Received: 24 April 2020 | Revised: 1 July 2020 | Accepted: 13 July 2020

DOI: 10.1002/bit.27501

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

Machine learning and statistical analyses for extracting and characterizing "fingerprints" of antibody aggregation at container interfaces from flow microscopy images

Austin L. Daniels[1] | Christopher P. Calderon[1,2] | Theodore W. Randolph[1]

[1]Department of Chemical and Biological Engineering, Center for Pharmaceutical Biotechnology, University of Colorado Boulder, Boulder, Colorado

[2]Ursa Analytics, Denver, Colorado

Correspondence
Christopher P. Calderon, Ursa Analytics, 3609 Osceola St, Denver, CO 80212.
Email: Chris.Calderon@UrsaAnalytics.com

Theodore W. Randolph, Department of Chemical and Biological Engineering, 596 UCB, University of Colorado, Boulder, Boulder, CO 80309.
Email: theodore.randolph@colorado.edu

Funding information
National Institutes of Health, Grant/Award Number: R41GM130513

Abstract

Therapeutic proteins are exposed to numerous stresses during their manufacture, shipping, storage and administration to patients, causing them to aggregate and form particles through a variety of different mechanisms. These varied mechanisms generate particle populations with characteristic morphologies, creating "fingerprints" that are reflected in images recorded using flow imaging microscopy. Particle population fingerprints in test samples can be extracted and compared against those of particles produced under baseline conditions using an algorithm that combines machine learning tools such as convolutional neural networks with statistical tools such as nonparametric density estimation and Rosenblatt transform-based goodness-of-fit hypothesis testing. This analysis provides a quantitative method with user-specified type 1 error rates to determine whether the mechanisms that produce particles in test samples differ from particle formation mechanisms operative under baseline conditions. As a demonstration, this algorithm was used to compare particles within intravenous immunoglobulin formulations that were exposed to freeze-thawing and shaking stresses within a variety of different containers. This analysis revealed that seemingly subtle differences in containers (e.g., glass vials from different manufacturers) generated distinguishable particle populations after the stresses were applied. This algorithm can be used to assess the impact of process and formulation changes on aggregation-related product instabilities.

KEYWORDS
image analysis, machine learning, protein aggregation, protein formulation

1 | INTRODUCTION

Aggregation is a major challenge in the manufacturing of therapeutic proteins (Randolph & Carpenter, 2007; Roberts, 2014; Wang, 1999). Numerous stresses encountered during protein production cause aggregation. These different stresses (e.g., freeze-thawing; Arsiccio & Pisano, 2017; Barnard, Singh, Randolph, & Carpenter, 2011; Twomey, Less, Kurata, Takamatsu, & Aksan, 2013, interactions at air-water and container-water interfaces; Cordes, Carpenter, & Randolph, 2012; Ludwig, Carpenter, Hamel, & Randolph, 2010; Sethuraman, Morcone, & Belfort, 2004; Sluzky, Klibanov, & Langer, 1992; Webb, Cleland, Carpenter, & Randolph, 2002, exposure to excipient degradation products such as those from polysorbates; Ha, Wang, & Wang, 2002; Kerwin, 2008; Wasylaschuk et al., 2007, pH extremes; Chi, 2004; Thirumangalathu, Krishnan, Brems, Randolph, & Carpenter, 2006, and elevated temperatures) produce polydisperse distributions of aggregates (Joubert, Luo, Nashed-Samuel, Wypych, & Narhi, 2011). As a result, aggregates may be observed in protein formulations following purification (Arakawa, Ejima, & Akuta, 2017), filtration (Barnard, Kahn, Cetlin, Randolph, & Carpenter, 2014;

Liu, Randolph, & Carpenter, 2012; A. Sharma, Anderson, & Rathore, 2008), pumping (Saller et al., 2016; Tyagi et al., 2009; Tzannis, Hrushesky, Wood, & Przybycien, 1996), freezing (Barnard et al., 2011; Kolhe, Amend, & Singh, 2010; Kueltzo, Wang, Randolph, & Carpenter, 2008; Vlieland et al., 2018), vial filling (Nayak, Colandene, Bradford, & Perkins, 2011), viral clearance steps and shipping (Siska, Harber, & Kerwin, 2020). The potential role of these aggregates in provoking unwanted immune responses (Chisholm et al., 2017; Fradkin, Carpenter, & Randolph, 2009; Freitag et al., 2015; Jiskoot et al., 2016; Rosenberg, 2006) has generated interest in developing techniques to identify their root causes.

The root cause of protein aggregation is often elusive. However, the various stresses that promote protein aggregation each induce aggregation by somewhat different molecular mechanisms (Roberts, 2007; Wang & Roberts, 2018). These distinct mechanisms lead to particle populations whose size and morphology distributions comprise particle "fingerprints" that reflect the root cause of their formation. Better techniques for characterizing these particle fingerprints would provide methods to rapidly determine the root causes of particle formation in a sample.

Flow imaging microscopy (FIM) is a commonly used technique for analyzing size distributions of protein aggregates (Narhi et al., 2015; D. K. Sharma, King, Oma, & Merchant, 2010; D. K. Sharma, Oma, Pollo, & Sukumar, 2010; Zölls et al., 2013) and other particles. FIM uses light microscopy combined with microfluidics to capture digital images of particles larger than 1 μm in size contained within a sample. The output from this instrument is a set of digital images of individual particles in a small liquid sample (usually about 10^3–10^5 images per 200 μl sample). The images contain a large amount of morphological information. However, in common practice, most of the morphology information potentially available from FIM measurements is not utilized.

Convolutional neural networks (ConvNets) can be used to extract and analyze morphological information embedded in FIM images (Calderon, Daniels, & Randolph, 2018; Gambe-Gilbuena, Shibano, Krayukhina, Torisu, & Uchiyama, 2020). ConvNets are a family of neural networks capable of learning relevant features from a collection of images that are useful when performing tasks such as classification and dimension reduction (Calderon et al., 2018; Esteva et al., 2017; Krizhevsky, Sutskever, & Hinton, 2012; Schroff, Kalenichenko, & Philbin, 2015). ConvNets trained on FIM data sets can accurately classify protein aggregates produced by different stresses. In Calderon et al. (2018) and Gambe-Gilbuena et al. (2020), a set of single, well-defined stresses (e.g., freeze-thawing and heating) was applied to protein solutions, causing aggregates to form. ConvNets were then trained on FIM images of the resulting particles to train classifiers to recognize particle morphologies generated by one of these stresses. The resulting classifiers were then used to classify FIM images of particles from new samples that had been subjected the same set of stresses.

Although these previous approaches are useful for analyzing protein aggregates within formulations exposed to single stresses, protein aggregates encountered in practice are likely the result of a superposition of a variety of stresses, yielding more varied fingerprints. The potentially large number of different aggregate sources may mask subtle but relevant changes in particle populations due to minor changes in process conditions such as changes in container-closure systems. This issue is compounded by the inherent variability in particle morphology even under tightly controlled conditions (Gambe-Gilbuena et al., 2020). Thus, it can be difficult to determine if morphology differences within a particle population reflect different root causes of aggregation or merely sample-to-sample variability.

In the present study, we demonstrate a ConvNet algorithm that can be used to quantitatively determine if particle morphologies recorded in a small collection of FIM images are statistically different from those generated under a user-defined baseline condition. This analysis uses a combination of dimension reduction and hypothesis testing. Facial recognition strategies (Sun, Chen, Wang, & Tang, 2014; Taigman, Yang, Ranzato, & Wolf, 2014) such as triplet loss approaches (Schroff et al., 2015) can reduce the dimensionality of FIM image data sets, compressing the information contained in color FIM images to two-dimensional (2D) feature vectors (i.e., the fingerprints). The extreme information compression enables the use of nonparametric techniques such as kernel density estimates of the probability density of these low-dimensional representations for particles made under a single baseline condition. Goodness-of-fit hypothesis test with user-tunable false-positive rates can then be used to compare collections of particle images from other samples to this density.

One potential application of this approach is testing whether formulation design decisions (e.g., pH, excipient concentrations, container-closure types) affect protein aggregate populations. In this study, we focus on the impact of container-closure systems on protein aggregate morphology. The geometry and chemistry of the container can affect protein aggregation (Kiese, PappenBerger, Friess, & Mahler, 2008). Container-induced particles may come directly from the container (e.g., glass flakes from delamination in glass vials; Ennis et al., 2001) as well as from protein aggregates triggered by the container itself (Bee, Randolph, Carpenter, Bishop, & Dimitrova, 2011; Gerhardt et al., 2014). Aggregation may depend not only on the type of container (Krayukhina, Tsumoto, Uchiyama, & Fukui, 2015; Kumru et al., 2012; Teska, Brake, Tronto, & Carpenter, 2016) but also may vary between different lots of the same container from a given manufacturer.

