

Physical Biology



PAPER

Observation of two-step aggregation kinetics of amyloid- β proteins from fractal analysis

RECEIVED
24 January 2022

REVISED
21 March 2022

ACCEPTED FOR PUBLICATION
5 April 2022

PUBLISHED
29 April 2022

Soham Mukhopadhyay¹ , Subhas C Bera²  and Kabir Ramola^{1,*}

¹ Centre for Interdisciplinary Sciences, Tata Institute of Fundamental Research, Hyderabad 500046, India

² Interdisciplinary Center for Clinical Research (IZNF), Friedrich-Alexander-University, Erlangen-Nürnberg (FAU), Cauerstr. 3, 91058 Erlangen, Germany

* Author to whom any correspondence should be addressed.

E-mail: sohamm@tifrh.res.in and kramola@tifrh.res.in

Keywords: protein aggregation, fractal analysis, image analysis, TIRF microscopy

Abstract

Self-aggregation in proteins has long been studied and modeled due to its ubiquity and importance in many biological contexts. Several models propose a two step aggregation mechanism, consisting of linear growth of fibrils and secondary growth involving branch formation. Single molecule imaging techniques such as total internal reflection fluorescence (TIRF) microscopy can provide direct evidence of such mechanisms, however, analyzing such large data-sets is challenging. In this paper, we analyze for the first time, images of growing amyloid fibrils obtained from TIRF microscopy using the techniques of fractal geometry, which provides a natural framework to disentangle the two types of growth mechanisms at play. We find that after an initial linear growth phase, identified by a plateau in the average fractal dimension with time, the occurrence of branching events leads to a further increase in the fractal dimension, with a final saturation value ≈ 2 . This provides direct evidence of the two-step nature of the aggregation kinetics of amyloid- β proteins, with an initial linear elongation phase followed by branching at later times.

1. Introduction

Proteins are amongst the most ubiquitous biological molecules, responsible for controlling and catalyzing the many chemical processes that make life possible. An important property of proteins that governs the high degree of selectivity in their functionality, is their ability to fold into intricate and unique three dimensional structures. Since the proper folding of protein molecules is critical to their function, misfolding typically leads to many detrimental effects on their related biological processes. Of particular importance among such effects is the formation of filamentous aggregates termed amyloids [1–4]. Many proteins and peptide fragments can form amyloid aggregates [5, 6] which have been implicated in the pathology of several diseases with high morbidity and mortality, such as Alzheimer's, Parkinson's, type-II diabetes, etc [7–10].

In this context, understanding the mechanisms through which amyloid aggregation occurs is essential to design therapeutic and preventative strategies against such diseases. Over the past few decades the process of amyloid aggregation has been

studied widely, with the aim of understanding the microscopic steps behind it. Ensemble experiments that study amyloid aggregation through measuring changes in a macroscopic quantity such as the total fluorescence intensity, combined with global curve fitting over a range of initial concentrations have produced many insights into the mechanisms that drive amyloid aggregation [11]. Such analyses have allowed researchers to identify that the mechanism of amyloid aggregation of the A β 42 peptide—involved in the pathology of Alzheimer's disease—contains a secondary nucleation step [12]. Further research suggests that this secondary nucleation step is driven by the surface of the fibrils providing a nucleation site for monomers in the solution phase [13].

Owing to the small size of aggregation clusters, ensemble experiments alone are insufficient to elucidate and confirm the mechanistic details of the aggregation process of such proteins. It is therefore important to develop single molecule methods which can monitor the aggregation process with a much higher resolution—at the level of a fibril. Several such methods have been designed and employed to study aggregation; including techniques

such as epifluorescence microscopy [14], atomic force microscopy [15–18], and total internal reflection fluorescence microscopy (TIRFM) [19–23]. Superresolution microscopy techniques have also been applied to the problem of imaging the protein aggregation process for amyloid- β [24], α -synuclein [25] and huntingtin [26]. Studies have also focused on attempting to image protein aggregation inside cells [27, 28]. Features unavailable from conventional ensemble experiments—such as measurements of the shapes and sizes of individual growing fibrils—can be accessed from these single molecule techniques. However, since they monitor the aggregation process at such high resolution, single molecule techniques generate large amounts of data. Therefore, in order to truly take advantage of these techniques, methods that can sort through and analyze large volumes of such data and extract the relevant microscopic parameters need to be developed. A technique often utilized in past works to analyze total internal reflection fluorescence (TIRF) microscopy images is to count the number of fibrils present inside the field of view and calculate how the length of the fibrils changes from frame to frame. In this way, estimates for the rates of linear elongation of different fibrils can be obtained. However, it is difficult to obtain an estimate for the rates of surface-catalyzed secondary nucleation—the process that leads to the formation of new branches—from this method.

In this paper we develop an analysis method of experimental images that identifies clusters and measures their fractal characteristics. This allows us to distinguish between linear growth and branching growth. The technique of fractal analysis is widely used to study irregularities in geometric objects or in signals obtained from irregular natural phenomena. Fractal analysis has been used in the past to characterized aggregation in other areas such as nanoparticle aggregation [29] and the formation of soot aggregates during combustion of wood [30], as well as to study diffusion-limited aggregation [31]. Combined with TIRF microscopy and cluster analysis, which allows us to identify and analyse protein aggregates *in vitro* in an automated way, fractal geometry provides us the tools to measure and describe the physical properties of protein aggregates.

2. Materials and methods

Synthetic A β 42 was purchased from AAPPTec LLC (Louisville, KY, USA). All other chemicals were purchased from Sigma (USA) unless otherwise mentioned. Powder A β 42 (1 mg) was dissolved in 2 ml of ice-cold 5 mM NaOH and filtered through a 0.22 μ M syringe filter before injection to size exclusion chromatography for further purification using a Superdex peptide column (GE Healthcare, USA) in 5 mM NaOH containing 1 mM EDTA and 5

mM beta-mercaptoethanol (β ME). The aggregation experiments were performed following dilution of the freshly purified A β 42 solution to the desired concentration in the aggregation buffer (20 mM phosphate buffer, pH 7.4 containing 150 mM NaCl, 1 mM EDTA, 5 mM β ME) directly on a clean Petri dish. Lower concentrations were achieved by serial dilution i.e. diluting from the adjacent higher concentration. All samples were supplemented with 4 μ M thioflavin T (ThT) as a fluorescent marker.

A glass bottom Petri dish (4 well, Cellvis, USA) was cleaned with 10 M NaOH for 15 minutes followed by a thorough rinsing with MiliQ-water. The surface was further cleaned with 70% v/v EtOH/water solution afterwards and finally washed with MiliQ-water before addition of samples. A β 42 aggregation was started by dilution of the freshly purified sample in the aggregation buffer as mentioned above in the Petri dish and monitored on a home-built TIRF setup (see appendix A). The Petri dish was enclosed inside a customized aluminium block with proper humidification to prevent the sample from drying out and to maintain uniform temperature. The progress of aggregation was recorded by imaging the ThT fluorescence at multiple positions every 30 minutes with optimal laser exposure (200 msec) and power (\sim 10 mW). ThT shows very strong fluorescence in the presence of A β 42 aggregates and the resultant images were used for further analysis. Figure 1 shows example images of the aggregation process from different stages of aggregation obtained using TIRF microscopy.

