the existence of nitrogen cast doubt on this idea and the improvement of detection methods proved their proteinaceous content. Later

X-ray fiber diffraction patterns helped to reveal the structures, once separated from the blurring effect of tissues. Invention of material analysis methods later proved that amyloid fibrils are found in many diseases. This discovery motivated biologists to investigate all kind of diseases and identify particular amino acids that cause aggregation using statistical approaches. The defining characteristic of amyloid fibril is an ordered structure of cross-β conformation in their polypeptide backbone. Statistical analysis has defined sequence dependent characteristic affecting the propensity to form amyloid [5].

Understanding the mechanism of formation of amyloid fibril from the first steps to completely mature insoluble form, has motivated many theoretical models and experimental studies. Investigations showed the mechanism is a delicate function of the physicochemical condition of system. In the other words, the system undergoes different trajectories with different conditions. These trajectories determine which final form amyloid fibril will have. These steps are important, because they are points for formation of metastable structure and start of new more complicated structure[6–10].

Many models have been proposed for the aggregation sequence leading from soluble proteins to fibrils. These include various states such as folded and unfolded monomers, dimer, trimer, protofibril and fibril. Many of these states are metastable, but most evidence implicates some form of metastable aggregate as the toxic species responsible for disease progression. Still there is much debate on the nature of the toxic species and the aggregation pathway.

It is important to determine whether the insoluble form of β-sheets are the start of disease, or are actually the end point. Research showed there is correlation between Alzheimer's disease and soluble Aβ structures [14]. In other words, this correlation says soluble non-fibrilar intermediates are causing diseases. Also, attempts for more accurate ways to understand which part of polypetide strands are aggregation prone, started with investigating the structural information banks of amyloid fibrils. These attempts resulted in the creation of bioinformatic tools for predicting aggregation prone regions of strands. The principle of these bioinformatic methods are different. Some of them like TANGO and PAGE tried to predict aggregation prone regions based on physicochemical properties of amino acids. The TANGO algorithm emphasizes physicochemical rules behind β-sheet formation and defines a probability score for finding particular peptide segments in different conformations such as α-helix, β-strand, turn, random coil, and β-aggregate[17]. PAGE computes the aggregation

score for a given amino acid sequence based on propensity and absolute aggregation rates. The PASTA model emphasizes on statistical information based on the propensity of particular residues in aggregates and defining energy functions [5]. These three are just a few among many attempts. There are several important insights that have come from these data analyses: First, aggregation prone regions avoid charged or polar residues. In other words, charged residues favor the soluble state. Second, the net charge of motifs is reported to have an inverse relation with their aggregation propensity. Third, another important result is that aggregation-prone regions often are full of aromatic and hydrophobic β branched residues [14–17].

One question that arises is: how reliable are available detection techniques for the investigation of these structures? There is a debate that detection techniques can only give some aspect of the species and some other aspects remain hidden. For example NMR and X-ray techniques are efficient methods for the detection of crystalline structure, but before the creation of crystalline structure there are many initial important steps in amyloid fibril formation that are inaccessible to NMR and X-ray studies.

In this work we tried to model the protein aggregation process. Experiments, have shown that aggregate growth is a function of concentration and structure of polypeptides [5]. Although many physical factors other than those mentioned previously are also effective, in this work we attempted to focus on these two factors. This thesis is structured as follows: In chapter (2) I present the random-walk model of aggregation and develop the transfer matrix method for calculating residence times. Next, dimer and trimer structures are defined and residence times for peptides with these symmetries are calculated. Chapter (3) includes simulations of the model with stochastic aggregation process. Also, growth rate and average length of registry versus concentration graphs are shown for uniform chains. The non-uniform structures are left for future work.

Chapter 2

Residence Time Models for Uniform and Non-uniform Structures
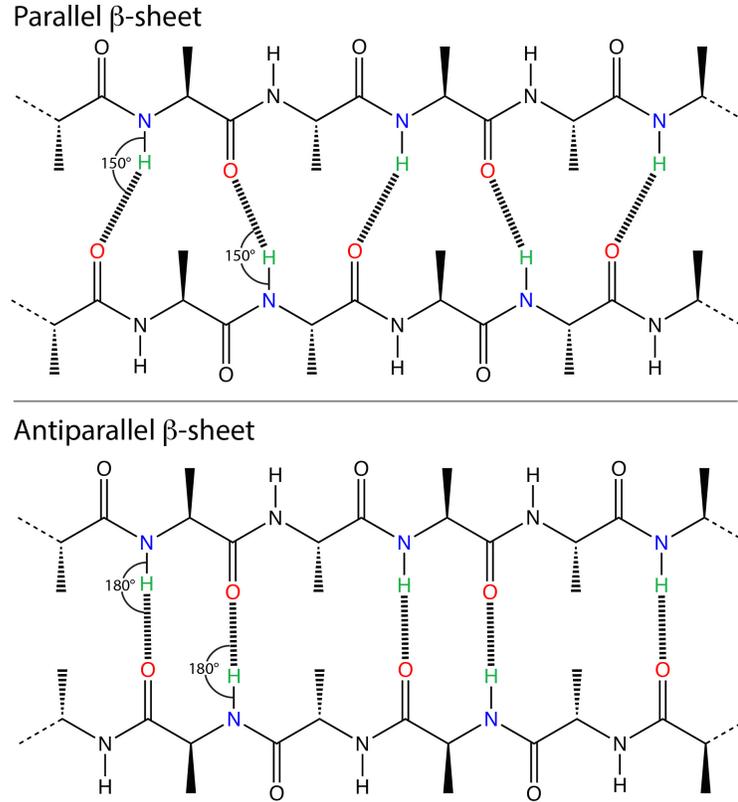
Figure 2.1: A schematic diagram of β-sheet structure, parallel and antiparallel [1].

2.1 Uniform Structure

2.1.1 Calculation of Residence Time in Uniform Structures

Our theory of amyloid aggregation is based on the model first introduced by Schmit [2]. In that work, the formation and breakage of hydrogen bonds (H-bond) between strands is mapped to a random-walk. To accomplish this mapping, the rates of bond formation and breakage are related to the free energy using detailed balance [1, 2]. According to the principle of detailed balance, for a system with states with different free energies, the transition rates are related by:

$$k_2/k_1 = e^{-(\epsilon_2 - \epsilon_1)/k_\mathrm{B}T}. \tag{2.1}$$

Where k_B is Boltzmann constant, T temperature, ϵ_2 and ϵ_1 are free energies of states and k_2 and k_1 are rates that system moves to states 2 and 1. Schmit assumed one polypeptide,
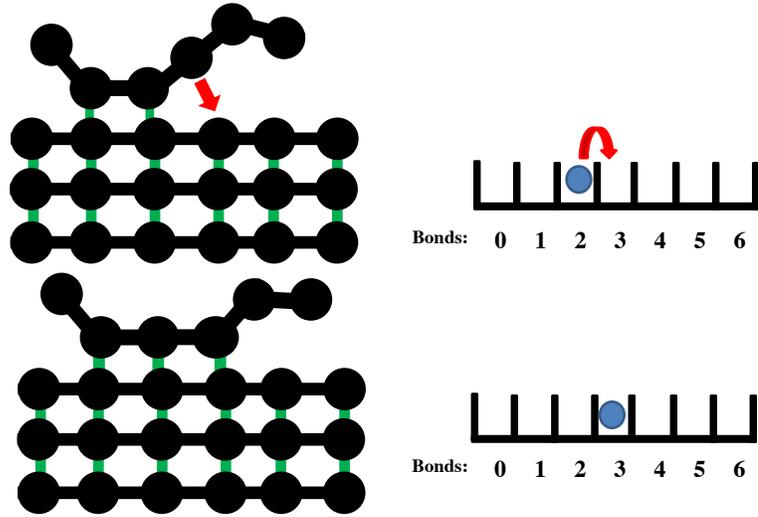
Figure 2.2: The number of H-bond is a good reaction coordinate. We model the number of H-bond (left) as a random walk (right). The number of bonds can change from x to x-1 or x+1. The number of bonds change between 0 (a dissociated molecules) to N (a fully bounded state).

after having contact with fibril, starts forming bonds between its backbone and the exposed peptide groups on the fibril. In uniform chains, all the amino acids have the same sidechains, so all the bonds have the same strength. Assuming that polypeptide makes x H-bond with the fibril, (x=2) at the top panel of the Figure. 2.2. Then system can make additional bonds or break bonds, leading to the situation with x+1 or x-1 bonds. So the variable x, would perform a random walk [23], between 0 to N, where N is the maximum number of possible bonds, (this number depends on the length of the molecule and the alignment between the incoming molecule and the fibril template). Hereafter we refer to the alignment as the "registry". Starting with the situation of x bonds, the peptide can form an additional bond with the ratio:

$$k_+/k_- = e^{-\epsilon_b/kT}. \tag{2.2}$$

Where $-\epsilon_b$ is the thermodynamic stability of H-bond. Defining P(x,t) as the probability that a system with x H-bond at time =0 has neither made or broken any additional bonds at time t, obeys the equation:

$$\frac{dP(x,t)}{dt} = -(k_- + k_+)P(x,t) \tag{2.3}$$

Which has the trivial solution $P(x,t) \propto e^{-t(k_- + k_+)}$. The mean residence time before any change in x is given by:

$$\tau = \int_0^\infty tP(x,t)dt = \frac{1}{k_- + k_+} \tag{2.4}$$

It also can be shown that τ, is the average residence time for both systems that eventually reach x+1 or x-1 bond. The probability of these two outcomes are:

$$p_+ = \int_0^\infty k_+ P(x,t)dt = \frac{k_+}{k_- + k_+} \tag{2.5}$$

$$p_- = \int_0^\infty k_- P(x,t)dt = \frac{k_-}{k_- + k_+} \tag{2.6}$$

The mean time required for a particle starting at x=0 to reach an absorbing boundary at x=0 satisfies the recursion relation:

$$t(x) = p_+ t(x+1) + p_- t(x-1) + \tau \tag{2.7}$$

reflecting that after waiting one time step τ the particle that has started at x, starts a new random walk at x+1 with the probability p_+ or at x-1 with the probability p_-. This inhomogeneous equation is hard to solve, so we transform it to a homogenous form.

$$\Theta(x) = p_+ \Theta(x+1) + p_- \Theta(x-1) \tag{2.8}$$

Where the transformed variable is given by:

$$t(x) = \Theta(x) - x\frac{\tau}{p_+ - p_-} \tag{2.9}$$

Equation 2.8 can be written as the vector product:

$$V(x+1) = MV(x) \tag{2.10}$$

Where V is given by:

$$V(x) = \begin{pmatrix} \Theta(x) \\ \Theta(x-1) \end{pmatrix} \qquad (2.11)$$

And the transfer matrix (M) is given by:

$$M = \begin{pmatrix} \frac{1}{p_+} & -1 + \frac{1}{p_+} \\ 0 & 1 \end{pmatrix} \qquad (2.12)$$

We can calculate the residence time on the Nth site by repeating the multiplication:

$$V(N) = M^N V(1) \qquad (2.13)$$

Which can be done

$$V(N) = U(U^{-1}MU)^N U^{-1}V(1) \qquad (2.14)$$

The diagonal form of transfer matrix $(M_{\mathrm{d}} = U^{-1}MU)$ is [1]:

$$U^{-1}MU = \begin{pmatrix} -1 + \frac{1}{p_+} & 0 \\ 0 & 1 \end{pmatrix} \qquad (2.15)$$

And the matrix U is:

$$U = \begin{pmatrix} -1 + \frac{1}{p_+} & 1 \\ 1 & 1 \end{pmatrix} \qquad (2.16)$$