2 | MATERIALS AND METHODS

2.1 | Materials

Intravenous immunoglobulin (IVIg, Gammagard Liquid) was obtained from Takeda International (Lexington, MA). Phosphate-buffered saline (PBS) containing 144 mg/L potassium phosphate monobasic, 795 mg/L potassium phosphate dibasic, and 9,000 mg/L sodium chloride at pH 7.4 was obtained from Gibco (Waltham, MA). Polypropylene, 2 ml microcentrifuge tubes ("Plastic") were from Fisher Scientific (Waltham, MA). FIOLAX Clear 3 ml type 1 borosilicate glass

vials ("Glass 1") were obtained from Schott (Elmsford, NY). A second 3 ml type 1 borosilicate glass vial ("Glass 2") was obtained from Duran Wheaton Kimble (Mainz, Germany). Micro-90 was obtained from International Products Corp. (Burlington, NJ). Polystyrene 20-μm calibration beads were from Thermo Fisher Scientific (Waltham, MA).

2.2 | Generation of protein aggregates

IVIg aggregates were made using combinations of two aggregation-inducing stresses in three container types. Five experimental replicates were made per combination of container and stress. In each replicate, two containers were cleaned by filling the container with ultrapure water generated using a PURELAB flex 1 water deionization system from ELGA Labwater (Wycombe, UK), shaking the filled container, then emptying the container and allowing the container to air dry for 1 hr. IVIg stock solution was made by centrifuging the as-received drug product containing 100 mg/ml IVIg at 15,000*g* for 20 min at 4°C. in. The supernatant was then diluted to 0.5 mg/ml using filtered PBS, and 1.5 ml of this solution was filled into each container. Samples were then exposed to either freeze-thawing or shaking stresses as described below.

2.3 | Freeze-thaw stress

Samples stressed by freeze-thawing underwent four freeze-thaw cycles. During each cycle, the samples were suspended in a fixed orientation in liquid nitrogen for 4 min and then suspended in a 30°C water bath for 10 min. FIM analysis was performed immediately after the final freeze-thaw cycle was completed.

2.4 | Shaking stress

Samples were taped in a horizontal orientation onto an orbital plate shaker and shaken at 800 rpm for 4 hr. FIM analysis was performed immediately after the shaking was completed.

2.5 | FIM

FIM images were recorded with a FlowCam® VS instrument (Fluid Imaging Technologies, Inc., Scarborough, ME) with a ×10 objective, a field-of-view flow cell with a depth of 80 μm and width of 700 μm, and color imaging. The instrument was focused using the built in autofocus protocol for optimal image quality using 20-μm calibration beads. 1% Micro-90 solution followed by filtered ultrapure water were flushed through the instrument before and between measurements. The flash duration of the instrument was adjusted between replicates to achieve a constant background intensity of 150. Three 0.2 ml aliquots were analyzed from each replicate vial. Images were collected at a flow rate of 0.05 ml/min using 15 light and 17 dark pixel thresholds for particle segmentation.

2.6 | Image postprocessing

FIM images of particles were imported into Python 2.7. Before further analysis, the size of each image was adjusted to 24 × 24 pixels. Smaller images were padded with pixels sampled from a normal distribution with the same mean and variance as the border of the image and smoothed using Gaussian smoothing. For larger images a centered 24 × 24 crop of the image was used. Three experimental replicates for each combination of container and stress were used to train the algorithm, while the remaining two independent replicates were retained for use in subsequent testing. Fourteen thousand images were randomly selected from each of the three training replicates to be used as training data for the algorithm described in the next section. The remaining two replicates for each condition were not shown to the algorithm at all during training. Two thousand images from each replicate, including those not included in algorithm training, were set aside during algorithm training and used to test the performance of the trained algorithm.

2.7 | Algorithm overview

An algorithm was developed to determine if FIM images from a test sample were statistically consistent with those in a baseline sample. Hereafter, we refer to these FIM images as "particles" since each FIM image is recorded on a single particle. Figure 1 shows the process of training the algorithm to identify particles in a baseline sample. First, a ConvNet is trained on the collection of FIM images (Figure 1, first row, first column) to compress information within these images into a low-dimensional (2D here) point cloud of embeddings (Figure 1, second row, first column). A nonparametric kernel density estimate is then constructed from this low-dimensional point cloud to estimate the probability density of embeddings in the baseline sample (Figure 1, second row, second column). The estimated probability density is subsequently used to define a Rosenblatt transform which maps an embedding to a new random vector having the same dimensions of the embedding point (Rosenblatt, 1952; Figure 1, second row, third column). Goodness-of-fit hypothesis tests can be applied in conjunction with this Rosenblatt transform to determine if sets of FIM image embedding points are consistent with the estimated baseline density. The hypothesis test exploits the following mathematical fact: if a collection of embedding points are distributed according to the probability density associated with the baseline sample, the Rosenblatt transform yields multivariate random vectors whose components are independent and identically distributed with each component being a uniformly distributed random variable between 0 and 1. Goodness-of-fit hypothesis testing can formally check sets

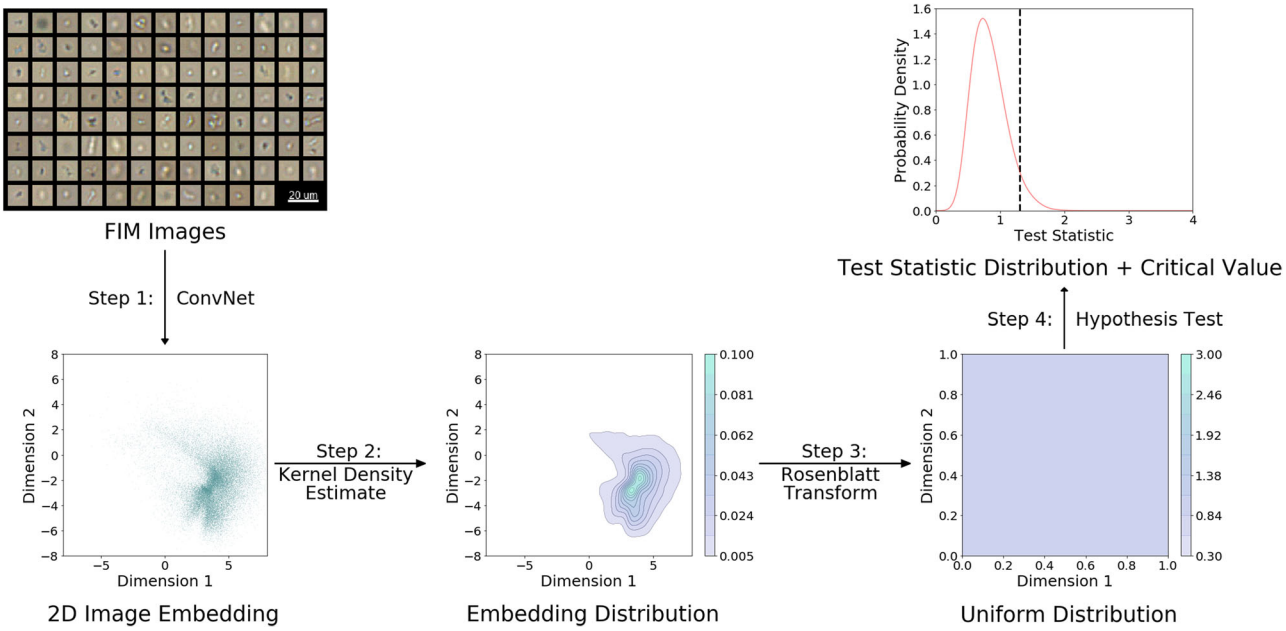

FIGURE 1 Flowchart showing how the algorithm is trained to detect particle populations similar to those made under some baseline condition. FIM images of particles made under the baseline condition (first figure) are used to train a ConvNet capable of compressing the image into a two-dimensional (2D) embedding (second figure). The probability density of these embeddings is then estimated using a kernel density estimate (third figure). A Rosenblatt transform defined using this distribution can then be used to map embeddings from the baseline sample onto a uniform distribution (fourth figure). The transformed embeddings can then be used to set up goodness-of-fit hypothesis tests by estimating the distribution of a test statistic for the baseline sample (fifth figure, curve) and using the distribution to determine an appropriate critical value for the test (fifth figure, dashed line). ConvNet, convolutional neural networks; FIM, flow imaging microscopy [Color figure can be viewed at wileyonlinelibrary.com]

of transformed embeddings for this property. Critical values for this hypothesis test are set by repeatedly subsampling Rosenblatt-transformed embedding points from the baseline sample, calculating test statistics for each subsample, selecting a value based on the resulting test statistic distribution to obtain a user-specified type I error (i.e., false-positive) rate (Figure 1, first row, third column).

Once trained, the algorithm can be used to quantify how statistically similar particle populations in test samples are to that in the baseline sample. Figure 2 shows the application of the trained algorithm to test samples containing either similar or different particle populations. To analyze a test sample, a small number of FIM images (e.g., 5–200) are subsampled from the test sample, converted to 2D embeddings with the trained ConvNet, and transformed using the Rosenblatt Transform defined by the baseline density. Goodness-of-fit hypothesis tests using the critical values from the baseline sample are then used to test if the transformed embeddings are consistent with a uniform distribution. Applying this algorithm to particles that resemble those in the baseline sample (Figure 2, top row) results in embeddings that are both visually and statistically consistent with those in the baseline sample. Conversely, particles that do not resemble the baseline sample (Figure 2, bottom row) yield embeddings less consistent with the baseline sample and are thus identified through goodness-of-fit hypothesis testing as a different particle population from the baseline sample.