To study the growth mechanisms that govern the dynamics of the clusters, it is necessary to isolate them from the image background. We accomplish this by thresholding followed by clustering using the density-based clustering algorithm DBSCAN (see supplementary information). Once the clusters have been isolated and identified, we classify them by their fractal dimension d_f . We use a box-counting algorithm detailed below to obtain an estimate of d_f for each cluster. Figure 2 shows a schematic diagram outlining our analysis pipeline.

2.1. Image registration

Image registration is the process of aligning images taken from different viewpoints, or at different times. Especially in microscopy experiments that are meant to acquire long time series data, the field of view under observation can shift slightly over time. Image registration, therefore, is a necessary first step that must be carried out before any other kind of processing or analysis can be done on the images.

For registering the images of the amyloid fibrils in our studies, we used the `pyStackReg` Python library. This library is a Python port of the ImageJ plugin `StackReg`, which in turn is based on the image registration method presented in [32].

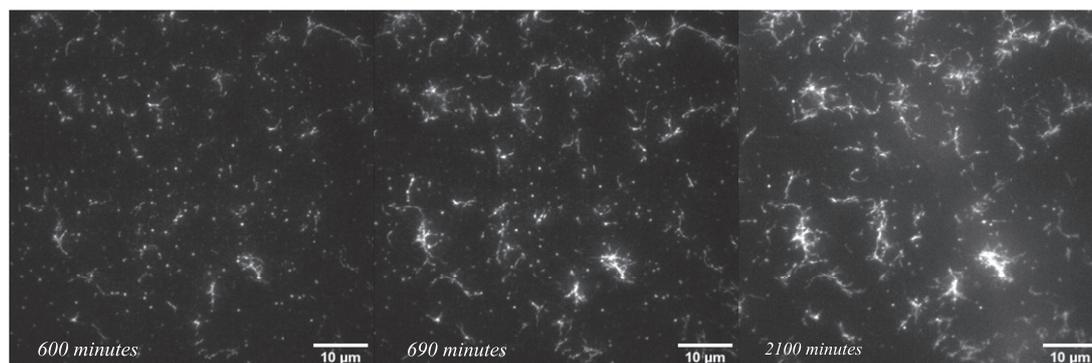


Figure 1. Images from the aggregation process at three different stages. (Left) This image is at 600 minutes, the very end of the initial lag phase, and the beginning of the explosive growth phase. Some of the larger aggregates visible in the later images have started forming. (Center) This image is at 690 minutes, in the middle of the exponential growth phase. The larger aggregates seen in the final image have already taken on the basic structure. (Right) This image is at 2100 minutes, at the very end of the saturation phase, where all aggregates have fully formed.

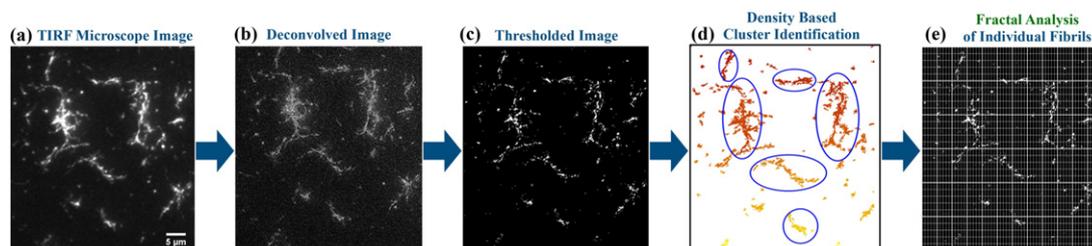


Figure 2. Schematic diagram describing the analysis methodology developed in this paper. After obtaining the raw images from TIRF microscopy (a), we deconvolve the images with the fitted PSF (point spread function) of the TIRF microscope (b). Following that, we threshold the images in order to separate the clusters from the background (c). Once the thresholded clusters are obtained, we employ the clustering algorithm DBSCAN to extract the location of the clusters from the image (d). Finally, we employ the box-counting algorithm in order to calculate the fractal dimension of each localized cluster at every frame (e).

2.2. Deconvolution

The widths of amyloid fibrils are of the order of 5–20 nm [33–35]; far below the diffraction limit of modern microscopes and imaging systems, which have a minimum resolvable distance of ≈ 100 nm. As a result, images of amyloid fibrils are blurred and deformed according to the point-spread function (PSF) of the microscope and imaging system.

The PSF describes the response of an imaging system to a point source. In a fluorescent microscope, which is a non-coherent imaging system, the imaging process is linear, which is to say:

$$\text{Image}(\text{object}_1) + \text{Image}(\text{object}_2) = \text{Image}(\text{object}_1 + \text{object}_2). \quad (1)$$

The linearity property means that an image from a fluorescent microscope can be treated as a sum of the images of multiple point sources, each of which has undergone blurring in line with the PSF of the microscope. In essence, the obtained image is a convolution of the ground truth image with the PSF. This suggests that in principle, it should be possible to de-blur the image and recover the ground truth image by deconvolving with the PSF. In practice, however, deconvolving a sub-resolution image with the PSF to restore it is a highly challenging and ill-posed problem. Like

other image restoration tasks, infinite solutions to this problem are possible in the high-dimensional space of all feasible images.

The Richardson–Lucy algorithm [36, 37] is an iterative scheme to recover an image that has been blurred by a known PSF. In the current study, the PSF of the microscopy system was estimated by fitting a two-dimensional Gaussian function to the image of fluorescent beads. The beads were 100 nm in diameter, their small size enabling them to serve as sub-resolution point sources. The deconvolution was carried out frame-by-frame on every frame of an image stack by using this fitted PSF as the input to the implementation of the Richardson–Lucy algorithm available in the scikit-image Python module [38].

2.3. Thresholding

Before cluster analysis can be carried out on aggregation images, it is necessary to threshold the images in order to separate the fibrils which make up the foreground of our images from the background. Without this step, there is no way to signify to the clustering algorithm which pixels make up the fibrils and should be considered for clustering. Furthermore, since the illumination in our images is non-uniform—as the illumination comes from the fibrils themselves, and

the intensity of each fibril varies depending upon its mass—an adaptive thresholding algorithm that can calculate a separate threshold for each local region of the image must be used.

For the purpose of our study, we have employed the adaptive thresholding algorithms available in the scikit-image module in Python [38]. The scikit-image algorithm operates by first calculating the threshold for an individual pixel as the Gaussian weighted mean of the intensity values in the local neighbourhood of the pixel, then subtracting a constant, called the **offset**, from that value. The ‘local neighbourhood’ of a pixel is defined as all the pixels falling inside a square matrix centered on the pixel of interest. The size of this square matrix, in terms of numbers of pixels, as well as the **offset** are both user-definable parameters in the algorithm. In our studies, we set the offset and size parameters manually for stacks of images obtained from different experiments. We use the same offset and size parameter values for all image frames within a single stack. Generally it was found that a neighbourhood of size $\approx 150 \times 150$ pixels and an offset of ≈ 120 worked for images obtained from most experimental conditions. The thresholding algorithm outputs a binary image with all the pixels classified as ‘true’, or ‘false’. For the purposes of displaying the image, the ‘true’ pixels are defined to have the maximum intensity for the image type (255 for a eight-bit image, 65 535 for a 16-bit image and so on), while the ‘false’ pixels are defined to have 0 intensity.

