Protein Aggregation Review & Entropy Trends in a Model Independent Protein Aggregation Process at Various Stages Using Fokker-Planck Dynamics

Leslie Cook, Preet Sharma[1]

Non-Linear Science Research Group
Center for Theoretical Research
Midwestern State University

E-mail: lncook0501@my.msutexas.edu, preet.sharma@msutexas.edu

Abstract:
The dynamics of protein aggregation is of extreme importance due to its direct connection with neuro-degenerative diseases such as Parkinson's, Alzhiemer's and Motor Neuron Disease, just to name a few. These diseases are very complex in nature and require a deep understanding of the interactions between protein-protein structures and the dynamics of the aggregation process. The complexity of protein aggregation increases as the aggregate progresses. This progression and the complexity of the molecules increases at various levels. There have been numerous theoretical and experimental studies on protein aggregation and we have a good enough understanding, but there is still a big gap in understanding the dynamics of the process of protein aggregation. We are showing the protein aggregation process from the point of view of an entropy increase trend.

Keywords: Protein Aggregation, Langevin Equation, Fokker-Planck Equation, Entropy Generation, Entropy Production

[1]Corresponding author.

1 Introduction

The seventh leading cause of death is dementia with the most common form Alzheimer's disease (AD) and second-most common is Parkinson's disease (PD). Dementia is hypothesized as a result of noxious protein buildup that obstructs communication between brain cells. As the disease progresses over time, irreversible brain damage considerably impairs cognitive and/or motor function. The risk of developing neurodegenerative diseases substantially increases with age and the socio- economic impact is suffered on a personal, communal, and global level, as the universal cost of care is estimated in trillions of dollars.

Currently, there is no cure to inhibit or reverse the progression of neurodegenerative diseases and treatment options are limited and somewhat ineffective. Despite the exponential advancements in modern medical technology, our understanding of neurodegenerative diseases is severely lacking. At this time, there is no laboratory test to detect when neurodegeneration begins in the brain. It can take years before a patient begins exhibiting symptoms to determine a proper diagnosis and by then neuron and tissue death has already taken its hold. And so the need to identify and develop models of the systems that cause neurodegeneration is of great importance. The brain is continually generating, delivering, and digesting proteins. It is a crucial cellular resource for the storage and processing of information. Cellular degradation due to aging assists in the inability to fix, change or digest away certain proteins. At neuron synapses, these structurally dynamic biomolecules are in high demand and mostly regulated through allosteric conformational changes. As neurodegeneration occurs, some of the proteins become corrupted [1], [2], [3], [4], [5]. The exact cause of this malfeasance is unknown but it is a popular topic among researchers with many new emerging theories. The majority of biological proteins do not aggregate. Some proteins are supersaturated, meaning they are expressed in an abundance that exceeds their solubility. This requires a significant amount of metabolic energy to ensure the protein are properly folded, or digested away when no longer needed.

Aggregation of intrinsically disordered proteins (IDPs) has been identified as a key factor in neurodegenerative diseases. Toxic accumulation occurs as soluble proteins sequester to form insoluble oligomers that shut down proper signals in the brain and eventually destroy neurons [6], [7], [8]. The environmental, chemical forces that initiate protein aggregation are not well known. An understanding of what drives aggregation is imperative for the development of treatments and therapies to slow the progression or better yet, cure the disease. Protein aggregation is a nuanced phenomenon where IDPs synchronize and self assemble into structured amyloid fibrils or form amorphous aggregates with a granular morphology. It is increasingly understood that specific regions in the polypeptide drive the aggregation process. When these specific regions are removed from the sequence, in vitro aggregation is inhibited [9], [10]. Proteins with specific sequences for amyloids, in vivo, are believed to have a hydrophobicity, strong inclination to form beta-sheets, and low net charge [11], [12], [13].The same rules apply as in the most basic of chemistry, "like attracts like." Monomeric proteins have a natural inclination to sequester in the form of dimers,

trimers, tetramers, etc. In recent studies, the phenomena of seeding, also known as cross seeding has been observed, which demonstrates that when preformed fibrils aggregates are introduced into a solution monomeric proteins, oligomers and fibril formations progresses at a much faster rate than otherwise observed. IDPs misfold into beta sheet confirmations and when two or more beta sheets aggregate laterally to form fibrils it is known as cross-beta confirmations [14], [15], [16].

Aggregation is heavily dependent on hydrophobic interactions that drive nonpolar regions towards one another, similar to micelle formation[17], [18]. Conformational changes of allosteric proteins can cause resistance to degradation. The process involved in the removal of waste may lose its effectiveness, which would be contributed to chaperone failure.The inability to digest away aggregate prone proteins has the potential to create an environment ripe for aggregation. Chaperones are responsible for guiding the way a protein species folds and protects against misfolding. The nature of chaperone vehicles generate a bistable system between protein aggregation and chaperone production and the potential for failure during stress response [19], [20]. Molecular chaperones are typically over expressed during cell stress in order to suppress anomalous processes and recognized as inhibitors of steps in the conversion of soluble proteins into amyloid fibrils. Demonstrating the binding interactions between these proteins presents challenges due to the dynamic nature of the systems [21].

The first step in process of converting soluble proteins into insoluble oligomers or amyloid fibrils requires production of an unfolded peptide. After overcoming the free energy barrier, IPDs or misfolded proteins rapidly advance to a stable amyloid form. The kinetic process of fibrilogenesis is shown to exhibit three stages; a lag phase, growth phase, and a final plateau, like a sigmoidal curve. The lag phase starts as monomers combine and form oligomers and fibrils, but remain at a low concentration. During the growth phase, nucleation-dependent aggregation substantially increases in concentration, before leveling off, and ending with highly structured amyloid fibrils during the plateau phase. In addition to structured fibril formation, aggregation may also follow a different path of assembly that leads to the formation of amorphorous deposits. Different pathway intermediates may drive differing aggregate morphologies for the same protein [11], [22], [23]. Soluble oligomers formed during aggregation has been found to be more cytotoxic. The oligomers are very unstable, present in low concentrations, and have a strong proclivity to interconvert with one another [14].

