Computational and Structural Biotechnology Journal 18 (2020) 1458–1465

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Electron microscopy-based semi-automated characterization of aggregation in monoclonal antibody products

Mohit Kumar [a], Apoorv Pant [b], Rohit Bansal [c], Ashutosh Pandey [a], James Gomes [a], Kedar Khare [b], Anurag Singh Rathore [c], Manidipa Banerjee [a,*]

[a] Kusuma School of Biological Sciences, Indian Institute of Technology – Delhi, Hauz Khas, New Delhi 110016, India
[b] Department of Physics, Indian Institute of Technology – Delhi, Hauz Khas, New Delhi 110016, India
[c] Department of Chemical Engineering, Indian Institute of Technology – Delhi, Hauz Khas, New Delhi 110016, India

ARTICLE INFO

Article history:
Received 17 March 2020
Received in revised form 3 June 2020
Accepted 3 June 2020
Available online 11 June 2020

Keywords:
Antibodies
Heterogeneity
Aggregation
Electron microscopy
Connected component labelling

ABSTRACT

Aggregation is a critical parameter for protein-based therapeutics, due to its impact on the immunogenicity of the product. The traditional approach towards characterization of such products is to use a collection of orthogonal tools. However, the fact that none of these tools is able to completely classify the distribution and physical characteristics of aggregates, implies that there exists a need for additional analytical methods. We report one such method for characterization of heterogeneous population of proteins using transmission electron microscopy. The method involves semi-automated, size-based clustering of different protein species from micrographs. This method can be utilized for quantitative characterization of heterogeneous populations of antibody/protein aggregates from TEM images of proteins, and may also be applicable towards other instances of protein aggregation.

1. Introduction

Complex protein-based therapeutic drugs are prone to structural and functional alterations [1]. Changes in the primary sequence, expression system, post-translational modification, pH and temperature variations can result in loss-of-function or aggregation, and consequently unwanted immune response when administered clinically [1,2]. The complexity inherent in therapeutic proteins suggest that even minor changes to the protein structure can significantly affect the safety and efficacy of biotherapeutics [3–7]. Since the development of biotherapeutics is witnessing an enormous growth worldwide, it has become imperative to establish effective protocols for quickly analyzing their stability and purity [8,9].

Protein aggregation is an important phenomenon in biopharmaceutical formulations due to the immunogenicity associated with aggregates. Protein aggregates are generally classified into two categories – 1) reversible, and 2) irreversible, based on the strength or reversibility of the association amongst their constituent monomers [10,11]. Reversibility of aggregates is generally defined on the basis of the bonding between different constituent monomers or oligomeric species. Weaker molecular interactions includes non-covalent Van der Waal's, electrostatic, hydrophobic interactions, hydrogen bonding etc. [11], while strong molecular interactions such as covalent bonding requires higher energy for dissociation. Irreversible aggregation is generally attributed to covalent interactions between monomeric and oligomeric molecules [11], and tends to form in proteins over long storage or transportation conditions. Stages in this process include nuclei formation, protein unfolding and formation of metastable states [11].

A myriad of techniques are available for characterizing biomolecules and their aggregates in varied size range, from a few nanometres to several micro and millimetres [12]. A few such techniques with their projected size range of detection are – size

Abbreviations: CD, Circular Dichroism; DPBS, Dulbecco's phosphate-buffered saline; DLS, Dynamic Light Scattering; EM, Electron Microscopy; FEG, field emission electron gun; GUI, Graphical User Interface; HDX-MS, Hydrogen Deuterium Exchange Mass Spectroscopy; MS, Mass Spectroscopy; *SEC*, Size Exclusion Chromatography; *SEC*-MALS, Size Exclusion Chromatography Multi Angle Light Scattering; mAb, monoclonal Antibody; TEM, Transmission Electron Microscopy; UV, Ultra Violet; TV, Total Variation.

* Corresponding author: Kusuma School of Biological Sciences, Block 1A, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India.

E-mail addresses: jgomes@bioschool.iitd.ac.in (J. Gomes), kedark@physics.iitd.ac.in (K. Khare), arathore@chemical.iitd.ac.in (A. Singh Rathore), mbanerjee@bioschool.iitd.ac.in (M. Banerjee).

https://doi.org/10.1016/j.csbj.2020.06.009
2001-0370/

exclusion-high performance liquid chromatography (SE-HPLC) (1 nm-50 nm), asymmetric flow field flow fractionation (AF4) (2 nm−1 μm), microscopic methods (light (0.8 μm–150 μm), electron (1 nm–2 μm), atomic force microscopy), static and dynamic light scattering (SLS and DLS) (1 nm–5 μm), analytical ultracentrifugation (AUC) (1 nm–100 nm), light obscuration (LO) (1 μm–200 μm), coulter counter (CC) (0.4 μm–1.6 mm), nanoparticle tracking analysis (NTA) (30 nm–1 μm) and micro flow imaging (MFI) (2 μm–300 μm) [13].

These techniques vary widely in their mode of detection, sample quantity and preparation method, robustness, and sensitivity [13].

Among these variables, the sample preparation step has an outsized effect on the susceptibility of proteins towards aggregation, and the ability of various methods to detect aggregation. There is always the possibility of introducing artefacts in samples during preparation steps which could either cause new aggregate formation or destruction of existing aggregates, thus affecting the overall accuracy of the detection method [10,13]. Some common steps of sample preparation include centrifugation, sample dilution and concentration. All of these steps have the potential to drastically change the aggregate content in the sample. *SEC*, which is considered as the gold standard for aggregate detection, can be difficult to interpret if the aggregate size is larger than the size range of the stationary phase matrix. Larger insoluble aggregates might be trapped in the column prefilter and impact the overall sample mass balance and analysis [13,14]. Similarly, heating or reducing conditions during sample preparation for polyacrylamide gel electrophoresis can cause aggregate dissociation due to breakage of covalent bonds [15,16]. Therefore it is always recommended that multiple orthogonal tools be utilized for analysing aggregated samples to ensure the accuracy and reliability of the results obtained [15,16]. Converging results from varied techniques are essential for inferring the status of biotherapeutics, which may be highly immunogenic at aggregated states.