2.8 | ConvNets

ConvNets are used in this analysis to extract and compress information in FIM images into a set of image features. While previously these image features were used as the input to a classifier that predicted the stress to which a sample had been exposed (Calderon et al., 2018), in the current analysis nonparametric techniques were used to estimate the distribution of these features. To apply these techniques, the ConvNet needed to be trained to learn extremely low-dimensional (i.e., 2–3 image features) representations of FIM images to avoid the exponential decrease in accuracy of these techniques with each additional dimension in the data (Scott, 2015).

The ConvNet used in this analysis was trained using a triplet loss algorithm, an approach that was developed for facial recognition to learn highly compressed image representations (Schroff et al., 2015). In this algorithm, a ConvNet is trained to learn a low-dimensional representation or embedding of images that acts to cluster together images from similar sources (e.g., faces of the same person, or protein aggregates made by the same stress and in the same container). During training, triplets (i.e., sets of three images) are assembled from the training data consisting of an image of one particle type (the anchor image), another image of the same particle type (a positive image), and a third image of a different particle type (the negative image). These triplets are fed through the neural network to calculate embeddings for each of the three images. The network's parameters

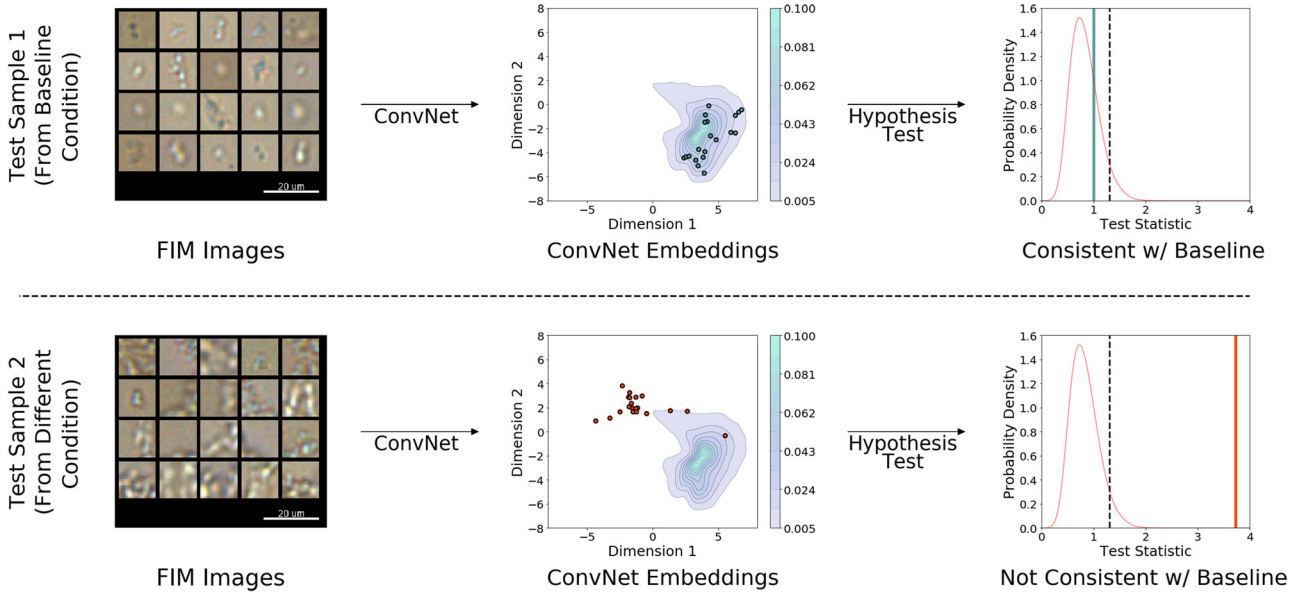

FIGURE 2 Flow chart showing the application of the algorithm to test samples that either resemble the baseline sample (top row) or do not resemble the baseline sample (bottom row). To perform the analysis small sets of images are selected from each sample (first column) and analyzed with the ConvNet to obtain two-dimensional embeddings for the images (second column, points). These embeddings are then compared against the distribution of embeddings for the baseline sample (second column, contour) using a combination of Rosenblatt transforms and hypothesis testing (third column). If the test statistic for the test sample (third column, solid line) is less than critical value for the baseline sample (third column, dashed line), the algorithm identifies that the particles in the test sample are consistent with the baseline sample. ConvNet, convolutional neural networks; FIM, flow imaging microscopy [Color figure can be viewed at wileyonlinelibrary.com]

are then adjusted to minimize a modified triplet loss function (Hermans, Beyer, & Leibe, 2017):

$$l = \ln(e^{-(d_{ap}-d_{an}+\alpha)} + 1), \tag{1}$$

where l is the triplet loss, d_{ap} is the Euclidian distance between the representations of the anchor and positive images returned by the ConvNet, d_{an} the distance between the representations of the anchor and negative image, and α is the margin, a small number used to scale the distances between dissimilar particle types in the embedding. This loss function is minimized when particles from a common source are close to each other in the embedding space and far apart from particles from other sources. In addition to allowing nonparametric density estimation techniques to be used, the resulting ConvNet can also be used to effectively analyze FIM image types not shown to the network during training. The algorithm is trained on a modest number (e.g., >10,000) of particles made under a small set conditions, but once trained, the network can then be used to analyze conditions not shown to the network during training using a smaller number of particles.

A ConvNet was trained to compress the particle information in 24 × 24 × 3 preprocessed FIM images into a 2D representation of the image. The network structure used in this study is shown in Table 1 . The Visual Geometry Group (VGG)-inspired network (Simonyan & Zisserman, 2015) uses convolutional layers with rectified linear unit (ReLU) activations in conjunction with max-pooling and dropout layers. Unlike the VGG structure, the first dense layer of the network used softplus activations (Dugas, Bengio, Bélisle, Nadeau, & Garcia, 2001) as the sparsity introduced by ReLU activations was found to cause nonsmooth features in embedding point clouds which subsequently complicated obtaining accurate kernel density estimates of the probability density of the points.

The ConvNet was trained on FIM images of particles produced in Plastic and Glass 1 vials after applying either freeze-thaw or shaking stresses. Particles generated within Glass 2 vials were not used to train the ConvNet, but instead were used to test the network's generalization to unseen particle types. The network was trained with a margin of 0.5 using minibatches of 64 triplets using an Adam optimizer (Kingma & Ba, 2015) with a 0.001 learning rate. Triplet minibatches were generated by assembling minibatches of 64 anchor images from the training images and calculating image embeddings for each training image at the start of each epoch. Positive and negative images for each anchor image were then randomly selected from all training images until a triplet was found that met semihard triplet mining criteria (Schroff et al., 2015) based on the most recentlycalculated embeddings. This approach filters out triplets that have low and high values of the loss function which can prevent the network from learning effective image representations. The current value of the triplet loss function, as well as the variance in embeddings from each condition, was monitored during training at the end of each epoch. The network was trained for 100 epochs and the network parameters that minimized the triplet loss was used in subsequent steps of the analysis.

TABLE 1 ConvNet structure used in this study

Layer no.	Layer type	No. of features	Feature size	Activation	Input shape	Output shape
1	Convolutional	32	3 × 3	ReLU	24 × 24 × 3	22 × 22 × 32
2	Convolutional	32	3 × 3	ReLU	22 × 22 × 32	20 × 20 × 32
3	Dropout (10% rate)	–	–	–	20 × 20 × 32	20 × 20 × 32
4	Convolutional	32	3 × 3	ReLU	20 × 20 × 32	18 × 18 × 32
5	Convolutional	64	3 × 3	ReLU	18 × 18 × 32	16 × 16 × 64
6	Max pooling (2 × 2)	–	–	–	16 × 16 × 64	8 × 8 × 64
7	Dropout (10% rate)	–	–	–	8 × 8 × 64	8 × 8 × 64
8	Convolutional	64	3 × 3	ReLU	8 × 8 × 64	6 × 6 × 64
9	Convolutional	64	3 × 3	ReLU	6 × 6 × 64	4 × 4 × 64
10	Flatten	–	–	–	4 × 4 × 64	1,024
11	Dropout (10% rate)	–	–	–	1,024	1,024
12	Dense	64	n/a	Softplus	1,024	64
13	Dense	2	n/a	None	64	2

Abbreviations: ConvNet, convolutional neural network; ReLU, rectified linear unit.

2.9 | Kernel density estimation

Kernel density estimation is a nonparametric technique for estimating the probability density function (PDF) of a data set using data sampled from this distribution (Scott, 2015). This technique was used to estimate the distribution of the low-dimensional FIM image embeddings for the baseline sample directly from the embeddings. Embedding sets from test samples were then compared against this distribution to decide if the particles in the test sample were consistent with those in the baseline sample.

Kernel density estimates of the distribution of embeddings for the baseline sample were constructed using a product kernel and using normal distributions as the kernel in each dimension. This kernel function was chosen so that the estimated PDF has an infinite support, which was helpful in obtaining meaningful evaluations of the Rosenblatt transform on particles that embedded far away from the mode of the PDF. The bandwidth of the kernel in each dimension was calculated using a normal reference rule (Scott, 2015):

$$h_i = \left(\frac{4}{d+2}\right)^{1/d+4} \sigma_i n^{-1/d+4}, \quad (2)$$

where h_i is the bandwidth in dimension i, d is the number of dimensions of the embeddings (2 in this study), n is the number of datapoints used to construct the density estimate, and σ_i is the standard deviation of the embeddings in dimension i.