1.1 Protein Aggregation Diseases

Dementia: Dementia is hypothesized as a result from noxious protein buildup that obstructs communication between brain cells. As the disease progresses over time, irreversible brain damage considerably impairs cognitive and/or motor function. The risk of developing neurodegenerative diseases substantially increases with age. Currently, there is no cure to inhibit or reverse the progression of protein aggregation diseases and treatment options are limited and somewhat ineffective. Despite the exponential advancements in modern

medical technology, our understanding of neurodegenerative diseases is severely lacking. At this time, there is no laboratory test to detect when neurodegeneration begins in the brain. It can take years before a patient begins exhibiting symptoms to determine a proper diagnosis and by then, neuron and tissue death has already taken its hold. And so it's of great importance to develop models of the systems that cause neurodegeneration. The brain is continually generating, delivering, and digesting proteins. It is a crucial cellular resource for the storage and processing of information. Cellular degradation due to aging assists in the inability to fix, change or digest away certain proteins. At neuron synapses, these structurally dynamic biomolecules are in high demand and mostly regulated through allosteric conformational changes. As neurodegeneration occurs, some of the proteins become corrupted [2].. The exact cause of this malfeasance is unknown but it is a popular topic among researchers with many new emerging theories. ‘ The seventh leading cause of death is dementia with the most common form being Alzheimer's disease (AD). The second-most common is Parkinson's disease (PD). Approximately 1 in 6 people are affected by dementia year [24]. The socio-economic impact is suffered on a personal, communal, and global level, as the universal cost of care is estimated in trillions of dollars. An emergent protein aggregation disease of interest is Chronic Traumatic Encephalopathy (CTE). CTE is caused by repeated traumatic injuries to the head. Frequent concussions from high contact, physical sports over a persons lifetime poses great risk to professional athletes as it is estimated that a minimum 1 in 10 players will develop CTE.

Alzheimer's Disease:
AD is the most common type of dementia. The dominating protein associated with this disease is Amyloid-beta, often referred to as A-β. A-β is normally a soluble protein that can aggregates into insoluble fibers, called Amyloid fibrils. These fibrils are rich in β-sheet secondary structures and share a cross beta core structure. The highly ordered core of the amyloid fibrils is thought to be the basis of its strength for robust resistance to degradation. [25]

Parkinson's disease: Parkinson's is the second most common neurodegenerative disease. The predominant protein is a-synuclein. Alpha-synuclien has been reported to aggregate into poorly organized organized granulofibrillary structures, commonly referred to as Lewy bodies (LBs). LBs are said to have rounded structures with a halo around densely packed filaments. Several reports suggest that a region of the α-Synuclein protein termed NAC (non-amyloidβ-component) is a critical determinant of the fibrillation process of α-synuclein. [26, 27]

Chronic Traumatic Encephalopathy:
CTE is trending a topic in professional sports because of its rapid progression and acute symptoms of extreme personality changes, violently aggressive behavior, and high suicide rate. The primary protein believed to cause CTE is the highly soluble Tau protein. Abnormal Deposits of insoluble, misfolded, and hyperphosphorylated tau proteins collect inside neurons in the form of Neurofibrillary tangles (NFT). NFTs are highly neurotoxic and

synaptoxic. They have been shown appear as flame shaped or globulose lesions. These neurofibrillary fibers have a characteristic morphology of paired helical filaments. [28, 29].

In the following sections we will describe various aspects of protein aggregation and then show the entropy trend of the aggregation of proteins.

2 Protein Aggregation Theory

Protein aggregation is a process which is widely know to be of a dynamical origin that follows the evolution and growth over space and time[30]. Protein aggregation follows the dynamics of self-assembly of normal proteins into large structures which then result in diseases because either the protein is not functioning effectively or has become toxic because of the the proteins getting aggregated [30]. Protein aggregation is a problematic situation in cells because the physical, chemical and biological properties of aggregated proteins is not similar to the original protein. This often results in toxicity and in neurogenerative diseases [30, 31].

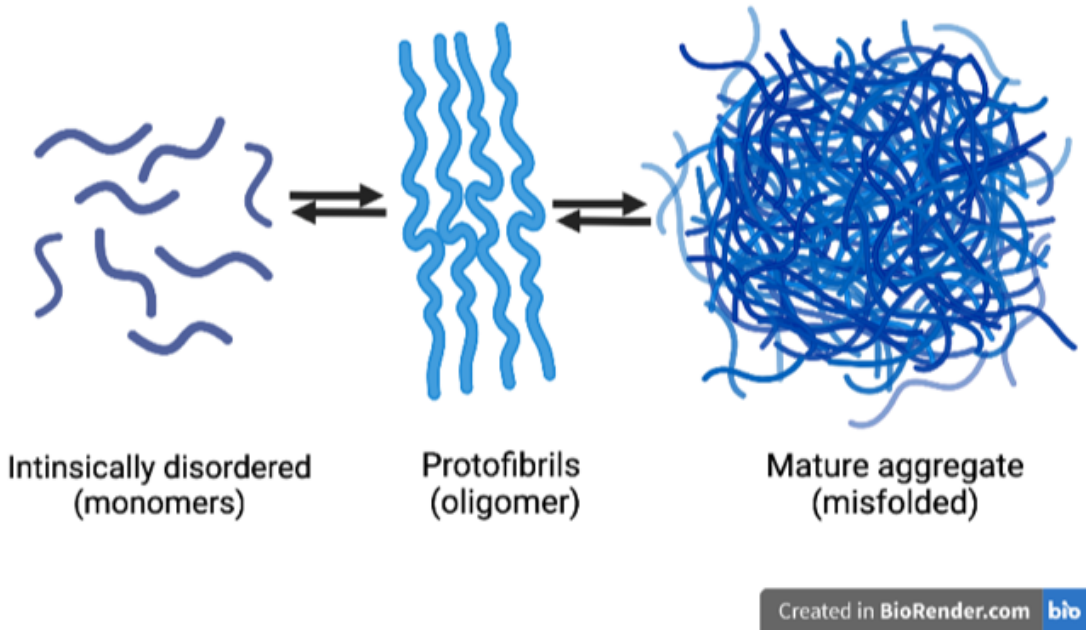


Figure 1. Diagram showing protein aggregation

The proteins in our system undergo various phenomena from its surrounding environment and because of its own properties and this creates an inefficient functioning protein

