



Article

# Assembly Mechanism for Aggregation of Amyloid Fibrils

Lingyun Zhang

Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China; lyzhang@iphy.ac.cn

Received: 15 June 2018; Accepted: 19 July 2018; Published: 23 July 2018



**Abstract:** The assembly mechanism for aggregation of amyloid fibril is important and fundamental for any quantitative and physical descriptions because it needs to have a deep understanding of both molecular and statistical physics. A theoretical model with three states including coil, helix and sheet is presented to describe the amyloid formation. The corresponding general mathematical expression of  $N$  molecule systems are derived, including the partition function and thermodynamic quantities. We study the equilibrium properties of the system in the solution and find that three molecules have the extreme value of free energy. The denaturant effect on molecular assemble is also discussed. Furthermore, we apply the kinetic theories to take account of the nucleation and growth of the amyloid in the solution. It has been shown that our theoretical results can be compared with experimental results.

**Keywords:** amyloid fibrils; assembly mechanism; partition function; free energy

## 1. Introduction

The aggregation of amyloid fibrils in biological processes is associated with neurodegenerative diseases such as Alzheimer's, Parkinson's and Huntington's or prion diseases [1–3]. Despite the specificity of the proteins related with each individual neurodegenerative disease, these kinds of diseases are a nucleation process [4–9]. Under experimental conditions *in vitro*, the aggregation pathway can be obtained. However, it is not clear how to extrapolate these results to identify the dominant pathway and timescales under physiological conditions.

The current computation technique is unable to access even the accelerated timescales of the *in vitro* systems. Some computational techniques could be used to predict the assembly of amyloid in solution and their secondary structure changes [10,11]. However, these computational simulations are only feasible for millisecond time scale. The most simplistic physical description of proteins is analogy with colloidal particles. The random-coil like proteins exist in an unfolded state and the helix is very similar to folded state. These simple models have been used to explore the nucleation processes of amyloid fibrils. There are two classes of nucleation theories, one is the mass action theories, the other is nucleation models [12–21]. However, these methods missed the internal dynamics of the protein molecules. In this paper, a microscopic model of the assembly process is developed to explore the mechanism of amyloid fibrils and explain the transitions between the various assembly pathways as well as how side chain interactions determine the sheet structure in the aggregate phase. Based on the microscopic model, molecular equilibrium states and kinetic equations have been constructed to probe the physical properties of amyloid formation. The paper is organized as follow. Firstly, a single peptide molecule in the solution can be transited from a coil state to a helix state. Then two molecules will be formed by concentration in the solution that supplies a driving force. The combining force is resulted from the hydrogen bond. Based on the structure of two molecules, three molecules can be constructed with the aid of solution concentration. Furthermore,  $N$  molecule systems are able to be constructed,

which is called  $\beta$ -sheet. The corresponding general mathematical description of  $N$  molecule systems are derived, including the partition function and thermodynamic quantities. Then we study the equilibrium properties of the system in the solution. The phase diagrams of assembly structures are depicted. Furthermore, we employ the kinetic theories to study the amyloid formation. Free energy landscape and side chain effect can be illustrated. The theoretical predictions are in agreement with the experimental results.

## 2. Results

### 2.1. Single Molecule

The conformations in single molecule have sequences of helix and coil units [22]. The coil state is an ensemble of disordered conformations which often exists for higher temperature, otherwise, the helix state exists for lower temperature [23]. Therefore, the temperature determines the transition between coil and helix. The experimental temperature ranges from 10 to 60 °C [24]. This structural transition is a cooperative behavior [25,26].

In order to provide a valuable intuition into the nucleation process, we use a similar lattice model and suppose a single molecular chain has  $N$  units, each unit along the chain can be in either of the two states,  $H$  (helix) or  $C$  (coil). According to the ZB model [27,28], only the nearest neighbor interaction is considered, which is a modified one-dimensional Ising model. Therefore, there are four statistical weights, their matrix elements are  $m_{11} = m(C|C)$ ,  $m_{12} = m(H|C)$ ,  $m_{21} = m(C|H)$ ,  $m_{22} = m(H|H)$ . A propagation parameter  $s$  can be defined by the free energy change,

$$s = \exp[-\beta(F_{helix} - F_{coil})] \quad (1)$$

where  $s$  is slightly larger than 1, so the matrix element  $m(H|H) = s$ . The other parameter  $\sigma$  is as a nucleation parameter. In general,  $\sigma \ll 1$  and takes from  $10^{-3}$  to  $10^{-4}$ , so  $\sigma s$  is for helix next to coil, i.e.,  $m(H|C) = \sigma s$ . On the other hand, one uses the statistical weight 1 for any coil, and  $m(C|C) = m(C|H) = 1$ . Then a matrix form of statistical weight can be defined as

$$M = \begin{pmatrix} 1 & \sigma s \\ 1 & s \end{pmatrix} \quad (2)$$

Furthermore, the partition function can be obtained by summing over all the possible sequences in the one-dimensional chain with  $N$  monomers,

$$Z_N = \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot M^N \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (3)$$

Let  $T$  be a transformation matrix to get a diagonalize matrix  $\Lambda$ ,

$$T^{-1}MT = \Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \quad (4)$$

where  $\lambda_1, \lambda_2$  are the eigenvalues of  $M$ .

$$\lambda_{1,2} = \frac{1}{2}[(1+s) \pm \sqrt{(1-s)^2 + 4s\sigma}] \quad (5)$$

By employing  $M^N = (TT^{-1}MTT^{-1})^N = T\Lambda^N T^{-1}$ , the partition function can be rewritten as

$$Z_N = \frac{1}{\lambda_1 - \lambda_2} [\lambda_1^{N+1}(1 - \lambda_2) - \lambda_2^{N+1}(1 - \lambda_1)] \quad (6)$$

Because of  $\lambda_1 > \lambda_2$ , we have  $\lambda_1^{N+1} \gg \lambda_2^{N+1}$  for large  $N$ , then

$$Z_N \approx \lambda_1^N \quad (7)$$

Based on  $\ln Z_N = N \ln \lambda_1$ , the average number of helix states can be computed by

$$\langle h \rangle = \frac{\partial \ln Z_N}{\partial \ln s} = N \frac{\partial \ln \lambda_1}{\partial \ln s} \quad (8)$$

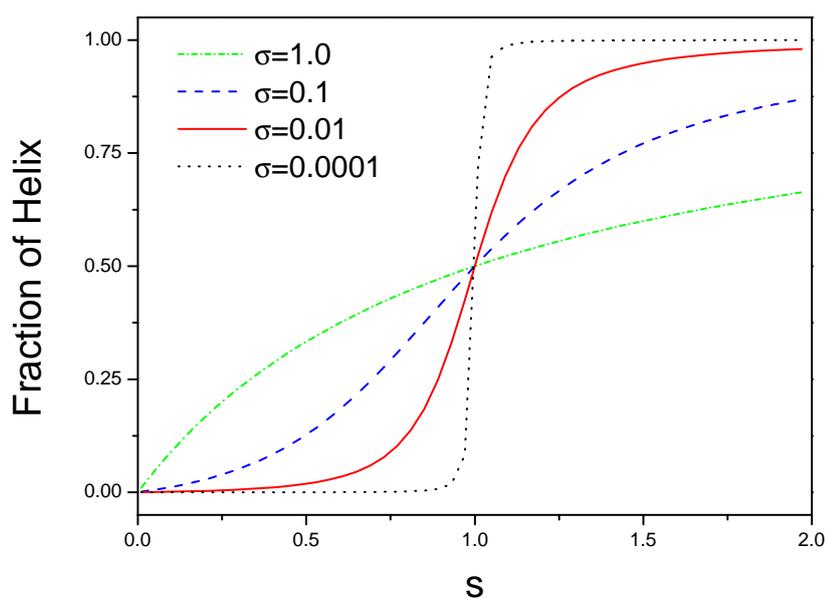
and the fraction of helix states  $\theta_h$  is defined as

$$\theta_h = \frac{\langle h \rangle}{N} = \frac{s}{2\lambda_1} \left\{ 1 + \frac{[(s-1) + 2\sigma]}{\sqrt{(1-s)^2 + 4\sigma s}} \right\} \quad (9)$$