---

[1]Appendix 1,Part A

2.1.2 Boundary Condition in Uniform Structure

To calculate the residence time we require two boundry conditions, when all bonds are broken, the residence time is zero, so $t(0) = \Theta(0) = 0$. At the end of strand, we don't have possibility to form more bonds. Therefore we have:

$$p_+ = \frac{k_+}{k_+ + k_-} = 0 \rightarrow k_+ = 0 \tag{2.17}$$

$$p_- = \frac{k_-}{k_+ + k_-} = 1 \tag{2.18}$$

based on these two equations, the waiting time for the last site (see Eq 2.7) is given by:

$$\tau_{\text{end}} = \frac{1}{k_-} \tag{2.19}$$

We calculate this τ_{end} to use Eq (2.9) with n or n-1 jumps and calculate the amount of Θ_1:

$$V(N) = M^N V(1) = U M_{\text{d}}^n U^{-1} V(1) \tag{2.20}$$

$$V(N-1) = M^{N-1} V(1) = U M_{\text{d}}^{n-1} U^{-1} V(1) \tag{2.21}$$

$$V(N) = \begin{pmatrix} \Theta(N) \\ \Theta(N-1) \end{pmatrix} \tag{2.22}$$

$$V(N-1) = \begin{pmatrix} \Theta(N-1) \\ \Theta(N-2) \end{pmatrix} \tag{2.23}$$

$$t(N) = t(N-1) + \tau_{\text{end}} \tag{2.24}$$

$$\Theta(N) - (N)\frac{\tau}{p_+ - p_-} = \Theta(N-1) - (N-1)\frac{\tau}{p_+ - p_-} + \tau_{\text{end}} \tag{2.25}$$

$$\rightarrow \Theta(1) = \frac{(\frac{1}{p_+} - 1)^n (p_+ - 1)(2p_+ - 1 + \tau k_-)}{p_+(2p_+ - 1)k_-} \tag{2.26}$$

2.2 Dimer Structure

2.2.1 Calculation of Residence Time in Amyloid Fibril with Dimer Structures, Definition of Structure and Interactions

In this part we consider we have two kind of sidechaines, hydrophobic and polar (or charged) shown in Figure.2.3 with red (R) and green (G) respectively. We also assume if two hydrophobic sidechaines, H and H (hydrophobic and hydrophobic, RR in the figure) meet each other and form bond, that bond would be stronger than if two sidechaines with different kind meet, H and P or P and P (GR in the figure). As it mentioned before, polar residues favor high solubility and hydrophobic residues are more prone for aggregation [4, 5]. Dill also based on lattice statistical mechanics, suggested a theory for the folding of heteropolymer molecules. Folding considered to be driven by the association of solvophobic monomers to avoid solvent and opposed by the chain configurational entropy [13].
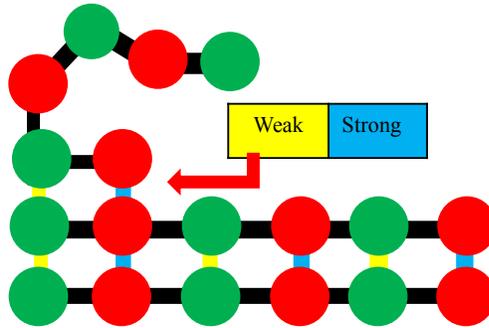
Figure 2.3: A schematic diagram of aggregation of peptid with dimer periodicity. Sidechaines are are color coded as green (polar) or red (hydrophobic). Also, strong interactions are shown with blue and weak interactions are shown with yellow color.

First, consider starting from strong bond at x, so the next bond which we have to form is a weak bond. We assume that the rate of bond formation is limited by diffusion through the solvent and therefore is independent of the bond strength. Therefore, the site dynamics only differ in the bond breakage ratio. We have "slow" sites for which bond breakage requires rapturing a strong bond, and "fast" sites for which x denotes a weak bond. We have two forms of recursion relation in new "RG" structure (for simplicity we label the site-specific quantities as "s" for strong and "w" for weak), $(p_{+s}), (p_{-s})$ and $(p_{+w}), (p_{-w})$ in these equations represent the probability for motion from strong and weak bonds to right and left respectively:

$$t(x-1) = p_{+\text{w}}t(x) + p_{-\text{w}}t(x-2) + \tau_1, \tag{2.27}$$

$$t(x) = p_{+\text{s}}t(x+1) + p_{-\text{s}}t(x-1) + \tau_2, \tag{2.28}$$

$$t(x+1) = p_{+\text{w}}t(x+2) + p_{-\text{w}}t(x) + \tau_1. \tag{2.29}$$

Where:

$$p_{+\text{w}} = \frac{k_{+\text{w}}}{k_{+\text{w}} + k_{-\text{w}}}, \tag{2.30}$$

$$p_{-\text{w}} = \frac{k_{-\text{w}}}{k_{+\text{w}} + k_{-\text{w}}}, \tag{2.31}$$

$$p_{-\text{s}} = \frac{k_{-\text{s}}}{k_{+\text{s}} + k_{-\text{s}}}, \tag{2.32}$$

$$p_{+\text{s}} = \frac{k_{+\text{s}}}{k_{+\text{s}} + k_{-\text{s}}}, \tag{2.33}$$

Eqs. 2.27, 2.28 and 2.29 τ_1 and τ_2 are the average time for random walk steps starting from weak and strong sites respectively. Using Eq. 2.27 and Eq. 2.29 to eliminate $t(x-1)$ and $t(x+1)$ from Eq. 2.28 give us an equation including $t(x-2)$ and $t(x+2)$ which can be written in the form of:

$$t(x) = \frac{p_{+\text{w}}p_{+\text{s}}t(x+2)}{(1 - p_{-\text{w}}p_{+\text{s}} - p_{+\text{w}}p_{-\text{s}})} + \frac{p_{-\text{w}}p_{-\text{s}}t(x-2)}{(1 - p_{-\text{w}}p_{+\text{s}} - p_{+\text{w}}p_{-\text{s}})} + \frac{\tau_1 + \tau_2}{(1 - p_{-\text{w}}p_{+\text{s}} - p_{+\text{w}}p_{-\text{s}})}. \tag{2.34}$$

Now defining:

$$p'_+ = \frac{p_{+\text{w}}p_{+\text{s}}}{(1 - p_{-\text{w}}p_{+\text{s}} - p_{+\text{w}}p_{-\text{s}})}, \tag{2.35}$$

$$p'_- = \frac{p_{-\text{w}}p_{-\text{s}}}{(1 - p_{-\text{w}}p_{+\text{s}} - p_{+\text{w}}p_{-\text{s}})}, \tag{2.36}$$

$$\tau' = \frac{\tau_1 + \tau_2}{(1 - p_{-\text{w}}p_{+\text{s}} - p_{+\text{w}}p_{-\text{s}})}, \tag{2.37}$$

And considering jumps with two steps (we named every two steps a supercell in our calcu-

lation, in other words supercell includes the symmetry of the system and get repeated with the dimer pattern) with equation:

$$V_2(x + 2) = M'V_2(x) \tag{2.38}$$

Where

$$V_2(x) = \begin{pmatrix} \Theta(x) \\ \Theta(x - 2) \end{pmatrix}$$

In dimer structure the transformation to a homogenous recursion relation is accomplished with: [2]

$$t(x) = \Theta(x) - \frac{x}{2}\frac{\tau'}{p'_+ - p'_-}. \tag{2.39}$$

---

[2]For additional calculations refer to appendix, part A

2.2.2 Boundary Condition in Dimer Structure

The boundary conditions are derived as in the calculation of section 1.1.2 for uniform structure. Here, we also need to account for the dimer periodicity of the lattice. If we start from a strong bond and try to break a weak bond, is faster than starting from a weak bond and try to break the strong bond. We have two kind of motion, fast to go forward or backward from strong to weak bond with the probabilities respectively $p_{+\mathrm{s}}$ and $p_{-\mathrm{s}}$ and slow to go forward or backward from weak to strong bond with the probabilities respectively $p_{+\mathrm{w}}$ and $p_{-\mathrm{w}}$.

Assume that we have started from one weak bond and the strand is formed of pairs, so the last dimer unit there would be one strong bound, but the strand is finished, so we only can have $k_{-\mathrm{w}}$. But this condition doesn't exist for motion from the previous bond and we still are allowed to jump one more time. In this condition our boundary condition determine the τ' in this form:

$$p_{+\mathrm{s}} = 0 \tag{2.40}$$

$$p_{-\mathrm{s}} = 1 \tag{2.41}$$

Based on these two equations:

$$\tau_2 = \frac{1}{k_{-\mathrm{s}}} \tag{2.42}$$

$$\tau_1 = \frac{1}{k_{+\mathrm{w}} + k_{-\mathrm{w}}} \tag{2.43}$$

$$\tau'_{\mathrm{end}} = \frac{\frac{1}{k_{-\mathrm{s}}} + \frac{1}{k_{-\mathrm{w}} + k_{-\mathrm{w}}}}{1 - \frac{k_{+\mathrm{w}}}{k_{+\mathrm{w}} + k_{-\mathrm{w}}}} \tag{2.44}$$

Where this τ'_{end} is the relation for τ_{end} for the structure with the "Dimer" symmetry.

Similarly we can find the value of Θ_1 with the Eq. 2.39:

$$\Theta(1) = \frac{(\frac{1}{p'_+} - 1)^{n/2}(p'_+ - 1)(k_{-w}(2p'_+ - 1 + \tau'k_{-s}) + (2p'_+ - 1)(k_{-s} + k_{+s})}{p'_+(2p'_+ - 1)k_{-s}k_{-w}}. \qquad (2.45)$$

2.3 Trimer Structure

2.3.1 Calculation of the Residence Time in Amyloid Fibril with Trimer Structures
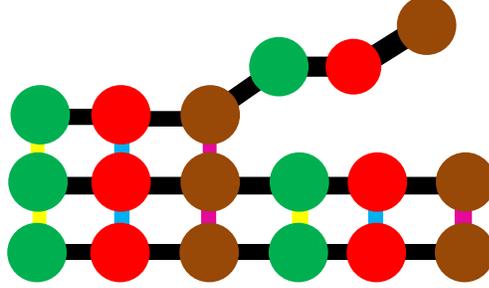
Figure 2.4: A schematic diagram of the aggregation of peptides with trimer periodicity. Three types of amino acids are represented with green, red and brown colors.