2.4. The DBSCAN clustering algorithm

In our study, we employ clustering algorithms for the purpose of detecting fibrils, which show up as clusters of ‘true’ pixels against background of ‘false’ pixels after thresholding an image. These cluster can be of arbitrary shapes and sizes. In addition, the number of clusters is not fixed, and may change arbitrarily from the beginning to the end of an experiment. As one of the aims of this study has been to automate the method of analysis as far as possible, density-based clustering methods appear to be the more logical choice for the clustering task at hand, under the constraints described. Of the density-based clustering methods, the DBSCAN algorithm [39], is the most well-known and most used. We therefore decided on DBSCAN as the algorithm of choice for detection of fibrils.

The DBSCAN algorithm requires two parameters—a distance cutoff *eps* and a minimum number of points within that cutoff, *min_samples*—to define a cluster. These two parameters together define a global density threshold. The algorithm then checks every point in the database and labels them whether as part of a cluster or as a ‘noise’ point. In this study, we have applied the implementation of DBSCAN available in the scikit-learn module [40] of the Python programming language.

2.5. Estimation of fractal dimension

For the empirical estimation of fractal dimensions, especially in computational applications, the box-counting dimension, also referred to as the Minkowski–Bouligand dimension, is very widely used. Mathematically, the box-counting dimension is defined as follows [41]: let F be a non-empty bounded subset of the n -dimensional Euclidean space \mathbb{R}^n . (For the specific case of the analysis of images, such as the ones in the current study, the relevant space is the two-dimensional Euclidean space, corresponding to the flat surface of a piece of paper or the computer monitor. As a corollary, for the case of image analysis, F will be the region of that two-dimensional space that is occupied by the image, in other words, F is the image of an entire field-of-view, when talking about the average fractal dimension. When talking about the fractal analysis of individual fibrils, F is the portion of the image where that fibril resides.) Let $N_\delta(F)$ be the smallest number of sets of diameter at most δ that can cover F . Then, the box-counting dimension of F is defined as:

$$\dim_B F = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta}. \quad (2)$$

The above definition, while mathematically exact, does not lend itself to an easy implementation for empirical estimation of the dimension. For such cases, there are other equivalent definitions of the box-counting dimension that are easy to implement computationally. Let us consider a mesh of n -dimensional cubes in \mathbb{R}^n , such that the side of each cube is of the length δ . So the cubes will be of the form:

$$[m_1\delta, (m_1 + 1)\delta] \times \cdots \times [m_n\delta, (m_n + 1)\delta], \quad (3)$$

where $m_1 \dots m_n$ are integers. Again, for the two-dimensional case, the mesh of n -dimensional δ —cubes is simply a square grid of grid size δ . Now if we take $N'_\delta(F)$ as the number of δ -mesh cubes that intersect F , then at the limit $\delta \rightarrow 0$, we can replace $N_\delta(F)$ in equation (2) with $N'_\delta(F)$.

This definition for the box-counting dimension is the one most widely used for empirical estimation of the ‘fractal’ dimension of a set F . Essentially, it corresponds to laying the fractal set F out on an evenly spaced grid, calculating how many boxes are required to cover the set, and then plotting how the number of boxes changes as the grid is made finer and finer. This definition also suggests a way to connect the value given by the box-counting dimension to a tangible feature of the set. One can interpret the number of the δ -mesh cubes that intersect with the set F as a mark of how irregular the set is when examined at the scale δ . So the box-counting dimension of an irregular set F indicates how rapidly irregularities develop as $\delta \rightarrow 0$. The box-counting dimension is most commonly estimated by some version of the box-counting algorithm, originally introduced as the ‘reticular

cell counting' method by Gagnepain and Roques-Carnes [43].

For applying the box-counting method in practice, one of the first steps is to decide upon the sizes of the boxes (δ)—in terms of pixels, for image data—that will be used. The most common strategy in this regard is to use the geometric-step (GS) method [44]. In the GS method, box sizes are chosen as powers of 2—thus, for example, a possible set of box sizes (in terms of pixels) for an image that is 1000×1000 pixels could be: $2 \times 2, 4 \times 4, 8 \times 8, 16 \times 16, 32 \times 32, 64 \times 64, 128 \times 128, 256 \times 256$ and 512×512 pixels. However, when this method is applied to images of an arbitrary size $M \times N$ pixels, where M and N may not necessarily be powers of 2, there will be an unavoidable loss of information from the regions of the image that are close to the edges. This issue assumes particular importance in our case when one considers that we are attempting to calculate the fractal dimensions of individual clusters, whose linear dimensions are $\sim 0.1 \times$ the dimensions of the full image. For such clusters, which are of much smaller size than the image, the amount of information available is already limited, and further loss of information can lead to greater errors in the estimation.

To mitigate the above issue, for this study we employ the enhanced box-counting algorithm developed by So *et al* [42]. The method outlined by So *et al* has two advantages over the standard box-counting method; firstly, it allows the selection of a greater number of box sizes than the GS method, which in turn means a larger number of data points that can be used for regression, which leads to a more robust estimate. Secondly, $N(\delta)$, the number of boxes at a particular box size δ , is allowed to take on positive real values instead of being limited to whole numbers. This makes use of the pixels that are close to an edge of the image, which might fall into a fractional box if the dimensions of the image are not a multiple of the box size. Figure 3 displays the $N(\delta)$ obtained from this fractional box-counting procedure for four different δ on an example image.

3. Results

We have performed a fractal characterization of TIRF images of growing fibrils obtained from experiments (see supplementary information for details of TIRF setup). The plots in panel (a) of figure 4 compare the change in the average fractal dimension over time for two different concentrations—8 and $1 \mu\text{M}$. At the initial stages, clusters will be small in size and show up as tiny points of light in the view of the microscope. This visual observation is borne out by the fractal dimension analysis. As figure 4 shows, for both concentrations the plot of the average fractal dimension starts off at a value of 0. For $8 \mu\text{M}$, as the clusters grow in size, the average fractal dimension also rises, until

