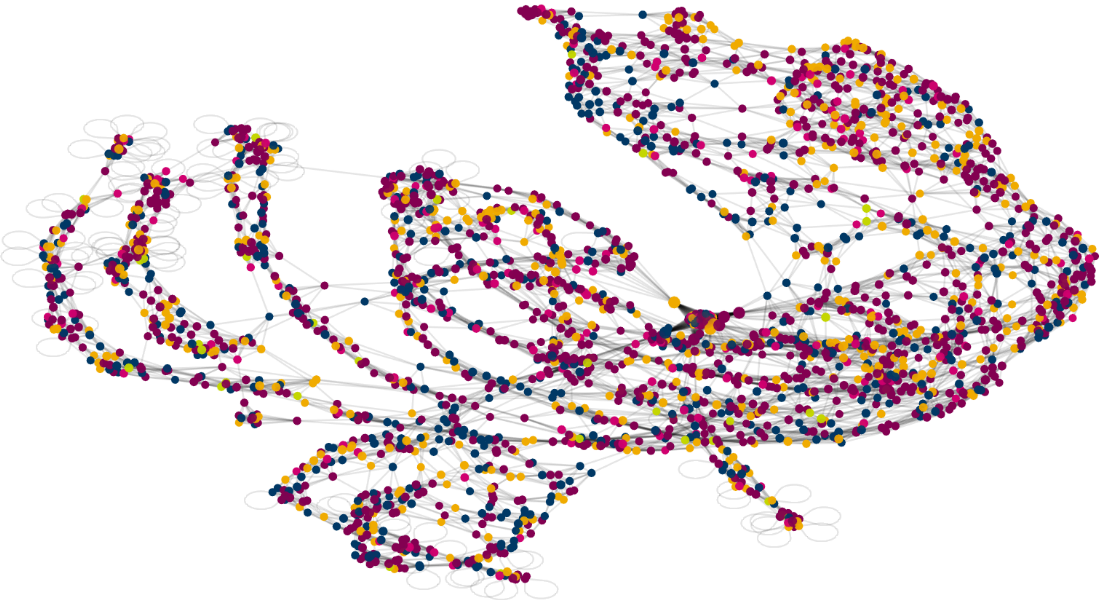


Advancing Sub-Visible Particle Characterization in Inhaled Biologics

Micro-Flow Imaging for Particle Analysis and Sub-Population Differentiation

Master's Thesis in Biomedical Engineering

SOPHIE YOU

DEPARTMENT OF PHYSICS

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Master's Thesis 2025

Advancing Sub-Visible Particle Characterization in Inhaled Biologics

Micro-Flow Imaging for Particle Analysis and Sub-Population Differentiation

SOPHIE YOU

Department of Physics
Division of Nano- and Biophysics
Chalmers University of Technology
Gothenburg, Sweden 2025

Pharmaceutical Technology & Development
Analytical Sciences, Global Product Development
AstraZeneca
Mölndal, Sweden 2025

Advancing Sub-Visible Particle Characterization in Inhaled Biologics
Micro-Flow Imaging for Particle Analysis and Sub-Population Differentiation
SOPHIE YOU

Supervisors: Noemi Gaglianone, AstraZeneca
Gerard Masdeu, AstraZeneca
Examiner: Jan Swenson, Department of Physics

Master's Thesis 2025
Department of Physics
Division of Nano- and Biophysics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Force-directed graph of MFI particle data, arranged using the Fruchterman-Reingold algorithm to reflect feature-based similarity.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

Advancing Sub-Visible Particle Characterization in Inhaled Biologics
Micro-Flow Imaging for Particle Analysis and Sub-Population Differentiation
SOPHIE YOU
Department of Physics
Chalmers University of Technology

Abstract

Protein-based therapeutics represent one of the fastest growing sectors in the pharmaceutical industry, offering highly specific and effective treatment options for a range of diseases. However, their structural complexity makes them particularly vulnerable to instability and aggregation, especially in formulations intended for inhalation. Protein aggregation not only compromises therapeutic efficacy but also poses a significant immunogenicity risk. Therefore, quantifying and characterizing sub-visible particles, including protein aggregates, is essential to understanding these risks and improving the overall safety and effectiveness of biologic drug products.