Considering a poly-peptide with a repeating pattern of three amino acids (Figure 2.4), we have three forms of recursion relations. If we start from a bond located at x, the probability for forward and reverse steps are p_{+1} and p_{-1} respectively and average time for moving from this step τ_1. Similarly if we start from the bond located at $x+1$, we can go forth and back with the possibilities p_{+2} and p_{-2}, and time constant τ_2 and finally for the bond located at $x+2$, we can go forth and back with the possibilities p_{+3} and p_{-3}, and time constant τ_3. The rest of calculations is briefly in this form:

$$t(x) = p_{+1}t(x+1) + p_{-1}t(x-1) + \tau_1, \tag{2.46}$$

$$t(x+1) = p_{+2}t(x+2) + p_{-2}t(x) + \tau_2, \tag{2.47}$$

$$t(x+2) = p_{+3}t(x+3) + p_{-3}t(x+1) + \tau_3. \tag{2.48}$$

Where:

$$p_{+1} = \frac{k_{+1}}{k_{+1} + k_{-1}}, \tag{2.49}$$

$$p_{-1} = \frac{k_{-1}}{k_{+1} + k_{-1}}, \tag{2.50}$$

$$p_{+2} = \frac{k_{+2}}{k_{+2} + k_{-2}}, \tag{2.51}$$

$$p_{-2} = \frac{k_{-2}}{k_{+2} + k_{-2}}, \tag{2.52}$$

$$p_{+3} = \frac{k_{+3}}{k_{+3} + k_{-3}}, \tag{2.53}$$

$$p_{-3} = \frac{k_{-3}}{k_{+3} + k_{-3}}, \tag{2.54}$$

And τ_1, τ_2 and τ_3 are average waiting time for starting motion from x, x+1 or x+2. Substituting equations 2.46 and 2.48 into Eq. 2.47 will give us the relation in term of $t(x-3)$ and $t(x+3)$:

$$
\begin{aligned}
t(x) = {} & \frac{p_{+1}p_{+2}p_{+3}t(x+3)}{(1 - (p_{-2}p_{+1} + p_{-3}p_{+2} + p_{-1}p_{+3}))} + \\
& \frac{p_{-1}p_{-2}p_{-3}t(x-3)}{(1 - (p_{-2}p_{+1} + p_{-3}p_{+2} + p_{-1}p_{+3}))} + \\
& \frac{\tau_1 + \tau_3 p_{-1} + \tau_2 p_{-1}p_{-3} + \tau_2 p_{+1} - \tau_1 p_{-3}p_{+2} + \tau_3 p_{+1}p_{+2}}{(1 - (p_{-2}p_{+1} + p_{-3}p_{+2} + p_{-1}p_{+3}))}.
\end{aligned}
\tag{2.55}
$$

Also similar to the dimer structure we can define p'_+ and p'_- for this trimer structure. A straightforward calculation confirms that $p'_+ + p'_- = 1$ [3].

$$p'_+ = \frac{p_{+1}p_{+2}p_{+3}}{(1 - (p_{-2}p_{+1} + p_{-3}p_{+2} + p_{-1}p_{+3}))} \tag{2.56}$$

$$p'_- = \frac{p_{-1}p_{-2}p_{-3}}{(1 - (p_{-2}p_{+1} + p_{-3}p_{+2} + p_{-1}p_{+3}))} \tag{2.57}$$

$$\tau' = \frac{\tau_1 + \tau_3 p_{-1} + \tau_2 p_{-1}p_{-3} + \tau_2 p_{+1} - \tau_1 p_{-3}p_{+2} + \tau_3 p_{+1}p_{+2}}{(1 - (p_{-2}p_{+1} + p_{-3}p_{+2} + p_{-1}p_{+3}))} \tag{2.58}$$

In jumping over three steps we have:

$$V_3(x+3) = M'' V_3(x) \tag{2.59}$$

Where:

$$V_3(x) = \begin{pmatrix} \Theta(x) \\ \Theta(x-3) \end{pmatrix}$$

---

[3] Appendix 1, Part B

16

In this case:

$$t(x) = \Theta(x) - \frac{x}{3}\frac{\tau'}{p'_+ - p'_-} \tag{2.60}$$

We have three kind of motion, with possibility to go forward or backward $p_{+1}, p_{-1}, p_{+2}, p_{-2}$, p_{+3} and p_{-3} from different positions A, B or C. Assume that we start from the bond number 1 (A) so the last bond would be the bond number 3 (C type), after the last bond forms no more forward steps are possible so, we only can have $p_{-3} = 1$. In this condition the boundary condition determines the τ'_{end} in this form:

$$p_{+3} = 0$$

$$k_{+3} = 0$$

$$p_{-3} = 1$$

based on these equations:

$$\tau_3 = \frac{1}{k_{-3}}$$

but still we have:

$$\tau_1 = \frac{1}{k_{+1} + k_{-1}}$$

$$\tau_2 = \frac{1}{k_{+2} + k_{-2}}$$

$$\begin{aligned}
\tau'_{\text{end}} = {}& \left[1 - \left[(\frac{k_{-2}}{k_{+2} + k_{-2}})(\frac{k_{+1}}{k_{+1} + k_{-1}}) + (\frac{k_{+2}}{k_{+2} + k_{-2}})\right]\right]^{-1} \\
& \times \left[(\frac{1}{k_{+1} + k_{-1}}) + (\frac{1}{k_{-3}})(\frac{k_{-1}}{k_{+1} + k_{-1}}) + (\frac{1}{k_{+2} + k_{-2}})(\frac{k_{+1}}{k_{+1} + k_{-1}})\right. \\
& + (\frac{1}{k_{+2} + k_{-2}})(\frac{k_{+1}}{k_{+1} + k_{-1}}) - (\frac{1}{k_{+1} + k_{-1}})(\frac{k_{+2}}{k_{+2} + k_{-2}}) \\
& \left. + \frac{1}{k_{-3}}\frac{k_{+1}}{(k_{+1} + k_{-1})}\frac{k_{+2}}{(k_{+2} + k_{-2})}\right].
\end{aligned} \tag{2.61}$$

After finding Θ_1 in the form of:

$$\Theta(1) = \frac{1}{k_{-1}k_{-2}k_{-3}(2p'_+ - 1)p'_+} \times$$

17

$$[(\frac{1}{p'_+} - 1)^{n/3}(p'_+ - 1)(k_{-2}k_{-3} + k_{+1}(k_{-3} + k_{+2})(2p'_+ - 1)+$$

$$k_{-1}(k_{-3} + k_{+2})(2p'_+ - 1) + k_{-2}(2p'_+ - 1 + k_{-3}\tau')]. \tag{2.62}$$

Using this initial function and equation 2.60 would give us the residence time in any part of strand.[4].

---

[4]Appendix 1 part B

2.4 Residence Time for Uniform and Non-uniform Structures with Different Length of Registries

We have calculated the residence time for different structures and varying registries. First we consider a uniform strand with the length of 26 molecules with $E_{\text{bind}} = 0.28k_{\text{B}}T$. The residence time for molecules attached by a single bond ($x = 1$) is short, and exponentially increases with increasing the number of bonds. However, after particular number of bonds we observe saturation for the amount of time. This is because the residence time is a
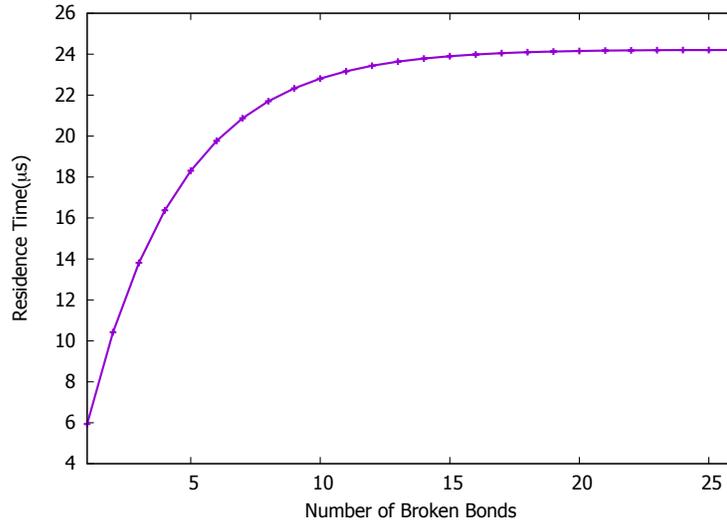
Figure 2.5: Residence time of two strands with the length of 26 molecules with $E_{\text{bind}} = 0.28K_{\text{B}}T$ as a function of the initial number of bonds. System with more than $\simeq 10$ will eventually reach the fully bound state before detaching.

weighted average of two outcomes: (1) a quick detachment before many bonds form, and (2) breakage of all bonds from a (nearly) fully bound state. The reason we get a saturating behavior is that outcome (1) gets less likely as more bonds form. Therefore, above a certain number of bonds, outcome (2) is the only option.

It is trivial that if we compare the saturation time (keeping the energy of binding constant) and change the total length of strand, the saturation time for different lengths shows an exponential increase. Also if we compare final residence time for all 26 bonds for different strands with different E_{bind} we expect for higher binding energy, the longer saturation time.
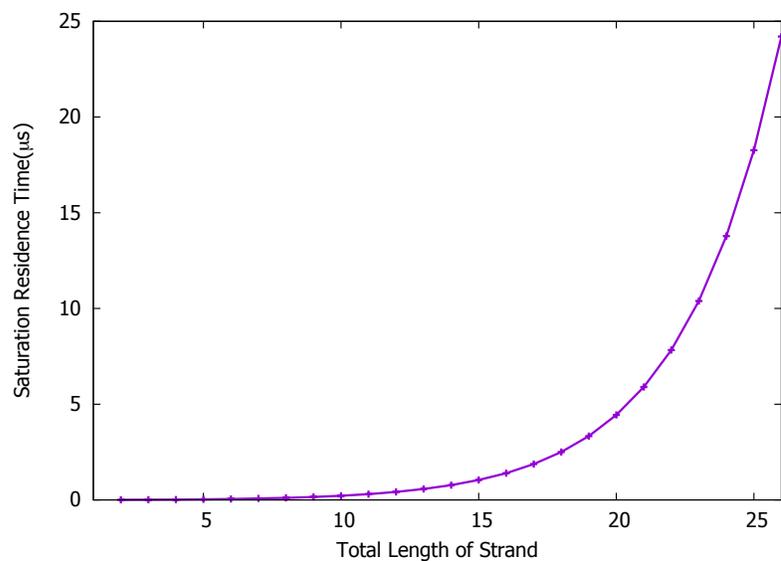
Figure 2.6: The residence time depends exponentially to the molecule length($E_{\text{bind}} = 0.28K_{\text{B}}T$), longer molecule needs more time resulting in an Arhenius dependence.