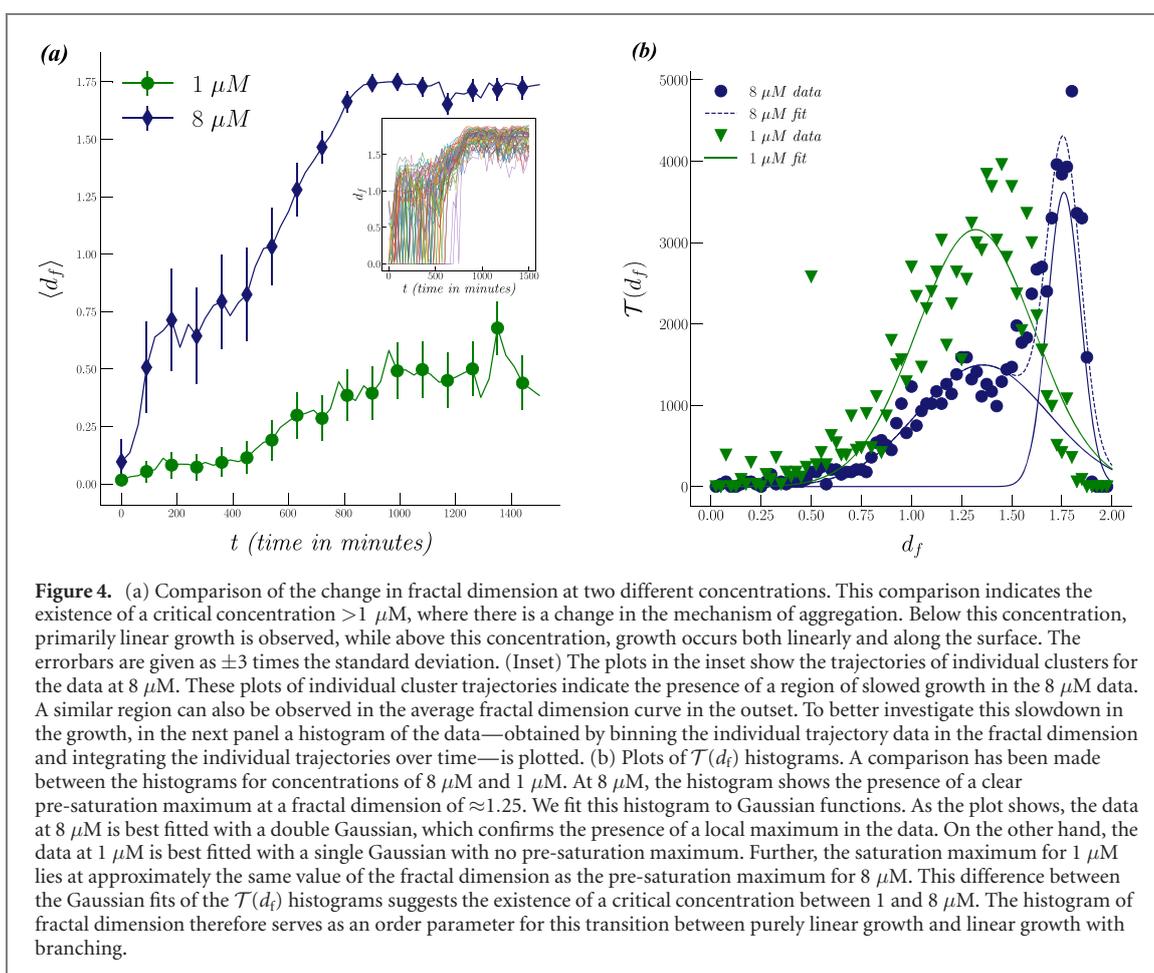
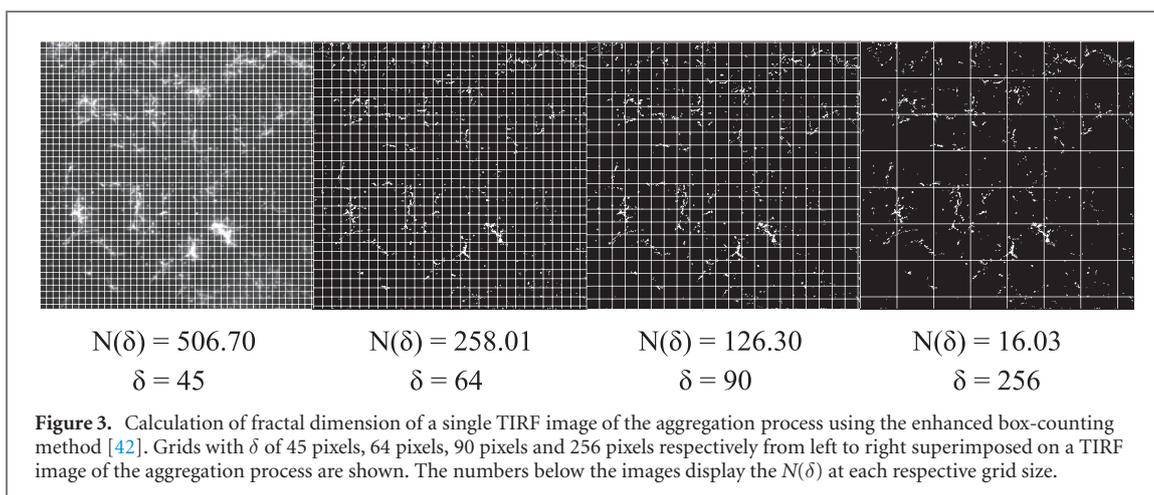
it finally saturates at a value close to 1.75. At $1 \mu\text{M}$ however, the average fractal dimension does not grow beyond 1, saturating at a final value of around 0.6. This contrast indicates that there are different concentration dependent growth mechanisms at work, which in turn leads to the difference in outcomes of the aggregation process. Further, the inset in the same panel displays the trajectories of individual clusters in the fractal dimension-time space for $8 \mu\text{M}$. The plot of these individual trajectories indicates the presence of a region in the fractal dimension-time space where the rate of growth of the fractal dimension temporarily slows down. This slowdown can also be observed in the averaged fractal dimension plot (see inset).

The nature of the evolution of the fractal dimension at $8 \mu\text{M}$ suggests that initially, aggregation begins through a mechanism that causes linear growth of the clusters. After a period of time, a different mechanism—one that causes fibrils to branch out—takes over, as a result of which the average fractal dimension starts to take on values larger than 1. Our analysis indicates that for monomer concentrations $\leq 1 \mu\text{M}$, this second mechanism does not seem to dominate in the same way as it does for higher monomer concentrations. This conclusion is supported by the behavior of the average fractal dimension at the two different concentrations. The conclusions drawn from this analysis are in line with the two-step model—primary nucleation followed by secondary nucleation catalyzed by fibril surface—of aggregation that has long been proposed in the literature [45–51].

Although the presence of a plateau in the evolution of the average fractal dimension $\langle d_f(t) \rangle$ around the value $\langle d_f \rangle \approx 1$ clearly indicates that the clusters grow primarily through linear elongation in this phase, our localization of single clusters allows us to characterize the kinetics of this two step aggregation further. Since clusters can appear at any time in the system, each of them may acquire branches at different times depending on their age, which is directly related to their length through a linear elongation rate. These effects naturally contribute to the evolution of $\langle d_f(t) \rangle$, making the average plateau at the linear elongation dimension less discernible. In this context it is useful to look at the *individual trajectories* of the clusters through fractal dimension space. We plot these trajectories for the experiments at $8 \mu\text{M}$ concentrations in the inset of figure 4(a). These trajectories clearly display a prominent plateau in their evolution through fractal dimension space. In order to quantify the time dependence of this evolution further, we monitor

$$\mathcal{T}(d_f) = \int_0^{t_{\max}} dt' \int_{d_f}^{d_f+\delta} dx n(x, t'), \quad (4)$$

where $n(d_f, t)$ represents the un-normalized distribution of the fractal dimensions of the individual clusters, present at a given time t . Here t_{\max} represents the maximum or cutoff time which for the



experiments analyzed here was 25 hours. $\mathcal{T}(d_f)$ therefore represents the *total time* spent by all the clusters between the fractal dimensions d_f and $d_f + \delta$ where δ represents the binwidth (we choose $\delta = 0.025$). The time-integrated histogram of fractal dimensions $\mathcal{T}(d_f)$ obtained from the two sets of experiments at concentrations $8 \mu\text{M}$ and $1 \mu\text{M}$ are plotted in panel (b) of figure 4. This $\mathcal{T}(d_f)$ histogram will have a trivial maximum located around the saturation value of the average fractal dimension; what is of interest, however, is the presence of any pre-saturation peaks.

The presence of such pre-saturation peaks in this histogram suggests that the two processes controlling the growth of the aggregate dominate at different fractal dimensions.

The histogram for $8 \mu\text{M}$ clearly indicates the presence of a pre-saturation maximum, confirmed further by fitting a double Gaussian to this data. The data at $1 \mu\text{M}$, on the other hand, fits best to a single Gaussian, which shows there are no pre-saturation maxima in this data. In addition, the plots indicate that the saturation maximum for $1 \mu\text{M}$ and the pre-saturation

maximum for $8 \mu\text{M}$ are at similar values of the fractal dimension. This further supports our conclusion that the secondary nucleation mechanism—responsible for the formation of branches—is not dominant at the lower concentration. Taken together, our observations naturally motivate the simulation of two step aggregation to more quantitatively estimate the microscopic parameters governing the aggregation process observed in the experiments.