resulting in protein misfolding. This process of misfolding is irreversible and causes the aggregates which leads to the diseases as mentioned in the previous sections. The process of protein aggregation can be summarized bybits properties [32]. A general protein aggregation model can be described by the process given by,

$$N \xrightarrow{k_a} U \ ; \ U \xrightarrow{k_b} N \ ; \ U \xrightarrow{k_c} A \ ; \ A \xrightarrow{k_d} U \tag{2.1}$$

where N→ natured protein, U→ denatured protein and A$rightarrow$ aggregated protein states. k_a, k_b, k_c, k_d are known as the reaction rates. Eqn. (2.1) can be converted into differential equations using chemical kinetics and expressed as;

$$\frac{dN}{dt} = -k_a N + k_b U \tag{2.2}$$

$$\frac{dU}{dt} = k_a N - k_b U - k_c U + k_d A \tag{2.3}$$

$$\frac{dA}{dt} = k_c U - k_d A \tag{2.4}$$

if we take the sum of eqns. (2.2), (2.3), (2.4) we get,

$$\frac{dN}{dt}\frac{dU}{dt}\frac{dA}{dt} = 0 \tag{2.5}$$

which gives us $A+U+N = constant$. Studies have also shown that the protein folding and misfolding phenomena also depend on the temperature and the time [33]. This is shown in the **Figure 2.** taken from [33].

We can gain a deeper understanding of how the unfolded protein U distribution occurs during protein aggregation as the interaction is happening in the following way [30].
Let x be the population of U at instant t. Let P(x,t) be the probability distribution of x at time t, then P(x,t) can be calculated using eqn. (2.1). The equations using eqn. (2.1) can be constructed as shown in the studies [34, 35] as,

- initially the rate constants are calculated using the equation [35]

$$c_i = V^{(1-\nu_i)} k_i \tag{2.6}$$

  where ν_i is the i-th state change.

- in the next step, the process is assumed to be a Markovian process and the transition probabilities of each reaction is calculated.

- then all the interactions in a molecule are taken into account over a time interval of $(t, t + \Delta t)$.

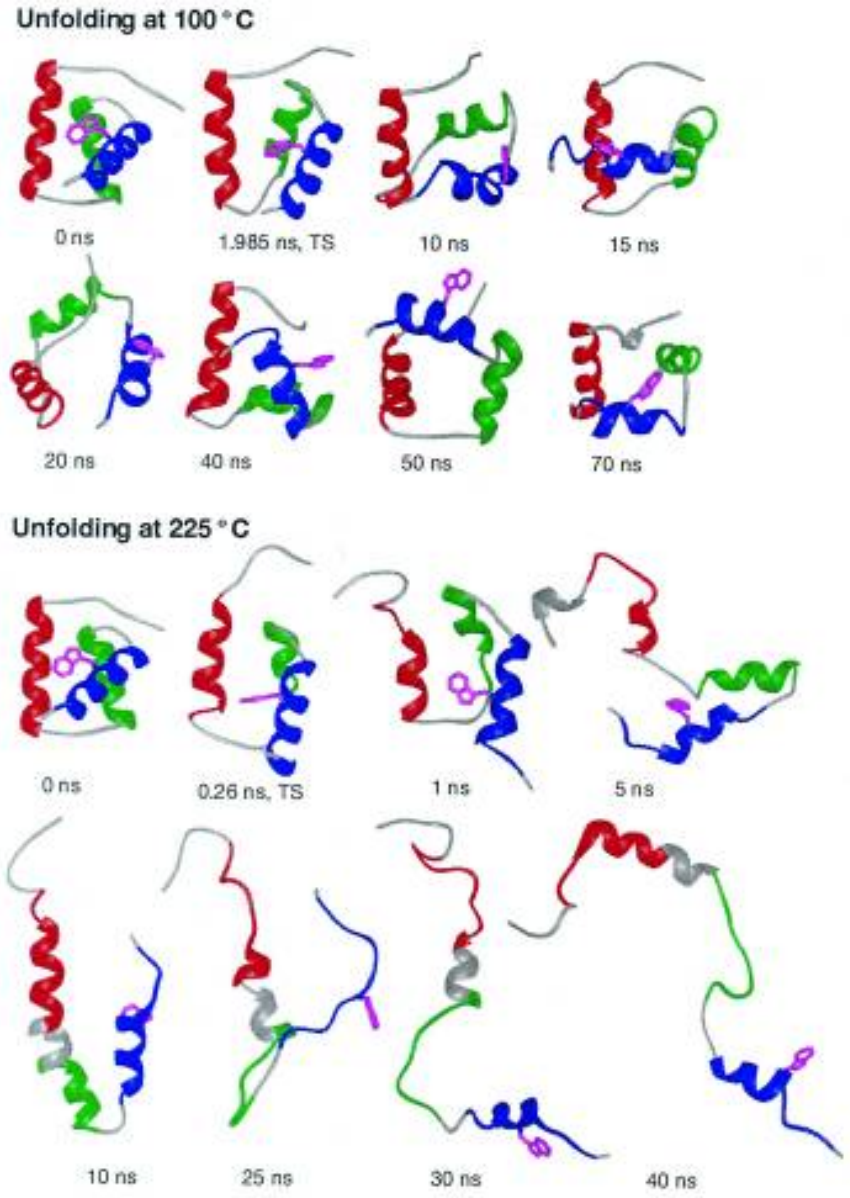


Figure 2. Temperature & time dependence of protein folding & misfolding [33]

- the first order spproximation is assumed such that,

$$\omega_i = f_i(x, N, A)\Delta t + O(\Delta^2) \tag{2.7}$$

where f_i is the function obtained using the molecular interactions [34, 35].