Also we can have the same plot in logarithmic time scale versus different length of registry 2.7 that clearly shows exponential increase at longer length of registry.
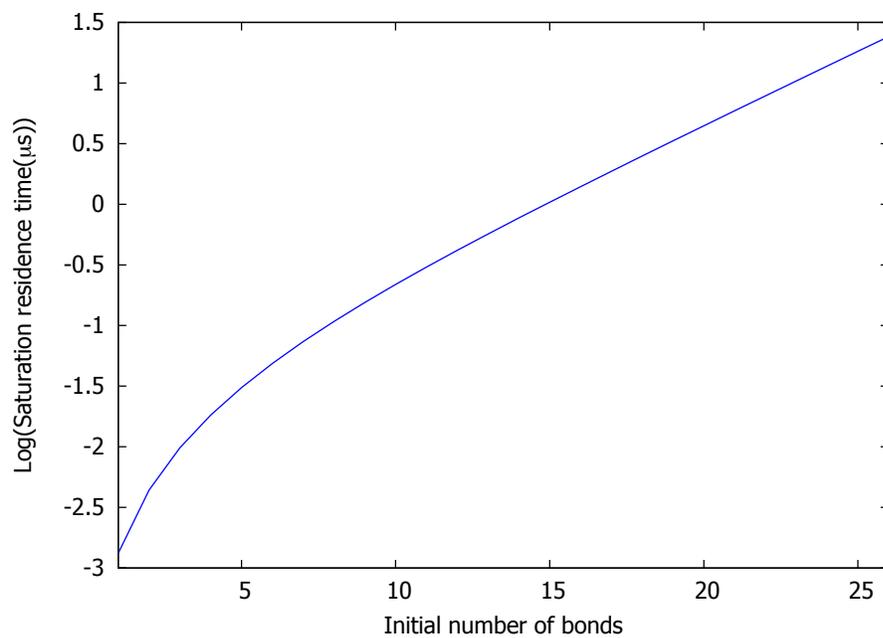
Figure 2.7: Log plot of the residence time showing the exponential dependence on molecule length ($E_{\mathrm{bind}} = 0.28K_{\mathrm{B}}T$).
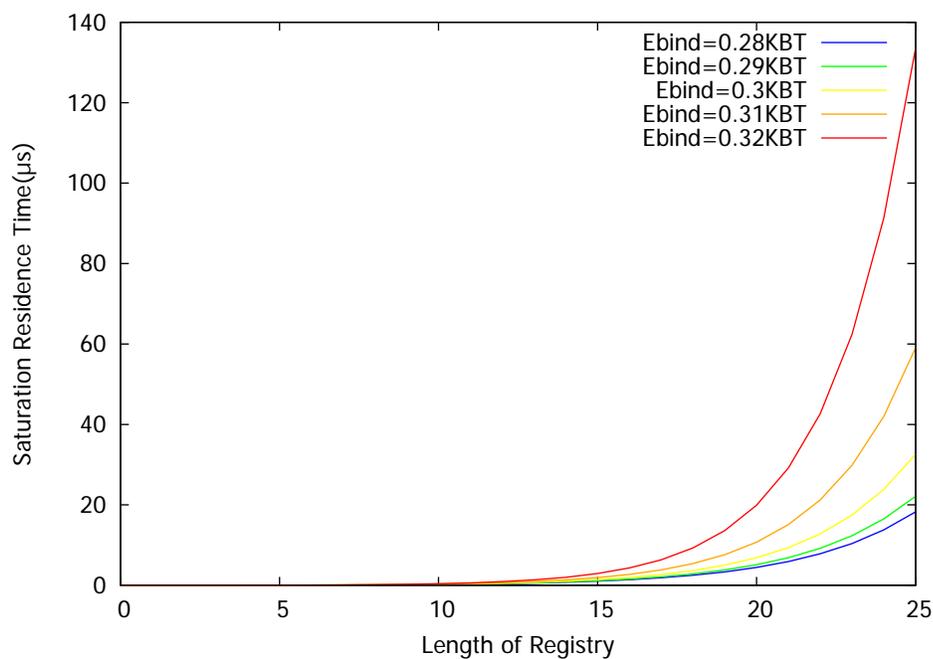
Figure 2.8: Saturation residence time for different molecule lengths and different energies.
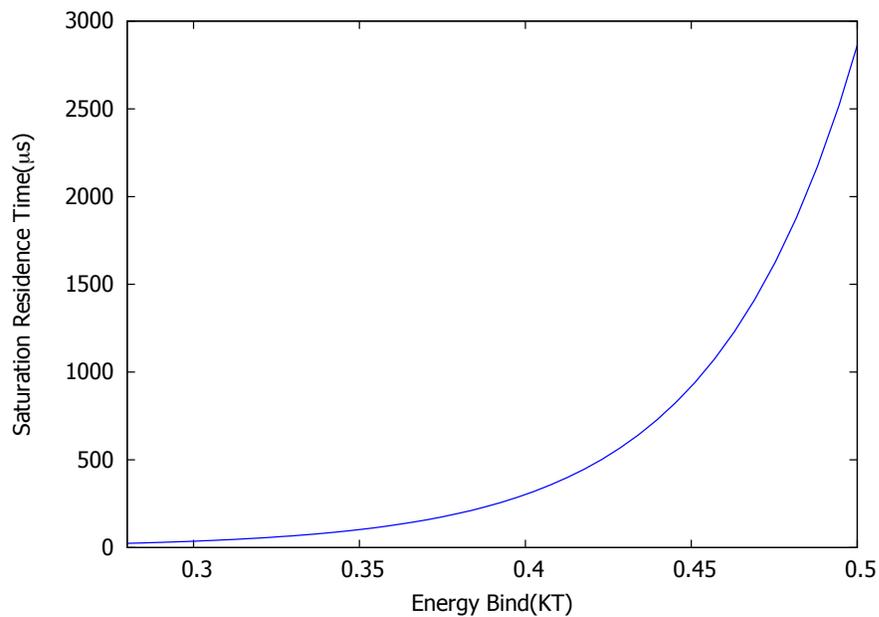
Figure 2.9: Residence time for fully bonded molecules of length L=26 as a function of E_{bind}.

Also we can have the same plot in logarithmic time scale versus different length of registry 2.10 that clearly shows exponential increase at longer length of registry.
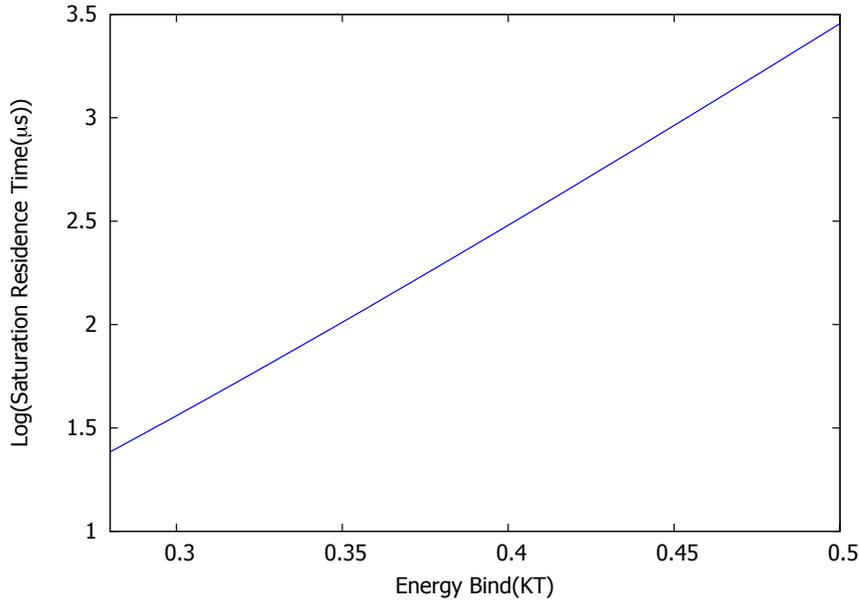
Figure 2.10: Log plot of saturation residence time, showing the exponential dependence on E_{bind}.

For the non-uniform structure, we assumed there are different kind of interactions between three kind of side chains. Strong between hydrophobic and hydrophobic (named HH), and weak between polar and polar (named PP). Figure 2.11 shows how the shift in registry position causes a weak structure with all weak interaction between strands or a stronger structure which is built with combination of weak and strong interactions. In the figure, the red color is representative for hydrophobic (H) and green color for polar (P) and for the interactions blue color is representative for strong interactions and yellow color for the weak interactions. It can be seen that for length of registry equal to three (attachment starts from the head), strands make three weak bonds, but for length of registry equal to two bonds (attachment starts from the head), strands make one strong and one weak bond. This important feature shows longer length of registry doesn't necessarily cause an increase in the time of residence if the structure is more complicated. This claim can be proved with looking at saturation residence times for different length of registries in Figure. 2.11.

There is a strong even-odd effect in the residence time. This is the result of the fact that odd registries prevent the formation of HH contacts. This means that all of the bonds will be weak, resulting in short residence time. For this example we used $E_{\text{bind}} = 0.2k_{\text{B}}T$ for PP and HP interactions and $E_{\text{bind}} = 0.8k_{\text{B}}T$ for HH interaction.
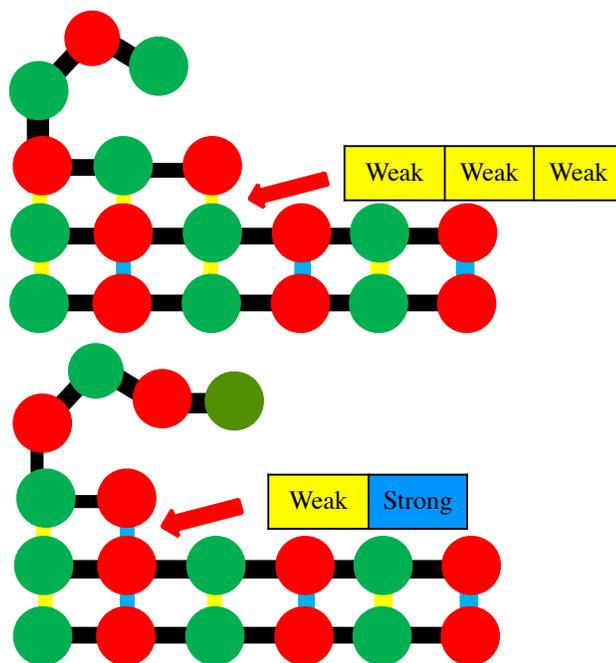
Figure 2.11: Binding energy depends on molecule alignment for patterned sequence. Red color is representative for hydrophobic (H) and green color for polar or charged (P), also for the interactions blue color is representative for strong interactions and yellow color for weak interactions. In the top panel miss-alignment of the H and P residues results in the weak binding energies. In the bottom panel alignment of the H residues results in strong binding.
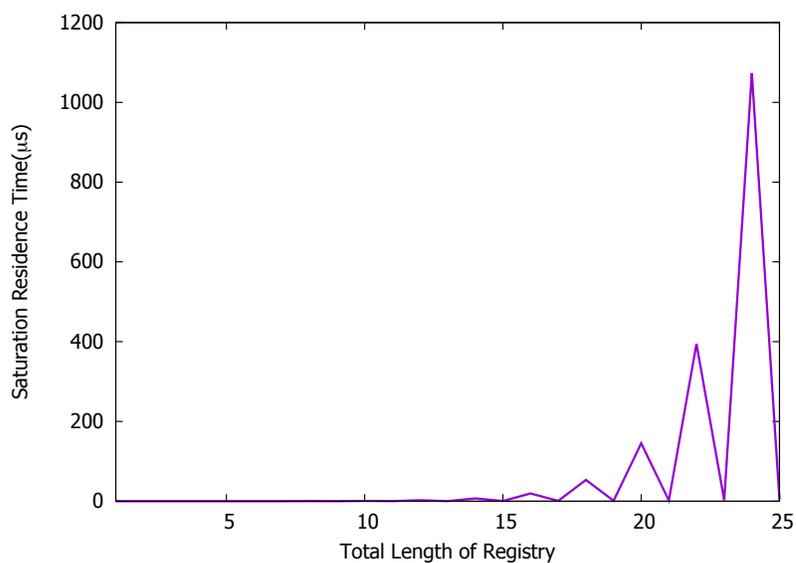
Figure 2.12: Plot of the saturation residence time for dimer structure as a function of the length of registry for structure of figure 2.11. The even-odd effect is a result of the lack of strong bonds with odd registries.