The second maximum in the histogram for the $8 \mu\text{M}$ concentration appears at a fractal dimension of ~ 1.75 . From panel (a) of figure 4, it can be observed that this is approximately equal to the final saturation value for the average fractal dimension at this concentration. The small size of the error bars in the same plot also suggests that the standard deviation of the distribution of the fractal dimension of clusters at the saturation stage is very small. This observation is further corroborated by the small standard deviation of the Gaussian function that fits the second maximum of the $8 \mu\text{M}$ data in the $\mathcal{T}(d_f)$ histogram displayed in figure 4. From these observations we conclude that in the $8 \mu\text{M}$ dataset, most clusters acquire a fractal dimension $d_f \approx 1.75$ by the time they reach the saturation stage, with a few clusters saturating at lower values, as observed from the individual cluster trajectories in the inset of figure 4(a).

Our Gaussian fits of $\mathcal{T}(d_f)$ histogram for 8 and $1 \mu\text{M}$ therefore suggest a transition between two different aggregation pathways as the concentration is decreased—a two-step aggregation mechanism consisting of linear and branching growth at the higher concentration, and purely linear growth at the lower concentrations. Consequently, we posit the existence of a critical concentration between 1 and $8 \mu\text{M}$ —where the aggregation process transitions over from one mechanism to the other. The histogram of fractal dimension therefore serves as an order parameter for this transition between purely linear growth and linear growth with branching. Additionally, we can rule out other space-filling mechanisms—such as fibrils from bulk attaching to the surface, or space-filling linear growth of the fibrils—as possible origins for the observed evolution of $\langle d_f(t) \rangle$.

4. Discussion

In this paper, we have demonstrated that single-molecule techniques such as TIRF microscopy allow for a microscopic characterization of the kinetics of protein aggregation. The large amounts of data generated in such techniques is hard to analyze directly, and therefore to utilize their full potential it is necessary to develop methods to conveniently organize and analyze large datasets. In this study, we have combined approaches from different domains of image processing, computer science and mathematics to develop a semi-automated analysis procedure for images of protein aggregates obtained from TIRF microscopy. Our

fractal analysis technique provides *direct* evidence of a two step mechanism in the aggregation of the A β 42 peptide. Our analysis methodology is quite general and can easily be adapted to analyze the aggregation of other proteins from TIRF microscopy, as well as other single molecule imaging techniques. The average fractal dimension of the images from TIRF data, without clustering and isolation of single fibrils, also shows a plateau in the fractal dimension, and therefore our technique helps to extract such behavior from noisy data, where isolation of fibrils is not always possible. However, we have not included this in the present manuscript as we have insufficient experimental runs to draw a conclusive quantification that the single cluster data allows.

The main advantages of our methodology is that our technique works when isolation of linear fibrils is not available. The clustering algorithm is robust to noise, and therefore can be used in situations where intensity of data is low. It is in general hard to classify the irregular shapes of aggregation clusters, and therefore a single quantifiable characteristic such as the fractal dimension allows for a quantitative characterization of the growth process in real time.

However, our method is useful only for data with good signal-to-noise ratio. For data with poor signal-to-noise ratio, thresholding the image becomes very difficult. Without proper segmentation, the determination of the fractal dimension in the later stages becomes less robust. Similarly, our method benefits greatly from higher-resolution data. If the size of aggregates within the field of view is too small, this reduces the number of box sizes available to the box-counting algorithm, and makes the estimation of the fractal dimension correspondingly less robust.

Further, the measurements of the average fractal dimension over time, shown in panel (a) of figure 4, shows that there is a fundamental difference in the growth behavior between the aggregates at concentrations of $1 \mu\text{M}$ and $8 \mu\text{M}$, with the lower concentrations demonstrating primarily linear growth. This observation suggests the existence of a critical concentration below which the dominant mechanism of aggregation is linear elongation, and above which both linear elongation and branching are present. The existence of a critical concentration for amyloid- β 42 aggregation has been suggested previously in the literature based on ensemble experiments [52]. However, the critical concentrations observed in ensemble experiments are in the nanomolar range. We attribute the difference in the critical behavior observed in the data analyzed in this study to the two-dimensional nature of the experimental setup. TIRF microscopy utilizes the evanescent field formed due to total internal reflection of light at the boundary between two media to excite fluorophores in the experimental medium. As the evanescent field decays exponentially with distance from the boundary, we can only observe a slice of the experimental medium that lies along this

boundary. Such a confinement leads to changes in the aggregation kinetics that manifest as a shift in the rate constants as well as the critical concentration for linear growth with branching. It would be interesting to perform a systematic study of the effect of confinement on the aggregation kinetics of the amyloid proteins, to better understand the nature of the two-step mechanism present in live cells that exhibit different confinement geometries [53, 54]. Another important future direction for our work would be to expand our analysis methodology and apply it to image data from microscopy sources other than TIRF microscopy. Of particular interest would be the application of this fractal analysis technique to 3D protein aggregation data. Furthermore, it would be interesting to apply this technique to study the aggregation process in complex proteins which are not well-characterized, such as many intrinsically disordered proteins [55].

Author contributions

SM and KR conceived of the project; SCB designed the TIRF setup, performed experiments and collected data; SM and KR performed the data analysis; and SM, SCB and KR wrote the paper.

Acknowledgments

We thank Kanchan Garai for providing experimental data. We thank Vishnu V Krishnan, Chaitanya Athale, Saroj Nandi, Aprotim Mazumder, Mustansir Barma, Stephy Jose and Roshan Maharana for useful discussions. This project was funded by intramural funds at TIFR Hyderabad from the Department of Atomic Energy (DAE).