While pulmonary drug delivery offers a non-invasive and direct route for administering the active pharmaceutical ingredient (API), the powders must meet specific size and property requirements. Excipients are therefore used to enhance these characteristics and optimize the performance of the API. Conventional methods, such as Light Obscuration, are commonly used to quantify sub-visible particles, but often face challenges in detecting small and translucent particles.

This thesis explores how continuous flow microscopy, specifically micro-flow imaging (MFI), can be used to differentiate and characterize sub-visible particles. By analyzing morphological and light intensity data of particles detected by the MFI, this study aims to identify sub-populations within drug products and their accompanying excipients. Understanding the composition and distribution of these particles and how they contribute to the overall particulate environment in the formulations is in focus.

The results reveal that different formulations exhibit particles with distinct characteristics. Intensity distributions have been established, resulting in unique profiles that could potentially be used to identify the presence of specific formulations in unknown samples. A deeper understanding of the interactions and complexities among various substances in the formulations might be required to enhance the accuracy of the identification test.

Keywords: micro-flow imaging, sub-visible particle, sub-population, protein aggregation, inhaled biologics.

Acknowledgements

Firstly, I would like to thank AstraZeneca and the Analytical Sciences team for making this thesis possible and for giving me a once-in-a-lifetime opportunity. You welcomed me as part of the Analytical family and have continued to inspire me every single day. I couldn't have asked for a more rewarding experience. To Alisija, for being a great office roomie, Rodolfo, for the Spanish lessons, Aida, who seems to know where to find everything, and the rest of the team, thank you.

Many thanks to Noemi, who made me feel welcome from day one. Noemi, you truly have a wonderful team. Wonderful team, you truly have a great manager.

My gratitude also extends to those outside the Analytical team who have supported me - Magnus Fransson, Tatiana Starciuc, and Maria Lindkvist, thank you for taking the time out of your busy schedules to share your valuable knowledge and insights.

To my examiner and former supervisor Jan Swenson - thank you for opening my eyes up for biophysics. Had it not been for your great tutelage, I wouldn't be where I am today. My thanks also extend to Kajsa Ahlgren.

Virginia, thank you for brightening my days with endless laughter and chitchat, for all your support and help. Everyday at office has been a joy thanks to you. Without you, this journey wouldn't be the same. Crazy.

Special thanks to my dear friends Anni, John, Olivia and Agnes, to name a few, and to my family. Further thanks goes to Anni for making me believe in myself, and to John for being a true Latex expert and helping me, even though it may not have been voluntarily every time.

Lastly, Gerard, I am deeply grateful for your guidance and support. For answering all questions I have bothered you with every single day, for being able to make the most irrational things still make sense. For your valuable comments on my work, for all interesting lunch discussions, and for you being willing to listen to my gibberish. Moltes gràcies per tot!

Sophie You, Gothenburg, June 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

ADA	Anti-Drug Antibody
API	Active Pharmaceutical Ingredient
DP	Drug Product
DPI	Dry Powder Inhaler
DS	Drug Substance
ECD	Equivalent Circular Diameter
ETFE	Ethylene Tetrafluoroethylene
LO	Light Obscuration
MDI	Metered Dose Inhaler
MFI	Micro-Flow Imaging
PCA	Principal Component Analysis
PDD	Pulmonary Drug Delivery
SMI	Soft Mist Inhaler
SVP	Sub-Visible Particle
XRPD	X-Ray Powder Diffraction

Contents

List of Figures

List of Tables

1

Introduction

This chapter introduces the background of the project as well as presenting previous work that has been conducted on the subject. It also outlines the aims and limitations of the project.

1.1 Background

In recent years, the pharmaceutical industry has witnessed rapid growth in the development of protein-based drugs, owing to their ability to target disease mechanisms with high precision. Respiratory distress syndrome, prevention of thrombosis, treatments for diabetes, cancer and autoimmune diseases, are just some areas where protein-based pharmaceuticals are used today [1], [2]. The development and manufacturing process of such protein-based pharmaceuticals presents challenges, such as abnormal protein folding, stability and immunogenicity [3]. Like other biological molecules, proteins are susceptible to changes in their environment, where adverse effects like misfolding, unfolding and aggregation, pose a great risk to the efficacy of the drugs.