We can repeat the analysis for peptides that have a repeating sequences with every three amino acids. Considering a strand with a "HPP" (hydrophobic-polar-polar) pattern along the strand, see Figure 2.13). Clearly if two strands meet each other in the form that every HPP face with another HPP, the bonds will have a strong-weak-weak pattern. This is more stable than the condition when a HPP is paired with PPH or PHP resulting in all weak interactions.
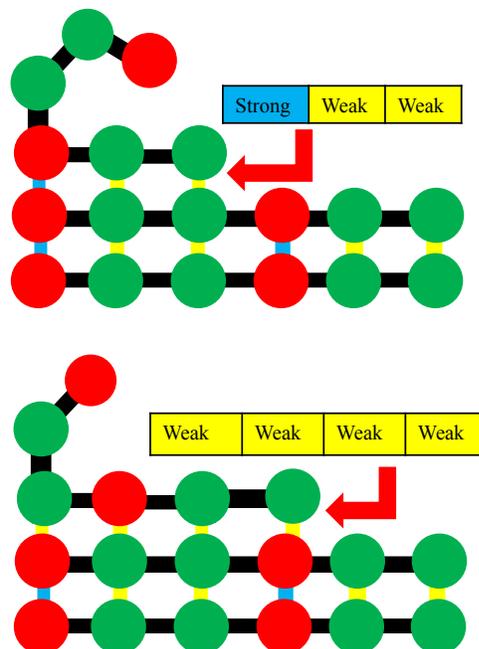
Figure 2.13: Comparison of different lengths of registry for trimer structure. The alignment between the molecule and fibril determines the sequence of strong and weak bonds formed.
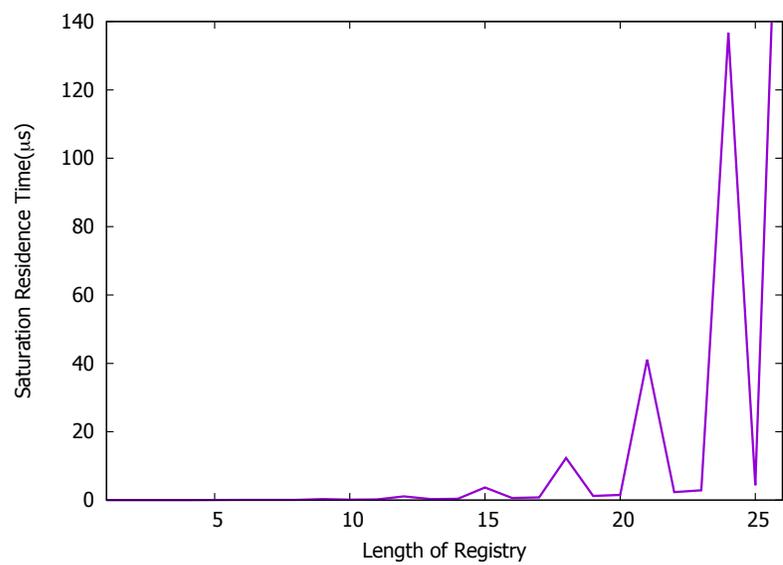
Figure 2.14: Plot of the saturation residence time for trimer structure as a function of the length of registry for structure of figure 2.13.

2.5   Discussion

Our calculations show that polypeptides dissimilar to fibril structure disturb the growth of amyloid fibril. In other words, the existence of polypeptides with the same symmetry as the fibril, gives a better chance for fibril growth. Also, it is obvious that for those fibrils which have more hydrophobic amino acid side chains, there is more chance for growth.

Chapter 3

Results of Simulation for

Protein Aggregation Process

3.1 Definition of the Physical Characteristics and Principles in Simulation

3.1.1 The assumption about diffusion time and concentration

Using the Stoke-Einstein expression, the diffusion constant of a sphere is given by:

$$D = \frac{RT}{6\pi\eta R_{\mathrm{h}}} \tag{3.1}$$

Where RT is the gas constant multiplied by temperature ($R = k_{\mathrm{B}}N_{\mathrm{A}}, N_{\mathrm{A}}$ is Avogadro's number), η is the viscosity of the solvent in a dilute solution and R_{h} is hydration radius of the sphere. Hong L and Lei J, based on an analysis of 37,000 protein structures, found that the radius of native proteins scales with the chain length, according to [26, 27]:

$$R_{\mathrm{n}}(N) = 2.24N^{0.392} \tag{3.2}$$

Where N is the number of amino acids in protein and R_{n} has units of angstrom. However in a different study, Tyn and Gusek demonstrated that the diffusion properties of proteins are approximated by $R_{\mathrm{h}} = 1.45R_{\mathrm{n}}$ [26, 28]. This model is for free and naturally folded protein in

dilute environment. In dense in-vivo environment molecules have more obstacles around for moving in their trajectories. At high protein concentration the crowding effect will increase the effective viscosity of the media, leading to an increase in the diffusion time. This effect is initially linear in the concentration becomes strongly non-linear at higher concentration[29]. To calculate the diffusion time from the diffusion constant we employ the Smoluchowski result for the protein creating a spherical absorbing surface.

$$t_{\text{diff}} = \frac{1}{4\pi a c D_{\text{p}}} \tag{3.3}$$

Where a is the radius of the absorbing surface (the protofilament radius $\simeq 2nm$), c is the concentration far from the surface, and D_{p} is the diffusion constant of the particles [2, 31, 32], In our simulation we use t_{diff} as the independent variable which contains the combined effects of protein concentration on the reduction of diffusion caused by molecular crowding. We model the aggregation process as a series of steps in which the fibril can either add a molecule or lose one. The probability for adding another strand to the aggregate, is

$$P_{\text{add}} = \frac{K_{\text{diff}}}{K_{\text{diff}} + K_{\text{res}}} \tag{3.4}$$

Where K_{res} is $\frac{1}{t_{\text{res}}}$ and t_{res} is the residence time from section 2.1.1 . In other words, if time of residence is long enough and diffusion time is short enough, the chance of association is more likely to grow than shrink.

3.1.2 Assumptions for Considering the Stochastic Nature of the Process

Whether we are investigating formation of hundreds of different kinds of molecules from initial materials or formation of aggregation from similar proteins, the Gillespie algorithm is a reasonable efficient method for determining the outcomes of a "well-stirred" reaction [33]. In this work we consider that there are only two reactions in the experiment, aggregation and dissociation. We assume that the system is large enough that aggregation or dissociation of proteins don't change the concentration of proteins. However as we consider the effect of stochasticity with factor C in the process, so random nature of process still is kept. It is also important to mention that volumetric concerns in Gillespie algorithm for simplification is not taken into account.

Making decision for conditional selection of time steps in simulation can accelerate simulation and lower time costs. This means that considering big time steps when nothing happens or small time steps when the process is happening can make the simulation work much closer to real process and more efficient in simulation. If particular reaction happens with the rate of K it means that the time step should be selected not bigger than this time step. For the case that several reactions are happening selecting sum of all rates can make the smallest leap time and safest selection. Considering that we have two rates K_{res} and K_{diff} (that later will be explained profoundly) will give time step:

$$\tau = \frac{1}{K_{\text{res}} + K_{\text{diff}}} \tag{3.5}$$

However it is better to consider stochasticity in selecting time steps too. Therefore time step is considered as a exponential distribution around this amount:

$$\tau = -\frac{1}{K_{\text{res}} + K_{\text{diff}}} \log(r). \tag{3.6}$$

Where r is uniform random number in the region (0,1]. By this selection the leap time in the simulation is a fraction of the average time for the reaction which derives by two diffusion and residence processes.

3.1.3 Energy Scales

All energy scales are given in units of thermal energy at room temperature (300°K), it has assumed that the strength of H-bonds in the aqueous environment is less than 0.1 of H-bond strength in the gas phase The strength of H-bond, in case of having bond between amino group (N-H) and Oxygen in dilute environment is 0.8 $kJ/mole$ [35].

$$800\frac{\text{J}}{mole} \times \frac{1mole}{6.02 \times 10^{-23}molecule} \times \frac{k_{\text{B}}T}{4.14 \times 10^{-21} \text{ J}} = 3.2\frac{k_{\text{B}}T}{molecule}. \tag{3.7}$$

Experimental data also shows that the stability of $A\beta_{40}$ fibrils with N=24 is $N\epsilon_{\text{b}} = 13.1k_{\text{B}}T$ [36]. So I chose the energy scales to be around $0.32k_{\text{B}}T$ and as the interaction of residues are smaller than hydrogen bonds, it is considered the interactions are weakening or strengthening bonds by 0.1 of this energy scale.

3.2 Definitions of the Physical Assumptions in the Simulation and Results

In this simulation we start with uniform structure. As shown in Figure 3.1, blue arrows show the registry coordinate and green arrows show the direction which aggregation grows. With
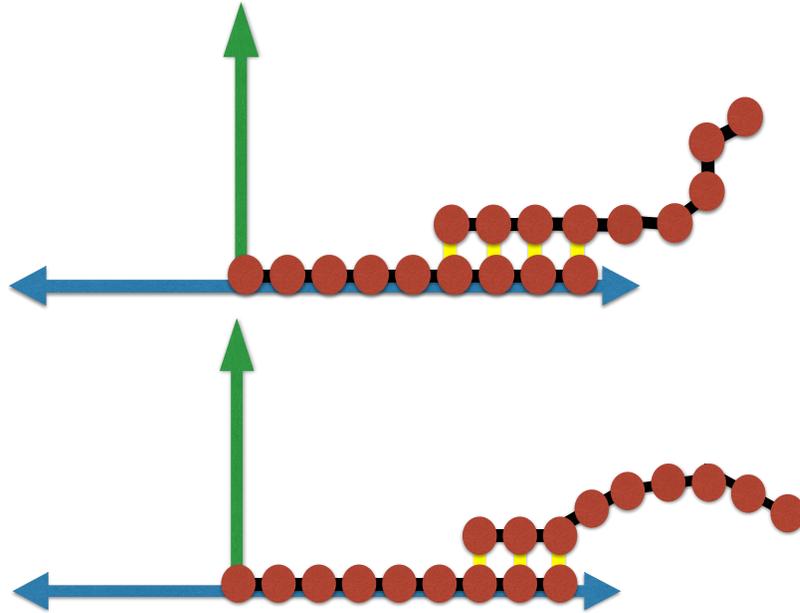
Figure 3.1: Schematic of aggregation process and two important direction in simulation. "Y" direction which is shown with green color and shows the growth of fibril, the "X" direction shows the position of registry. The next strand at the upper section is attached at the position of registry=5 and length of registry=4. In the bottom section of figure, the next strand is attached at the position of registry=6 and the length of registry=3.