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

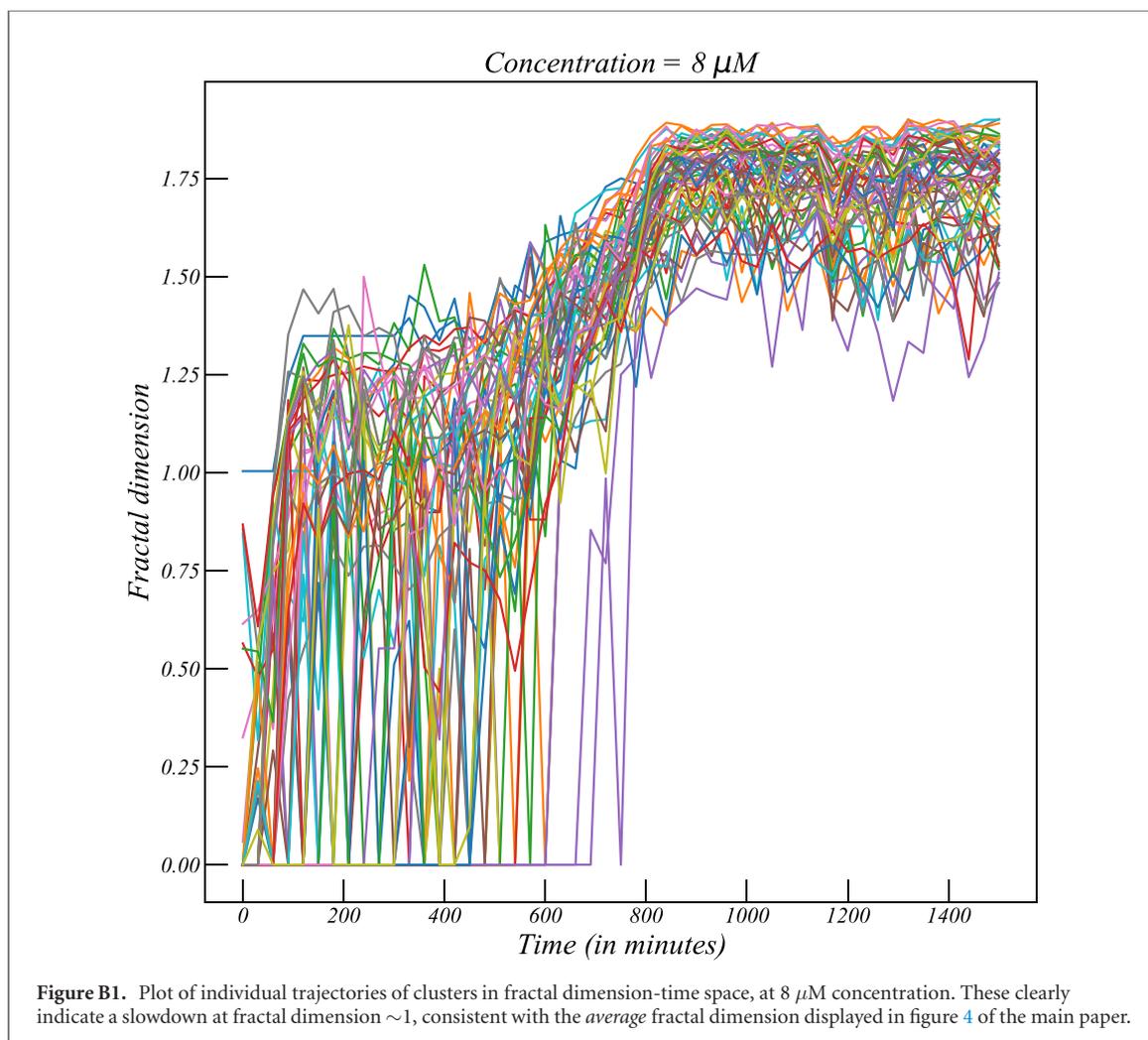
Appendix A. TIRF microscopy

When electromagnetic waves—including light—undergo total internal reflection at the boundary between two media, an evanescent field that oscillates with the same frequency as the original wave is formed in the medium across the boundary. The energy of the evanescent field does not propagate like a wave, instead, it stays concentrated around

the vicinity of the origin. In terms of fluorescence microscopy, using the evanescent field to excite fluorophores instead of direct illumination means that only the fluorophores very close to the boundary, such as surface-bound fluorophores, are excited. This provides TIRF microscopy with two advantages over standard epifluorescence microscopy:

- (a) As the entire volume of the solution is not illuminated, background fluorescence is greatly diminished, which improves the signal-to-noise ratio of observations.
- (b) Only the fibrils lying along the glass slide are selectively monitored, thus the lengths of the fibrils as obtained from these images are close to their exact length.

The TIRFM setup was built on an inverted Nikon (model No. Ti-E) microscope. The evanescent field was generated using a high NA ($NA = 1.49$) oil immersion objective (Nikon) using objective-type TIR. For excitation of ThT (fluorescent marker) a solid-state laser ($\lambda = 450$ nm) was used. An excitation filter (450 ± 10 nm) was used to clean up the laser. The beam expander, mounted on a micrometre translation stage, was used for translating the laser beam to attain the critical position of incidence required in TIRFM. The excitation beam was focused on the back focal plane of the objective after being reflected off a dichroic mirror. The objective is mounted on a piezo stage (PI, Germany) and was used to generate the evanescent field and collect the resultant fluorescence. The fluorescence was then transmitted through the dichroic and detected using a sCMOS camera (PCO, Germany) after being focused with a tube lens. An emission filter (510 ± 40 nm) was used to separate ThT fluorescence signal from any other light contamination. An infrared diode laser ($\lambda = 980$ nm), the objective piezo stage, a quadrant photodiode (QPD, Thorlabs) and a PID controller were used for building the auto-focus system. The sample was mounted on a motorized XY stage (Thorlabs, USA) on top of the objective. The temperature of the stage and the objective was controlled at $23 \pm 0.1^\circ\text{C}$ using a PID temperature controller (SELEC, India). The dichroic mirror and the optical filters were procured from Chroma, USA. All the other components were procured from Thorlabs, USA. Multiple position imaging was achieved by the grid creation feature available in multi-dimensional acquisition in micro-manager [56, 57]. The images were stored in the computer HDD as separate stacks for each position during data acquisition and analysed afterward.



Appendix B. Following individual clusters

(Figure B1) Figure B2 plots the evolution of the average fractal dimension for four different concentrations— $8 \mu\text{M}$, $4 \mu\text{M}$, $2 \mu\text{M}$ and $1 \mu\text{M}$, with two datasets at each concentration. This comparison across concentrations shows that there exists a critical concentration $>1 \mu\text{M}$ and $\leq 2 \mu\text{M}$, where there is a change in the dominant aggregation mechanism. Below this critical concentration, aggregation happens primarily through the linear pathway, as shown by the average fractal dimension plots at concentration $1 \mu\text{M}$, which saturate at a value <1 . Above the critical concentration, secondary nucleation, which causes branching, seems to become more dominant, and the average fractal dimension plots from experiments at concentrations $\geq 2 \mu\text{M}$ grow beyond 1, and go on to saturate at value ~ 1.75 . Figure B1 shows the fractal dimension curves for individual clusters from a dataset at concentration $8 \mu\text{M}$. The slowdown in the

growth of the fractal dimension can be observed in this plot of individual cluster traces.

B.1. Analysis of mean t_1 distribution

Our fractal analysis technique gives us access to another important quantity that can be used to characterize the aggregation process—an aggregate-level timescale t_1 , defined as the time required for the aggregate to acquire a fractal dimension ≥ 1 . The aggregation process is highly stochastic at the scale of a single aggregate, thus the timescale t_1 can also be expected to be a highly stochastic variable.

In figure B3, the mean t_1 as obtained from TIRF experiments at multiple concentrations has been plotted against the respective concentrations, along with errorbars. At lower concentrations, the aggregation process is more stochastic, as shown by the larger errorbars at those concentrations. In addition, these plots further show that at concentrations $<20 \mu\text{M}$, the mean t_1 does not depend upon concentration.