2.10 | Rosenblatt transform

The Rosenblatt Transform is a statistical normalization that maps a d-dimensional random variable onto a new d-dimensional random variable using the PDF believed to govern the original variable (Rosenblatt, 1961). If the PDF does govern the initial random variable, the resulting random variable will be governed by a d-dimensional uniform distribution in which each dimension is independent and identically distributed. Goodness-of-fit hypothesis testing can then be used to check if the transformed variable is consistent with the uniform distribution. Performing this testing on the transformed variable allows us to use simple goodness-of-fit hypothesis test statistics with known distributions for a uniform null PDF—regardless of the initial null PDF that would be used if the testing was performed on the untransformed embeddings. The Rosenblatt transform also ensures that the two dimensions of the embeddings are independent under the null hypothesis. This feature of the transform simplifies setting critical values for the hypothesis test as the critical value does not need to be adjusted to account for dependencies in the embedding dimensions under the null hypothesis. Embedding sets from test samples were transformed based on the embedding PDF for the baseline sample before performing goodness-of-fit hypothesis testing.

2.11 | Goodness-of-fit hypothesis testing

After applying the Rosenblatt Transform defined by the density estimate for the baseline sample to image embeddings from the test sample, goodness-of-fit hypothesis tests were used to test the null hypothesis that the transformed embeddings are consistent with a uniform distribution. Rejection of this null hypothesis indicated that the particles in the test sample were not consistent with those in the baseline sample and thus potentially formed under a different set of conditions.

All goodness-of-fit hypothesis testing was performed using a Kolmogorov–Smirnov (KS) test (Darling, 1957), a one-dimensional

(1D) goodness-of-fit hypothesis test that compares the cumulative distribution function (CDF) of the embeddings to the CDF of a 1D uniform distribution. Two 1D KS tests were performed on each dimension of the transformed embeddings to test the null hypothesis as the dimensions of the transformed embeddings are independent under the null hypothesis. The null hypothesis was rejected if either dimension was not consistent with a uniform distribution. The desired overall type I error (i.e., false-positive) rate of the test can be used to set error rates for the two individual tests obtained using the Bonferroni correction. The overall null hypothesis was rejected if either of the two tests rejected the null hypothesis. A 5% overall type I error rate was used in this analysis.

Test statistics for the hypothesis test were calculated using sets of a small number (e.g., 5–200) of particles randomly selected from the test sample. Since the statistical power of these hypothesis tests scales with the number of data points used in the analysis, restricting the number of particles that are analyzed at once helped control the sensitivity of the analysis. In this study sets of either 20 or 200 particles were used to compare test samples to the baseline sample.

Monte Carlo simulations were used to select appropriate critical values (aiming at obtaining a user-specified type I error rate) of the test statistic to account for the bias introduced by both the nonparametric density estimate and subsampling scheme. The test statistic distribution for sets of 20 particles was estimated by randomly subsampling 10,000 sets of 20 training particles from the baseline condition and evaluating the test statistics for each subsample. These distributions were then used to select critical values at the appropriate significance level for each test. This process was repeated using sets of 200 particles, resulting in a second test statistic distribution and critical value for these larger particle sets.

Test samples were compared against baseline samples by repeatedly subsampling sets of particles from the test sample and using the algorithm to identify the fraction of these subsamples that were consistent with the baseline sample. 2,500 sets of either 20 or 200 particles were subsampled from the test sample. After computing the Rosenblatt transformed embeddings for each subsample, the hypothesis test was used to determine if each subsample was consistent with the baseline sample. The similarity between the test sample and the baseline sample using a given number of particles was recorded as the fraction of the 2,500 subsamples that did not contain particles consistent with the baseline sample.

2.12 | Particle morphology comparison

The algorithm described above was used to compare samples of aggregated IVIg formed under different stress conditions and in different containers. A ConvNet was trained on FIM images of particles made in Plastic and Glass 1 containers with a triplet loss approach. The remaining steps of the algorithm were then trained to identify sets of either 20 or 200 particles that resemble sets of the corresponding number of particles made in one container after exposure to one stress. These later steps were separately trained 12 times to cover the six possible baseline classes (particles made by one stress in one container) and the two-particle set sizes (20 or 200) that were used during testing.

The trained algorithms were used to investigate the impact of different stresses and different containers on particle morphology. This comparison was performed by comparing small sets of test particles from each sample to all the training particles from a single container and stress. To investigate the impact of stresses on particle populations, the algorithm was used to compare particles generated by freeze-thaw stress (the baseline stress class) in each of the three containers to particles made in the same container type after exposure to shaking and freeze-thaw stresses. Similarly, the effect of container on particle populations was investigated by comparing particles made in Glass 1 containers after exposure to each stress to samples of those made in each of the three containers after exposure to the same stress.

2.13 | Surface characterization

The two glass vial types used in this analysis were characterized using contact angles and surface profilometry. Each of these measurements was performed by cutting off the bottom of the vial and cleaning the inner surface with ethanol, water, and nitrogen before measurements.

To assess the hydrophobicity of the glass containers, contact angles were measured for each vial using a ramé-hart Model 210 goniometer/tensometer with DROPimage Pro software (Succasunna, NJ). This instrument was used to measure static, advancing, and receding contact angles on each of the three surfaces. These measurements were performed in triplicate. Between measurements, the surfaces were cleaned with ethanol, water, and nitrogen gas.

Surface profilometry was performed with a Dektak 3030 Profilometer (Billerica, MA) to measure the roughness of the two vials. Surface profiles were measured along a flat 1 mm length along the inner surface of the vial. These profiles were fitted to a second-degree polynomial which was then subtracted from the raw data to account for the macroscopic curvature of these surfaces. The flattened surface profiles were then used to calculate the arithmetic average roughness R_a of each container which is calculated using:

$$R_a = \frac{1}{n}\sum_{i=1}^{n}|h_i|, \tag{3}$$

where n is the total number of locations along the 1 mm length that the height was measured, i indexes the different height measurements, and h_i is the height measured at point i.

3 | RESULTS

3.1 | FIM

Figure 3 shows collections of randomly selected FIM images obtained from each of the six conditions compared in this analysis.

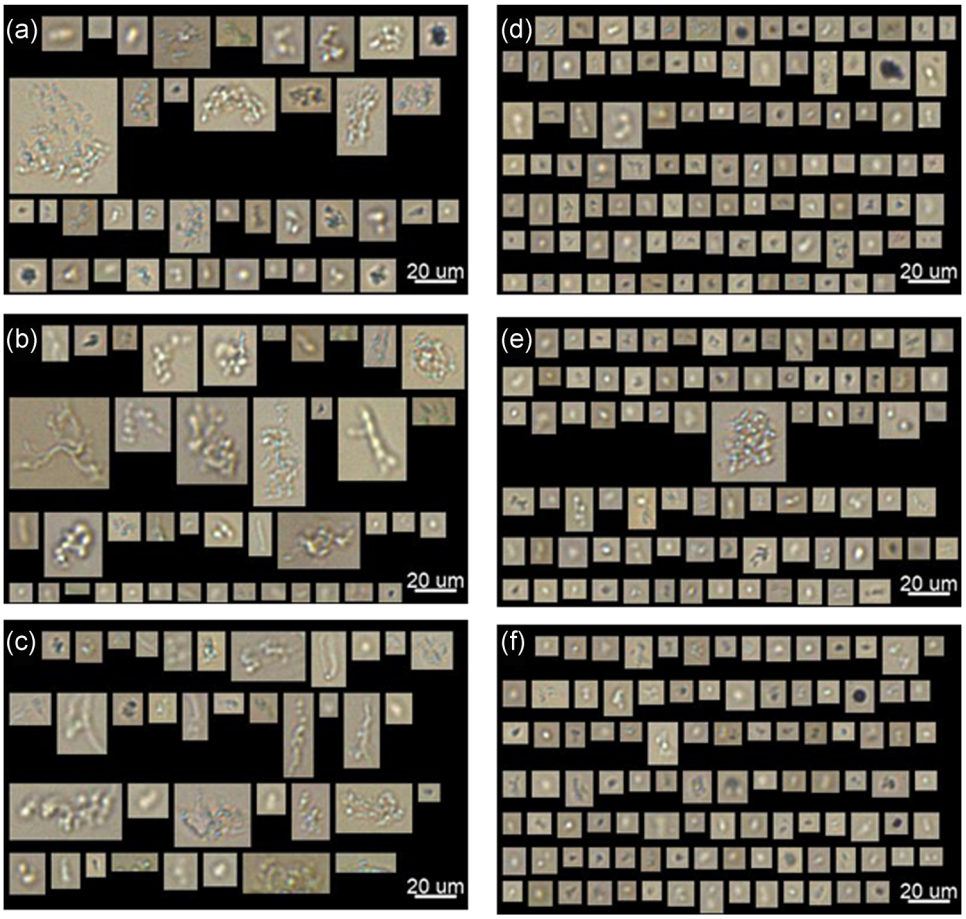

FIGURE 3 Sample collages of intravenous immunoglobulin aggregates generated by (a) shaking in plastic microcentrifuge tubes, (b) freeze-thawing in plastic microcentrifuge tubes, (c) shaking in Glass 1 vials, (d) freeze-thawing in Glass 1 vials, (e) shaking in Glass 2 vials, (f) freeze-thawing in Glass 2 vials [Color figure can be viewed at wileyonlinelibrary.com]

These images reveal obvious differences between particles generated by freeze-thawing and shaking stresses; particles observed after shaking (Figure 3a–c) are typically large and exhibit complex morphologies while particles imaged after freeze-thaw cycling (Figure 3d–f) are much smaller with simple morphologies. Conversely, the effect of different containers on particle morphologies generated by these stresses is not visually obvious from the images.