Then the final equation of the probability distrubution is constructed which is given as,

$$\frac{\partial P(x,t)}{\partial t} = (k_a N + k_d A)P_- + (k_b + k_c)(x+1)P_+ + (k_a N + k_b x + k_c x + k_d A)P_0 \tag{2.8}$$

where $P_- \equiv P(x-1,t)$, $P_+ \equiv P(x+1,t)$ and $P_0 \equiv P(x,t)$.
This equation can be solved using the generating function technique. However, the aim

of this study is not to solve eqn.(2.8) but to understand the entropy trend of a protein aggregation.

3 Non-Equilibrium Fokker-Planck Equations

Consider a set of n interacting particles. Let the particles evolve with time through the Langevin equations given by

$$\frac{dx_i}{dt} = f_i(\mathbf{x}) + r_i(t) \tag{3.1}$$

where x_i is the position of the ith particle, $\mathbf{x} = \{x_i\}$, $f_i(\mathbf{x})$ is the force acting on the ith particle, r_i is the noise that is mathematically considered to be a stochastic variable such that

$$\langle r_i(t) \rangle = 0 \tag{3.2}$$

$$\langle r_i(t) r_j(t') \rangle = 2D_i \delta_{ij} \delta(t - t') \tag{3.3}$$

with $D_i \geq 0$, different constants for each particle.The associated Fokker-Planck equations describe how the probability distribution, $P(\mathbf{x}, t)$ evolves with time [36]. This can be written as

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\sum \frac{\partial}{\partial x_i}[f_i(\mathbf{x}) P(\mathbf{x}, t)] + \sum D_i \frac{\partial^2}{\partial x_i^2} P(\mathbf{x}, t) \tag{3.4}$$

We can write down the Fokker-Planck equation in a more convenient way as a continuity equation,

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\sum \frac{\partial}{\partial x_i} J_i(\mathbf{x}, t) \tag{3.5}$$

$$J_i(\mathbf{x}, t) = [f_i(\mathbf{x}) - D_i \frac{\partial}{\partial x_i}] P(\mathbf{x}, t) \tag{3.6}$$

where J_i is the ith component of the current of probability The condition of irreversibility can be expressed as

$$D_i \neq D_j, i \neq j$$

or

$$D_i = D_j = D, i \neq j$$

but

$$\frac{\partial f_j}{\partial x_i} \neq \frac{\partial f_i}{\partial x_j} \tag{3.7}$$

The Fokker-Planck equation has to be solved inside a given region of the space spanned by the set of variables x_i subject to a prescribed boundary condition which governs the behavior of $P(\mathbf{x},t)$ and $J_i(\mathbf{x},t)$. In the thermodynamic equilibrium case the Langevin equation and the associated Fokker-Planck equations which describe a system where

$$\frac{\partial f_j}{\partial x_i} = \frac{\partial f_i}{\partial x_j}$$

for any pair i and j and

$$D_i = D_j \tag{3.8}$$

4 Entropy Production and Fokker-Planck Equations

Cancer cell growth can be considered as an irreversible system and there is continuous production of entropy in such systems. The rate of change of the entropy S of a system can be written as [37]

$$\frac{dS}{dt} = \varsigma - \Omega \tag{4.1}$$

where ς is the entropy production due to the irreversible processes in the system and Ω is the entropy flux from the system to the environment. In an equilibrium system entropy is a well defined quantity but in non-equilibrium systems the entropy as well as the production of entropy is not well defined. Since a non-equilibrium system is defined by the Fokker-Planck equations, hence we have attempted to calculate the production of entropy in such systems. The Gibbs entropy of a system at any time t is given by [36, 38–40]

$$S(t) = -\int P(\mathbf{x},t)\ ln[P(\mathbf{x},t)]\ d\mathbf{x} \tag{4.2}$$

where $d\mathbf{x} = dx_1 dx_2 .....dx_n$. Using eqn. (2.5) we can express the derivative of the entropy as

$$\frac{d}{dt}S(t) = -\int [lnP(\mathbf{x},t) + 1] \sum \frac{\partial}{\partial x_i} J_i(\mathbf{x},t)\ d\mathbf{x} \tag{4.3}$$

Integrating we get

$$\frac{d}{dt}S(t) = -\int \sum J_i(\mathbf{x},t)\ \frac{\partial}{\partial x_i} lnP(\mathbf{x},t)\ d\mathbf{x} \tag{4.4}$$

using eqn. (2.6) we can write

$$\frac{d}{dt}S(t) = -\int \sum \frac{1}{D_i}\ J_i(\mathbf{x},t)\ f_i(\mathbf{x})\ d\mathbf{x} + \int \sum \frac{[J_i(\mathbf{x},t)]^2}{D_i P(\mathbf{x},t)}\ d\mathbf{x} \tag{4.5}$$

comparing this with eqn. (3.1) we see that

$$\Omega = \int \sum \frac{1}{D_i}\ J_i(\mathbf{x},t)\ f_i(\mathbf{x})\ d\mathbf{x} \tag{4.6}$$

and

$$\varsigma = \int \sum \frac{[J_i(\mathbf{x},t)]^2}{D_i P(\mathbf{x},t)}\, d\mathbf{x} \tag{4.7}$$

Using eqn. (2.6) we can write eqn. (3.6) as

$$\Omega = \int \sum \left\{ \frac{1}{D_i}\, [f_i(\mathbf{x})]^2 + f_{ii}(\mathbf{x}) \right\}\, P(\mathbf{x},t)\, d\mathbf{x} \tag{4.8}$$

where $f_{ii}(\mathbf{x}) = \frac{\partial f_i(\mathbf{x})}{\partial x_i}$. This can be expressed as an average over the probability distribution.

$$\Omega = \langle \sum \left\{ \frac{1}{D_i}\, [f_i(\mathbf{x})]^2 + f_{ii}(\mathbf{x}) \right\} \rangle \tag{4.9}$$

There is another study of the total entropy production [41, 42]. The authors have clearly mentioned that the the total entropy production, $\dot{S}_{tot}$ the sum of two constitutive parts, namely so called adiabatic $\dot{S}_a$ and nonadiabatic $\dot{S}_{na}$ contribution. Each of these entropies cannot be less than zero.