The current, widely used orthogonal techniques for aggregate detection have certain drawbacks. *SEC*-based separation, as mentioned before, suffers from lack of resolution in the higher molecular weight range, trapping of larger aggregate species in the column pre-filter or eluting in column's void volume, aggregated species adsorbing to stationary phase, aggregates breaking or forming during analysis, etc., all of which can skew molecular weight analysis [17–19]. DLS, which predicts the average hydrodynamic size of the sample [20,21], is a high throughput technique, but lacks the ability to accurately separate or quantify aggregated species. Circular Dichroism (CD) is an important tool for secondary and tertiary structure determination [22], however it cannot identify reversible aggregates, that could have expected secondary structure. While *SEC*-MALS and hydrogen deuterium exchange mass spectroscopy (HDX-MS), are more effective for analysing altered conformational and aggregated states [23], they do not allow for direct visual examination of aggregated species in a population. Given these limitations in existing techniques, it is imperative to develop additional analytical methods, ideally for both visual and quantitative analysis of aggregates in therapeutic protein samples.

Electron microscopy is arguably the only technique that can provide visual information about aggregated species in submicron range present in therapeutic proteins. Negative stain electron microscopy has been utilized for investigating the effect of extraneous factors such as storage, temperatures, time, and stress on the aggregation status of biotherapeutics like monoclonal antibodies [13,24,25]. However, the output has been limited to visual assessment of aggregation, in order to supplement data from techniques such as DLS and *SEC* [26,27]. Typically, negative stain electron microscopy has been useful in studying morphology of protein complexes, nanoparticles, and biological ultrastructures. Most of the studies have been based on visual or qualitative parameters, however, there have been very few efforts to quantify information obtained from low resolution electron micrographs. Some examples include quantification of additive dispersion [28]; and that of nanoparticle uptake into cells from TEM images [29]. In case of therapeutic proteins, numerical analysis of images has the potential to deliver an accurate physical distribution of aggregates in a wide size range and might be useful as an orthogonal method for effective monitoring of samples. Aggregated species in the submicron range or larger are expected to be fairly non-uniform and heterogeneous in terms of shapes and sizes. Size-based analysis of aggregates from micrographs requires automated procedures since manual quantification involves significant time input and may incorporate substantial user-specific variations in the output [30]. Although automated particle picking and classification techniques are available for 3D reconstruction procedures from electron cryo-micrographs, categorizing an entirely heterogeneous population requires different approaches.

In this work, we attempted to combine automated detection and numerical size distribution of aggregates in images from conventional electron microscopy. The micrographs were subjected to de-noising, low-pass filtering and background normalization, converted to binary images, and aggregates of heterogeneous sizes and shapes were automatically picked and clustered to obtain a distribution of higher order species present in the sample. Data from visual or manual characterization of micrographs were found to correlate well with the size histogram of heterogeneous units. When applied towards samples of therapeutic monoclonal antibodies, this method was found to correlate with, but provide information above and beyond that is generated from traditional analytical methods like *SEC*. To the best of our knowledge, automated quantification of aggregates from electron micrographs has not been attempted before, and we expect that this method will be applicable to the study of aggregation in biotherapeutics as well as protein complexes in other systems.

2. Materials and methods

2.1. Reagents

Three monoclonal antibodies, designated as mAb A, B and C, from subclass IgG1 were procured for EM-based analysis. Both mAb B and C are biosimilars of an anti CD-20 molecule. While mAb B is a commercialised, marketed product, mAb C was donated by a major Indian biosimilar manufacturer. mAb A, an antibody against flagellin of *Salmonella typhi,* was a kind gift from a colleague. All antibodies were stored at 4 °C. All buffers used in the study were filtered using a 0.22-μm filter and degassed.

2.2. Sample preparation

mAb B is a commercialised anti CD-20 molecule as mentioned previously. It was procured in its commercial formulation buffer at pH 6.5 (polysorbate 80 (0.7 mg/mL), sodium chloride (9 mg/mL), sodium citrate dihydrate (7.35 mg/mL)). mAb C is a biosimilar of anti CD-20, provided by a major Indian manufacturer, in 20 mM acetate buffer at pH 5.0. Both mAbs were dialysed into Dulbecco's phosphate-buffered saline (DPBS) using Nanosep® Centrifugal Devices with Omega™ Membrane 10 K (Pall Corporation, NY, US) at 4 °C, followed by size exclusion chromatography for an estimate of the aggregate content in the sample. *SEC* was carried out at 25 °C using a Superdex 200 column (10/30, GE Healthcare, USA) in conjunction with a Dionex Ultimate 3000 UHPLC unit (Thermo Scientific, USA). Each mAb was eluted isocratically over 30 min, at a

constant flow rate of 0.5 ml/min, with a mobile phase containing 50 mM phosphate, pH 6.8, and 300 mM NaCl. UV absorbance at 280 nm was monitored to detect the eluted protein. Monomer and aggregate contents were estimated by calculating the percentage area under the corresponding peaks using Chemstation® (Agilent). No sample precipitation was observed during sample preparation and analysis by *SEC*, and total protein peak area was conserved during repeated sample injections. mAb A is an antibody against the flagellin of *Salmonella typhi*, produced in mice by an academic laboratory. It was stored and provided in DPBS, and was used as a control in our studies after optimum dilution with DPBS.

2.3. Negative staining and electron microscopy

4 μl of therapeutic mAbs, at a concentration of 400 μg/ml, was placed on the carbon face of 200 mesh glow discharged carbon coated copper grids (Agar Scientific) and allowed to adsorb for 1 min. Grids were glow discharged for 60 s at 20 mA (SC7620 Mini Sputter Coater/Glow Discharge system, Quorum Technologies), prior to usage. Grids were washed 3 times with 10 μl of double distilled water, followed by blotting with Whatmann filter paper. For staining, 4 μl of 2% uranyl acetate, pH ~ 7.0 was applied onto the grid twice, for a time period of 20 s and 40 s, respectively, and the surplus stain was removed by blotting in each case. The grid was dried at room temperature for 2 min and visualized at a magnification of 100,000x in a FEG-TEM (FEI Tecnai) operating at 200 kV. Three micrographs from different grid squares were collected for each sample.