the histograms below the effect of concentration K_{diff} on the incorporation of molecules with various registries can be seen. At low concentration only we can see random lengths of registry find the chance of growth and some of them never find this chance, however in higher concentration we see all lengths of registry happens almost with the same amount. This simulation is done with accumulation time $t = 0.01s$. It can be seen that during the increment of concentration (K_{diff}) at low concentration we have only stochastic behavior for attachment of strands at different lengths of registry. However we see that at high concentration, all lengths happen with the same chance.
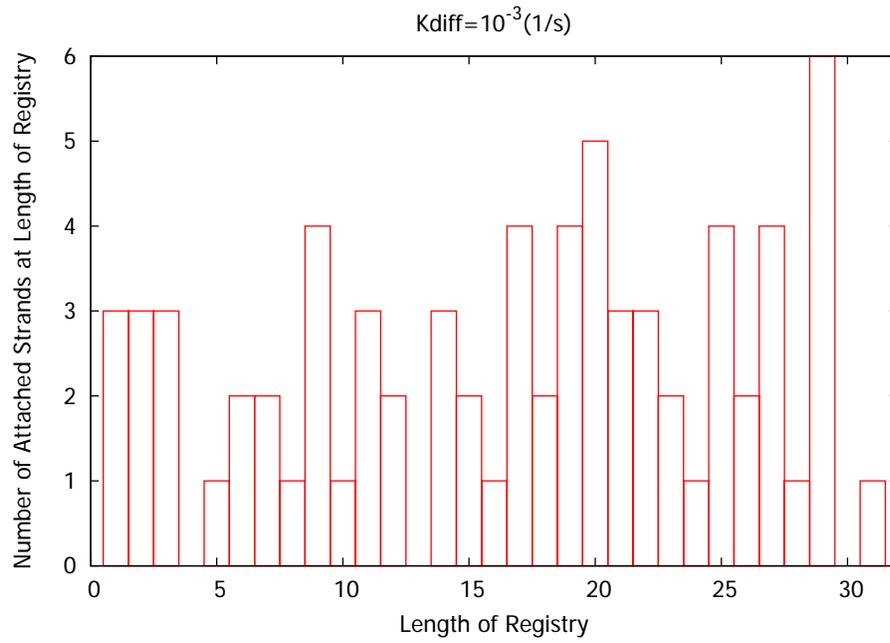
Figure 3.2: Histogram of the number of strands attached at different length of registry at the concentration corresponding to $K_{\text{diff}} = 10^{-3}(\frac{1}{s})$.
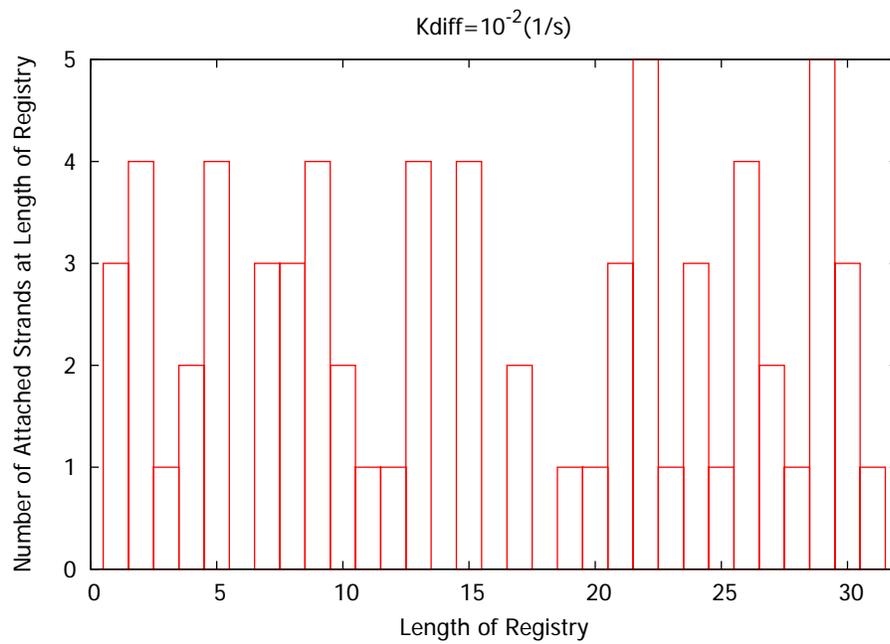
Figure 3.3: Histogram of the number of strands attached at different length of registry at the concentration corresponding to $K_{\text{diff}} = 10^{-2}(\frac{1}{s})$.
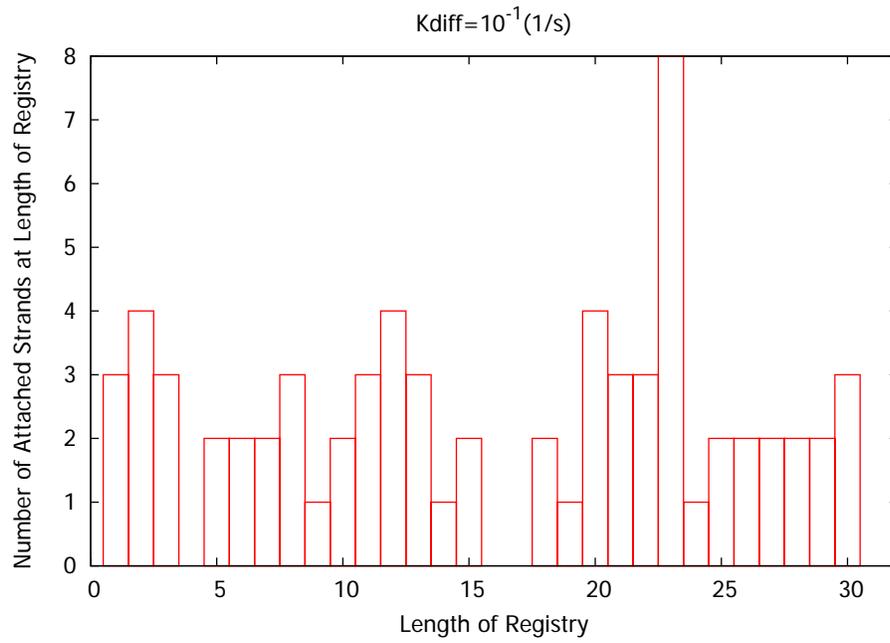
Figure 3.4: SHistogram of the number of strands attached at different length of registry at the concentration corresponding to $K_{\text{diff}} = 10^{-1}(\frac{1}{s})$.
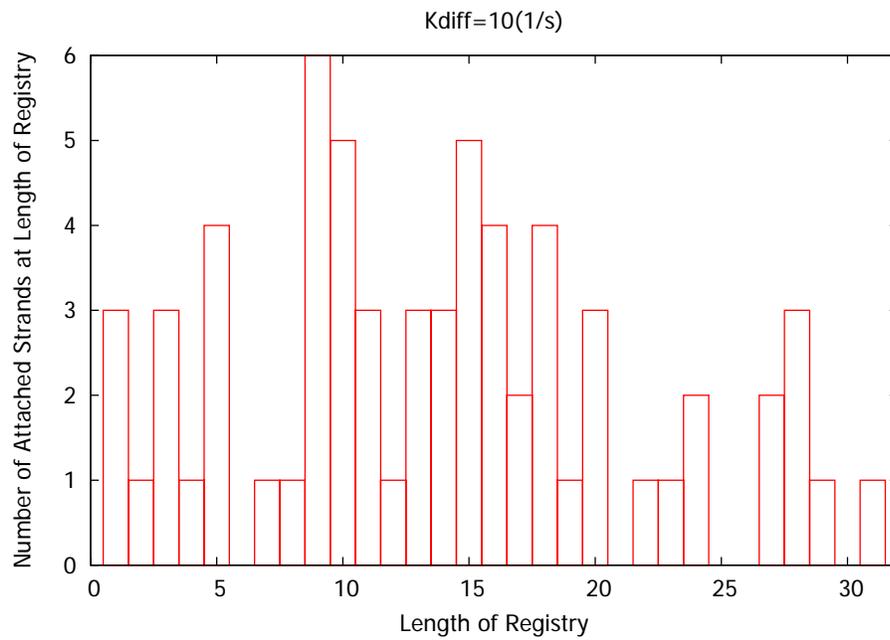
Figure 3.5: Histogram of the number of strands attached at different length of registry at the concentration corresponding to $K_{\text{diff}} = 10(\frac{1}{s})$.
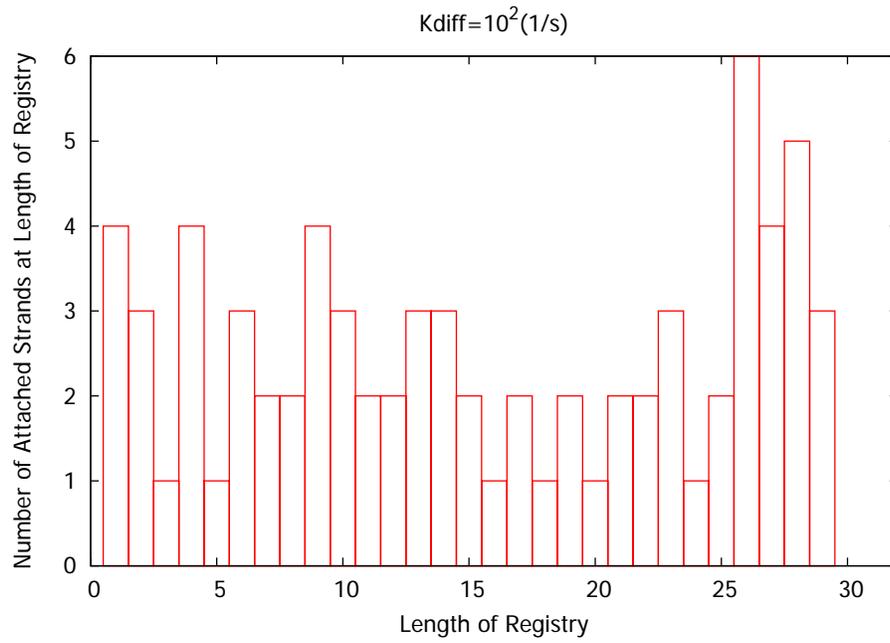
Figure 3.6: Histogram of the number of strands attached at different length of registry at the concentration corresponding to $K_{\mathrm{diff}} = 10^2 (\frac{1}{s})$.
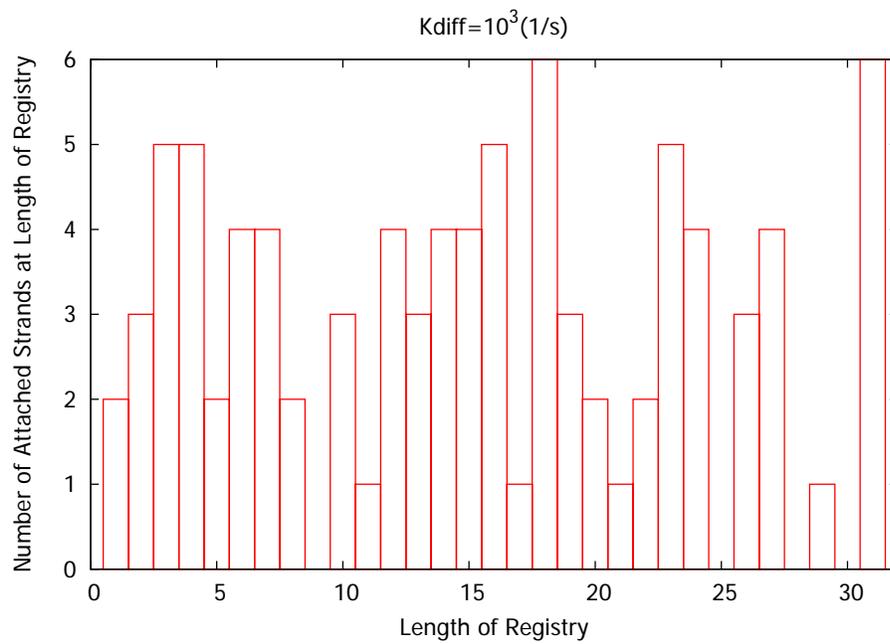
Figure 3.7: Histogram of the number of strands attached at different length of registry at the concentration corresponding to $K_{\mathrm{diff}} = 10^3 (\frac{1}{s})$.
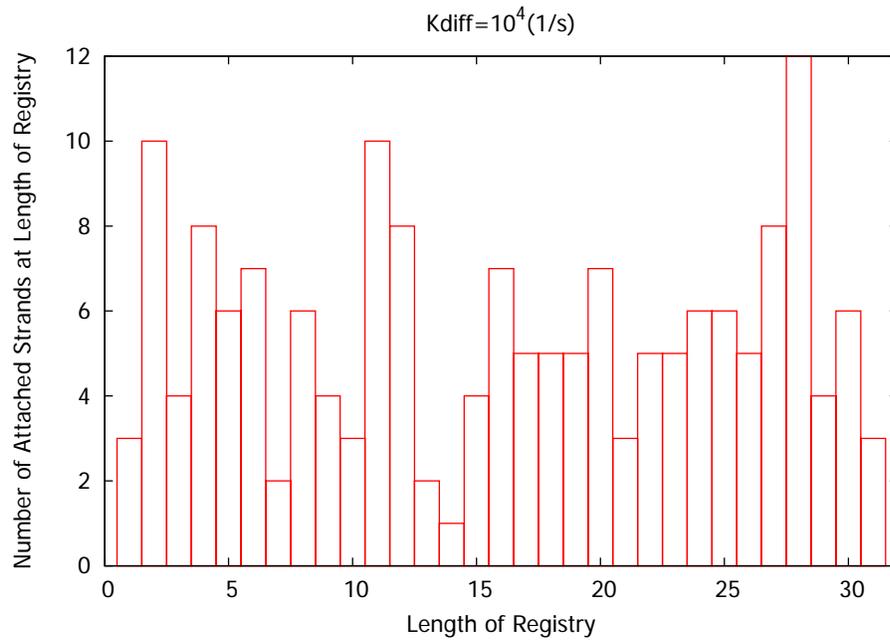
Figure 3.8: Histogram of the number of strands attached at different length of registry at the concentration corresponding to $K_{\text{diff}} = 10^4(\frac{1}{s})$.
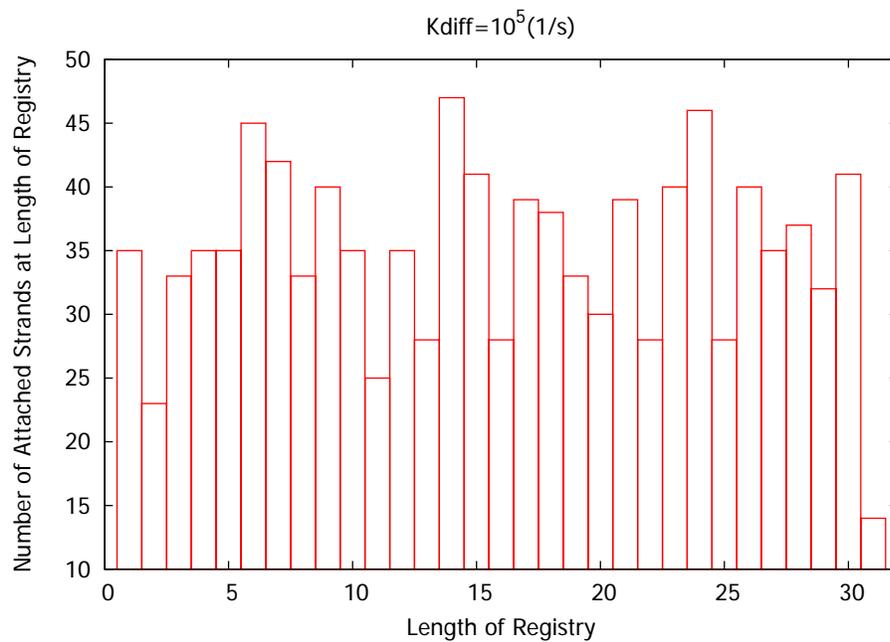
Figure 3.9: Histogram of the number of strands attached at different length of registry at the concentration corresponding to $K_{\text{diff}} = 10^5(\frac{1}{s})$.
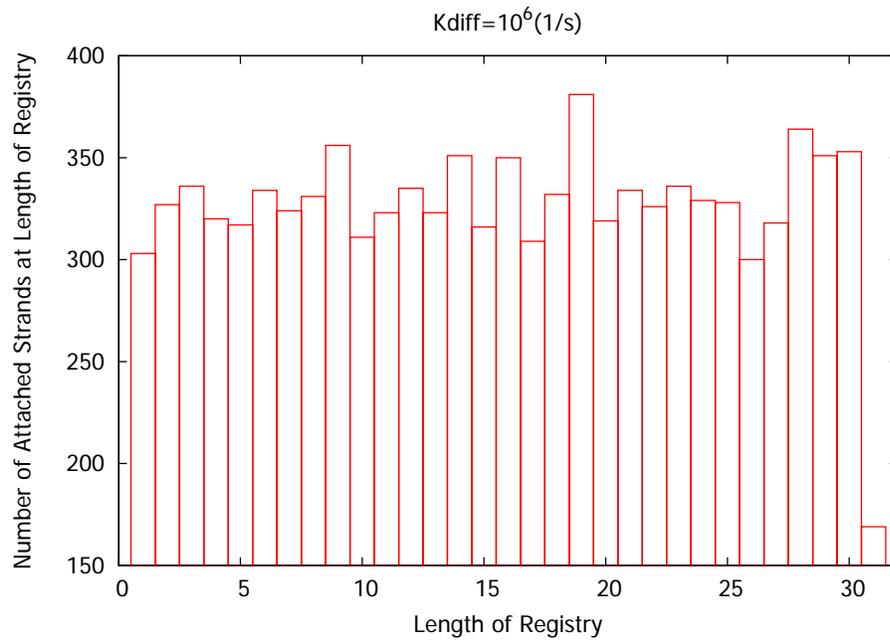
Figure 3.10: Histogram of the number of strands attached at different length of registry at the concentration corresponding to $K_{\text{diff}} = 10^6(\frac{1}{s})$.