5 Entropy Generation and Fokker-Planck Equations

It has been discussed by Jaynes that Gibbs formalism for statistical physics of systems under equilibrium can be understood as a generalized form in a statistical inference theory for non-equilibrium systems [43]. Jaynes developed non-equilibrium statical physics for the stationary state constraint on the basis of maximum entropy, and his approach consisted of maximizing the path. The Shannon information entropy for the path can be written as

$$S = -\sum_{\gamma} p_\gamma ln(p_\gamma) \tag{5.1}$$

with respect to p_γ of the path γ. According to Shannon, the information entropy can be written as the logarithm of the number of outcomes i with non negligible probability p_i, while in non-equilibrium statistical physics it is the given as the logarithm of the number of microscopic phase-space paths γ having non negligible probability p_γ [43, 44] Following this approach, we know that the information entropy for open systems is related to their entropy generation by [45–47]

$$S_g = \kappa_B S = -\kappa_B \int P_\gamma(\mathbf{x},t) ln[P_\gamma(\mathbf{x},t)] d\mathbf{x} \tag{5.2}$$

with $p_\gamma = P_\gamma(\mathbf{x},t)$. This relation is the statistical definition of entropy generation. This can also be explained as the missing information which is necessary for predicting which path a system of the ensemble takes during the transition from one state to another. The Guoy-Stodola theorem [44] gives

$$\bar{W} = T_0 S_g \tag{5.3}$$

where $\bar{W}$ is work lost due to internal irreversibility in a system. By definition, the entropy generation can be related to the power lost, P due to irreversibility,

$$S_g = \frac{1}{T_0}\int_0^\tau P dt \tag{5.4}$$

wihere T_0 is the environmental temperature, considered constant and τ is the time duration of a physical process. The power lost by definition is given as,

$$P = \langle \sum f_i(\mathbf{x}) \frac{dx_i}{dt} \rangle \tag{5.5}$$

Using the Langevin equation we can write this as

$$P = \langle \sum f_i(\mathbf{x})[f_i(\mathbf{x}) + r_i(t)] \rangle \tag{5.6}$$

and so S_g can be written as

$$S_g = \frac{\tau}{T_0} \langle \sum ([f_i(\mathbf{x})]^2 + D_i f_{ii}(\mathbf{x})) \rangle \tag{5.7}$$

where $f_{ii} = \frac{\partial f_i}{\partial x_i}$ Considering the mean value, we can finally write this as

$$S_g = \frac{\tau}{T_0} \int \sum ([f_i(\mathbf{x})]^2 + D_i f_{ii}(\mathbf{x})) P_\gamma(\mathbf{x}, t) d\mathbf{x} \tag{5.8}$$

and hence

$$S_g = \frac{\tau}{T_0} \int \sum f_i(\mathbf{x}) J_i(\mathbf{x}, t)\ d\mathbf{x} \tag{5.9}$$

where the last term is related with the Fokker-Planck equation.

6 Model Independent Protein Aggregation

Based on the understanding that the aggregation in protein we have constructed a growth function. This growth function is given by,

$$f_i(r) = \frac{r^3(1-r^2)(1-r)\exp(-\alpha r^2)}{(\beta r^6 + \delta)} \tag{6.1}$$

where α, β and δ are constants related to the strength of the force.
For the entropy production approach, using the force term we can write eqn. (3.9) as

$$\Omega = \langle \left\{ \frac{1}{D}\ \frac{r^6(1-r^2)^2(1-r)^2 \exp(-2\alpha r^2)}{(\beta r^6 + \delta)^2} + f_{ii} \right\} \rangle \tag{6.2}$$

where f_{ii} is given by

$$f_{ii} = \frac{(2\alpha\beta)(r^2)(\exp^{\alpha r^2})}{(\beta r^6 + \delta)} \Big[r^{11} - r^{10} - r^9 + (\frac{1}{2\alpha} - 1) r^8 + (\alpha) r^7 - (\frac{3\alpha}{2}) r^6 + (\frac{\delta}{\beta}) r^5 - (\frac{\delta}{\beta}) r^4 - (\frac{\delta}{\beta})(3-\alpha) r^3 - (\frac{\delta}{2\alpha\beta})(2\alpha - 5) r^2 - (\frac{2\delta}{\alpha\beta}) r + (\frac{3\delta}{2\alpha\beta}) \Big]$$

also, eqn. (2.6) can be written as

$$J(r,t) = [\frac{r^3(1-r^2)(1-r)\exp(-\alpha r^2)}{(\beta r^6+\delta)} - D\frac{\partial}{\partial r}]P(r,t) \tag{6.3}$$

and we can write eqn. (3.6) as

$$\varsigma = \frac{1}{D}\int \frac{[[\frac{r^3(1-r^2)(1-r)\exp(-\alpha r^2)}{(\beta r^6+\delta)} - D\frac{\partial}{\partial r}]P(r,t)]^2}{P(r,t)}\, dr \tag{6.4}$$

and finally we can express eqn. (3.1) as

$$\frac{dS}{dt} = \frac{1}{D}\int \frac{[[\frac{r^3(1-r^2)(1-r)\exp(-\alpha r^2)}{(\beta r^6+\delta)} - D\frac{\partial}{\partial r}]P(r,t)]^2}{P(r,t)}\, dr \; - \; \langle\left\{\frac{1}{D}\,\frac{r^3(1-r^2)(1-r)\exp(-\alpha r^2)}{(\beta r^6+\delta)} - f_{ii}]\right\}\rangle \tag{6.5}$$