While pulmonary delivery of therapeutics offers a non-invasive alternative to other administration routes, it also presents unique challenges. Where in the respiratory tract the drug is being deposited is dependent on the particles' behavior in airflow, which in turn depends on properties like the size, shape and density of the particles [4]. In order for the drug, and consequently the protein, to reach the lungs, the drug must be in the form of very small liquid or solid particles, e.g., between 1-5 µm to reach the lower respiratory tract and between 1-3 µm to reach the alveolus tissue [5], [6]. Unlike parenteral formulations, inhaled protein drugs must remain stable in aerosolized form and withstand mechanical stress during nebulization or dry powder dispersion. Proteins in these formulations are particularly susceptible to aggregation, degradation and denaturation [4]. To achieve these characterstics, inert ingredients named excipients are added to the formulation to enhance the properties of the active pharmaceutical ingredient (API) [7].

Typically, proteins need to be folded in their native state to have their original characteristics and to function effectively. The folding of the protein is induced by forces and interactions between amino acids. The same forces and interactions also exist between neighboring amino acids in adjacent proteins [2]. This means that proteins have a tendency to aggregate, where the monomers form dimers, trimers, oligomers, etc. Reversible aggregation is in a small number of cases beneficial to the product, as the monomer may have lower stability and solubility than the oligomeric counterpart. However, aggregation is of major concern in most cases, of which some include a limited solubility and a decreased viscosity. Moreover, there is also the concern of how irreversible aggregates persisting in the patient's blood stream can lead to the patient developing an immune response to the drug. Mildly seen, the patient can form an immunity to the drug, but at worst case develop an autoimmune disease.

Given the potential risks associated with protein aggregation, particularly the formation of sub-visible particles (SVPs), understanding their characteristics is critical. While guidelines regarding quantification of SVPs exist today, such as the <787> and the <788> general chapters in the United States Pharmacopeia [8], [9], further sub-population differentiation is of interest. This differentiation is crucial for advancing the understanding of particle characteristics, origins, and their potential impacts on product safety and efficacy. Light obscuration (LO) techniques are standard for particle quantification, and have well established methods in pharmacopeial guidelines. These methods enable particle counting and therefore a way to control the levels of SVPs. However, the emerging micro-flow imaging (MFI) method has showed its potential in providing additional information on the particle itself. The instrument provides images of individual particles, offering morphological and light intensity data upon which sub-populations should be able to be distinguished. However, MFI is a relatively new technique, and particle differentiation is still evolving.

1.2 Previous Work

Method optimization of the MFI has been performed prior to the work presented in this thesis, including guidelines for sample volume, running of replicates and run set up. Initial validation of the method, including evaluation of linearity of measured particle concentration compared to its theoretical linear value, and recovery testing has also been performed. Preliminary work on the separation of sample populations was conducted by sorting and filtering morphological data using the vendor's software. This approach had some limitations, as the filters that can be applied themselves are limited, suggesting the need for an external software and coding in order to improve the separation.

1.3 Aim

This project aims to advance the development of MFI technology by building a method able to distinguish and differentiate between proteinaceous and non-proteinaceous populations within mixtures. Limitations of existing methods will be addressed by enhancing the precision of particulate matter identification. A specific objective throughout this process is to gain a comprehensive understanding of the data provided by the MFI system, assess its accuracy, and identify ways to address potential inaccuracies.

1.4 Limitations

While there are numerous ways to analyze particles in solutions, this project only aims to use MFI. LO was also planned to be used to compare the particle count between MFI and LO, but was downprioritized in order to address the main aim of the project. A limitation to this project is the optical resolution for the MFI instrument, which limits how the particles can be analyzed and how the data can be used.

Another limitation of this project is the constraints on the available materials and substances. This restriction applies to the spray-drying step - only materials that had been previously spray-dried were available for this work. Formulations and substances that already exist were utilized. Producing new spray-dried materials for this project was not feasible due to time constraints.