The free energy for the single molecule can approximately be expressed as

$$F^{(1)} = -k_B T \ln Z_N = -N k_B T \ln \lambda_1 \quad (10)$$

We plot  $\theta_h$  as a function of  $s$  for different  $\sigma$  as shown in Figure 1. There are two limitation cases for  $\sigma$ -value. For small  $\sigma$ , a sudden transition can occur for a narrow range of  $s$ , this is a helix-coil structural transition. For large  $\sigma$ , there is a non-cooperative behavior because  $\theta_h$  is equal to  $s/(1+s)$  for  $\sigma = 1$ .



**Figure 1.** Phase diagrams of structural transition from coil to helix for different interactions. The number of helical state in the block is plotted against  $s$ , where  $s$  is an equilibrium constant for a coil state converting into a helical state.

## 2.2. Two Molecules

There are two cases for two molecules, one is no interaction between two molecules, and the other case is the interaction between two molecules. The partition function can be written as

$$Z^{(2)} = \sum_{\{m_1, m_2\}} \Omega^{(2)}(m_1, m_2) Z^{(2)}(m_1, m_2) \quad (11)$$

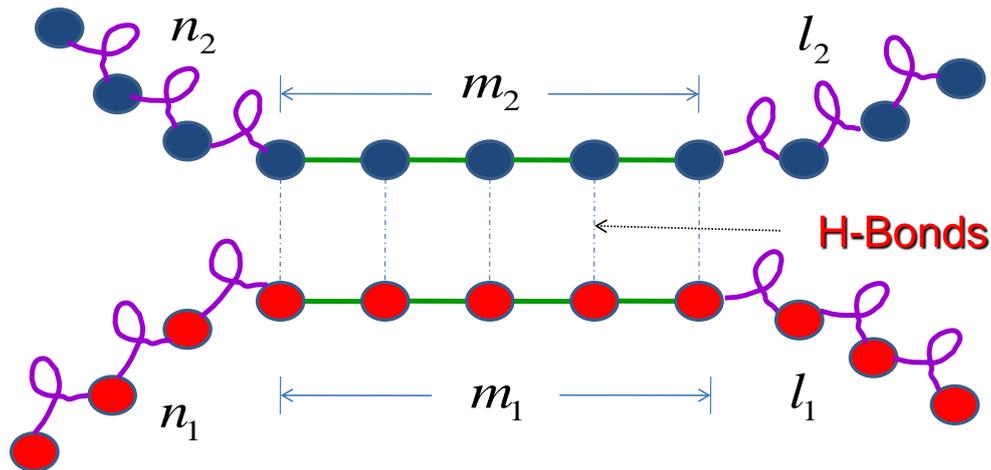
where  $\Omega^{(2)}(m_1, m_2)$  is the degeneracy factor,

$$\Omega^{(2)}(m_1, m_2) = (N - m_1)(N - m_2) \quad (12)$$

which represents that  $N - m_1$  hydrogen bonds in the first molecular chain and  $N - m_2$  hydrogen bonds in the second molecular chain. The other factor in the above  $Z^{(2)}(m_1, m_2)$  is

$$Z^{(2)}(m_1, m_2) = Z_{n_1} Z_{l_1} g_2^{m_2} Z_{n_2} Z_{l_2} \tag{13}$$

where  $g_2$  is the interaction parameter between the first molecule and the second molecule, which results from the hydrogen bond,  $m_1$  and  $m_2$  are the total number of hydrogen bond of molecule 1 and molecules 2 respectively.  $Z_{n_1}$ ,  $Z_{l_1}$ ,  $Z_{n_2}$  and  $Z_{l_2}$  are the partition functions for each segment as shown in Figure 2.



**Figure 2.** Schematic illustration of the monomer attachment. Two molecules through intermolecular hydrogen bond form a simplistic helix-coil-sheet model.  $m_1$  and  $m_2$  are the total number of hydrogen bonds for molecule 1 and molecule 2 respectively.

By defining the parameters  $\theta_1$  and  $\theta_2$  to demonstrate the ratio of non-interaction,

$$\begin{aligned} \theta_1 &= \frac{Z_{n_1} Z_{l_1}}{Z_N} \\ \theta_2 &= \frac{Z_{n_2} Z_{l_2}}{Z_N} \end{aligned} \tag{14}$$

then the partition function of the two molecules can be rewritten as

$$Z^{(2)} = \sum_{\{m_1, m_2\}} (N - m_1)(N - m_2) Z_N^2 \theta_1 g_2^{m_2} \theta_2 \tag{15}$$

Actually, we are able to discuss how chain length influences the partition function. According to the derivation for the single molecule, the partition function can be simplified as

$$Z_{n_1} \approx \lambda_1^{n_1} \quad Z_{l_1} \approx \lambda_1^{l_1} \tag{16}$$

then  $\theta_1$  is

$$\theta_1 = \lambda_1^{n_1 + l_1 - N} = \lambda_1^{-m_1} \tag{17}$$

Due to the same reason,  $\theta_2$  is

$$\theta_2 = \lambda_1^{-m_2} \tag{18}$$

thus we have

$$Z^{(2)}(m_1, m_2) = Z_N^2 \lambda_1^{-m_1 - m_2} g_2^{m_2} \tag{19}$$

In terms of  $m_1 = m_2 = m$ , we obtain

$$Z^{(2)}(m) = Z_N^2 \left(\frac{g_2}{\lambda_1^2}\right)^m \quad (20)$$

Therefore, the partition function of two molecules can be rewritten as

$$Z^{(2)} = Z_N^2 \sum_{m=1}^M (N-m)^2 \left(\frac{g_2}{\lambda_1^2}\right)^m \quad (21)$$

where  $M$  is the total number of the hydrogen bond, its maximum value is  $M_{max} = N$ .

A function can be defined as

$$G_2(m, x_2) = x_2^m \quad (22)$$

where  $x_2 = g_2/\lambda_1^2$ . The other function can be defined as

$$f^{(2)}(M, x_2) = \frac{Z^{(2)}}{Z_N^2} = \sum_{m=1}^M (N-m)^2 G_2(m, x_2) \quad (23)$$

The free energy of two molecules is obtained by

$$F^{(2)} = -k_B T \ln Z^{(2)} = 2F^{(1)} - k_B T \ln f^{(2)}(M, x_2) \quad (24)$$

### 2.3. Three Molecules

Now we add the third molecule to the dimer to form a trimer. The partition function for this system can be represented as

$$Z^{(3)} = \sum_{m_3 \leq m_2} \sum_{m_2} \Omega^{(3)}(m_2, m_3) Z^{(3)}(m_2, m_3) \quad (25)$$

where the degeneracy factor for three molecules

$$\Omega^{(3)}(m_2, m_3) = 2(m_2 - m_3)(N - m_2)^2 \quad (26)$$

and

$$\begin{aligned} Z^{(3)}(m_2, m_3) &= Z^{(2)}(m_2) g_3^{m_3} Z_{n_3} Z_{l_3} \\ &= Z_N^3 \left[ \frac{Z^{(2)}(m_2)}{Z_N^2} g_3^{m_3} \frac{Z_{n_3} Z_{l_3}}{Z_N} \right] \end{aligned} \quad (27)$$

where  $Z_{n_3}$  and  $Z_{l_3}$  are the partition function for the segments of  $n_3$  and  $l_3$  respectively.