Similarly, for the entropy generation approach we can express eqn. (4.9) as

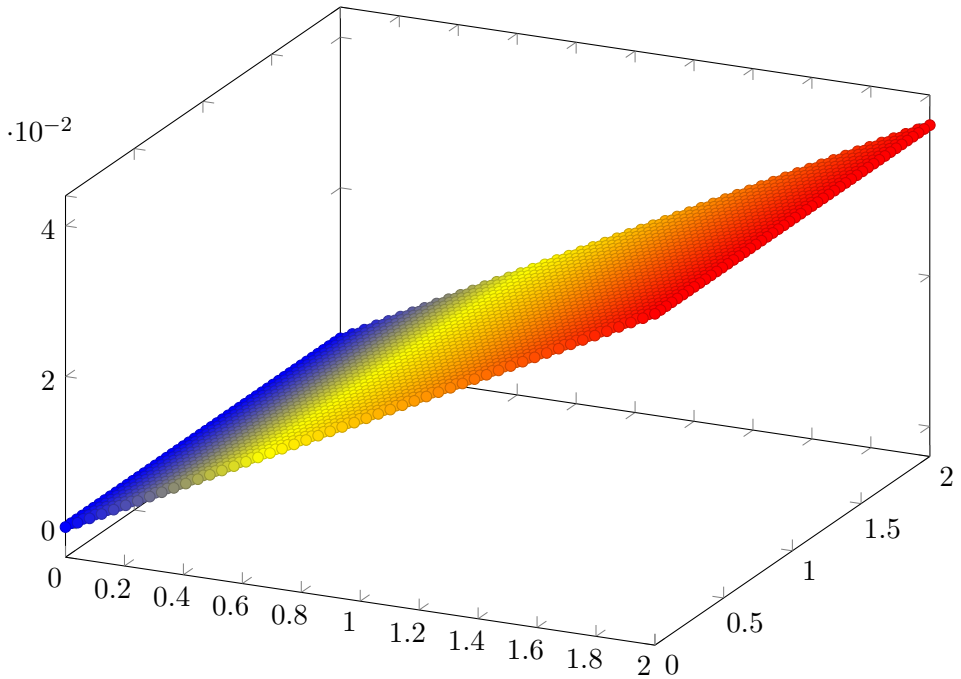
$$S_g = \frac{\tau}{T_0}\int \left\{\frac{r^3(1-r^2)(1-r)\exp(-\alpha r^2)}{(\beta r^6+\delta)} - D f_{ii}]\right\} P_\gamma(r,t)dr \tag{6.6}$$

where f_{ii} is given as shown above.

7 Entropy Trend in Model Independent Protein Aggregation

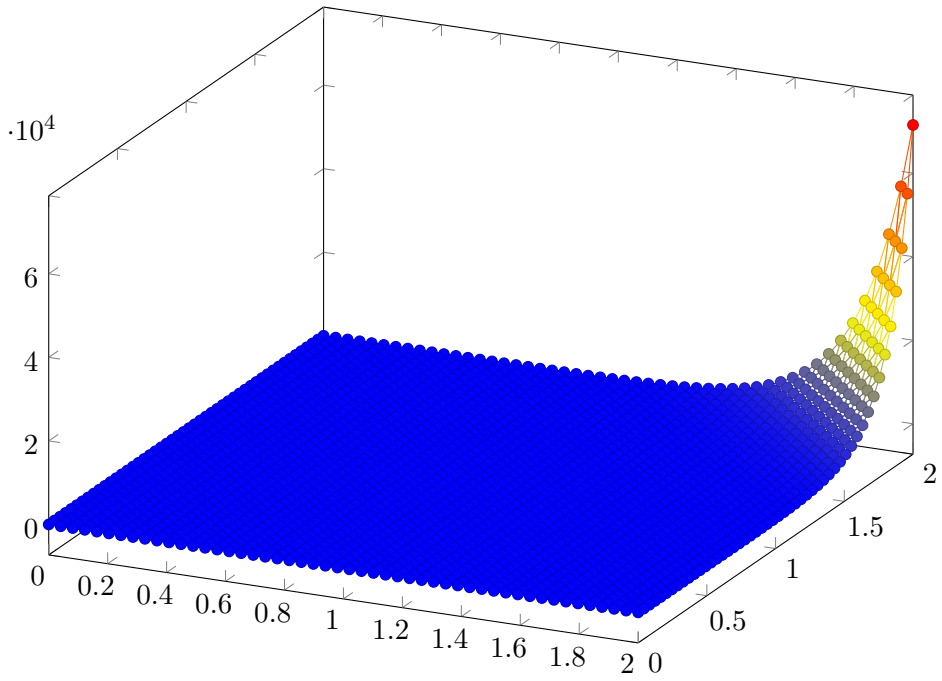
The results of the entropy trend in a model independent protein aggregation is shown below. This is divided into 3 parts.