2.4. Manual counting

A black isosceles triangle was placed as a fiducial marker in the center of each area of density corresponding to either a monomer or aggregate on the micrographs in order to prevent duplicate counting. Antibodies on the same micrographs were counted by a trained electron microscopist (User A) and a novice (User B), in order to estimate the accuracy of counting (Supplementary Fig. 1 (d)).

2.5. Implementing MATLAB

A MATLAB GUI application was developed to obtain the image mask with labelled particles and enable the user to plot histograms of radii and area of aggregates automatically, while visualizing the picked particles at the same time. The GUI application 'softEM' was developed in MATLAB using suitable functions from the Image Processing Toolbox. The application first de-noises the image iteratively by reducing the Huber penalty function, as described in the next section. The de-noised image is then low-pass filtered, in order to obtain a highly blurred background of the image. This background is used to normalize the image so that a single threshold can be applied to the whole image. This threshold, applied to determine the binary mask from the image, is slightly higher than the one estimated using Otsu's method (Supplementary Fig. 2). The user is provided the option to vary this threshold as well as the background filter size in order to ensure that the mask can be determined efficiently from micrographs with different number of aggregates and different extent of staining. The 8-connected components are labelled as particles and information such as area characteristics, equivalent radii, position and boundaries of aggregates can be easily determined from the mask image.

2.5.1. Size characterization

The algorithm for size characterization is described below:

(a) The micrograph is first de-noised by reducing the Huber penalty function [31], which is a well-known "edge-preserving" penalty, i.e., it de-noises the image while preserving the features of interest. The Huber penalty function is defined as:

$$H(g) = \sum_{k=all\ pixels} \left[\sqrt{1 + \frac{|\nabla g_k|^2}{\delta^2}} - 1 \right]$$

Here g is the 2D image and the parameter δ is determined using the statistics of gradient magnitudes $|\nabla g_k|$ for index k ranging over all pixels in the image. For image pixels with gradient magnitude $\ll \delta$, the penalty function is effectively a quadratic gradient penalty which is known to have smoothening property. On the other hand for pixels with gradient magnitude $\gg \delta$ the penalty is equivalent to the Total Variation (TV) which is known to be edge preserving. This penalty function is minimized using 20 iterations of gradient descent which lead to a de-noised image [32].

(b) The next step is to estimate the local background variations in the micrograph due to uneven presence of the negative stain. This is done by low-pass filtering the micrograph using a very small aperture window. This window size is provided as a parameter in the GUI which can be altered by the user in case the user feels the background estimation is not being done correctly, which may happen if the magnification is much higher than 100,000. In this case, the window size must be reduced further so that the particle information is not included in the background.

(c) After the background has been estimated, this background is used to normalize the de-noised micrograph image. This yields a background flattened image, where the variations in the micrograph due to the presence of stain have been removed. In this image, a single threshold can be applied to the whole image. This threshold is calculated using the well-known Otsu's method, which involves minimizing the variance of the background as well as foreground pixels. In practice, it was observed that a threshold slightly higher (β = 1.1) than the calculated threshold provided better results. The applied threshold (=β*Otsu's threshold) can be controlled by the user in case the default threshold does not yield satisfactory results.

(d) Once the binary mask has been obtained after applying the threshold, the white pixels are characterized as aggregates while the black pixels are treated as background. The statistics of the size distribution of these aggregates is determined using Blob (binary large object) analysis method, where 8-connected pixels are treated as a single aggregate. This algorithm is able to determine the number of aggregates and area of each of these aggregates in square pixels. The calculation of actual area of each pixel is described in the next section and using this value, the area of each aggregate can be calculated in square nano-meters.

2.5.2. Size-based clustering of aggregated species

The pixel size was converted to nm based on the scan parameters of the camera (FEI Eagle 4 k × 4 k CCD attached to a 200 KV FEI-Tecnai FEG-TEM). The characterized antibody aggregates were clustered based on their equivalent radius (r_e) defined as

$$r_e = \sqrt{\frac{A}{\pi}}$$

where the area $A = n_p \times$ area of each pixel for the system, where n_p is the number of pixels occupied by the aggregate. This is the equivalent radius of a circle having the same area. The aggregated species were clustered and distributed into bins of equivalent radii.

3. Results and discussion

3.1. Negative staining electron microscopy of mAbs show heterogeneous aggregates in varied quantities

Three monoclonal antibodies, mAb A, B, and C, were used in this study. mAb A showed discrete particles with limited aggregation in DPBS buffer (Fig. 1A). mAb B and C showed varied degrees of heterogeneous, aggregated species in buffers containing 15 mM Phosphate, pH 6.5 and 200 mM NaCl (Fig. 1B–C), while aggregation of mAb B appeared to be significantly reduced in DPBS (Fig. 1D). To quantify the degree of aggregation by other means, mAb B in DPBS buffer was subjected to size-exclusion chromatography, and the percentage of aggregation was determined by measuring the area under monomer and aggregate peaks. However, the percentage aggregation (~0.48%) obtained from *SEC* (Supplementary Fig. 1 (e)), did not correlate with visual analysis of aggregate sizes or patterns (Fig. 1D). This discrepancy could be attributed to the co-migration of different sized species in the same *SEC* peak, due to limitations in achievable resolution. We therefore attempted a manual, size-based classification of aggregated species from micrographs directly, with the goal of quantifying the degree of aggregation, independent of the information obtained from *SEC*.

3.2. Manual counting of aggregates from electron micrographs

For optimizing manual counting methods, mAb B in DPBS was subjected to multiple *SEC* runs to quantify the degree of aggregation, which remained at ~ 0.5% (data not shown). Concurrently with *SEC*, the samples were also visualized by electron microscopy multiple times, and the particles in representative micrographs were subjected to manual counting by a person trained in electron microscopy (user A), and a novice (user B), for the purpose of comparison of data generated by different users.