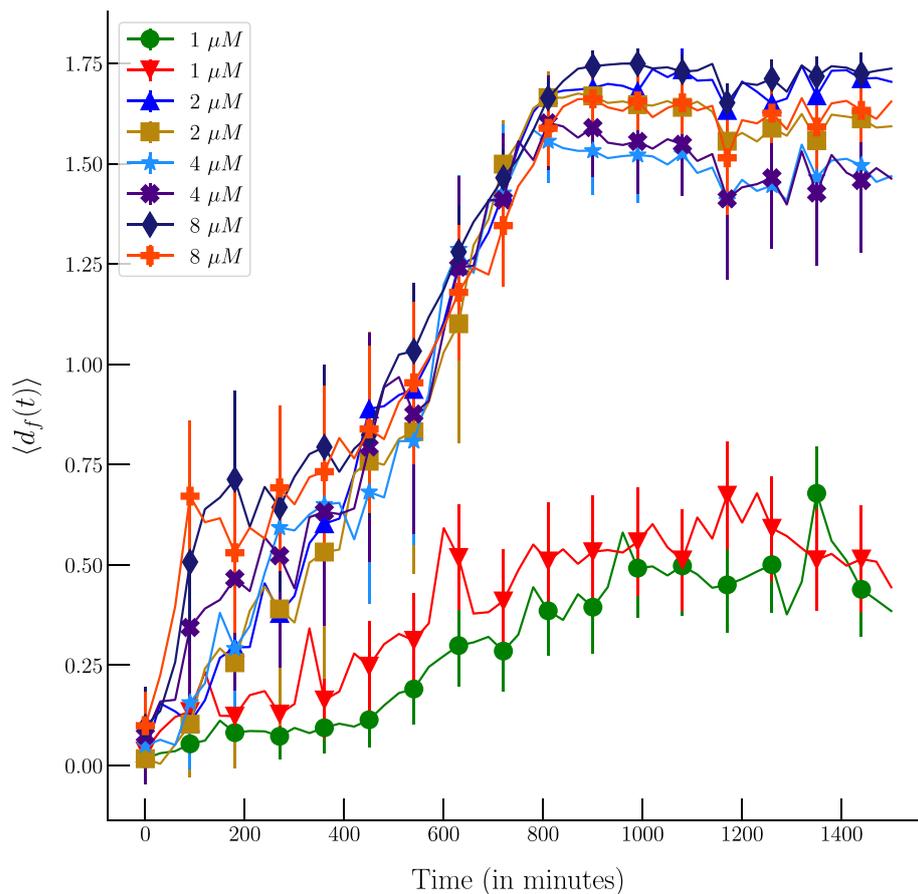


Figure B2. Plots of the average fractal dimension for four concentrations—1 μM , 2 μM , 4 μM and 8 μM . This comparison between different concentrations shows that there exists a critical concentration $>1 \mu\text{M}$ and $\leq 2 \mu\text{M}$ where there is a change in the dominant mechanism of aggregation. Below this concentration, primarily linear growth is observed, while above this concentration, growth occurs both linearly and along the surface.

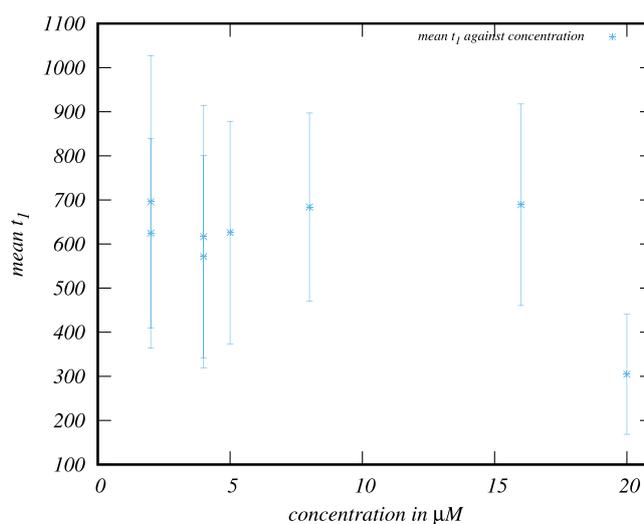


Figure B3. Plot of mean t_1 against concentration, for concentrations at 2, 4, 5, 8, 16 and 20 μM . For concentrations below 20 μM , the mean t_1 does not display any dependence on the concentration. Errorbars are ± 3 times the standard deviation.