Furthermore, we can define  $\theta_3 = Z_{n_3} Z_{l_3} / Z_N$ . In a similar manner,  $\theta_3 \approx \lambda_1^{-m_3}$ , we have

$$Z^{(3)}(m, m_3) = Z_N^3 G_2(m, x_2) \left(\frac{g_3}{\lambda_1}\right)^{m_3} \quad (28)$$

We define a new function

$$G_3(m_3, m, x_3, x_2) = G_2(m, x_2) x_3^{m_3} \quad (29)$$

where  $x_3 = g_3/\lambda_1$ . Then the partition function of three molecules can be rewritten as

$$Z^{(3)} = 2Z_N^3 \sum_{m_3 \leq m} \sum_m (m - m_3)(N - m)^2 G_3(m_3, m, x_3, x_2) \quad (30)$$

The free energy of three molecules can be obtained by

$$F^{(3)} = -k_B T \ln Z^{(3)} = 3F^{(1)} - k_B T \ln f^{(3)}(M_3, M, x_3, x_2) \tag{31}$$

### 2.4. General Expression of $n$ Molecules

Now we extend the expressions of  $n$  molecules with  $\beta$ -strand, the first expression is the partition function of  $n$  molecules,

$$Z^{(n)} = \sum_{m_n \leq m_{n-1}} \sum_{m_{n-1} \leq m_{n-2}} \cdots \sum_{m_3 \leq m_2} \sum_{m_2} \Omega^{(n)}(m_2, m_3, \cdots, m_{n-1}, m_n) \times Z^{(n)}(m_2, m_3, \cdots, m_{n-1}, m_n) \tag{32}$$

where the degeneracy factor of  $n$  molecules is

$$\Omega^{(n)}(m_2, m_3, \cdots, m_{n-1}, m_n) = 2 \prod_{i=3}^n (m_{i-1} - m_i)(N - m_2)^2 \tag{33}$$

and

$$Z^{(n)}(m_2, m_3, \cdots, m_{n-1}, m_n) = Z^{(n-1)}(m_2, m_3, \cdots, m_{n-2}, m_{n-1}) g_n^{m_n} Z_{nn} Z_{nl} \tag{34}$$

then  $G_n$  function can be defined as

$$G_n(m_2, m_3, \cdots, m_{n-1}, m_n, x_2, \cdots, x_n) = G_{n-1}(m_2, m_3, \cdots, m_{n-2}, m_{n-1}, x_2, \cdots, x_{n-1}) x_n^{m_n} \tag{35}$$

Let us define the function of  $f^{(n)}$  as

$$f^{(n)}(M_n, M_{n-1}, \cdots, M, x_n, x_{n-1}, \cdots, x_2) = 2 \sum_{m_n=1}^{M_n} \cdots \sum_{m_2=1}^M \prod_{i=3}^n (m_{i-1} - m_i)(N - m_2)^2 \times G_n(m_2, m_3, \cdots, m_{n-1}, m_n, x_2, \cdots, x_n) \tag{36}$$

where  $x_n = g_n / \lambda_1$ , and used  $\theta_n \approx \lambda_1^{-m_n}$ , so the partition function of  $n$  molecules can be rewritten as

$$Z^{(n)} = Z_N^n f^{(n)}(M_n, M_{n-1}, \cdots, M, x_n, x_{n-1}, \cdots, x_2) \tag{37}$$

The expression of free energy is

$$F^{(n)} = -k_B T \ln Z^{(n)} = nF^{(1)} - k_B T \ln f^{(n)}(M_n, M_{n-1}, \cdots, M, x_n, x_{n-1}, \cdots, x_2) \tag{38}$$

As we mentioned, the reference state is the coil state which has a statistical weight of 1 (free energy = 0). The helix state is favorable  $s > 1$  and the nucleation parameter is unfavorable  $0 < \sigma < 1$ . A single  $\beta$ -sheet is probably not stable in solution [29,30], so  $g \approx 1$  or slightly bigger than 1. However, a  $\beta$ -sheet bilayer is more stable than a helix, thus we need to introduce a new parameter,  $z$ , to describe side chain interactions, then  $gz > s$ . The free energy of  $\beta$ -sheet includes the conformation entropy (which supplies the repulsive force) and the interaction of hydrogen bond and the side-chain interactions (that is attractive force). Therefore,  $x_i$  can be replaced by  $x_{i,j} = x_i z_i^{j-1}$ , the corresponding general expression of physics quantities for fibril structure with  $j\beta$ -sheet can be written as

$$Z_j^{(n)} = Z_N^n f_j^{(n)}(z_n, M_n, M_{n-1}, \cdots, M, x_{n,j}, x_{n-1,j}, \cdots, x_{2,j}) \tag{39}$$

$$F_j^{(n)} = nF^{(1)} - k_B T \ln f_j^{(n)}(M_n, M_{n-1}, \cdots, M, x_{n,j}, x_{n-1,j}, \cdots, x_{2,j})$$

where  $j = 1, 2$  and

$$f_j^{(n)}(M_n, M_{n-1}, \dots, M, x_{n,j}, x_{n-1,j}, \dots, x_{2,j}) = 2 \sum_{m_n=1}^{M_n} \dots \sum_{m_2=1}^M \prod_{i=3}^n (m_{i-1} - m_i)(N - m_2)^2 \quad (40)$$

$$\times G_n^j(m_2, m_3, \dots, m_{n-1}, m_n, x_{2,j}, \dots, x_{n,j})$$

### 3. Discussion

#### 3.1. Molecular Interaction with Denaturant

The influence of denaturant on amyloid fibril is investigated by the interaction parameter  $x_2$ . When the denaturant urea is added in the system, the free energy is dependent on the denaturant concentration, and can be expanded as the first order approximation,

$$\Delta F_i = \Delta F_i(C_{urea} = 0) + \delta_i C_{urea} \quad (41)$$

where  $\delta_i > 0$ .

The parameters  $s$  and  $g_2$  are expressed as

$$s = \exp[-\Delta F_s(0) - \delta_s C_{urea}] = s(0) \exp(-\delta_s C_{urea}) \quad (42)$$

$$g_2 = \exp[-\Delta F_g(0) - \delta_g C_{urea}] = g_2(0) \exp(-\delta_g C_{urea})$$

where  $s(0) = \exp[-\Delta F_s(0)]$ ,  $g_2(0) = \exp[-\Delta F_g(0)]$ .  $s(0) > 1$  is for the helix state (folded state), and  $s(0) < 1$  is for the coil state (unfolded state).