Also it is interesting to compute the amount of average of growth versus concentration during 1 million of time steps defined at the section 3.1.2 for 100 runs. As shown in Figure 3.11, at low concentration the stochastic nature of the process can be seen with relative standard deviation (standard deviation/average), but its nature changes to completely deterministic phase at higher concentration.
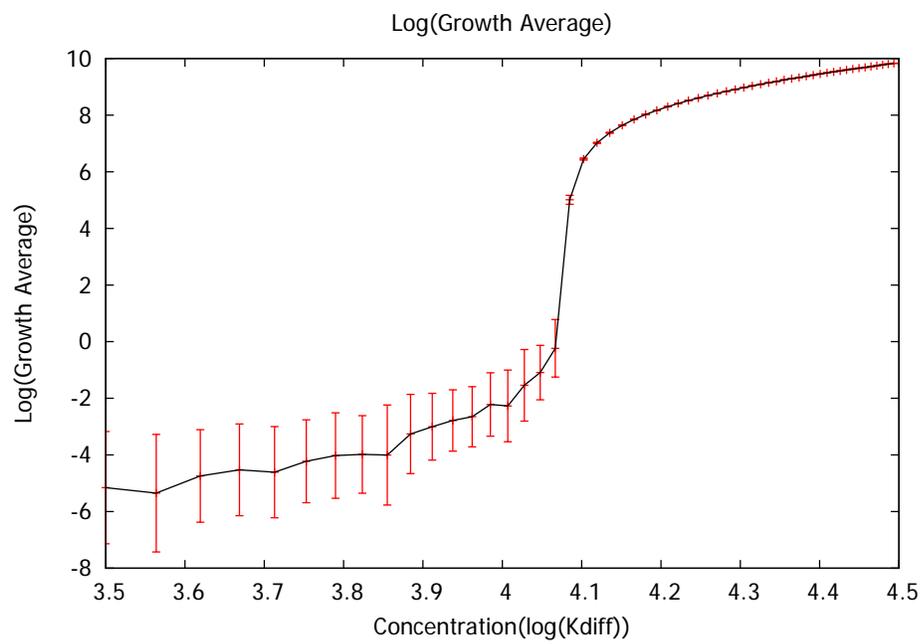
Figure 3.11: Growth average versus concentration for $L = 31$ and $E_{\mathrm{bind}} = 0.28k_BT$.

In giving an idea about the reason for growth, it can be said that two effective phenomena in growth at particular concentration arrive to the point that two probabilities of adding (P_{add}) and loosing (P_{loose}) for a fibril gets equal. This point is the start point for growth. Then at higher concentration, P_{add} is more than P_{loose} and so growth can proceed. For one single step we can show this schematic (figure 3.12). Referring to this figure 3.12, considering a molecule with the length of 30 peptides, and assuming all 30 peptides make bond, the concentration necessary for the start of growth is between 4.1- 4.2 in the graph where the top line corresponding to the residence time of molecule with the length of 30 cut the diffusion time graph. This result is consistent with the results shown in the figure 3.11
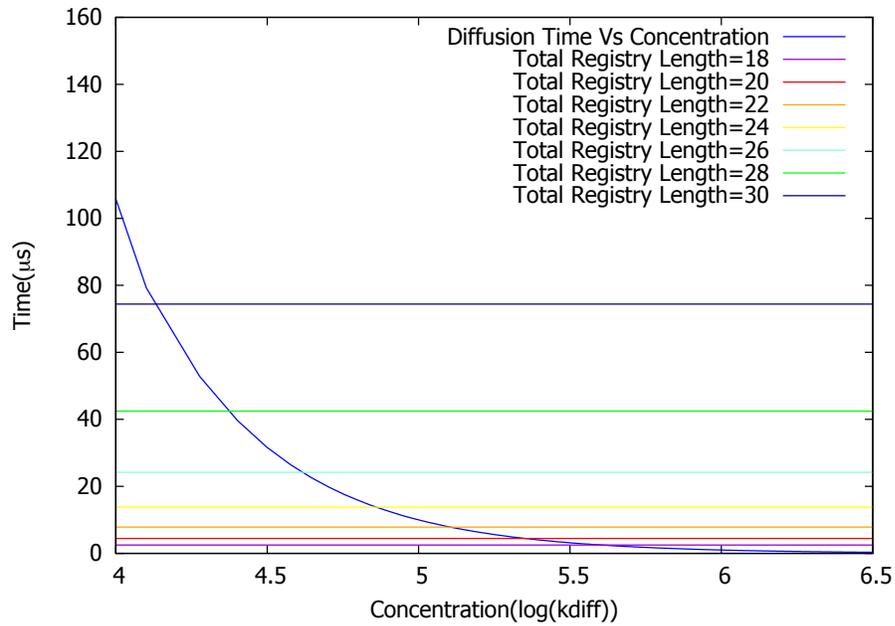
Figure 3.12: Concentration requirement for the start of growth for different length of proteins ($E_{\mathrm{bind}} = 0.28 k_B T$).

Chapter 4

Conclusion and Future Directions

The result of simulation shows completely deterministic behavior at higher concentration. However at low concentration that the process of aggregation starts the behavior is stochastic. This amount of concentration for the start of process depends on the amount of binding energy that we consider. At lower binding energy we need higher concentration for the start of the aggregation process, however the critical concentration decreases with increasing energy for the start of the aggregation process. This simulation is done for uniform structure that doesn't have any complexity and all the bonds have the same binding energy. It is obvious that the non-uniform structure will result in more complex behaviour.

Our suggestion for future works is continuing simulation with non-uniform structure. Although the model neglects many aspects of the aggregation process, we have captured the fundamental process of molecular diffusion and the sampling of incorrect binding states. Refinement of the model in my suggestion should be started from section 3.1.2 where we considered only constructive effect of the concentration. In this work we observed only a direct relationship between the concentration and growth rate, it is also possible to have an inverse relationship when increasing the concentration, it results to decrease in the growth rate. This is possible when proteins with wrong alignments inhibit the binding of properly bound proteins. Although it is said that protein aggregation and in our work amyloid have crystalline structure, it doesn't mean also the environment behaves like crystal and we can't

neglect the effect of alignment in the environment which provides molecules for aggregation. With defining three-dimensional coordinates for this process can be a good exercise that also can include the chance of having several aggregation process at different sides. We can also explore more realistic diffusion model using physical values for $4\pi a D_{\mathrm{p}}$ (taken to be equal to one) and including crowding effects.