ORCID iDs

Soham Mukhopadhyay  <https://orcid.org/0000-0002-6480-5081>

Subhas C Bera  <https://orcid.org/0000-0002-4168-1805>

References

- Wechalekar A D, Gillmore J D and Hawkins P N 2016 Systemic amyloidosis *Lancet* **387** 2641–54
- Stewart K L and Radford S E 2017 Amyloid plaques beyond A β : a survey of the diverse modulators of amyloid aggregation *Biophys. Rev.* **9** 405–19
- Luo J, Wärmländer S K T S, Gräslund A and Abrahams J P 2016 Cross-interactions between the alzheimer disease amyloid- β peptide and other amyloid proteins: a further aspect of the amyloid cascade hypothesis *J. Biol. Chem.* **291** 16485–93
- Eisenberg D and Jucker M 2012 The amyloid state of proteins in human diseases *Cell* **148** 1188–203
- Uversky V N and Fink A L 2004 Conformational constraints for amyloid fibrillation: the importance of being unfolded *Biochim. Biophys. Acta, Proteins Proteomics* **1698** 131–53
- Stefani M and Dobson C M 2003 Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution *J. Mol. Med.* **81** 678–99
- Ravichandran S, Lachmann H J and Wechalekar A D 2020 Epidemiologic and survival trends in amyloidosis, 1987–2019 *New Engl. J. Med.* **382** 1567–8
- Merlini G, Dispenzieri A, Sancharawala V, Schönland S O, Palladini G, Hawkins P N and Gertz M A 2018 Systemic immunoglobulin light chain amyloidosis *Nat. Rev. Dis. Prim.* **4** 38
- Bharadwaj P, Wijesekera N, Liyanapathirana M, Newsholme P, Ittner L, Fraser P and Verdile G 2017 The link between type 2 diabetes and neurodegeneration: roles for amyloid- β , amylin, and tau proteins *J. Alzheimer's Dis.* **59** 421–32
- Musiek E S and Holtzman D M 2015 Three dimensions of the amyloid hypothesis: time, space and 'wingmen' *Nat. Neurosci.* **18** 800–6
- Michaels T C T, Šarić A, Habchi J, Chia S, Meisl G, Vendruscolo M, Dobson C M and Knowles T P J 2018 Chemical kinetics for bridging molecular mechanisms and macroscopic measurements of amyloid fibril formation *Annu. Rev. Phys. Chem.* **69** 273–98
- Törnquist M, Michaels T C T, Sanagavarapu K, Yang X, Meisl G, Cohen S I A, Knowles T P J and Linse S 2018 Secondary nucleation in amyloid formation *Chem. Commun.* **54** 8667–84
- Cohen S I A et al 2013 Proliferation of amyloid- β 42 aggregates occurs through a secondary nucleation mechanism *Proc. Natl Acad. Sci. USA* **110** 9758–63
- Rajan R S, Illing M E, Bence N F and Kopito R R 2001 Specificity in intracellular protein aggregation and inclusion body formation *Proc. Natl Acad. Sci. USA* **98** 13060–5
- Ruggeri F S, Šneideris T, Vendruscolo M and Knowles T P J 2019 Atomic force microscopy for single molecule characterisation of protein aggregation *Arch. Biochem. Biophys.* **664** 134–48
- Lou Z, Wang B, Guo C, Wang K, Zhang H and Xu B 2015 Molecular-level insights of early-stage prion protein aggregation on mica and gold surface determined by AFM imaging and molecular simulation *Colloids Surf. B* **135** 371–8
- Zhou L and Kurouski D 2020 Structural characterization of individual α -synuclein oligomers formed at different stages of protein aggregation by atomic force microscopy-infrared spectroscopy *Anal. Chem.* **92** 6806–10
- Yu J, Warnke J and Lyubchenko Y L 2011 Nanoprobining of α -synuclein misfolding and aggregation with atomic force microscopy *Nanomed.: Nanotechnol. Biol. Med.* **7** 146–52
- Asanov A N, Delucas L J, Oldham P B and Wilson W W 1997 Interfacial aggregation of bovine serum albumin related to crystallization conditions studied by total internal reflection fluorescence *J. Colloid Interface Sci.* **196** 62–73
- Walder R and Schwartz D K 2011 Dynamics of protein aggregation at the oil–water interface characterized by single molecule TIRF microscopy *Soft Matter* **7** 7616–22
- Torrent M, Pulido D, Nogués M V and Boix E 2012 Exploring new biological functions of amyloids: bacteria cell agglutination mediated by host protein aggregation *PLoS Pathog.* **8** e1003005
- Hedayati M, Krapf D and Kipper M J 2021 Dynamics of long-term protein aggregation on low-fouling surfaces *J. Colloid Interface Sci.* **589** 356–66
- Young L J, Kaminski Schierle G S and Kaminski C F 2017 Imaging a β (1–42) fibril elongation reveals strongly polarised growth and growth incompetent states *Phys. Chem. Chem. Phys.* **19** 27987–96
- Kaminski Schierle G S et al 2011 *In situ* measurements of the formation and morphology of intracellular β -amyloid fibrils by super-resolution fluorescence imaging *J. Am. Chem. Soc.* **133** 12902–5
- Pinotsi D, Michel C H, Buell A K, Laine R F, Mahou P, Dobson C M, Kaminski C F and Kaminski Schierle G S 2016 Nanoscopic insights into seeding mechanisms and toxicity of α -synuclein species in neurons *Proc. Natl Acad. Sci. USA* **113** 3815–9
- Duim W C, Jiang Y, Shen K, Frydman J and Moerner W E 2014 Super-resolution fluorescence of huntingtin reveals growth of globular species into short fibers and coexistence of distinct aggregates *ACS Chem. Biol.* **9** 2767–78
- Lu M, Williamson N, Mishra A, Michel C H, Kaminski C F, Tunnacliffe A and Kaminski Schierle G S 2019 Structural progression of amyloid- β arctic mutant aggregation in cells revealed by multiparametric imaging *J. Biol. Chem.* **294** 1478–87
- Lu M, Banetta L, Young L J, Smith E J, Bates G P, Zaccane A, Kaminski Schierle G S, Tunnacliffe A and Kaminski C F 2019 Live-cell super-resolution microscopy reveals a primary role for diffusion in polyglutamine-driven aggresome assembly *J. Biol. Chem.* **294** 257–68
- Abdellatif M H, Abdelrasoul G N, Salerno M, Liakos I, Scarpellini A, Marras S and Diaspro A 2016 Fractal analysis of inter-particle interaction forces in gold nanoparticle aggregates *Colloids Surf. A* **497** 225–32
- Gwaze P, Schmid O, Annegarn H J, Andrea M O, Huth J and Helas G 2006 Comparison of three methods of fractal analysis applied to soot aggregates from wood combustion *J. Aerosol Sci.* **37** 820–38
- Tokuyama M and Kawasaki K 1984 Fractal dimensions for diffusion-limited aggregation *Phys. Lett. A* **100** 337–40
- Thevenaz P, Ruttimann U E and Unser M 1998 A pyramid approach to subpixel registration based on intensity *IEEE Trans. Image Process.* **7** 27–41
- Close W et al 2018 Physical basis of amyloid fibril polymorphism *Nat. Commun.* **9** 699
- Meinhardt J, Sachse C, Hortschansky P, Grigorieff N and Fändrich M 2009 A β (1–40) fibril polymorphism implies diverse interaction patterns in amyloid fibrils *J. Mol. Biol.* **386** 869–77
- Fitzpatrick A W P et al 2013 Atomic structure and hierarchical assembly of a cross- β amyloid fibril *Proc. Natl Acad. Sci. USA* **110** 5468–73
- Richardson W H 1972 Bayesian-based iterative method of image restoration *J. Opt. Soc. Am.* **62** 55–9
- Lucy L B 1974 An iterative technique for the rectification of observed distributions *Astron. J.* **79** 745

- [38] van der Walt S, Schönberger J L, Nunez-Iglesias J, Boulogne F, Warner J D, Yager N, Gouillart E and Yu T (The Scikit-Image Contributors) 2014 Scikit-image: image processing in Python *PeerJ* **2** e453
- [39] Ester M, Kriegel H-P, Sander J and Xu X 1996 A density-based algorithm for discovering clusters in large spatial databases with noise *KDD'96: Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining* vol 96 pp 226–31
- [40] Pedregosa F et al 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30
- [41] Falconer K 2004 *Fractal Geometry: Mathematical Foundations and Applications* (New York: Wiley)
- [42] So G-B, So H-R and Jin G-G 2017 Enhancement of the box-counting algorithm for fractal dimension estimation *Pattern Recognit. Lett.* **98** 53–8
- [43] Gagnepain J J and Roques-Carmes C 1986 Fractal approach to two-dimensional and three-dimensional surface roughness *Wear* **109** 119–26
- [44] Clarke K C 1986 Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method *Comput. Geosci.* **12** 713–22
- [45] Ferrone F A, Hofrichter J and Eaton W A 1985 Kinetics of sickle hemoglobin polymerization: II. A double nucleation mechanism *J. Mol. Biol.* **183** 611–31
- [46] Ferrone F A, Hofrichter J, Sunshine H R and Eaton W A 1980 Kinetic studies on photolysis-induced gelation of sickle cell hemoglobin suggest a new mechanism *Biophys. J.* **32** 361–80
- [47] Ruschak A M and Miranker A D 2007 Fiber-dependent amyloid formation as catalysis of an existing reaction pathway *Proc. Natl Acad. Sci. USA* **104** 12341–6
- [48] Collins S R, Douglass A, Vale R D and Weissman J S 2004 Mechanism of prion propagation: amyloid growth occurs by monomer addition *PLoS Biol.* **2** e321
- [49] Knowles T P J, Oppenheim T W, Buell A K, Chirgadze D Y and Welland M E 2010 Nanostructured films from hierarchical self-assembly of amyloidogenic proteins *Nat. Nanotechnol.* **5** 204
- [50] Hall D and Edskes H 2004 Silent prions lying in wait: a two-hit model of prion/amyloid formation and infection *J. Mol. Biol.* **336** 775–86
- [51] Xue W-F, Hellewell A L, Gosal W S, Homans S W, Hewitt E W and Radford S E 2009 Fibril fragmentation enhances amyloid cytotoxicity *J. Biol. Chem.* **284** 34272–82
- [52] Novo M, Freire S and Al-Soufi W 2018 Critical aggregation concentration for the formation of early amyloid- β (1–42) oligomers *Sci. Rep.* **8** 1783
- [53] Simpson L W, Good T A and Leach J B 2020 Protein folding and assembly in confined environments: implications for protein aggregation in hydrogels and tissues *Biotechnol. Adv.* **42** 107573
- [54] Michaels T C T, Dear A J and Knowles T P J 2018 Stochastic calculus of protein filament formation under spatial confinement *New J. Phys.* **20** 055007
- [55] Strodel B 2021 Energy landscapes of protein aggregation and conformation switching in intrinsically disordered proteins *J. Mol. Biol.* **433** 167182
- [56] Edelstein A, Amodaj N, Hoover K, Vale R and Stuurman N 2010 Computer control of microscopes using μ Manager *Curr. Protoc. Mol. Biol.* **92** 14–20
- [57] Edelstein A D, Tsuchida M A, Amodaj N, Pinkard H, Vale R D and Stuurman N 2014 Advanced methods of microscope control using μ Manager software *J. Biol. Methods* **1** e10