We make an approximation  $\delta_s \approx \delta_g = \delta$ , the expression of  $x_2$  is

$$x_2 = \frac{g_2}{\lambda_1^2} = \frac{4g_2(0) \exp(\delta C_{urea}) / s^2(0)}{\left\{ \left[ \frac{\exp(\delta C_{urea})}{s(0)} + 1 \right] + \sqrt{\left[ \frac{\exp(\delta C_{urea})}{s(0)} - 1 \right]^2 + 4 \frac{\sigma}{s(0)} \exp(\delta C_{urea})} \right\}^2} \quad (43)$$

When  $\sigma/s(0) \ll 1$ ,

$$x_2 = \frac{4g_2(0) \exp(\delta C_{urea}) / s^2(0)}{\left\{ \left[ \frac{\exp(\delta C_{urea})}{s(0)} + 1 \right] + \left| \frac{\exp(\delta C_{urea})}{s(0)} - 1 \right| \right\}^2} \quad (44)$$

If  $\exp(\delta C_{urea})/s(0) < 1$ , it is that the strong folded state

$$x_2(sf) = \frac{g_2(0)}{s^2(0)} \exp(\delta C_{urea}) \quad (45)$$

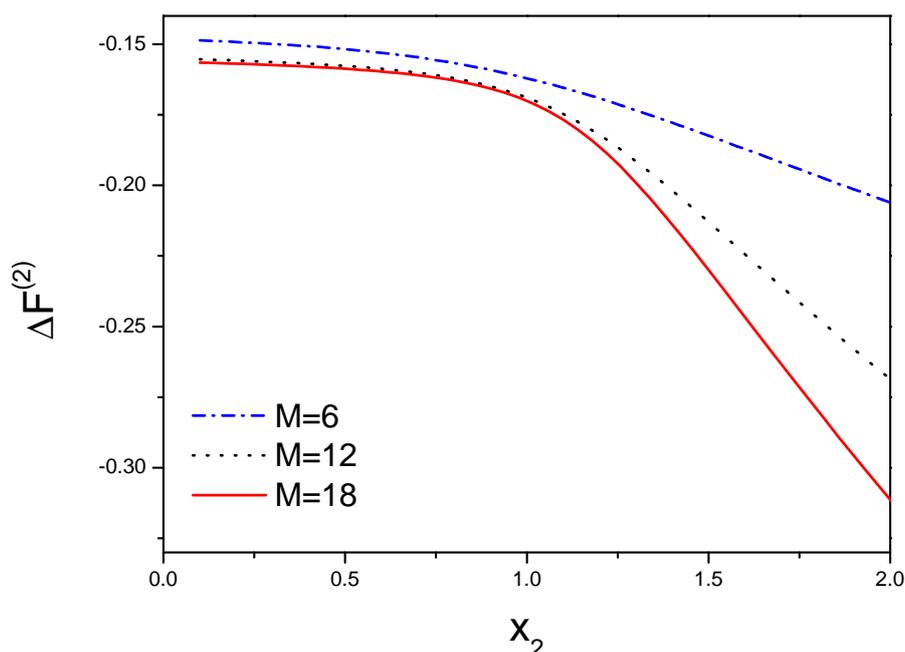
If  $\exp(\delta C_{urea})/s(0) > 1$ , it is that the strong unfolded state

$$x_2(suf) = g_2(0) \exp(-\delta C_{urea}) \quad (46)$$

Based on Equations (45) and (46), we can see how the denaturant affects the aggregation process. For strong folded state, the denaturant can strengthen the hydrophobic and hydrogen binding interactions. Otherwise, adding denaturant can weaken the interactions for strong unfolded state.

#### 3.2. Free Energy

As shown in Figure 3,  $\Delta F^{(2)}$  is a function of  $x_2$  that represents the effect of  $g_2$ ,  $\sigma$  and  $s$ . Here we choose  $N = 20$ ,  $M = 18$ ,  $s$  is from 0 to 2,  $\theta$  is from 0.001 to 1 and  $g_2$  is from 0 to 2. When  $g_2$  is zero, there is not any interaction between two molecular chains. Thus  $g_2$  is chosen as a smaller quantity, the interaction is a repulsive force, otherwise, the attractive force for larger  $g_2$ , so the range of  $x_2$  is from  $10^{-5}$  to 2.



**Figure 3.** The free energy is dependent on the interaction between the first molecule and the second molecule.  $x_2 = g_2/\lambda_1^2$ , where the free energy for each bond is  $-k_B T \ln g_2$  that illustrates the loss of conformational entropy from both chains, and  $\lambda_1$  indicates the contribution from the peptide tails not participating in the hydrogen bonds.  $\Delta F^{(2)} = F^{(2)} - 2F^{(1)}$  accounts for the free energy difference from both chains. The free energy will be decreased when the number of the hydrogen bond increases from 6 to 18.

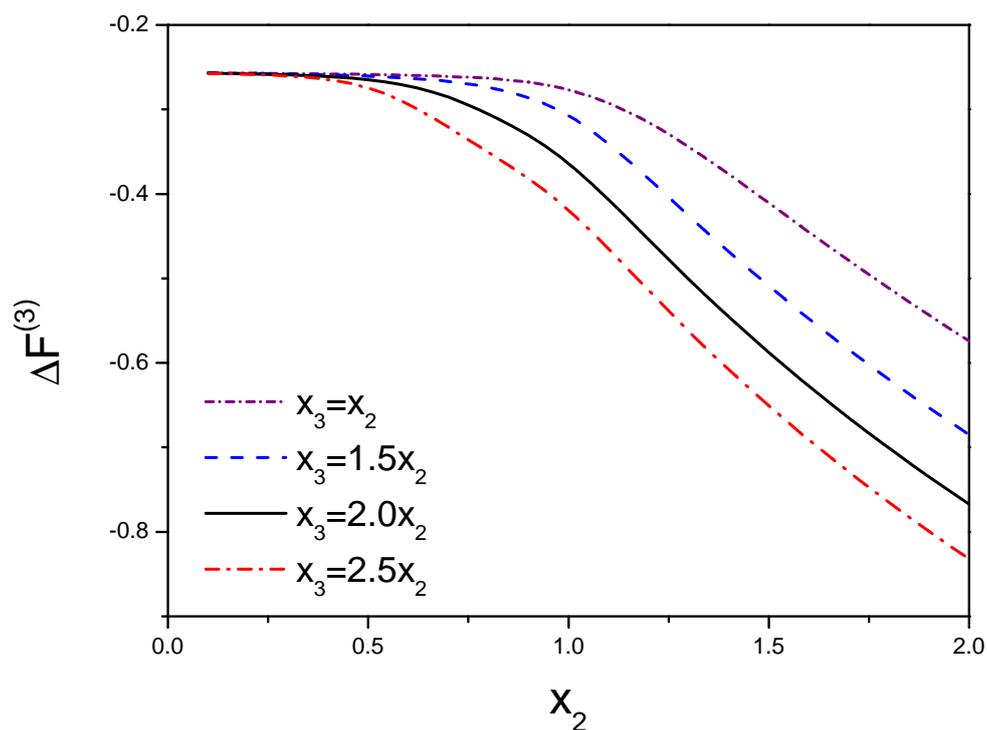
In Figure 4, we take  $M_3 = M - 3$  and  $g_2 = g_3$ . This figure demonstrates the free energy and the function of  $f^{(3)}$  are a function of  $x_2$  and  $x_3$ . Due to  $x_3 = \lambda_1 x_2$ , we choose a few of values for  $x_3 = x_2, 1.5x_2, 2x_2, 2.5x_2$  that are resulted from the range of  $\lambda_1$  from 1 to 3.

By using the above equations, we can obtain the landscape of free energy. The parameters such as  $g, \sigma, s$ , even  $z$  stand for the different peptide states, and corresponding to *coil, helix,  $\beta$ -strand,  $\beta$ -sheet* and fibrils.  $M_n = M - 3n$  is used to describe the number of hydrogen bond in the  $n$ -th chain. With the change of these parameters, the free energy will be a function as  $n$ . We have calculated the free energy for  $n \leq 6$ .