Appendix 1

(Part A) Dimer Structure Calculations (Checking important principles)

$$p'_+ + p'_- = 1$$

$$p'_+ = \frac{p_{+w}p_{+s}}{(1 - p_{-w}p_{+s} - p_{+w}p_{-s})}, \tag{1}$$

$$p'_- = \frac{p_{-w}p_{-s}}{(1 - p_{-w}p_{+s} - p_{+w}p_{-s})}, \tag{2}$$

$$p'_+ + p'_- = 1 \rightarrow p_{-w}p_{-s} + p_{+w}p_{+s} = 1 - p_{-w}p_{+s} - p_{+w}p_{-s} \rightarrow$$

$$p_{-w}p_{-s} + p_{+w}p_{+s} + p_{-w}p_{+s} + p_{+w}p_{-s} = 1$$

$$\rightarrow p_{+w}(p_{+s} + p_{-s}) + p_{-w}(p_{+s} + p_{-s}) = p_{+w} + p_{-w} = 1$$

Function x/2 in calculation is appropriate selection:

Having equations:

$$t(x + 2) = \Theta(x + 2) - ((x + 2)/2)\frac{\tau'}{p'_+ - p'_-} \tag{3}$$

$$t(x - 2) = \Theta(x - 2) - ((x - 2)/2)\frac{\tau'}{p'_+ - p'_-} \tag{4}$$

will we have $t(x) = p'_+ t(x + 2) + p'_- t(x - 2) + \tau'$?

$$\rightarrow t(x) = p'_+(\Theta(x + 2) - ((x + 2)/2)\frac{\tau'}{p'_+ - p'_-}) + p'_-(\Theta(x - 2) - ((x - 2)/2)\frac{\tau'}{p'_+ - p'_-}) + \tau'$$

$$\rightarrow t(x) = p'_+(\Theta(x + 2) - ((x + 2)/2)\frac{\tau'}{p'_+ - p'_-}) + p'_-(\Theta(x - 2) - ((x - 2)/2)\frac{\tau'}{p'_+ - p'_-}) + \tau'$$

$$t(x) = p'_+\Theta(x + 2) + p'_-\Theta(x - 2) - x/2\frac{\tau'}{p'_+ - p'_-} - (\frac{p'_+\tau'}{p'_+ - p'_-} - \frac{p'_-\tau'}{p'_+ - p'_-}) + \tau'.$$

where last two terms cancel each other and we have:

$$p'_+\Theta(x + 2) + p'_-\Theta(x - 2) - x/2\frac{\tau'}{p'_+ - p'_-} = t(x). \tag{5}$$

Diagonalization of Transfer Matrix

The matrix M can be brought into diagonal form with:

$$U^{-1}MU \qquad (6)$$

But finding U matrix can be done via mathematica, which gives U:

$$U = \begin{pmatrix} -1 + \frac{1}{p_+} & 1 \\ 1 & 1 \end{pmatrix} \qquad (7)$$

So we can calculate U^{-1}:

$$U^{-1} = \begin{pmatrix} \frac{1}{\frac{1}{p_+}-2} & \frac{-1}{\frac{1}{p_+}-2} \\ \frac{-1}{\frac{1}{p_+}-2} & (\frac{1}{p_+}-1)(\frac{1}{p_+}-2) \end{pmatrix} \qquad (8)$$

$$U^{-1}MU = \begin{pmatrix} -1 + \frac{1}{p_+} & 0 \\ 0 & 1 \end{pmatrix} \qquad (9)$$

And also for checking unitary matrix U:

$$UU^{-1} = \begin{pmatrix} -1 + \frac{1}{p_+} & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\frac{1}{p_+}-2} & \frac{-1}{\frac{1}{p_+}-2} \\ \frac{-1}{\frac{1}{p_+}-2} & (\frac{1}{p_+}-1)(\frac{1}{p_+}-2) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad (10)$$

(Part B)Trimer Structure Calculations(Checking important principles)

$p'_+ + p'_- = 1$

$$p'_+ + p'_- = 1,$$

$$p_{-1}p_{-2}p_{-3} + p_{+1}p_{+2}p_{+3} = 1 - (p_{-2}p_{+1} + p_{-3}p_{+2} + p_{-1}p_{+3}),$$

$$p_{-1}p_{-2}p_{-3} + p_{+1}p_{+2}p_{+3} + p_{-2}p_{+1} + p_{-3}p_{+2} + p_{-1}p_{+3} = 1,$$

$$(1 - p_{+1})p_{-2}p_{-3} + p_{+1}p_{+2}p_{+3} + p_{-2}p_{+1} + p_{-3}p_{+2} + p_{-1}p_{+3}$$

$$= p_{-2}p_{-3} - p_{+1}p_{-2}p_{-3} + p_{+1}p_{+2}p_{+3} + p_{-2}p_{+1} + p_{-3}p_{+2} + p_{-1}p_{+3}$$

$$= (p_{-2} + p_{+2})p_{-3} - p_{+1}p_{-2}p_{-3} + p_{+1}p_{+2}p_{+3} + p_{-2}p_{+1} + p_{-1}p_{+3}$$

$$= p_{-3} + (1 - p_{-3})p_{+1}p_{-2} + p_{-1}p_{+3} + p_{+1}p_{+3}(1 - p_{-2})$$

$$= p_{-3} + p_{+3}p_{+1}p_{-2} + p_{-1}p_{+3} + p_{+1}p_{+3} - p_{-2}p_{+1}p_{+3}$$

$$= p_{-3} + p_{+3}(p_{-1} + p_{+1}) = 1.$$

Function x/3 in calculation is appropriate selection:

Having equations:

$$t(x+3) = \Theta(x+3) - ((x+3)/3)\frac{\tau'}{p'_+ - p'_-} \tag{11}$$

$$t(x-3) = \Theta(x-3) - ((x-3)/3)\frac{\tau'}{p'_+ - p'_-} \tag{12}$$

will we have $t(x) = p'_+ t(x+3) + p'_- t(x-3) + \tau'$?

$$\rightarrow t(x) = p'_+(\Theta(x+3) - ((x+3)/3)\frac{\tau'}{p'_+ - p'_-}) + p'_-(\Theta(x-3) - ((x-3)/3)\frac{\tau'}{p'_+ - p'_-}) + \tau'$$

$$\rightarrow t(x) = p'_+(\Theta(x+3) - ((x+3)/3)\frac{\tau'}{p'_+ - p'_-}) + p'_-(\Theta(x-3) - ((x-3)/3)\frac{\tau'}{p'_+ - p'_-}) + \tau'$$

$$\rightarrow t(x) = p'_+\Theta(x+3) + p'_-\Theta(x-3) - x/3\frac{\tau'}{p'_+ - p'_-} - (\frac{p'_+\tau'}{p'_+ - p'_-} - \frac{p'_-\tau'}{p'_+ - p'_-}) + \tau'$$

where last two terms cancel each other and we have:

$$p'_+\Theta(x+3) + p'_-\Theta(x-3) - x/3\frac{\tau'}{p'_+ - p'_-} = t(x).$$

Bibliography

[1] http://en.wikipedia.org

[2] J. Schmit,"Kinetic theory of amyloid fibril templating", The journal of Chemical Physics, 138, 185102, 2013.

[3] R. S. Harrison, P. C. Sharpe, Y. Singh and D. P. Fairlie,"Reviews of Physiology, Biochemistry and Pharmacology", no 159 , 2007

[4] Joerg Gsponer and Michele Vendruscolo,"Theoretical Approaches to Protein Aggregation, Protein and Peptide Letters, no 13, 287-293, 2006

[5] Antonio Travato, Fabrizio Chiti, Amos Maritan, Flavio Seno, "Insight into the Structure of Amyloid Fibrils from the Analysis of Globular Proteins", Plos Computational Biology, Volume 2, Issue 12, e170, 2006.

[6] Roma N. Rambaran and Louise C. Serpell, "Review of Amyloid fibrils" , Landes Bioscience, 112-117, 2008

[7] Lomakin A, Chung DS, Benedek GB, et al On "the nucleation and growth of amyloid beta protein fibrils: detection of nuclei and quantitation of rate constants". Proc Natl Acad Sci USA 1996; 93:1125 9

[8] Ding TT, Harper JD. "Analysis of amyloid beta assemblies using tapping mode atomic force microscopy under ambient conditions". Vol 309: Academic press; 1999:510-25.

[9] Roher AE, Baudry J, Chaney MO, et al. "Oligomerizaiton and fibril asssembly of the amyloid beta protein". Biochim Biophys Acta 2000; 1502:31-43.

[10] Harper JD, Wong SS, Lieber CM, et al. "Assembly of A beta amyloid protofibrils: an in vitro model for a possible early event in Alzheimer's disease". Biochemistry 1999; 38:8972-80.

[11] Tom Walz, The Laboratory of Molecular Electron Microscopy, Harvard Medical School, Department of Cell Biology.

[12] Aneika C. Leney, "Insights into the role of the beta-2 microglobulin D-strand in amyloid propensity revealed by mass spectrometry", Mol. BioSyst., 2014, 10, 412

[13] Ken Dill, "Theory for the folding and stability of globular proteins", Bichemistry 1985 Mar 12;24(6):1501-9.

[14] TrevinoSR,ScholtzJM,PaceCN, "Amino Acid Contribution to Protein Solubility:Asp, Glu and Ser Contribute more Favorably than the other Hydrophilic Amino Acids in RNase", Biol 2007; 366:449-60.

[15] Chiti F, Dobson CM. Protein Misfolding,"Functional Amyloid and Human Disease. Annu Rev", Biochem 2006; 75:333-66.

[16] Creighton TE. "Proteins: Structures and Molecular Properties", W.H. Freeman and Company 1993.

[17] Xiaoling Wang, Tapan K. Das, Satish K. Singh and Sandeep Kumar, "Report on potential aggregation prone regions in biotherapeutics, A survey of commercial monoclonal antibodies", Landes Bioscience, mAbs 1:3, 254-267, 2009.

[18] Mario Pineda-Krch, "GillespieSSA: Implementing the Stochastic Simulation Algorithm in R", Volume 25, Issue 12. 2008

[19] Tae-Hyuk Ahn, Yang Cao, and Layne T. Watson, "Stochastic Simulation Algorithms for Chemical Reactions", Departments of Computer Science and Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0106

[20] Richard P Sear, "Nucleation: theory and applications to protein solutions and colloidal suspensions", J. Phys.: Condens. Matter 19 033101, (2007)

[21] http://www.studyblue.com

[22] S. Whitelam, R. Schulman, and L. Hedges, Phys. Rev. Lett. 109, 265506 26 (2012).

[23] Sidney Render, "A Guide to First Passage Processes", Section 2.4, Cambridge University Press, 2008

[24] http://pdb101.rcsb.org/motm/189

[25] University of Georgia, "Railsback's Some Fundamentals of Mineralogy and Geochemistry".

[26] Ken Dill, 2011, PNAS vol 108, no 44, 17876-17882, "Physical Limits of Cells and Proteomes".

[27] Hong L, Lei J (2009) Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity. J Poly Sci B 47:207?214.

[28] Tyn MT, Gusek TW, 1990, Biotechnol Bioeng 35:327-338, "Prediction of Diffusion Coefficients of Proteins".

[29] Grinnell Jones and Samuel Talley, 1934, Contribution from the mallinckrodt chemical laboratory of harvard university, "The Viscosity of Aqueous Solutions as a Function of the Concentration".

[30] Francisco J. Sevilla1 and Mario Sandoval, "Smoluchowski Diffusion Equation for Active Brownian Swimmers", arXiv:1501.07237v2, February 3, 2015

[31] J. L. Jimenez, E. J. Nettleton, M. Bouchard, C. V. Robinson, C. Dobson, 29and H. R. Saibil, Proc. Natl. Acad. Sci. U.S.A. 99, 9196 (2002).

[32] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, and R. Tycko, Proc. Natl. Acad. Sci. U.S.A. 99, 16742 (2002).

[33] Gillespie J. Phys. Chem. 81, 2340, 1977.

[34] Maic R. Roussel, "Lectures on Modeling Biochemical Reaction Network", Lecture 14, Department of Chemistry and Biochemistry, University of Lethbridge.

[35] http://www.wikipedia.org

[36] J. D. Schmit, K. Ghosh, and K. A. Dill, Biophys. J. 100, 450 (2011).