7.1 Early Stages



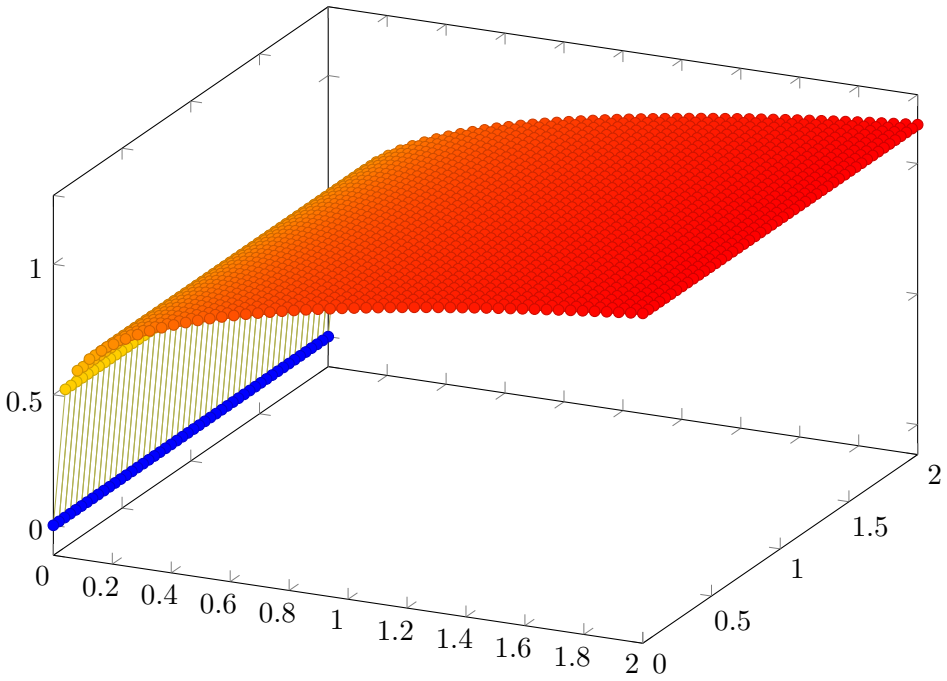
In the figure above we see that the initial stages of the protein aggregation process takes place very slow. This is clearly evident from the plot shown above.

7.2 Intermediate Stages



In the intermediate step, the aggregation speeds up and grows very fast. This is again evident from the plot shown which shows the trend in the entropy increase. This is a very interesting result because of the way protein aggregates in the intermediate stages. The significance of this can be very improtant from the the point of view of the dynamics, physics, chemical nature and also biomedical understanding of the aggregation.

7.3 Final Stages



This plot above shows the slowing down of the aggregation after reaching a saturation stage. We think that since most of the protein has already aggregated and the functioning of the aggregate is fulfilled, so the information is relayed to the aggregate to slow down since most of the aggregation process has been fulfilled.

8 Conclusion

In this article we have described the fundamentals of protein aggregation and how it affects in terms of various diseases. The sections also talk about Langevin equations and Fokker-Planck equations and how to calculate the entropy. Finally we have taken an example of a model independent protein aggregation growth and shown the entropy trend. We see that the entropy increase in the initial stages of the aggregation process is not very steep. This could be because at the start the protein is just starting to aggregate and does not yet reach a state where the aggregation speeds up.
In the intermediateb stage we see that the aggregation process speeds up. This is shown in the entropy plot. There is a very steep rise in the entropy when the aggregation is undergoing its maximum rate.
Finally, the entropy in the final stages is shown to reach almost saturation and the entropy increase is not very steep.
Our next step is understanding this dynamics of the protein aggregation and why it behaves in the way shown through the entropy trend.

Ethics and Consent to Participate:

No human or clinical trials were used in this study. This study is about applying physics principles to protein aggregation.

Consent to Publish:

The authors give consent to publish.

Conflict of Interest:

The authors declare no conflict of interest.

Funding Information:

The authors do not receive any external funding.

References

[1] K. Konstantoulea, P. Guerreiro, M. Ramakers, N. Louros, L. D. Aubrey, B. Houben et al., *Heterotypic amyloid β interactions facilitate amyloid assembly and modify amyloid structure*, *The EMBO Journal* (2021) e108591.

[2] A. G. Torres, M. Rodríguez-Escribà, M. Marcet-Houben, H. G. Santos Vieira, N. Camacho, H. Catena et al., *Human trnas with inosine 34 are essential to efficiently translate eukarya-specific low-complexity proteins*, *Nucleic Acids Research* (2021) .

[3] G. Heilbronner, Y. S. Eisele, F. Langer, S. A. Kaeser, R. Novotny, A. Nagarathinam et al., *Seeded strain-like transmission of -amyloid morphotypes in app transgenic mice*, *EMBO reports* **14** (2013) 1017 [https://www.embopress.org/doi/pdf/10.1038/embor.2013.137].

[4] C. Haass and D. J. Selkoe, *Soluble protein oligomers in neurodegeneration: lessons from the alzheimer's amyloid β-peptide*, *Nature reviews Molecular cell biology* **8** (2007) 101.

[5] R. F. Rosen, B. J. Ciliax, T. S. Wingo, M. Gearing, J. Dooyema, J. J. Lah et al., *Deficient high-affinity binding of pittsburgh compound b in a case of alzheimer's disease*, *Acta neuropathologica* **119** (2010) 221.

[6] M. P. Mattson and T. V. Arumugam, *Hallmarks of brain aging: adaptive and pathological modification by metabolic states*, *Cell metabolism* **27** (2018) 1176.

[7] S. Camandola and M. P. Mattson, *Brain metabolism in health, aging, and neurodegeneration*, *The EMBO journal* **36** (2017) 1474.

[8] Y. Hou, X. Dan, M. Babbar, Y. Wei, S. G. Hasselbalch, D. L. Croteau et al., *Ageing as a risk factor for neurodegenerative disease*, *Nature Reviews Neurology* **15** (2019) 565.

[9] C. P. Doherty, S. M. Ulamec, R. Maya-Martinez, S. C. Good, J. Makepeace, G. N. Khan et al., *A short motif in the n-terminal region of α-synuclein is critical for both aggregation and function*, *Nature structural & molecular biology* **27** (2020) 249.

[10] H. Shaykhalishahi, A. Gauhar, M. M. Wördehoff, C. S. Grüning, A. N. Klein, O. Bannach et al., *Contact between the β1 and β2 segments of α-synuclein that inhibits amyloid formation*, *Angewandte Chemie International Edition* **54** (2015) 8837.

[11] M. Carballo-Pacheco and B. Strodel, *Advances in the simulation of protein aggregation at the atomistic scale*, *The journal of physical chemistry B* **120** (2016) 2991.

[12] A. Morriss-Andrews and J.-E. Shea, *Simulations of protein aggregation: insights from atomistic and coarse-grained models*, *The Journal of Physical Chemistry Letters* **5** (2014) 1899.