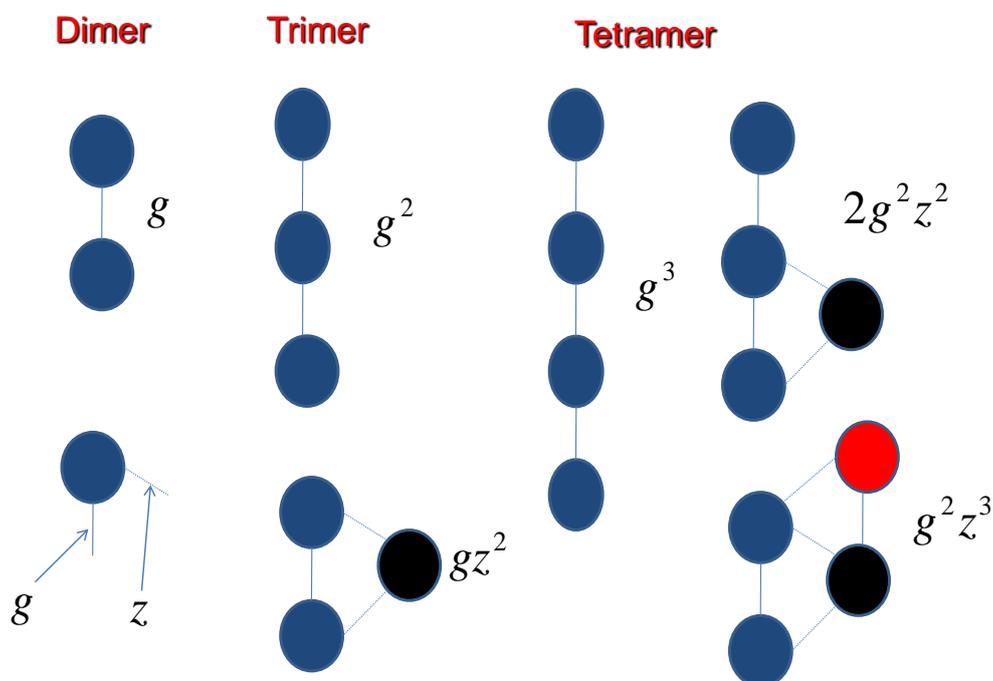
For counting the side chain effect, the interaction coefficients  $g_n$  have been demonstrated in Figure 5,

$$\begin{aligned}
 g_2 &= g \\
 g_3 &= gz \\
 g_4 &= g^2(z + z^2) \\
 g_5 &= g^3(z + z^2 + z^3)
 \end{aligned} \tag{47}$$

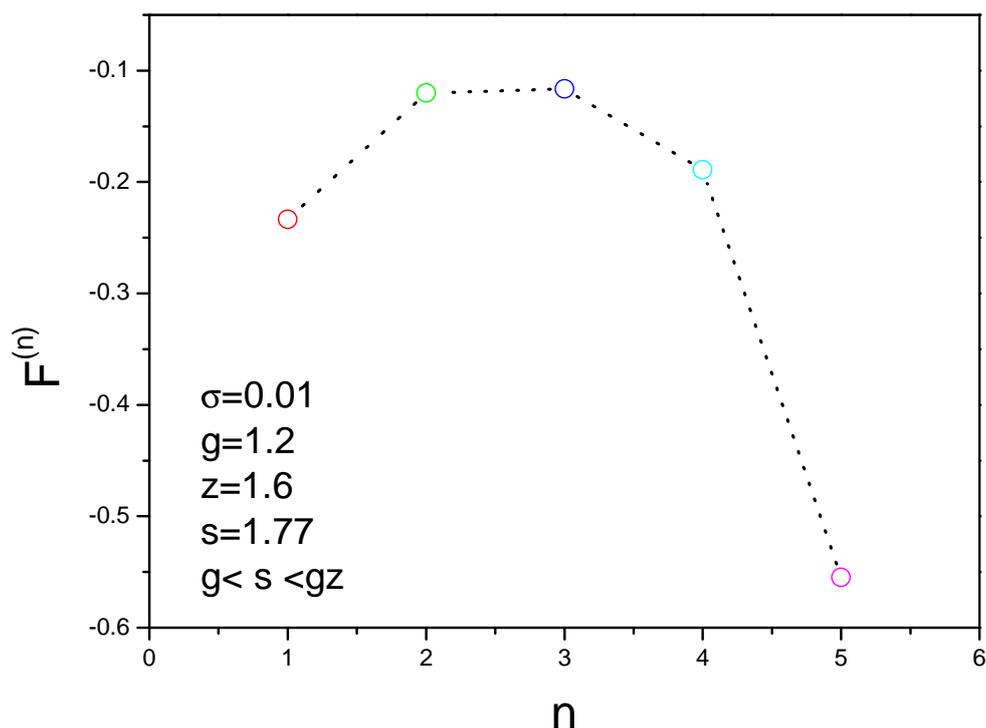
The energy landscape can be obtained as shown in Figure 6, where the parameters are  $\sigma = 0.01, g = 1.2, z = 1.6, s = 1.77$ . The nucleation can occur because the condition  $g < s < gz$  is satisfied. We find that the free energy has a peak value when  $n$  is equal to 3. In other words, three molecules have higher free energy and four molecules have lower free energy. This is an important result about the assemble mechanism for aggregation of amyloid fibril. Based on this result, we develop the kinetic formulas to study the nucleation processes of amyloid fibrils.



**Figure 4.** Variation of three molecules' free energy as a function of the molecular interactions.  $x_2$  denotes the interactions between molecule 1 and molecule 2.  $x_3$  is the interaction between the third molecule and dimer.  $\Delta F^{(3)}$  is defined as  $F^{(3)} - 3F^{(1)}$ . The competition between molecules results in the variation of the free energy.



**Figure 5.** The assembly process is dependent on the molecular interactions.



**Figure 6.** Free energy landscape for different molecules gives a nonmonotonic function that has a peak at three molecules.  $F^{(n)}$  is the free energy of an oligomer containing  $n$  molecules.

### 3.3. Peak Location in the 2D Nucleation Rate

A model of three molecules was presented for the dynamics of a nucleating trimer, the nucleation rate in 2D can be written as [31]

$$\begin{aligned}
 k_{nuc} &= \sum_{x=1}^L k_{diff} \varepsilon_+(x, 1) C_2(x) \\
 &= k_{diff} C_1^2 \sum_{x=1}^L \varepsilon_+(x, 1) \exp\left(-\frac{xf}{kT}\right)
 \end{aligned}
 \tag{48}$$

The analytic solution of  $\varepsilon_+(x, y)$  is

$$\varepsilon_+(x, y) = e_0 \exp[-(\alpha x + \beta y)] \tilde{B}(x, y)
 \tag{49}$$

where

$$\tilde{B}(x, y) = \frac{\sum_{m=2}^{\infty} B_m \sin[m \arctan(y/x)] I_m(v\sqrt{x^2 + y^2}/2D)}{\sum_{m=2}^{\infty} B_m \sin(m\pi/4) I_m(v\sqrt{2}L/2D)}
 \tag{50}$$

The peak location is determined by the extremum of the nucleation rate,

$$\frac{\partial k_{nuc}}{\partial x} = k_{diff} C_1^2 \sum_{x=1}^L \left[ \frac{\partial \varepsilon_+(x, 1)}{\partial x} - \frac{f}{kT} \varepsilon_+(x, 1) \right] \exp\left(-\frac{xf}{kT}\right) = 0
 \tag{51}$$

then

$$\frac{\partial \ln \varepsilon_+(x, 1)}{\partial x} = \frac{f}{kT}
 \tag{52}$$

By using the above equation,

$$\varepsilon_+(x, 1) = e_0' \exp(-\alpha x) \tilde{B}(x, 1)
 \tag{53}$$

where  $e_0' = e_0 \exp(-\beta)$ ,  $\alpha = \frac{v_x}{2D}$ ,  $\beta = \frac{v_y}{2D}$ .

$$\frac{\partial \tilde{B}(x, 1)}{\partial x} = (\varepsilon_2 + \alpha)\tilde{B}(x, 1) \tag{54}$$

where  $\varepsilon_2 = \frac{f}{kT}$ .