It was observed that the count varied widely, depending on the number and size of aggregated species present in the image. Particularly, images with a granular distribution of relatively smaller antibody aggregates were hard to quantify correctly (Supplementary Fig. 1(a–c)). This underlined the essentiality of introducing automation in the identification of small size proteins/particles.

3.3. Automated identification of aggregates from electron micrographs

Attempts were made to utilize the "particle picking" module from single particle reconstruction softwares for automated identification of various species from negatively stained micrographs of mAb B. Usually, these softwares use a specific box size for particle picking; and a set of particles of same size and shape but different orientations, are used for automated identification. Such methods are difficult to implement in cases where there are a significant amount of aggregated or oligomeric species of different sizes and shapes. When applied to our samples, we observed that these softwares identified particles randomly, probably due to the large variability in size and morphology of particles, and the counts varied widely. Other available tools for cell counting were also utilized with limited success (Fig. 2).

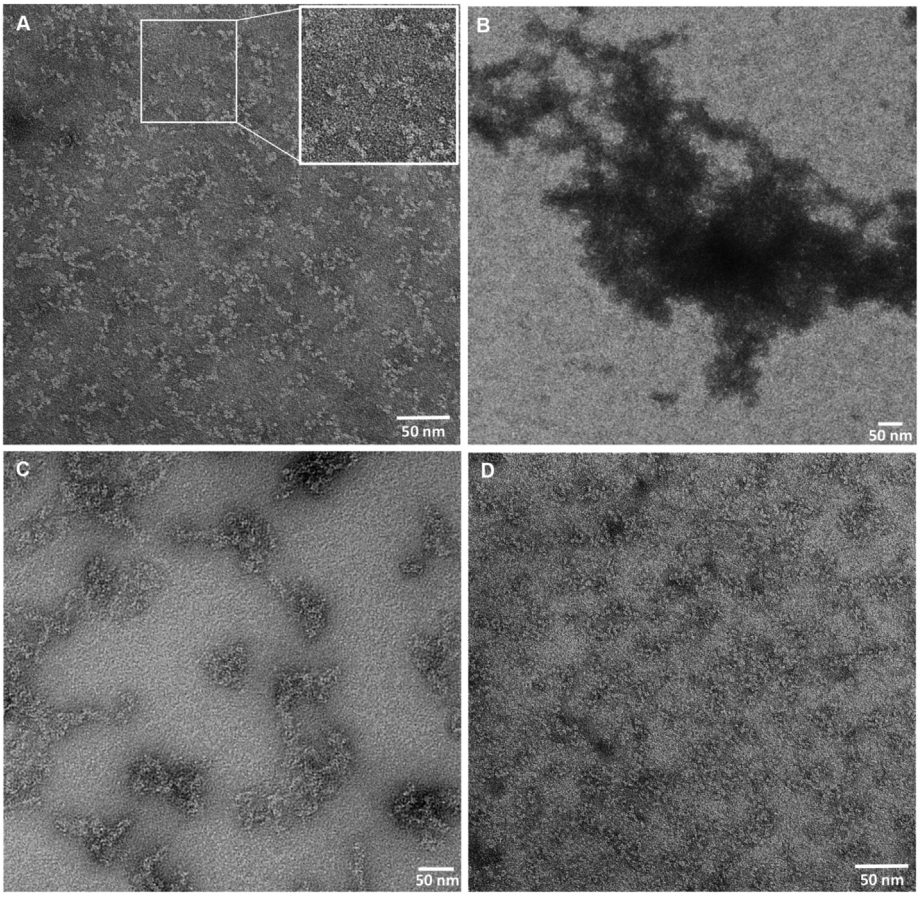

Fig. 1. TEM images of different mAbs showing individual antibody as well as varying patterns of aggregation: (A) mAb A showing discrete individual particle along with different heterogeneous aggregates. (B) (C) high level aggregation in different mAbs. (D) mAb B particles showing varied smaller aggregates. Magnification 100,000X. Scale Bar 50 nm.

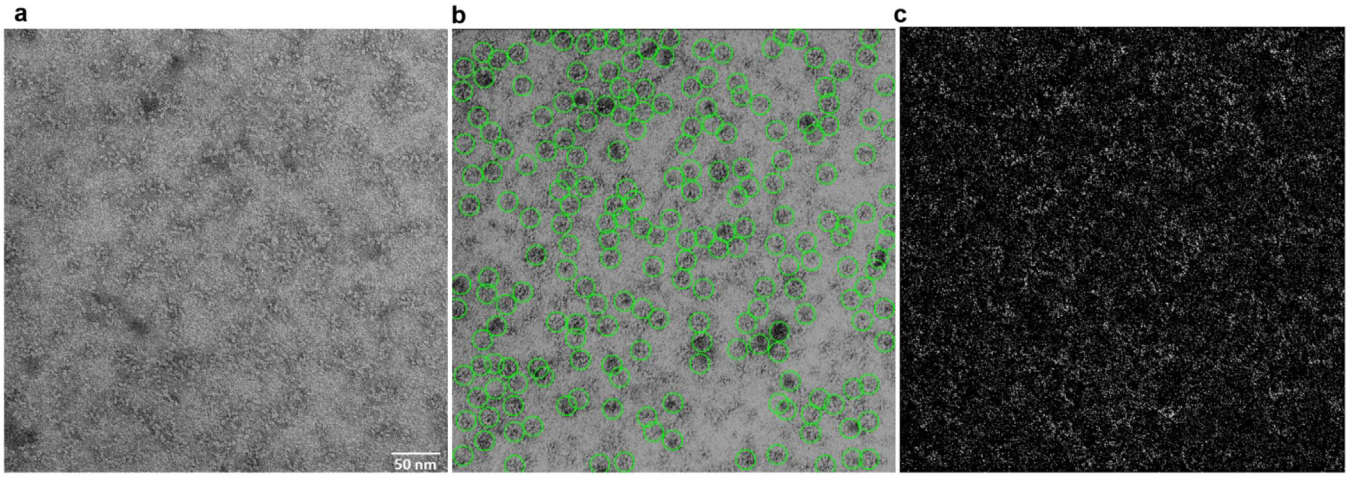

Fig. 2. Counting of monoclonal antibodies using available methods: An electron micrograph of a fraction of mAb B (panel a), subjected to automated counting of ''particles" using the 3D reconstruction program Relion, version 1.4 (panel b) and by Fiji, a cell counting extension of ImageJ (panel c).