3.2 | ConvNets

Figure 4 shows contour plots of the distribution of embeddings returned by the trained ConvNet for particles made by freeze-thawing and shaking stresses. The contours for the different stresses are visually separated within this embedding space, indicating that the network can distinguish between particles generated by shaking and freeze-thaw stresses. Figure 4 also shows sample particles that are mapped to different locations in the embedding space. In this embedding scheme, small particles with simple but common structures are mapped near the mode of the freeze-thaw distribution whereas large, complex heterogeneous particles are mapped near the mode of the shaking distribution.

Figure 5shows contour plots of the estimated PDF of embeddings returned by the trained ConvNet for all particles not included in the network—including those from the four samples per condition that were not used to train the network. Figure 5a shows the embeddings for particles formed in Plastic containers, Figure 5b shows those formed in Glass 1 containers, and Figure 5c shows those formed in Glass 2 containers. These contour plots indicate an observable difference in the particle morphologies produced in the three containers as a result of shaking stresses. Compared with particles produced by shaking Glass 2 or Plastic containers, particles produced by shaking samples in Glass 1 containers (Figure 5b) have a much tighter density in the embedding space than either of the other samples. Particles produced by freeze-thawing stress appear to be influenced to a lesser extent by the container in which they were formed; particles produced freeze-thaw cycling in Glass 1 containers exhibit a slightly more diffuse distribution than those produced in the other container types (Figure 5b).

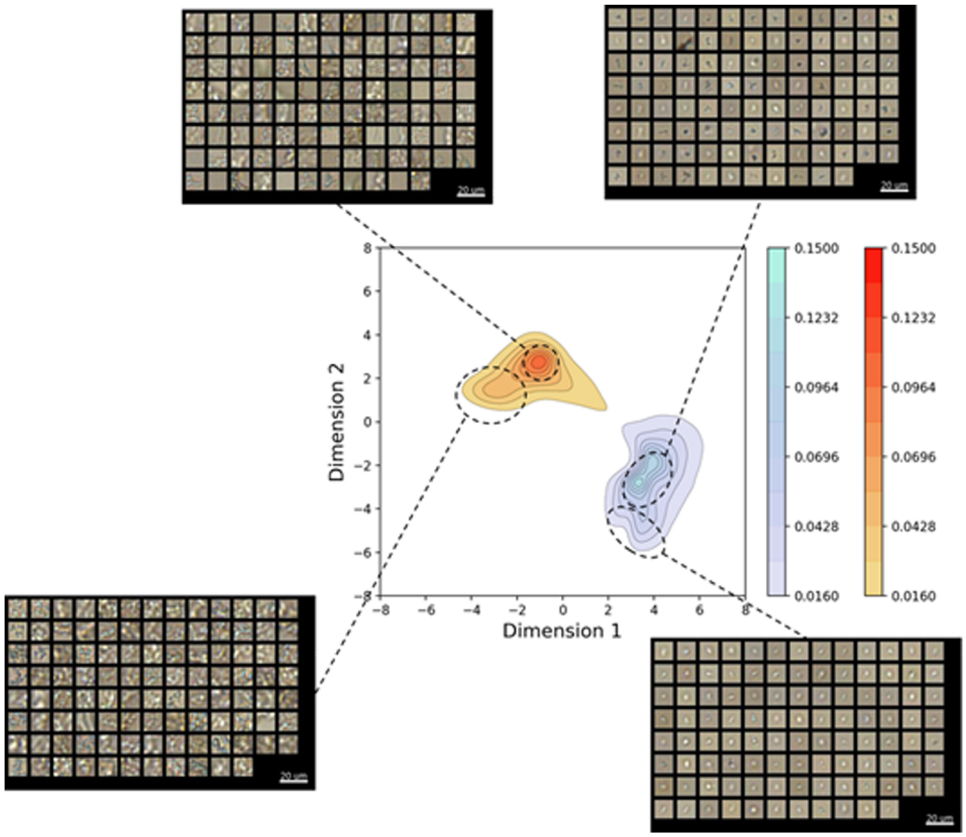

FIGURE 4 Contour plot of the FIM image embeddings of particles used to train the ConvNet. The colors of the plot indicate the density (i.e., PDF value) of image embeddings from a specific stress in that region of the embedding space. The darker orange-to-red contour corresponds to particles formed using shaking stress while the lighter blue-to-purple contour corresponds to particles formed using freeze-thaw stress. Regions of a single color have PDF values between the values indicated on the color bar. The values of the color transitions were manually selected to aid in viewing these PDFs. Also shown on each figure are typical FIM images that embed within specific regions of the contour plot. ConvNet, convolutional neural network; FIM, flow imaging microscopy; PDF, probability density function [Color figure can be viewed at wileyonlinelibrary.com]

3.3 | Particle comparisons

The remaining steps of the algorithm were used to compare the particle populations produced within different containers when exposed to different stresses. This comparison was done by choosing one of the samples being compared to be the baseline sample, subsampling small sets of particles from the other (test) sample, and testing the null hypothesis that each subsample contained particles that were consistent with those in the baseline sample. Table 2 shows the rejection frequencies when comparing sets of 20 particles from each of the three containers to those made by applying freeze-thaw stress to the same container. As was expected from both the raw flow imaging data shown in Figure 3 and the embeddings in Figure 5, the algorithm can easily identify morphology differences between particle populations that had been exposed to these two stresses using only a small number of particle images. Additionally, the algorithm only misidentified unseen test particles made under baseline conditions as being different from the baseline population around 5% of the time—the type I error rate that the test was designed to give.

Table 3 shows the rejection frequencies when comparing sets of either 20 or 200 test particles made by each stress to those produced in Glass 1 containers when exposed to the same stress. Interestingly, the ability of the algorithm to distinguish between particles produced in each of the three containers depended on the applied stress. Sets of 20 particles produced by exposing IVIG solutions in Plastic or Glass 2 containers to freeze-thawing stress were only able to be distinguished from those produced in Glass 1 containers at approximately the same rate as the type I error rate. In contrast, sets of 20 particles produced by shaking stress in each container were distinguishable from those made in Glass 1 containers approximately 40% of the time—eight times the type I error rate of the test. Increasing the size of the particle sets to 200 increased the fraction of shaking particle sets from Glass 2 and Plastic containers that were distinguished from those produced in Glass 1 containers with only a small increase in the false positive rate when the test was applied to held-out baseline samples. In addition, the larger particle sets allowed the algorithm to distinguish between particles made by freeze-thawing stress in Glass 1 and Glass 2 ~40% of the time and those produced in Glass 1 and Plastic 20% of the time.