Because denominator of  $\tilde{B}(x, 1)$  is a constant,

$$\frac{\partial \tilde{B}_{num}(x, 1)}{\partial x} = (\varepsilon_2 + \alpha)\tilde{B}_{num}(x, 1) \tag{55}$$

where  $\tilde{B}_{num}$  is a numerator of  $\tilde{B}$ .

$\tilde{B}_{num}(x, y)$  can be rewritten as

$$\tilde{B}_{num}(x, y) = \sum_{m=2}^{\infty} B_m \sin(m\theta) I_m(ar) \tag{56}$$

where  $r = \sqrt{x^2 + y^2}$ ,  $a = \frac{v_x}{2D}$ ,  $v = \sqrt{v_x^2 + v_y^2}$  and

$$\begin{aligned} B_2 &= 1 \\ B_3 &= 2 \frac{v_x - v_y}{v} \\ B_4 &= 2 \left( \frac{v_x - v_y}{v} \right)^2 \\ B_5 &= 2 \frac{v_x - v_y}{v} \\ I_m(x) &= \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+1)\Gamma(m+k+1)} \left(\frac{x}{2}\right)^{m+2k} \end{aligned} \tag{57}$$

We take the earlier terms including  $k = 0, 1, 2$ ,

$$\begin{aligned} \tilde{B}_{num}(x, y) &= \sum_{m=2} B_m \sin(m\theta) \sum_{k=0}^{\infty} \frac{(ar)^{2k+m}}{k! \Gamma(k+m+1)} \\ &= \sum_m B_m \sin(m\theta) \left[ \frac{(ar)^m}{\Gamma(m+1)} + \frac{(ar)^{m+2}}{\Gamma(m+2)} + \frac{(ar)^{m+4}}{2! \Gamma(m+3)} \right] \end{aligned} \tag{58}$$

By using  $\Gamma(n) = (n - 1)!$ , we have

$$\begin{aligned} \tilde{B}_{num}(x, y) &= \frac{a^2}{2!} \{ r^2 \sin(2\theta) + B_3 \frac{a}{3} r^3 \sin(3\theta) + [B_4 \frac{a^2}{3 \times 4} r^4 \sin(4\theta) + \frac{a^2}{3} r^4 \sin(2\theta)] + \\ & [B_5 \frac{a^3}{3 \times 4 \times 5} r^5 \sin(5\theta) + B_3 \frac{a^3}{3 \times 4} r^5 \sin(3\theta)] \} \end{aligned} \tag{59}$$

By using  $\sin(2\theta) = 2 \sin \theta \cos \theta$ ,  $\sin(3\theta) = 3 \sin \theta - 4 \sin^3 \theta$ ,  $\sin(4\theta) = 4 \sin \theta \cos \theta - 8 \cos \theta \sin^3 \theta$ ,  $\sin(5\theta) = 5 \sin \theta - 20 \sin^3 \theta + 16 \sin^5 \theta$ ,  $x = r \sin \theta$ ,  $y = r \cos \theta$ , we get

$$\begin{aligned} \tilde{B}_{num}(x, y) &= \frac{a^2}{2!} \left[ \left( B_5 \frac{a^3}{12} + B_3 \frac{a^3}{4} \right) x^4 y + \left( B_4 \frac{a^2}{3} + \frac{2a^2}{3} \right) x^3 y \right. \\ & + a B_3 x^2 y + \left( -B_5 \frac{a^3}{6} + B_3 \frac{a^3}{6} \right) x^2 y^3 + 2xy + \left( -B_4 \frac{a^2}{3} + \frac{2a^2}{3} \right) xy^3 \\ & \left. - B_3 \frac{a}{3} y^3 + \left( B_5 \frac{a^3}{60} - B_3 \frac{a^3}{12} \right) y^5 \right] \end{aligned} \tag{60}$$

The coefficients can be defined as

$$\begin{aligned}
 f_{41} &= (B_5 + 3B_3) \frac{a^3}{12} = (\alpha - \beta) \frac{a^2}{3} \\
 f_{31} &= (B_4 + 2) \frac{a^2}{3} = \frac{1}{3}(4a^2 - \alpha\beta) \\
 f_{21} &= aB_3 = \alpha - \beta \\
 f_{23} &= (-B_5 + B_3) \frac{a^3}{6} \\
 f_{11} &= 2 \\
 f_{13} &= (-B_4 + 2) \frac{a^2}{3} = \frac{\alpha\beta}{3} \\
 f_{03} &= -B_3 \frac{a}{3} = -\frac{\alpha - \beta}{3} \\
 f_{05} &= (B_5 - 5B_3) \frac{a^3}{60} = -\frac{a^2}{15}(\alpha - \beta)
 \end{aligned} \tag{61}$$

thus

$$\begin{aligned}
 \tilde{B}(x, y) &= f_{41}x^4y + f_{31}x^3y + f_{21}x^2y \\
 &+ f_{23}x^2y^3 + f_{11}xy + f_{13}xy^3 + f_{03}y^3 + f_{05}y^5
 \end{aligned} \tag{62}$$

When  $y = 1$ , the above expression can be written as

$$\tilde{B}_{num}(x, 1) = b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0 \tag{63}$$

where

$$\begin{aligned}
 b_4 &= f_{41} = (\alpha - \beta) \frac{a^2}{3} \\
 b_3 &= f_{31} = \frac{1}{3}(4a^2 - \alpha\beta) \\
 b_2 &= f_{21} + f_{23} = \alpha - \beta \\
 b_1 &= f_{11} + f_{13} = 2 + \frac{\alpha\beta}{3} \\
 b_0 &= f_{03} + f_{05} = -\frac{\alpha - \beta}{3} - \frac{a^2}{15}(\alpha - \beta) = -\frac{1}{15}(\alpha - \beta)(a^2 + 5)
 \end{aligned} \tag{64}$$

The expressions of  $v_x$  and  $v_y$  are employed by

$$\begin{aligned}
 v_x &= 2 \frac{[\exp(-E_{weak}) - 1]}{\exp(-E_{weak}) + 1} \\
 v_y &= 2 \frac{[\exp(-E_{strong}) - 1]}{\exp(-E_{strong}) + 1}
 \end{aligned} \tag{65}$$

where  $E_{weak} = \varepsilon_2$ .

We will discuss three cases for taking different  $\varepsilon_2$  to calculate the coefficients of polynomial,

**Case I:**  $\varepsilon_2 = 0.2$ ,  $E_{strong} = -0.5$

$$\begin{aligned}
 b_4 &= -0.00200776 \\
 b_3 &= 0.0314431 \\
 b_2 &= -0.344587 \\
 b_1 &= 1.99186 \\
 b_0 &= 0.115264
 \end{aligned} \tag{66}$$

then inserting these coefficients and obtain the peak value of  $x$ ,

$$x_m^{\varepsilon_2=0.2} = 3.5 \tag{67}$$

which is in agreement with the numerical result  $x_m^{\varepsilon_2=0.2} = 4$  [31].

**Case II:**  $\varepsilon_2 = 0.3$ ,  $E_{strong} = -0.5$

$$\begin{aligned} b_4 &= -0.00269598 \\ b_3 &= 0.0395389 \\ b_2 &= -0.393804 \\ b_1 &= 1.98785 \\ b_0 &= 0.131807 \end{aligned} \tag{68}$$

then inserting these coefficients and obtain the peak value of  $x$ ,

$$x_m^{\varepsilon_2=0.3} = 2.8 \tag{69}$$

which is in agreement with the numerical result  $x_m^{\varepsilon_2=0.2} = 3$  [31].