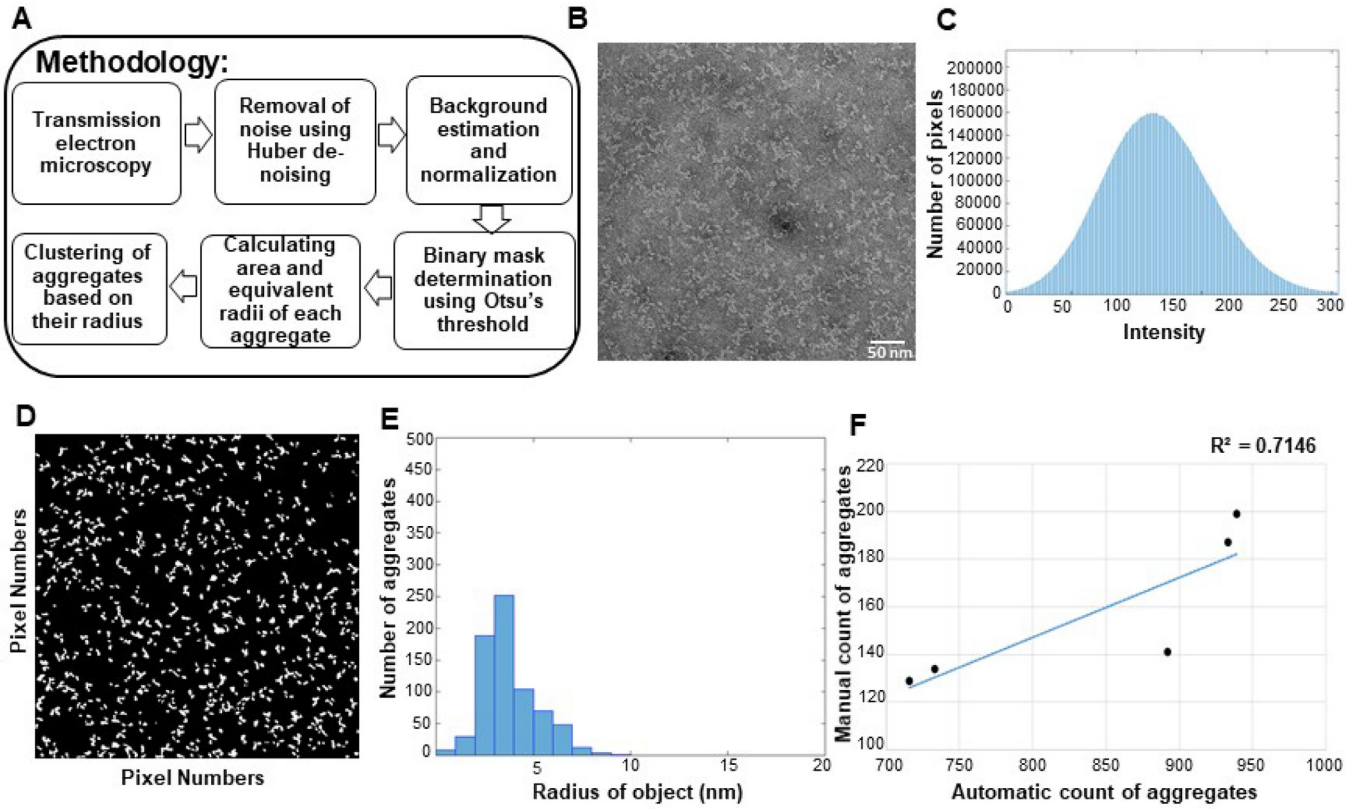

Fig. 3. Validation of algorithm: (A) Schematic depiction of the method utilized to count and cluster heterogeneous aggregates from electron micrographs. (B–F) Pictorial depiction of the method with panels representing the raw micrograph, histogram of the micrograph, and binary form of a micrograph (B–D), clustering of aggregates in groups based on equivalent radii (E), and a Y-Y plot showing agreement between manual and automated counting values of 5 such micrographs (F).

In order to identify and classify heterogeneous species present in micrographs, we developed a connected component labeling-based method, which was capable of automatically selecting a highly heterogeneous population of aggregates and performing a size-based classification (Fig. 3A–F).

First, efforts were made to reduce background noise in the micrographs, particularly in images containing a granular distribution of density. An examination of micrographs revealed that first, they have a high frequency of pixel noise and second, the grey-scale values for densities corresponding to antibodies are normally distributed, but skewed (Fig. 3B). On applying Gaussian filter directly to these images, the densities tended to lose information at their edges and became blurry. A Huber penalty was applied to preserve the information by smoothing the edges of these aggregates. About 20 gradient descent steps were used to reduce the Huber cost, which ultimately resulted in images where the background noise smoothened out and various species could be clearly visualized. To select the visualized antibody units from a de-noised image we applied Otsu thresholding [32] and the connected component labelling method [33]. The processed de-noised image was