[13] I. M. Ilie and A. Caflisch, *Simulation studies of amyloidogenic polypeptides and their aggregates*, *Chemical reviews* **119** (2019) 6956.

[14] S. Sharma, P. Modi, G. Sharma and S. Deep, *Kinetics theories to understand the mechanism of aggregation of a protein and to design strategies for its inhibition*, *Biophysical chemistry* **278** (2021) 106665.

[15] S. E. Youn, H.-C. Kang, H. D. Kim et al., *Long-term results of vagus nerve stimulation in children with dravet syndrome: Time-dependent, delayed antiepileptic effect*, *Epilepsy Research* **174** (2021) 106665.

[16] D. Samanta, *Changing landscape of dravet syndrome management: an overview*, *Neuropediatrics* **51** (2020) 135.

[17] M. Iqbal, C. Li, K. Wood, B. Jiang, T. Takei, O. Dag et al., *Continuous mesoporous pd films by electrochemical deposition in nonionic micellar solution*, *Chemistry of Materials* **29** (2017) 6405.

[18] X. Jiang, J. Chen, A. Bajić, C. Zhang, X. Song, S. L. Carroll et al., *Quantitative real-time imaging of glutathione*, *Nature communications* **8** (2017) 1.

[19] T. R. Rieger, R. I. Morimoto and V. Hatzimanikatis, *Bistability explains threshold phenomena in protein aggregation both in vitro and in vivo*, *Biophysical journal* **90** (2006) 886.

[20] T. R. Rieger, *Mathematical modeling of the eukaryotic heat shock response and associated ultrasensitive signaling cascades*, Ph.D. thesis, Northwestern University, 2005.

[21] T. Scheidt, J. A. Carozza, C. C. Kolbe, F. A. Aprile, O. Tkachenko, M. M. Bellaiche et al., *The binding of the small heat-shock protein αb-crystallin to fibrils of α-synuclein is driven by entropic forces*, *Proceedings of the National Academy of Sciences* **118** (2021) .

[22] M. Carballo-Pacheco, A. E. Ismail and B. Strodel, *On the applicability of force fields to study the aggregation of amyloidogenic peptides using molecular dynamics simulations*, *Journal of chemical theory and computation* **14** (2018) 6063.

[23] J. A. Luiken and P. G. Bolhuis, *Primary nucleation kinetics of short fibril-forming amyloidogenic peptides*, *The Journal of Physical Chemistry B* **119** (2015) 12568.

[24] S. L. Murphy, K. D. Kochanek, J. Xu and E. Arias, *Mortality in the united states, 2020*, .

[25] W. J. Lennarz and M. D. Lane, *Encyclopedia of biological chemistry*. Academic Press, 2013.

[26] M. Bisaglia, A. Trolio, M. Bellanda, E. Bergantino, L. Bubacco and S. Mammi, *Structure and topology of the non-amyloid-β component fragment of human α-synuclein bound to micelles: Implications for the aggregation process*, *Protein science* **15** (2006) 1408.

[27] H. Winn, *Youmans and winn neurological surgery. ed*, 2017.

[28] R. B. Daroff and M. J. Aminoff, *Encyclopedia of the neurological sciences*. Academic press, 2014.

[29] S. A. Youssef, *Pathology of brain aging and animal models of neurodegenerative diseases*, in *Conn's Handbook of Models for Human Aging*, pp. 899–908, Elsevier, (2018).

[30] A. M. Miangolarra, A. Duperray-Susini, M. Coppey and M. Castellana, *Two timescales control the creation of large protein aggregates in cells*, *Biophysical Journal* **120** (2021) 2394.

[31] U. Upadhyay, C. Barua, S. Devi, J. P. Kumar and R. Singh, *Dynamics of unfolded protein aggregation*, *arXiv preprint arXiv:2111.04103* (2021) .

[32] R. Lumry and H. Eyring, *Conformation changes of proteins*, *The Journal of physical chemistry* **58** (1954) 110.

[33] U. Mayor, C. M. Johnson, V. Daggett and A. R. Fersht, *Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation*, *Proceedings of the National Academy of Sciences* **97** (2000) 13518.

[34] D. A. McQuarrie, *Stochastic approach to chemical kinetics*, *Journal of applied probability* **4** (1967) 413.

[35] D. T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, *The journal of physical chemistry* **81** (1977) 2340.

[36] T. Tome, *Entropy production in nonequilibrium systems described by a Fokker-Planck equation*, *Brazilian Journal of Physics* **36** (2006) 1285 .

[37] G. Nicolas and I. Prigogine, *Self-Organization in Nonequilibrium Systems*. (Wiley, New York), 1977.

[38] T. Tome and M. J. T. DeOliviera, , *Phys. Rev. E* **82** (2010) 021120.

[39] T. Tome and M. J. T. DeOliviera, , *Phys. Rev. Lett.* **108** (2012) 020601.

[40] T. Tome and M. J. T. DeOliviera, , *Phys. Rev. E* **91** (2015) 042140.

[41] M. Esposito and C. Van den Broeck, , *Phys. Rev. E* **82** (2010) 011143.

[42] C. Van den Broeck and M. Esposito, , *Phys. Rev. E* **82** (2010) 011144.

[43] R. Dewar, *Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states*, *J. Phys. A: Math. Gen.* **36** (2003) 631.

[44] U. Lucia, *Entropy generation and the fokker-planck equation*, *Physics A* **393** (2014) 256.

[45] U. Lucia, *Statistical approach of the irreversible entropy variation*, *Physica A: Statistical Mechanics and its Applications* **387** (2008) 3454.

[46] U. Lucia, *Irreversibility, entropy and incomplete information*, *Physica A: Statistical Mechanics and its Applications* **388** (2009) 4025.

[47] U. Lucia, *Maximum entropy generation and κ-exponential model*, *Physica A Statistical Mechanics and its Applications* **389** (2010) 4558.