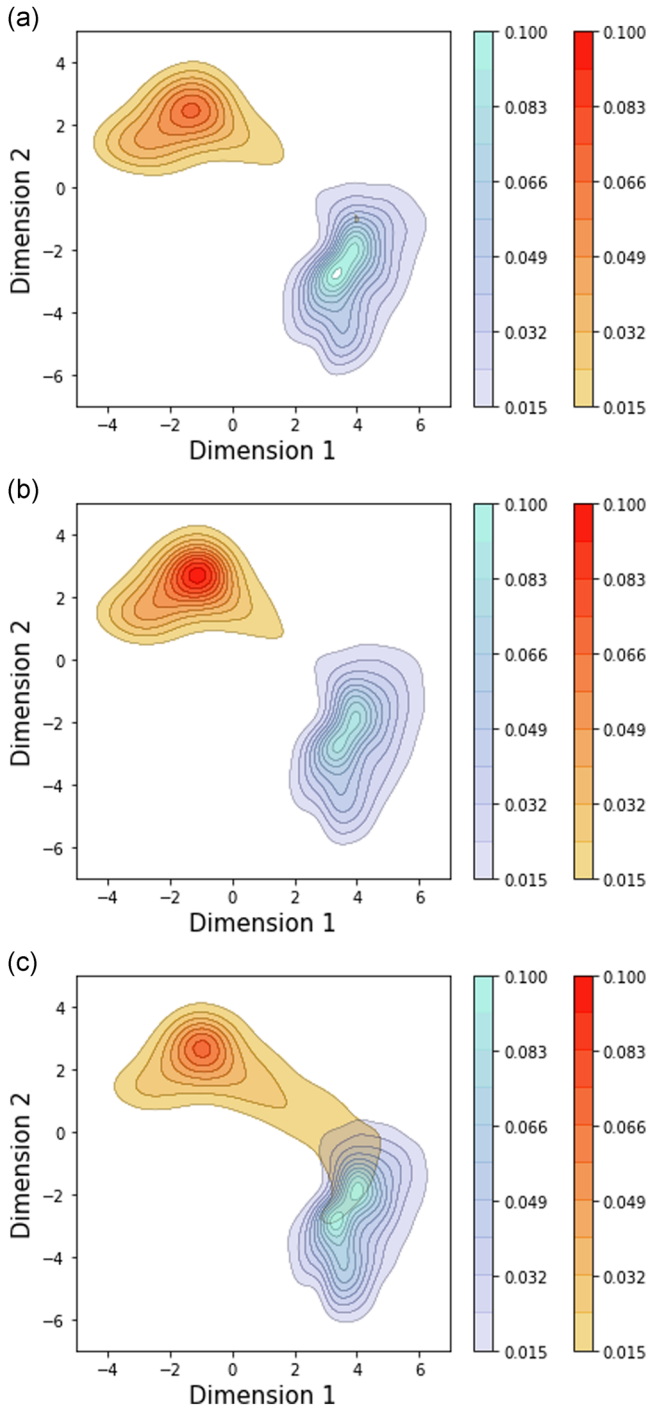

FIGURE 5 Contour plots of the flow imaging microscopy image embeddings of particles not shown to train the algorithm made in (a) Plastic microcentrifuge tubes (b) Glass 1 vials and (c) Glass 2 vials. The darker orange-to-red contours in each plot correspond to particles made using shaking stress while the lighter blue-to-purple contours in each correspond to particles made using freeze-thaw stress. Colors in this figure are interpreted as described for Figure 4 [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Probability that a set of 20 images of particles produced by freeze-thaw or by shaking stress in a container will be distinguishable from a baseline population of particles made by freeze-thaw stress in the same container

Baseline sample: particles made by freeze-thaw stress in denoted container	Probability of rejecting test particle sets (20 particles)	
	Aggregates made by freeze-thaw stress	Aggregates made by shaking stress
Glass 1	5.0%	100.0%
Glass 2	5.9%	100.0%
Plastic	5.0%	100.0%

3.4 | Surface characterization

Table 4 shows the surface characterization results for the two types of glass vials. The contact angle measurements suggest that the two glasses have similar hydrophobicities, with Glass 1 being slightly more hydrophilic than Glass 2. Both surfaces were also found to have similar roughnesses.

4 | DISCUSSION

This study presents and demonstrates a novel algorithm designed to compare FIM images of protein aggregates and other particles obtained from one sample to those obtained in some baseline sample. This approach is a departure from previous techniques used to predict to which of a small set of conditions a sample was exposed (Calderon et al., 2018; Gambe-Gilbuena et al., 2020). The primary advantage of this new approach is its ability to determine, *using only a small number of FIM images*, if a new sample exhibits significantly different particle populations than those found under baseline conditions. The combination of traditional statistical tools with powerful machine learning algorithms can be used to determine if two samples exhibit a morphology difference that cannot be explained by sample-to-sample variance in particle morphology under a single root cause. This approach is effective at identifying (statistically) significant differences in particle morphology occurring due to different root causes such as manufacturing changes or process upsets that could warrant further investigation.

The use of statistical tools in this algorithm also give users control over the sensitivity of the analysis to changes in particle morphology. Decreasing the type I error rate or increasing the number of particles used in the hypothesis test increases the sensitivity of the test so that smaller deviations in particle morphology from the baseline condition are identified as significant. This feature allows the sensitivity of the algorithm to be tuned for a specific

TABLE 3 Probability that a set 20 or 200 random particles formed in Glass 1, Glass 2, or Plastic containers by freeze-thaw or shaking stress will be distinguishable from a baseline population of particles made in the Glass 1 containers by the respective stress

Baseline sample: particles made in Glass 1 containers by denoted stress	Number of particles	Probability of rejecting test particle sets		
		Particles formed in Glass 1 containers (%)	Particles formed in Glass 2 containers (%)	Particles formed in plastic containers (%)
Freeze-thaw	20	5.0	6.8	5.1
	200	6.6	40.6	21.5
Shaking	20	5.1	36.0	13.4
	200	4.0	100.0	70.6

TABLE 4 Contact angle and surface roughness measurements for Glass 1 and Glass 2 containers

Container	Static contact angle	Advancing contact angle	Receding contact angle	Contact angle hys-teresis	Ra (nm)
Glass 1	16	23	<5	<18	28
Glass 2	27	31	<5	<26	12

application. For instance, the sensitivity of the algorithm can be increased for formulation development to better detect subtle changes in particle morphology between possible formulations. In contrast, in process monitoring applications the sensitivity of the algorithm can be decreased to minimize the chances that a false positive difference between a baseline "normal" reference batch and a new batch of product triggers unnecessary process shutdowns.

The techniques used to learn low-dimensional FIM image representations, calculate density estimates, and perform goodness-of-fit hypothesis testing were chosen to demonstrate the algorithm using relatively simple techniques. While the techniques used here were effective in this analysis, in practice other techniques for these analyses could be considered to further improve the performance of the algorithm. For example, different goodness-of-fit hypothesis tests (Anderson & Darling, 1954; Hong & Li, 2005; Justel, Peña, & Zamar, 1997) may provide better statistical power against deviations in particle morphology than the test used here.

The performance of this algorithm was demonstrated on particles made by subjecting IVIg solutions in three types of containers to two different stresses. Freeze-thawing and shaking stresses produced particle populations that would be easy to distinguish by visual inspection of the FIM images (Figure 3). These stresses produced visually resolved embeddings in the learned embedding space (Figures 4 and 5) which can then be easily distinguished using hypothesis testing (Table 2). It should be noted that this approach can still be used to classify samples by the stresses that they were exposed to as was done in previous papers (Calderon et al., 2018; Gambe-Gilbuena et al., 2020). If FIM images of different suspected stresses are available, this algorithm can be used to check if the particles in a sample are consistent with those produced by one of these stresses.

The algorithm was also able to identify the impact of container surfaces on particle populations. The results are shown in Table 3 suggest that the particle populations produced by freeze-thaw and shaking stress are influenced by the container, but that the effect is more obvious when shaking stress is used to create particles. The larger impact of the container on the particles produced by shaking stress agrees with the mechanistic understanding of these stresses. Agitation-induced aggregation likely occurs at interfaces including the air-water and container-water interfaces (Gerhardt et al., 2014; Teska et al., 2016). In contrast, during freeze-thawing aggregation due to adsorption to ice-water interfaces and cryoconcentration effects (Bhatnagar, Bogner, & Pikal, 2007) may occur at locations removed from container interfaces. Thus, the container interfaces might be expected to impact particle populations more when shaking stresses rather than freeze-thawing stresses are used to cause aggregation.

While the effect of container type was more subtle for particles made via freeze-thaw stress, the distribution of FIM image embeddings showed increased density near the mode of the distribution for Glass 2 vials as compared to that for Glass 1 vials (Figure 5). This difference was statistically discernible when sets of 200 particle images were analyzed. This result is somewhat surprising given the expected limited role of the container-water interface on aggregation induced by freeze-thawing. One possible explanation for the different particle fingerprints observed following freeze-thaw cycling in the two types of glass vials is differences in heat transfer through the vial walls. Differences in the thickness or geometry of the glass between the two vial types could cause a difference in the heat transfer rate through the container walls. Higher heat transfer rates would accelerate the growth of ice crystals from the walls of the container that occurs during liquid nitrogen-induced freezing (Searles, Carpenter, & Randolph, 2001). This faster growth results in a larger amount of ice interfacial area (Sarciaux, Mansour, Hageman, & Nail, 1999) and increased protein inclusion within growing ice crystals (Dong, Hubel, Bischof, & Aksan, 2009; Twomey et al., 2013) which can induce protein unfolding and aggregation (Strambini & Gabellieri, 1996; Strambini & Gonnelli, 2007).

The algorithm was not explicitly trained to detect the observed differences between particle populations produced in Glass 1 and 2 containers as particles generated in Glass 2 were not used to train the ConvNet embedding step. The ability to compare unseen particle types against those in a user-defined baseline allows new samples to be analyzed using a fraction of the FIM images (20–200) that would be required to retrain a ConvNet on a new particle type (>10,000).

The required number of FIM images can be recorded rapidly using small volumes of sample.

The algorithm revealed that different types of particles can form in a single protein formulation when stressed in different types of containers, even when the containers are as similar as the two borosilicate glass container types tested here. The container-dependent formation of different particles would have been difficult to predict using simple surface characterization techniques, since the glasses have similar roughness and hydrophobicity (Table 4).

The analysis presented here can be used to compare the effect of changes in container types (e.g., new lots of glass vials) on protein stability using an approach that incorporates standard accelerated stability protocols. Before any change, a baseline set of FIM images should be obtained after subjecting the protein formulation to accelerated stability conditions (e.g., agitation and freeze-thawing), capturing images of the resulting particles using FIM and training the algorithm to recognize the imaged particles. The accelerated stability protocol can then be repeated on a small number of containers in the new lot, and then the trained algorithm can be used to analyze whether new types of particles are associated with the new container lot. If the new container lot is found to produce statistically different particle populations, the lot may require additional characterization before use with the drug product.