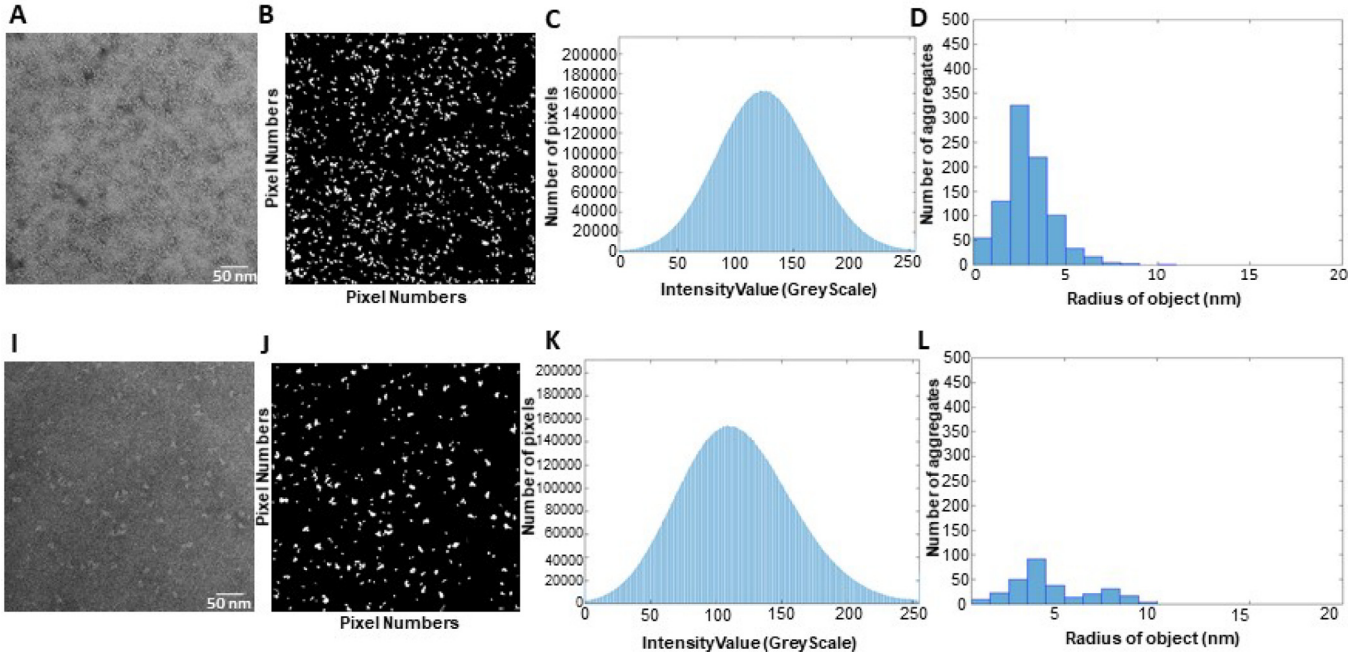

Fig. 4. Application of algorithm to antibody samples: Application of the algorithm to fractions of mAb B and mAb C (top and bottom), with aggregation levels of 0.48% and 0.63%, computed from SEC profiles. Upper panels A, B, C, D and below panels I, J, K and L represent the stages of the algorithm (raw micrograph (A/I), binary image (B/J), intensity histogram of raw micrograph (C/K), and clustering (D/L)) as applied to the fractions respectively.

divided by its highly blurred low-pass version that flattened the background of the whole micrograph so that the same threshold could be applied across the whole image. The binary thresholded image was subjected to a series of morphological image operations applying the connected component labelling method to create a masked image which represented the required clustered areas in the original image (Fig. 3D). For counting individual units, the mean cluster area of the unit was determined and a user-defined threshold was set to include area values higher or lower than the mean (Fig. 3C, E). The method was applied to more than 23 micrographs corresponding to different fractions of mAbs A, B, and C. Of these, the number of aggregates in 5 micrographs were counted manually in triplicates and compared with semi-automated picking defined by the algorithm. A comparison of the number of units manually counted by User A and calculated by the algorithm demonstrated satisfactory agreement, with the Y-Y plot providing an R^2 value of 0.7146 (Fig. 3F). This suggested that the method is successful in the semi-automated picking of heterogeneous particles, and can closely mimic the trajectory of manual picking by a trained electron microscopist.

3.4. Clustering of aggregates

As the aggregates picked were irregular in shape, the number of pixels associated with each aggregate, which is a function of the width of the filter chosen by the user, was utilized to calculate its equivalent radius (Fig. 3E). The distribution of aggregated species present in a sample (Fig. 3E) could be qualitatively verified by reverting to the binary image corresponding to the original micrograph.

3.5. Generating a numerical distribution of aggregation in samples

Our method produced size-distribution profiles of mAb units that quantitatively represented micrographs. The method was applied to different fractions of mAbs B and C containing similar levels of aggregation (0.48% and 0.63% respectively), as measured by *SEC* (Fig. 4).

However, negative staining and electron microscopy showed that the aggregate sizes in the two fractions substantially (Fig. 4A and I). While size-distribution histograms for the first sample indicated primarily smaller size particles, which can correspond to monomers or smaller oligomeric species (Fig. 4B, D), the corresponding analysis for the second sample showed several species present in relatively larger size ranges, which could indicate larger oligomers or aggregates (Fig. 4J, L). This identification of larger sized aggregates in a therapeutic mAb formulation constitutes information over and above that obtained through gold-standard *SEC*, and may direct further characterization of the formulation.

4. Discussion

Since the development of biotherapeutics is witnessing an enormous growth worldwide, it has become imperative to establish effective protocols for quickly analyzing their stability and purity [8,9].

Characterization of biosimilars or other therapeutic proteins is often accomplished by using a multitude of high resolution, orthogonal analytical tools. Mass spectroscopy, CD spectroscopy, HPLC, electrophoresis-based assays etc. are utilized for generating a detailed physicochemical profile for pre-clinical assessment [34], however, none of these techniques provide a direct visualization-based size categorization. On the other hand, 3D structure determination techniques such as X-ray crystallography, Nuclear Magnetic Resonance, and cryo-electron microscopy, while ideal for molecular-level comparisons, are not suitable for analysis of a highly heterogeneous population of morphological units. We propose an electron microscopy based semi-quantitative method, which can potentially fill the gap between non-visual, indirect techniques and direct, high-resolution structure determination by providing the means to obtain a numerical distribution of heterogeneous aggregates present in protein-based drugs, and provide

requisite information to reduce aggregation under various conditions of manufacture and storage.