**Case III:**  $\varepsilon_2 = 0.4$ ,  $E_{strong} = -0.5$

$$\begin{aligned} b_4 &= -0.00364679 \\ b_3 &= 0.0490944 \\ b_2 &= -0.442294 \\ b_1 &= 1.98389 \\ b_0 &= 0.148161 \end{aligned} \tag{70}$$

then inserting these coefficients and obtain the peak value of  $x$ ,

$$x_m^{\varepsilon_2=0.3} = 2.2 \tag{71}$$

which is in agreement with the numerical result  $x_m^{\varepsilon_2=0.2} = 2$  [31].

It is reasonable to neglect the higher terms from the above data, i.e.,  $b_4 = 0$  and  $b_3 = 0$ , so the polynomial is changed to

$$\tilde{B}_{num}(x, 1) = b_2x^2 + b_1x + b_0 \tag{72}$$

Due to  $b_1^2 \gg 4b_0b_2, b_2^2$ , and consider  $x$  is a positive number, so the root of equation is simplified as

$$x_m \approx \frac{1}{\varepsilon_2 + \alpha} \tag{73}$$

The nucleation flux is a nonmonotonic function of the number of the hydrogen bonds in dimer when the third molecule is added to form trimer. The peak values are a significant result of the deviation of the lowest free energy pathway.

### 3.4. Scaling Behavior of Nucleation Rate in 3D

The system of three molecules is unstable, so we have developed a stable 3D model. The nucleation rate in 3D model can be written as [32]

$$k_{nuc} = \sum_{x=1}^L k_{diff} \varepsilon_{mol}(x) C_2(x) \tag{74}$$

where

$$\varepsilon_{mol}(x) = \frac{\left(\frac{1}{p_{+mol}} - 1\right)^x - 1}{\left(\frac{1}{p_{+mol}} - 1\right)^{n^*} - 1} \tag{75}$$

and

$$C_2(x) = C_1^2 \exp(-xf/kT) \tag{76}$$

$p_{+mol}$  can be defined as

$$p_{+mol} = \frac{k_{diff}}{k_{diff} + k_{loss}} \tag{77}$$

Inserting the above equation into  $k_{nuc}$ , we have

$$k_{nuc} = \sum_{x=1}^L k_{diff} \frac{(\frac{k_{loss}}{k_{diff}})^x - 1}{(\frac{k_{loss}}{k_{diff}})^{n^*} - 1} C_1^2 \exp(-xf/kT) \tag{78}$$

furthermore,

$$k_{nuc} = \frac{k_{diff} C_1^2}{(\frac{k_{loss}}{k_{diff}})^{n^*} - 1} \sum_{x=1}^L \{ [\frac{k_{loss}}{k_{diff}} \exp(-\frac{f}{kT})]^x - [\exp(-\frac{f}{kT})]^x \} \tag{79}$$

With the help of

$$\sum_{x=1}^L y^x = \frac{y^{L+1} - y}{y - 1} \tag{80}$$

then

$$k_{nuc} = \frac{k_{diff} C_1^2}{(\frac{k_{loss}}{k_{diff}})^{n^*} - 1} \exp(-\frac{f}{kT}) \{ \frac{k_{loss}}{k_{diff}} \frac{(\frac{k_{loss}}{k_{diff}})^L \exp(-\frac{fL}{kT}) - 1}{\frac{k_{loss}}{k_{diff}} \exp(-\frac{f}{kT}) - 1} - \frac{\exp(-\frac{fL}{kT}) - 1}{\exp(-\frac{f}{kT}) - 1} \} \tag{81}$$

Based on the numerical result [32], we will discuss  $f > 0$ .

When  $f > 0, 1 > \exp(-\frac{f}{kT}) \gg \exp(-\frac{fL}{kT})$ . In this case, the first approximation is  $\exp(-\frac{fL}{kT}) \rightarrow 0$

$$k_{nuc} = \frac{k_{diff} C_1^2}{(\frac{k_{loss}}{k_{diff}})^{n^*} - 1} \{ \frac{1}{1 - \exp(\frac{f}{kT})} - \frac{k_{loss}/k_{diff}}{k_{loss}/k_{diff} - \exp(\frac{f}{kT})} \} \tag{82}$$

In the condition of  $\frac{k_{loss}}{k_{diff}} > 1$

$$k_{nuc} \approx \frac{k_{diff}^{n^*} C_1^2}{k_{loss}^{n^* - 1}} \exp(-\frac{f}{kT}) \tag{83}$$

Due to  $k_{diff} \propto C_1$ , so the scaling behavior in this case is

$$k_{nuc} \propto C_1^{n^* + 2} \tag{84}$$

This condition matches the numerical result, so the analytic result can be written as

$$\text{Log}k_{nuc} = (n^* + 2)\text{Log}C_0 + \text{const} \tag{85}$$

where  $C_1 = C_0$  and  $\text{const}$  is a constant including  $f$  and other terms.

In the experimental measurement of the nucleation process [33–35], kinetics were monitored by ThT fluorescence. The solution conditions for fibril formation were 100 mM KCl, 50 mM potassium phosphate, PH 7.4, 25 °C. Fiber sample were prepared as for transmission electron microscopy. The experimental result is  $k_{nuc} \propto 4$ . When we take  $n^* = 2$  in the analytic result, the theoretical result is in agreement with the experimental result.

#### 4. Conclusions

In summary, a microscopic model with three states including coil, helix and sheet is constructed to explore the mechanism of amyloid formation. The partition function and thermodynamic quantities of many molecule systems are obtained by considering the repulsive and attractive interactions. The equilibrium properties of the system in the solution have been investigated. Free energy landscape and side chain effect are illustrated. It is found that the system of three molecules has higher free energy. The kinetic properties of molecules related with amyloid formation are also studied. By using the random walk model in 2D and 3D, the nucleation processes of amyloid fibrils are quantitatively demonstrated. The microscopic theoretical model and results can be in agreement with numerical and experimental results. These theoretical approaches of the microscopic model could be used to improve the computational simulations in new timescale.