2

Theory

This chapter presents proteins and how they are used today in the pharmaceutical industry, protein structure as well as protein aggregation and its associated risks. The chapter also covers other topics regarding inhaled drugs. Quantification techniques for sub-visible particles are included, as well as methods used for differentiation of sub-populations.

2.1 Proteins

Proteins consist of amino acids joined by peptide bonds that together form polymers [10]. Proteins are fundamental components of various biological systems and take on a wide variety of roles, such as hormones, catalysis of metabolic processes as well as gene expression. Their ability to interact specifically with biological targets makes them valuable as pharmaceuticals in treating various diseases. This section describes protein structure and their pharmaceutical applications, as well as protein aggregation and associated immunogenicity risks.

2.1.1 Proteins in Pharmaceuticals

Therapeutic proteins have become a central focus in pharmaceutical research, and have emerged as compelling, and sometimes superior alternatives, to the traditional small-molecule based medicines. Due to their selective interaction with specific biological targets, proteins are less likely to cause side effects [11]. As versatile biomolecules, proteins perform various functions, including signaling, acting as receptors, catalysts, cellular and tissue scaffolds, and molecular transporters. They can also execute more complex functions as a result of their intricate, three-dimensional structures.

Some common classes of protein-based therapeutics include antibodies, enzymes, coagulation factors, protein hormones and cytokines [11]. Antibodies target specific antigens, blocking signaling pathways or inducing cell death, often in cancer therapies or immune system modulation. Enzymes are utilized in enzyme replacement therapy to substitute missing or deficient enzymes, as well as in targeted catalytic actions to treat specific conditions. Coagulation factors, vital for blood clotting, treat various bleeding-related disorders, like hemophilia. Protein hormones, such as insulin, are essential for regulating metabolic processes and are used to manage conditions like diabetes and other diseases. Cytokines function as immunomodulatory agents for a variety of diseases, such as multiple sclerosis and leukemia.

2.1.2 Protein Structure and Folding

Proteins can consist of few to several thousands of amino acids in a single chain [12], where the three-dimensional structure determines their function. This structure is divided into different hierarchical levels [10], see Figure 2.1. The primary structure of a protein is its linear sequence of amino acids. The secondary structure arises from local hydrogen bonding, where the most common structures include α-helices and β-sheets. These secondary structures further interact through e.g., hydrogen bonds, hydrophobic interactions and salt bridges, forming the tertiary structure - a three-dimensional conformation of a single protein molecule. Some proteins may also form a quaternary structure, involving the assembly and interaction of multiple protein subunits.

Proteins adopt their functional form when they fold into specific structures, referred to as their native structures [10]. Loss of this native structure typically results in a loss of function, as seen in diseases like Alzheimer's and cystic fibrosis that are associated with misfolded proteins [10], [12]. While the exact mechanisms of protein folding are not yet fully understood, the integrity of the folding process is crucial for protein function.