Conventional electron microscopy is a useful tool to detect morphological units present in a heterogeneous sample by direct visualization [24]. Apart from the visualization of oligomerization and aggregation, micrographs can also provide information regarding the physical nature of aggregates like shape, size and indicate potential resolvability. Our method involves de-noising micrographs by minimizing the Huber penalty to compensate for edge blurring effect, flattening of background, followed by Otsu thresholding, and obtaining a numerical map of morphological units present in the micrograph by application of the connected component labelling method. The various species are eventually clustered into groups based on their equivalent radii to generate a size histogram, and in our hands has provided the best match with manual characterization, compared to other available methods (Fig. 3). A drawback of the method at this point is determination of equivalent radii of species, and not their exact dimensions; however, this method does provide an understanding of the categories of higher order species present in a sample, which probbaly cohabit in *SEC* chromatograms (Fig. 4D and Supplementary Fig. 1(e)). The application of this algorithm to different samples of mAbs adds significantly to data obtained from traditional techniques like *SEC*. Unlike *SEC* and DLS, this method also offers the opportunity for side-by-side, visual evaluation of micrographs, and their corresponding binary images.

This method for size characterization of a heterogeneous population is distinctly different from currently available particle picking routines for cryo-electron microscopy and 3D reconstruction, which have completely different targets and purpose [35,36]. These methods are trained to pick similar particles at different orientations, and are thus not applicable to largely heterogeneous samples (Fig. 2). As protein aggregates are fairly diverse in terms of shape and size, we found it essential to develop an automated particle picking method, which could identify objects that are widely dissimilar in terms of shape and size. Connected component labelling has been applied towards particle picking previously for heterogeneous populations [35,37], but additional denoising and thresholding produced comparatively better results for studying smaller heterogeneous particles (Fig. 3). Although we have applied this method to negatively stained electron micrographs, it is equally applicable to images of diverse particles generated through other visual techniques. Cryoelectron microscopy and 3D reconstruction can be applied to understand high resolution moleular structures of monomeric proteins and discrete oligomeric species. However, the necessity for a simpler method without extensive sample processing or computational requirements, to accurately identify the presence and gross dimensions of aggregated species in samples, exists in parallel.

Traditionally, techniques such as mass spectroscopy (MS), circular dichroism (CD) spectroscopy, dynamic light scattering (DLS), and size exclusion chromatography (*SEC*) have been utilized to monitor the primary sequence, amino acid modifications, secondary structure, and aggregation state of therapeutic proteins [19,22,34]. In our study, *SEC* of mAb B and C indicated the presence of similar, negligible degrees of aggregation (Supplementary Fig. 1 (e)). The DLS size distribution profiles showed peak maxima at ~ 10 nm of hydrodynamic diameter for both mAbs, and no other species (Supplementary Fig. 3). We attribute these discrepancies to the lack of adequate resolution in these methods resulting in co-migration of different species in same peaks. The CD spectra of mAb B and C were characteristic of typical antibody profiles, with a dip at 218 nm indicating significant presence of anti-parallel β sheet secondary structures (Supplementary Fig. 4). This is expected in cases of oligomerization and reversible aggregation, where unfolding of the protein secondary structure is not expected. However, our method combining visual and quantitative evaluation detects a difference in the size of species present in the mAbs indicating potential aggregation, which constitutes essential information in context of biosimilar production.

Avoiding aggregation during production, transport, and storage is essential in order to reduce unwanted immune reactions towards complex protein therapeutics [38,39]. Although traditional 3D structure determination techniques can provide comprehensive data about secondary and tertiary structures as well as conformational alterations and posttranslational modification of protein therapeutics, it is essential to have a simple, quantitative method to identify and characterize heterogeneity in proteins at submicron ranges. Our method here showed that picking of heterogeneous particles from softEM provide additional information in terms of size and distribution of mAbs that validates the requirement of present analysis for rationalized development and assessment of heterogeneous proteins.

5. Conclusion

We have developed a GUI based software "softEM" for automated picking and size-based clustering of all species present in negatively stained electron micrographs of protein samples. This method combines visual and quantitative methods for identification of aggregates from image data, and can be applied to multiple types of images containing heterogeneous populations of biomolecules. It is hoped that this freely available method can serve as an orthogonal technique for analysis of aggregation in therapeutic proteins.

Author contributions

M.K. and R.B. designed and conducted the experiments; K.K., J. G., A.P. and A.P. devised the method for automated counting and size-based clustering; M.B. and A.S.R. wrote the paper.

CRediT authorship contribution statement

Mohit Kumar: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation, Validation, Writing - review & editing. **Apoorv Pant:** Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation, Validation, Writing - review & editing. **Rohit Bansal:** Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation, Validation, Writing - review & editing. **Ashutosh Pandey:** Conceptualization, Methodology, Software, Visualization, Investigation, Validation. **James Gomes:** Supervision. **Kedar Khare:** Supervision. **Anurag Singh Rathore:** Supervision. **Manidipa Banerjee:** Data curation, Writing - original draft, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge support from Grant no. BT/COE/34/SP15097/2015 of Department of Biotechnology, India, IIT-Delhi and Kusuma Trust, UK. The authors also thank Professor Steven Ludtke, and Muyuan Chen, Baylor College of Medicine, Houston, Texas, USA for providing valuable comments on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.06.009.