ACKNOWLEDGMENTS

Funding for this study was provided by AstraZenica and Department of Education GAANN Award P200A150211. C. C. was funded by NIH Award R41GM130513. The authors (A. D., C. C., and T. R.) are inventors on intellectual property related to this manuscript that is owned by the Regents of the University of Colorado and Ursa Analytics. This study was sponsored in part through an STTR grant to Ursa Analytics, a company in which C. C. has a controlling financial interest.

ORCID

Austin L. Daniels http://orcid.org/0000-0001-8334-7496
Christopher P. Calderon https://orcid.org/0000-0001-8024-9081

REFERENCES

Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, *49*(268), 765–769. https://doi.org/10.1017/CBO9781107415324.004

Arakawa, T., Ejima, D., & Akuta, T. (2017). Protein aggregation under high concentration/density state during chromatographic and ultrafiltration processes. *International Journal of Biological Macromolecules*, *95*, 1153–1158. https://doi.org/10.1016/j.ijbiomac.2016.11.005

Arsiccio, A., & Pisano, R. (2017). Stability of proteins in carbohydrates and other additives during freezing: The human growth hormone as a case study. *Journal of Physical Chemistry B*, *121*, 8652–8660.

Barnard, J. G., Kahn, D., Cetlin, D., Randolph, T. W., & Carpenter, J. F. (2014). Investigations into the fouling mechanism of parvovirus filters during filtration of freeze-thawed mAb drug substance solutions. *Journal of Pharmaceutical Sciences*, *103*, 890–899. https://doi.org/10.1002/jps.23881

Barnard, J. G., Singh, S., Randolph, T. W., & Carpenter, J. F. (2011). Subvisible particle counting provides a sensitive method of detecting and quantifying aggregation of monoclonal antibody caused by freeze-thawing: Insights into the roles of particles in the protein aggregation pathway. *Journal of Pharmaceutical Sciences*, *100*(2), 492–503.

Bee, J. S., Randolph, T. W., Carpenter, J. F., Bishop, S. M., & Dimitrova, M. N. (2011). Effects of surfaces and leachables on the stability of biopharmaceuticals. *Journal of Pharmaceutical Sciences*, *100*(10), 4158–4170. https://doi.org/10.1002/jps.22597

Bhatnagar, B. S., Bogner, R. H., & Pikal, M. J. (2007). Protein stability during freezing: Separation of stresses and mechanisms of protein stabilization. *Pharmaceutical Development and Technology*, *12*(5), 505–523. https://doi.org/10.1080/10837450701481157

Calderon, C. P., Daniels, A. L., & Randolph, T. W. (2018). Deep convolutional neural network analysis of flow imaging microscopy data to classify subvisible particles in protein formulations. *Journal of Pharmaceutical Sciences*, *107*, 999–1008. https://doi.org/10.1016/j.xphs.2017.12.008

Chi, E. Y. (2004). *Protein aggregation in aqueous solution—Mechanism, thermodynamics, and kinetics* (Dissertations & Theses), University of Colorado at Boulder, ProQuest Dissertations Publishing.

Chisholm, C. F., Soucie, K. R., Song, J. S., Strauch, P., Torres, R. M., Carpenter, J. F., … Randolph, T. W. (2017). Immunogenicity of structurally perturbed hen egg lysozyme adsorbed to silicone oil microdroplets in wild-type and transgenic mouse models. *Journal of Pharmaceutical Sciences*, *106*(6), 1519–1527.

Cordes, A. A., Carpenter, J. F., & Randolph, T. W. (2012). Accelerated stability studies of abatacept formulations: Comparison of freeze–thawing- and agitation-induced stresses. *Journal of Pharmaceutical Sciences*, *101*(7), 2307–2315.

Darling, D. A. (1957). The Kolmogorov-Smirnov, Cramer-von Mises tests. *The Annals of Mathematical Statistics*, *28*(4), 823–838.

Dong, J., Hubel, A., Bischof, J. C., & Aksan, A. (2009). Freezing-induced phase separation and spatial microheterogeneity in protein solutions. *Journal of Physical Chemistry B*, *113*(30), 10081–10087. https://doi.org/10.1021/jp809710d

Dugas, C., Bengio, Y., Bélisle, F., Nadeau, C., & Garcia, R. (2001). Incorporating second-order functional knowledge for better option pricing. In *Advances in neural information processing systems*. Denver, CO, USA.

Ennis, R. D., Pritchard, R., Nakamura, C., Coulon, M., Yang, T., Visor, G. C., & Lee, W. A. (2001). Glass vials for small volume parenterals: Influence of drug and manufacturing processes on glass delamination. *Pharmaceutical Development and Technology*, *6*(3), 393–405. https://doi.org/10.1081/PDT-100002248

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056

Fradkin, A. H., Carpenter, J. F., & Randolph, T. W. (2009). Immunogenicity of aggregates of recombinant human growth hormone in mouse models. *Journal of Pharmaceutical Sciences*, *98*(9), 3247–3264. https://doi.org/10.1002/jps.21834

Freitag, A. J., Shomali, M., Michalakis, S., Biel, M., Siedler, M., Kaymakcalan, Z., … Engert, J. (2015). Investigation of the immunogenicity of different types of aggregates of a murine monoclonal antibody in mice. *Pharmaceutical Research*, *32*(2), 430–444. https://doi.org/10.1007/s11095-014-1472-6

Gambe-Gilbuena, A., Shibano, Y., Krayukhina, E., Torisu, T., & Uchiyama, S. (2020). Automatic identification of the stress sources of protein aggregates using flow imaging microscopy images. *Journal of Pharmaceutical Sciences*, *109*(1), 614–623. https://doi.org/10.1016/j.xphs.2019.10.034

Gerhardt, A., McGraw, N. R., Schwartz, D. K., Bee, J. S., Carpenter, J. F., & Randolph, T. W. (2014). Protein aggregation and particle formation in prefilled glass syringes. *Journal of Pharmaceutical Sciences*, *103*(6), 1601–1612. https://doi.org/10.1002/jps.23973

Ha, E., Wang, W., & Wang, Y. J. (2002). Peroxide formation in polysorbate 80 and protein stability. *Journal of Pharmaceutical Sciences*, *91*, 2252–2264.

Hermans, A., Beyer, L., & Leibe, B. (2017). In defense of the triplet loss for person re-identification. Retrieved from: http://arxiv.org/abs/1703.07737

Hong, Y., & Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies*, *18*(1), 37–84. https://doi.org/10.1093/rfs/hhh006

Jiskoot, W., Kijanka, G., Randolph, T. W., Carpenter, J. F., Koulov, A. V., Mahler, H. C., … Narhi, L. O. (2016). Mouse models for assessing protein immunogenicity: Lessons and challenges. *Journal of Pharmaceutical Sciences*, *105*(5), 1567–1575. https://doi.org/10.1016/j.xphs.2016.02.031

Joubert, M. K., Luo, Q., Nashed-Samuel, Y., Wypych, J., & Narhi, L. O. (2011). Classification and characterization of therapeutic antibody aggregates. *Journal of Biological Chemistry*, *286*(28), 25118–25133. https://doi.org/10.1074/jbc.M110.160457

Justel, A., Peña, D., & Zamar, R. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics & Probability Letters*, *35*(3), 251–259. https://doi.org/10.1016/S0167-7152(97)00020-5

Kerwin, B. A. (2008). Polysorbates 20 and 80 used in the formulation of protein biotherapeutics: Structure and degradation pathways. *Journal of Pharmaceutical Sciences*, *97*(8), 2924–2935.

Kiese, S., PappenBerger, A., Friess, W., & Mahler, H. -C. (2008). Shaken, not stirred: Mechanical stress testing of an IgG1 antibody. *Journal of Pharmaceutical Sciences*, *97*(10), 4347–4366. https://doi.org/10.1002/jps

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. In *3rd international conference on learning representations, ICLR 2015—Conference track proceedings*, 1–15.

Kolhe, P., Amend, E., & Singh, S. K. (2010). Impact of freezing on pH of buffered solutions and consequences for monoclonal antibody aggregation. *Biotechnology Progress*, *26*(3), 727–733. https://doi.org/10.1002/btpr.377

Krayukhina, E., Tsumoto, K., Uchiyama, S., & Fukui, K. (2015). Effects of syringe material and silicone oil lubrication on the stability of pharmaceutical proteins. *Journal of Pharmaceutical Sciences*, *104*(2), 527–535. https://doi.org/10.1002/jps.24184

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 1–9. https://doi.org/10.1016/j.protcy.2014.09.007

Kueltzo, L. A., Wang, W., Randolph, T. W., & Carpenter, J. F. (2008). Effects of solution conditions, processing parameters, and container materials on aggregation of a monoclonal antibody during freeze–thawing. *Journal of Pharmaceutical Sciences*, *9*(5), 1801–1812. https://doi.org/10.1002/jps

Kumru, O. S., Liu, J., Ji, J. A., Cheng, W., Wang, Y. J., Wang, T., … Volkin, D. B. (2012). Compatibility, physical stability, and characterization of an IgG4 monoclonal antibody after dilution into different intravenous administration bags. *Journal of Pharmaceutical Sciences*, *101*(10), 3636–3650. https://doi.org/10.1002/jps

Liu, L., Randolph, T. W., & Carpenter, J. F. (2012). Particles shed from syringe filters and their effects on agitation-induced protein aggregation. *Journal of Pharmaceutical Sciences*, *101*, 2952–2959.