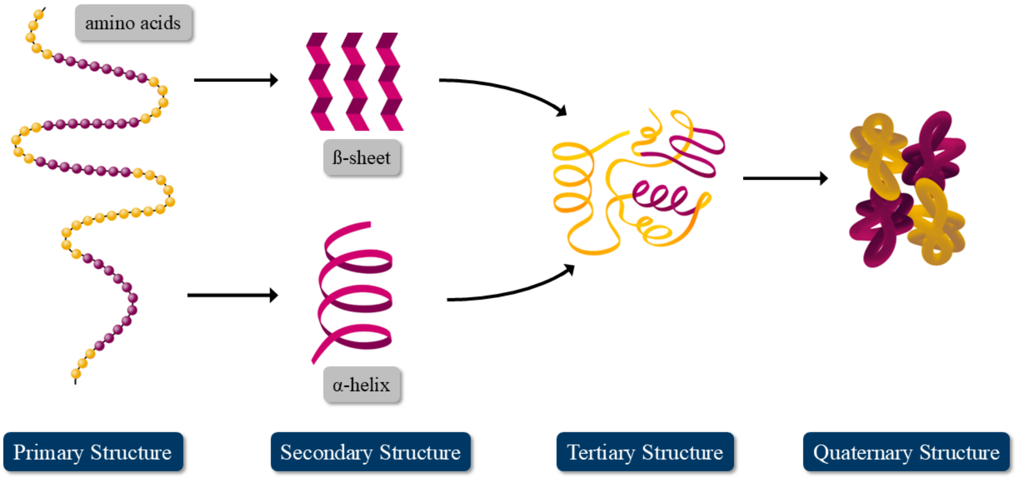


Figure 2.1: The different levels of protein structure, including primary, secondary, tertiary and quaternary structure.

The stability of a protein depends on the free energy difference between its folded native state and unfolded state. This energy difference indicates how thermodynamically favorable the folded state is and can be described using the Gibbs free energy equation

$$\Delta G = -RT \cdot \ln(K_{eq}) = \Delta H - T \cdot \Delta S, \tag{2.1}$$

where R is the gas constant, T the absolute temperature and K_{eq} the equilibrium constant between folded and unfolded states. ΔH and ΔS represent the changes in enthalpy and entropy, respectively [13]. A properly folded protein resides in its lowest energy state, often represented using the folding funnel hypothesis, illustrated in Figure 2.2 as an energy landscape. The folding funnel model illustrates how unfolded proteins follow pathways that lead to the global free energy minimum, which is the most stable state with the lowest energy. Along the way it can populate intermediate states, represented as local minima, that can either hinder and/or promote native structure formation depending on the funnel surface [14], [15].

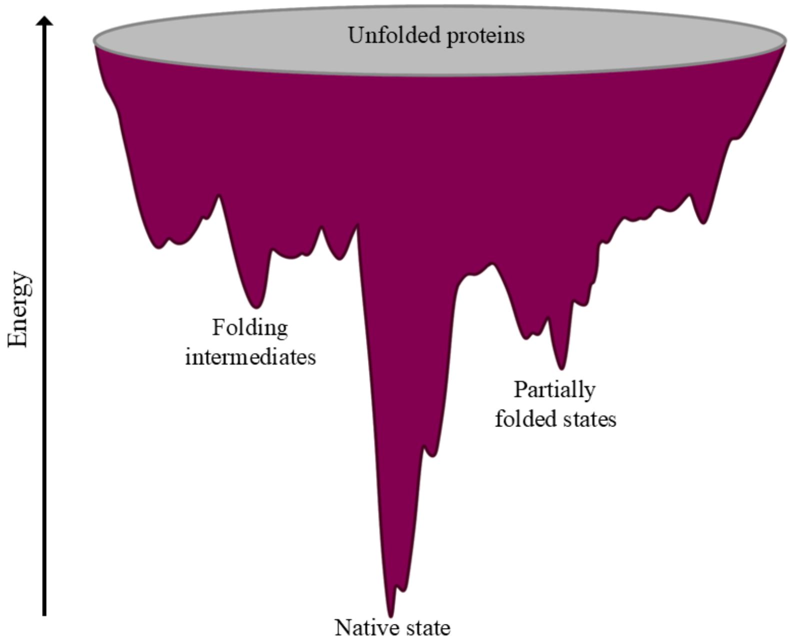


Figure 2.2: A simplified folding funnel depicting the energy landscape of protein folding. The further down, the lower the energy and the greater the stability of the state.

Furthermore, the protein can take on different conformations even within its native state, following several alternative folding trajectories [25]. Its functionality arises from their dynamic properties and conformational flexibility [26]. This flexibility is made possible by an aqueous environment - without it, the protein might not be able to function.

2.1.3 Protein Aggregation

Protein aggregation refers to the process where misfolded or partially folded proteins associate to form larger structures. These aggregates can vary in both structure and properties, typically forming either disordered amorphous aggregates or highly ordered amyloid fibrils [16]. Understanding how and why proteins aggregate is crucial in both pharmaceutical development and disease research, as aggregates can impact drug stability and trigger unwanted immune responses [17]. This section aims to further present the mechanics behind aggregation formation and their associated immunogenicity risks.