**Conflicts of Interest:** The author declares no conflict of interest.

#### References

1. Nelson, R.; Sawaya, M.R.; Balbirnie, M.; Madsen, A.Ø.; Riek, C.; Grothe, R.; Eisenberg, D. Structure of the cross- $\beta$  spine of amyloid-like fibrils. *Nature* **2005**, *435*, 773–778. [[CrossRef](#)] [[PubMed](#)]
2. Knowles, T.P.J.; Waudby, C.A.; Devlin, G.L.; Cohen, S.I.A.; Aguzzi, A.; Vendruscolo, M.; Terentjev, E.M.; Welland, M.E.; Dobson, C.M. An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science* **2009**, *326*, 1533–1537. [[CrossRef](#)] [[PubMed](#)]
3. Seeliger, J.; Estel, K.; Erwin, N.; Winter, R. Cosolvent effects on the fibrillation reaction of human IAPP. *Phys. Chem. Chem. Phys.* **2013**, *15*, 8902–8907. [[CrossRef](#)] [[PubMed](#)]
4. Hardy, J.; Selkoe, D.J. The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* **2002**, *297*, 353–356. [[CrossRef](#)] [[PubMed](#)]
5. Canchi, D.R.; Garcia, A.E. Cosolvent Effects on Protein Stability. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273–293. [[CrossRef](#)] [[PubMed](#)]
6. Miller, Y.; Ma, B.; Nussinov, R. Polymorphism in Alzheimer A $\beta$  amyloid organization reflects conformational selection in a rugged energy landscape. *Chem. Rev.* **2010**, *110*, 4820–4838. [[CrossRef](#)] [[PubMed](#)]
7. Dill, K.A. Theory for the folding and stability of globular proteins. *Biochemistry* **1985**, *24*, 1501–1509. [[CrossRef](#)] [[PubMed](#)]
8. Ghosh, K.; Dill, K.A. Theory for Protein Folding Cooperativity: Helix Bundles. *J. Am. Chem. Soc.* **2009**, *131*, 2306–2312. [[CrossRef](#)] [[PubMed](#)]
9. Dobson, C.M. Protein folding and misfolding. *Nature* **2003**, *426*, 884–890. [[CrossRef](#)] [[PubMed](#)]
10. Carballo-Pacheco, M.; Strodel, B. Advances in the Simulation of Protein Aggregation at the Atomistic Scale. *J. Phys. Chem. B* **2016**, *120*, 2991–2999. [[CrossRef](#)] [[PubMed](#)]
11. Cheon, M.; Hall, C.K.; Chang, I. Structural Conversion of A  $\beta_{17C42}$  Peptides from Disordered Oligomers to U-Shape Protofilaments via Multiple Kinetic Pathways. *PLoS Comput. Biol.* **2015**, *11*, e1004258. [[CrossRef](#)] [[PubMed](#)]
12. Kashchiev, D.; Auer, S. Nucleation of amyloid fibrils. *J. Chem. Phys.* **2010**, *132*, 215101. [[CrossRef](#)] [[PubMed](#)]
13. Cabriolu, R.; Auer, S. Amyloid fibrillation kinetics: insight from atomistic nucleation theory. *J. Mol. Biol.* **2011**, *411*, 275–285. [[CrossRef](#)] [[PubMed](#)]
14. Auer, S.; Meersman, F.; Dobson, C.M.; Vendruscolo, M. A Generic Mechanism of Emergence of Amyloid Protofilaments from Disordered Oligomeric Aggregates. *PLoS Comput. Biol.* **2008**, *4*, e1000222. [[CrossRef](#)] [[PubMed](#)]
15. Schreck, J.S.; Yuan, J.M. Exactly solvable model for helix-coil-sheet transitions in protein systems. *Phys. Rev. E* **2010**, *81*, 061919. [[CrossRef](#)] [[PubMed](#)]
16. Schmit, J.D. Kinetic theory of amyloid fibril templating. *J. Chem. Phys.* **2013**, *138*, 185102. [[CrossRef](#)] [[PubMed](#)]
17. Knowles, T.P.J.; Simone, A.D.; Fitzpatrick, A.W.; Baldwin, A.; Meehan, S.; Rajah, L.; Vendruscolo, M.; Welland, M.E.; Dobson, C.M.; Terentjev, E.M. Twisting transition between crystalline and fibrillar phases of aggregated peptides. *Phys. Rev. Lett.* **2012**, *109*, 158101. [[CrossRef](#)] [[PubMed](#)]

18. Lee, C.F. Self-assembly of protein amyloids: A competition between amorphous and ordered aggregation. *Phys. Rev. E* **2009**, *80*, 031922. [[CrossRef](#)] [[PubMed](#)]
19. Bellesia, G.; Shea, J.E. Diversity of kinetic pathways in amyloid fibril formation. *J. Chem. Phys.* **2009**, *131*, 111102. [[CrossRef](#)] [[PubMed](#)]
20. Cohen, S.I.A.; Linse, S.; Luheshi, L.M.; Hellstrand, E.; White, D.A.; Rajah, L.; Otzen, D.E.; Vendruscolo, M.; Dobson, C.M.; Knowles, T.P.J. Proliferation of amyloid- $\beta$ 42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 9758–9763. [[CrossRef](#)] [[PubMed](#)]
21. Zhang, J.; Muthukumar, M.J. Simulations of nucleation and elongation of amyloid fibrils. *J. Chem. Phys.* **2009**, *130*, 035102. [[CrossRef](#)] [[PubMed](#)]
22. Poland, D.; Scheraga, H. *Theory of Helix-Coil Transitions in Biopolymers*; Academic Press: New York, NY, USA, 1970.
23. Dill, K.; Bromberg, S. *Molecular Driving Forces*; Garland Science: New York, NY, USA, 2002.
24. Smirnovas, V.; Winter, R.; Funk, T.; Dzwolak, W. Thermodynamic properties underlying the  $\alpha$ -helix-to- $\beta$ -sheet transition, aggregation, and amyloidogenesis of polylysine as probed by calorimetry, densimetry, and ultrasound velocimetry. *J. Phys. Chem. B* **2005**, *109*, 19043–19045. [[CrossRef](#)] [[PubMed](#)]
25. Lifson, S.; Roig, A. On the Theory of Helix—Coil Transition in Polypeptides. *J. Chem. Phys.* **1961**, *34*, 1963–1974. [[CrossRef](#)]
26. Van Gestel, J.; de Leeuw, S.W. A statistical mechanical theory of fibril formation in dilute protein solutions. *Biophys. J.* **2006**, *90*, 3134–3145. [[CrossRef](#)] [[PubMed](#)]
27. Zimm, B.H.; Bragg, J.K. Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains. *J. Chem. Phys.* **1959**, *31*, 526–535. [[CrossRef](#)]
28. Qian, H.; Schellman, J.A. Helix-coil studies: a comparative study for finite length polypeptides. *J. Phys. Chem.* **1992**, *96*, 3987–3994. [[CrossRef](#)]
29. Bitan, G.; Kirkitadze, M.D.; Lomakin, A.; Vollers, S.S.; Benedek, G.B.; Teplow, D.B. Amyloid  $\beta$ -protein ( $A\beta$ ) assembly:  $A\beta$  40 and  $A\beta$  42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 330–335. [[CrossRef](#)] [[PubMed](#)]
30. Barghorn, S.; Nimmrich, V.; Striebinger, A.; Krantz, C.; Keller, P.; Janson, B.; Bahr, M.; Schmidt, M.; Bitner, R.; Harlan, J.; et al. Globular amyloid  $\beta$ -peptide oligomer—A homogenous and stable neuropathological protein in Alzheimer's disease. *J. Neurochem.* **2005**, *95*, 834–847. [[CrossRef](#)] [[PubMed](#)]
31. Zhang, L.; Schmit, J. Pseudo-one-dimensional nucleation in dilute polymer solutions. *Phys. Rev. E* **2016**, *93*, 060401(R). [[CrossRef](#)] [[PubMed](#)]
32. Zhang, L.; Schmit, J. Theory of amyloid fibril nucleation from folded proteins. *Isr. J. Chem.* **2017**, *57*, 738–749. [[CrossRef](#)] [[PubMed](#)]
33. Ferrone, F.A. Assembly of  $A\beta$  Proceeds via Monomeric Nuclei. *J. Mol. Biol.* **2015**, *427*, 287–290. [[CrossRef](#)] [[PubMed](#)]
34. Knowles, T.P.; White, D.A.; Abate, A.R.; Agresti, J.J.; Cohen, S.I.A.; Sperling, R.A.; De Genst, E.J.; Dobson, C.M.; Weitz, D.A. Observation of spatial propagation of amyloid assembly from single nuclei. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14746–14751. [[CrossRef](#)] [[PubMed](#)]
35. Ruschak, A.M.; Miranker, A.D. Fiber-dependent amyloid formation as catalysis of an existing reaction pathway. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12341–12346. [[CrossRef](#)] [[PubMed](#)]