References

[1] Wang W. Advanced protein formulations. Protein Sci 2015;24:1031–9.
[2] Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, Ramponi G, et al. Kinetic partitioning of protein folding and aggregation. Nat Struct Biol 2002;9:137–43.
[3] Amin S, Barnett GV, Pathak JA, Roberts CJ, Sarangapani PS. Protein aggregation, particle formation, characterization & rheology. Curr Opin Colloid Interface Sci 2014;19:438–49.
[4] Arakawa T, Philo JS, Ejima D, Sato N, Tsumoto K. Aggregation analysis of therapeutic proteins, part 3. Bioprocess Int 2007;5:52–70.
[5] Gramppand G, Ramanan S. The diversity of biosimilar design and development: implications for policies and stakeholders. BioDrugs 2015;29:365–72.
[6] Rathore AS. Follow-on protein products: scientific issues, developments and challenges. Trends Biotechnol 2009;27:698–705.
[7] Wang W. Protein aggregation and its inhibition in biopharmaceutics. Int J Pharm 2005;289:1–30.
[8] Crommelin DJ, Shah VP, Klebovich I, McNeil SE, Weinstein V, Fluhmann B, et al. The similarity question for biologicals and non-biological complex drugs. Eur J Pharm Sci 2015;76:10–7.
[9] Khraishi M, Stead D, Lukas M, Scotte F, Schmid H. Biosimilars: a multidisciplinary perspective. Clin Ther 2016;38:1238–49.
[10] Mahler HC, Friess W, Grauschopf U, Kiese S. Protein aggregation: pathways, induction factors and analysis. J Pharm Sci 2009;98:2909–34.
[11] Roberts CJ. Therapeutic protein aggregation: mechanisms, design, and control. Trends Biotechnol 2014;32:372–80.
[12] Bansal R, Gupta S, Rathore AS. Analytical platform for monitoring aggregation of monoclonal antibody therapeutics. Pharm Res 2019;36:152.
[13] den Engelsman J, Garidel P, Smulders R, Koll H, Smith B, Bassarab S, et al. Strategies for the assessment of protein aggregates in pharmaceutical biotech product development. Pharm Res 2011;28:920–33.
[14] Carpenter JF, Randolph TW, Jiskoot W, Crommelin DJ, Middaugh CR, Winter G. Potential inaccurate quantitation and sizing of protein aggregates by size exclusion chromatography: essential need to use orthogonal methods to assure the quality of therapeutic protein products. J Pharm Sci 2010;99:2200–8.
[15] Taylor FR, Prentice HL, Garber EA, Fajardo HA, Vasilyeva E, Blake Pepinsky R. Suppression of sodium dodecyl sulfate-polyacrylamide gel electrophoresis sample preparation artifacts for analysis of IgG4 half-antibody. Anal Biochem 2006;353:204–8.
[16] Thomasand TC, McNamee MG. Purification of membrane proteins. MethodsEnzymol 1990;182:499–520.
[17] Fekete S, Beck A, Veuthey JL, Guillarme D. Theory and practice of size exclusion chromatography for the analysis of protein aggregates. J Pharm Biomed Anal 2014;101:161–73.
[18] Li Y, Lubchenko V, Vekilov PG. The use of dynamic light scattering and Brownian microscopy to characterize protein aggregation. Rev Sci Instrum 2011;82:053106.
[19] Philo JS. A critical review of methods for size characterization of non-particulate protein aggregates. Curr Pharm Biotechnol 2009;10:359–72.
[20] Nobbmann U, Connah M, Fish B, Varley P, Gee C, Mulot S, et al. Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. Biotechnol Genet Eng Rev 2007;24:117–28.
[21] Sung JJ, Pardeshi NN, Mulder AM, Mulligan SK, Quispe J, On K, et al. Transmission electron microscopy as an orthogonal method to characterize protein aggregates. J Pharm Sci 2015;104:750–9.
[22] Joshi V, Shivach T, Yadav N, Rathore AS. Circular dichroism spectroscopy as a tool for monitoring aggregation in monoclonal antibody therapeutics. Anal Chem 2014;86:11606–13.
[23] Iacob RE, Krystek SR, Huang RY, Wei H, Tao L, Lin Z, et al. Hydrogen/deuterium exchange mass spectrometry applied to IL-23 interaction characteristics: potential impact for therapeutics. Expert Rev Proteomics 2015;12:159–69.
[24] Barnard JG, Kahn D, Cetlin D, Randolph TW, Carpenter JF. Investigations into the Fouling mechanism of parvovirus filters during filtration of freeze–thawed mAb drug substance solutions. J Pharm Sci 2014;103:890–9.
[25] De Carloand S, Harris JR. Negative staining and cryo-negative staining of macromolecules and viruses for TEM. Micron 2011;42:117–31.
[26] Fast JL, Cordes AA, Carpenter JF, Randolph TW. Physical instability of a therapeutic Fc fusion protein: domain contributions to conformational and colloidal stability. Biochemistry 2009;48:11724–36.
[27] Filipe V, Que I, Carpenter JF, Lowik C, Jiskoot W. In vivo fluorescence imaging of IgG1 aggregates after subcutaneous and intravenous injection in mice. Pharm Res 2014;31:216–27.
[28] Badami A, Beach M, Wood S, Rozeveld S, Marshall J, Heeschen W, et al. Evaluation of methods for quantification of transmission electron microscopy (TEM) dispersions. Microsc Microanal 2009;15:1088–9.
[29] Brown AP, Brydson RMD, Hondow NS. Measuring in vitro cellular uptake of nanoparticles by transmission electron microscopy. J Phys Conf Ser 2014;522:012058.
[30] Dieckmann Y, Colfen H, Hofmann H, Petri-Fink A. Particle size distribution measurements of manganese-doped ZnS nanoparticles. Anal Chem 2009;81:3889–95.
[31] Khare K. Fourier Optics and Computational Imaging. John Wiley & Sons; 2015.
[32] Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybernet 1979;9:62–6.
[33] Sametand H, Tamminen M. Efficient component labeling of images of arbitrary dimension represented by linear bintrees. IEEE Trans Pattern Anal Machine Intell 1988;10:579–86.
[34] Berkowitz SA, Engen JR, Mazzeo JR, Jones GB. Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. Nat Rev Drug Discov 2012;11:527–40.
[35] Nicholsonand WV, Glaeser RM. Review: automatic particle detection in electron microscopy. J Struct Biol 2001;133:90–101.
[36] Scheres SH. A Bayesian view on cryo-EM structure determination. J Mol Biol 2012;415:406–18.
[37] Harauzand G, Fong-Lochovsky A. Automatic selection of macromolecules from electron micrographs by component labelling and symbolic processing. Ultramicroscopy 1989;31:333–44.
[38] Pineda C, Hernández GC, Jacobs IA, Alvarez DF, Carini C. Assessing the immunogenicity of biopharmaceuticals. BioDrugs 2016;30:195–206.
[39] Wadhwa M, Knezevic I, Kang H-N, Thorpe R. Immunogenicity assessment of biotherapeutic products: an overview of assays and their utility. Biologicals 2015;43:298–306.