Ludwig, D. B., Carpenter, J. F., Hamel, J.-B., & Randolph, T. W. (2010). Protein adsorption and excipient effects on kinetic stability of silicone oil emulsions. *Journal of Pharmaceutical Sciences*, *99*, 1721–1733.

Narhi, L. O., Corvari, V., Ripple, D. C., Afonina, N., Cecchini, I., Defelippis, M. R., … Wuchner, K. (2015). Subvisible (2–100 μm) particle analysis during biotherapeutic drug product development: Part 1, considerations and strategy. *Journal of Pharmaceutical Sciences*, *104*(6), 1899–1908. https://doi.org/10.1002/jps.24437

Nayak, A., Colandene, J., Bradford, V., & Perkins, M. (2011). Characterization of subvisible particle formation during the filling pump operation of a monoclonal antibody solution. *Journal of Pharmaceutical Sciences*, *100*, 4198–4204.

Randolph, T. W., & Carpenter, J. F. (2007). Engineering challenges of protein formulations. *AIChE Journal*, *53*(8), 215–228.

Roberts, C. J. (2007). Non-native protein aggregation kinetics. *Biotechnology and Bioengineering*, *98*(5), 927–938. https://doi.org/10.1002/bit.21627

Roberts, C. J. (2014). Protein aggregation and its impact on product quality. *Current Opinion in Biotechnology*, *30*, 211–217. https://doi.org/10.1016/j.copbio.2014.08.001

Rosenberg, A. S. (2006). Effects of protein aggregates: An immunologic perspective. *The AAPS Journal*, *8*(3), E501–E507. https://doi.org/10.1208/aapsj080359

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, *23*(3), 470–472. https://doi.org/10.1214/aoms/1177729394

Rosenblatt, M. (1961). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, *32*(1), 12–40. https://doi.org/10.1214/aoms/1177705148

Saller, V., Hediger, C., Matilainen, J., Grauschopf, U., Bechtold-Peters, K., Mahler, H. C., & Friess, W. (2016). Influence of particle shedding from silicone tubing on antibody stability. *Journal of Pharmacy and Pharmacology*, *70*(5), 675–685. https://doi.org/10.1111/jphp.12603

Sarciaux, J. M., Mansour, S., Hageman, M. J., & Nail, S. L. (1999). Effects of buffer composition and processing conditions on aggregation of bovine IgG during freeze-drying. *Journal of Pharmaceutical Sciences*, *88*(12), 1354–1361. https://doi.org/10.1021/js980383n

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Computer Society Conference on Computer Vision and* Pattern Recognition, 815–823. https://doi.org/10.1109/CVPR.2015.7298682

Scott, D. W. (2015). Multivariate density estimation *theory, practice, and visualization*. Hoboken, NJ: Wiley.

Searles, J. A., Carpenter, J. F., & Randolph, T. W. (2001). The ice nucleation temperature determines the primary drying rate of lyophilization for samples frozen on a temperature-controlled shelf. *Journal of Pharmaceutical Sciences*, *90*(7), 860–871. https://doi.org/10.1002/jps.1039

Sethuraman, A., Morcone, T., & Belfort, G. (2004). Protein aggregation at interfaces. *Protein Science*, *13*, 223.

Sharma, A., Anderson, S., & Rathore, A. S. (2008). Filter clogging issues in sterile filtration. *Biopharm International*, *2008*(21), 53–57.

Sharma, D. K., King, D., Oma, P., & Merchant, C. (2010). Micro-flow imaging: Flow microscopy applied to sub-visible particulate analysis in protein formulations. *The AAPS Journal*, *12*(3), 455–464. https://doi.org/10.1208/s12248-010-9205-1

Sharma, D. K., Oma, P., Pollo, M. J., & Sukumar, M. (2010). Quantification and characterization of subvisible proteinaceous particles in Opalescent mAb formulations using micro-flow imaging. *Journal of Pharmaceutical Sciences*, *99*, 2628–2642.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *3rd international conference on learning representations,ICLR 2015—Conference track proceedings*, 1–14.

Siska, C., Harber, P., & Kerwin, B. A. (2020). Shocking data on parcel shipments of protein solutions. *Journal of Pharmaceutical Sciences*, *109*, 690–695.

Sluzky, V., Klibanov, A. M., & Langer, R. (1992). Mechanism of insulin aggregation and stabilization in agitated aqueous-solutions. *Biotechnology and Bioengineering*, *1992*(40), 895–903.

Strambini, G. B., & Gabellieri, E. (1996). Proteins in frozen solutions: Evidence of ice-induced partial unfolding. *Biophysical Journal*, *70*(2 I), 971–976. https://doi.org/10.1016/S0006-3495(96)79640-6

Strambini, G. B., & Gonnelli, M. (2007). Protein stability in ice. *Biophysical Journal*, *92*(6), 2131–2138. https://doi.org/10.1529/biophysj.106.099531

Sun, Y., Chen, Y., Wang, X., & Tang, X. (2014). Deep learning face representation by joint identification-verification. In *Advances in neural information processing systems*, 3.

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In *IEEE conference on computer vision and pattern recognition (CVPR)*. https://doi.org/10.1109/CVPR.2014.220

Teska, B. M., Brake, J. M., Tronto, G. S., & Carpenter, J. F. (2016). Aggregation and particle formation of therapeutic proteins in contact with a novel fluoropolymer surface versus siliconized surfaces: Effects of agitation in vials and in prefilled syringes. *Journal of Pharmaceutical Sciences*, *105*(7), 2053–2065. https://doi.org/10.1016/j.xphs.2016.04.015

Thirumangalathu, R., Krishnan, S., Brems, D. N., Randolph, T. W., & Carpenter, J. F. (2006). Effects of pH, temperature, and sucrose on benzyl alcohol-induced aggregation of recombinant human granulocyte colony stimulating factor. *Journal of Pharmaceutical Sciences*, *95*(7), 1480–1497.

Twomey, A., Less, R., Kurata, K., Takamatsu, H., & Aksan, A. (2013). In situ spectroscopic quantification of protein-ice interactions. *Journal of Physical Chemistry B*, *117*(26), 7889–7897. https://doi.org/10.1021/jp403267x

Tyagi, A. K., Randolph, T. W., Dong, A., Maloney, K. M., Hiitscherich, C. J., & Carpenter, J. F. (2009). IgG particle formation during filling pump omperation: A case study of heterogeneous nucleation on stainless steel nanoparticles. *Journal of Pharmaceutical Sciences*, *98*, 94–104. https://doi.org/10.1002/jps

Tzannis, S. T., Hrushesky, W. J. M., Wood, P. A., & Przybycien, T. M. (1996). Irreversible inactivation of interleukin 2 in a pump-based delivery environment. *Proceedings of the National Academy of Sciences*, *93*, 5460–5465.

Vlieland, N. D., Nejadnik, M. R., Gardarsdottir, H., Romeijn, S., Sediq, A. S., Bouvy, M. L., … Jiskoot, W. (2018). The impact of inadequate temperature storage conditions on aggregate and particle formation in drugs containing tumor necrosis factor-alpha inhibitors. *Pharmaceutical Research*, *35*(2), 1–11. https://doi.org/10.1007/s11095-017-2341-x

Wang, W. (1999). Instability, stabilization, and formulation of liquid protein pharmaceuticals. *International Journal of Pharmaceutics*, *185*, 129–188.

Wang, W., & Roberts, C. J. (2018). Protein aggregation—Mechanisms, detection, and control. *International Journal of Pharmaceutics*, *550*(1–2), 251–268. https://doi.org/10.1016/j.ijpharm.2018.08.043

Wasylaschuk, W. R., Harmon, P. A., Wagner, G., Harman, A. B., Templeton, A. C., Xu, H., & Reed, R. A. (2007). Evaluation of hydroperoxides in common pharmaceutical excipients. *Journal of Pharmaceutical Sciences*, *96*, 106–116.

Webb, S., Cleland, J. L., Carpenter, J. F., & Randolph, T. W. (2002). Protein aggregation at interfaces formed during lyophilization. In *Abstracts of papers of the American Chemical Society*, 224, 163.

Zölls, S., Weinbuch, D., Wiggenhorn, M., Winter, G., Friess, W., Jiskoot, W., & Hawe, A. (2013). Flow imaging microscopy for protein particle analysis—A comparative evaluation of four different analytical instruments. *The AAPS Journal*, *15*(4), 1200–1211. https://doi.org/10.1208/s12248-013-9522-2

How to cite this article: Daniels AL, Calderon CP, Randolph TW. Machine learning and statistical analyses for extracting and characterizing “fingerprints” of antibody aggregation at container interfaces from flow microscopy images. *Biotechnology and Bioengineering*. 2020;1–14. https://doi.org/10.1002/bit.27501