2.1.3.1 Aggregate Formation

Protein instability can generally be divided into two classes, chemical instability and physical instability [16]. Deamidation, oxidation, and disulfide exchange are examples of chemical processes that generate new chemical entities by making or breaking covalent bonds. Physical instabilities, however, maintain the chemical structure of the protein while altering the physical state instead. Denaturation, precipitation and adsorption are included in this. Several techniques exist today to address these issues, such as using sugars and polyols to improve protein stability [18], [19], [20], and surfactants to prevent adsorption [21].

Protein aggregation can occur when the forces between amino acids that drive folding within a single protein molecule also promote interactions between amino acids in neighboring proteins [2]. Therefore, at finite concentrations, proteins have a tendency to aggregate, whereas at infinite dilution, they would remain in the monomeric state. It is widely accepted that protein folding/unfolding intermediates facilitate the aggregation process [22]. Completely folded and unfolded proteins have their hydrophobic side chains buried within the protein or randomly scattered throughout. The intermediates however, expose more of these hydrophobic patches, where attraction between these patches enables aggregation. It is important to note, even in native proteins, that not all hydrophobic residues are buried in the protein [3].

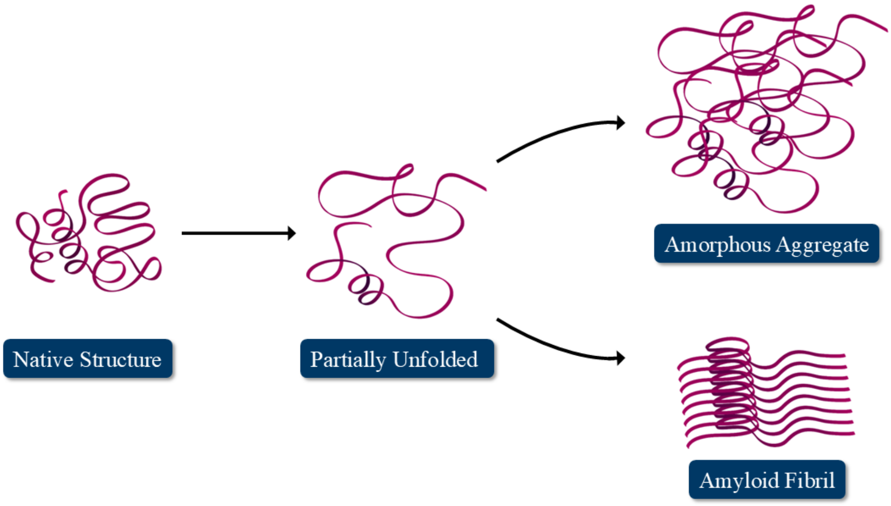


Figure 2.3: Schematic model of the aggregation process from native protein to different aggregated states of the protein.

There are different types of protein aggregates, such as amorphous aggregates or more structured, amyloid fibrils [2], [23]. Figure 2.3 shows a brief overview of the aggregation process. Aggregates are generally classified as either soluble or insoluble. Soluble aggregates usually fall in the size range of 1-100 nm and have a lower molecular mass, while insoluble aggregates can be both subvisible (0.1-100 µm), and visible (>100 µm) [17], [24]. Newly formed aggregates are often soluble, but can become more insoluble with time as they exceed solubility and size limits [22]. The degree to which proteins aggregate depends on various factors, broadly divided into two categories - intrinsic, depending on the protein's primary, secondary, tertiary and quaternary structure, and extrinsic, depending on the protein's environment [22]. The reversibility of protein aggregation depends on the stage of the aggregation process, where later stages are often less reversible.

While the folding funnel concept accounts for all potential folding conformations of a protein, it is limited in that it only represents the behavior of a single polypeptide chain under conditions of infinite dilution [15]. The model does not include interactions between multiple chains, and as such cannot fully describe misfolding events that would lead to self-association and aggregation. However, the concept can be extended to include aggregation by representing it as a second funnel, as illustrated in Figure 2.4. While the original funnel represents intramolecular contacts, the aggregation funnel represents intermolecular contacts instead.

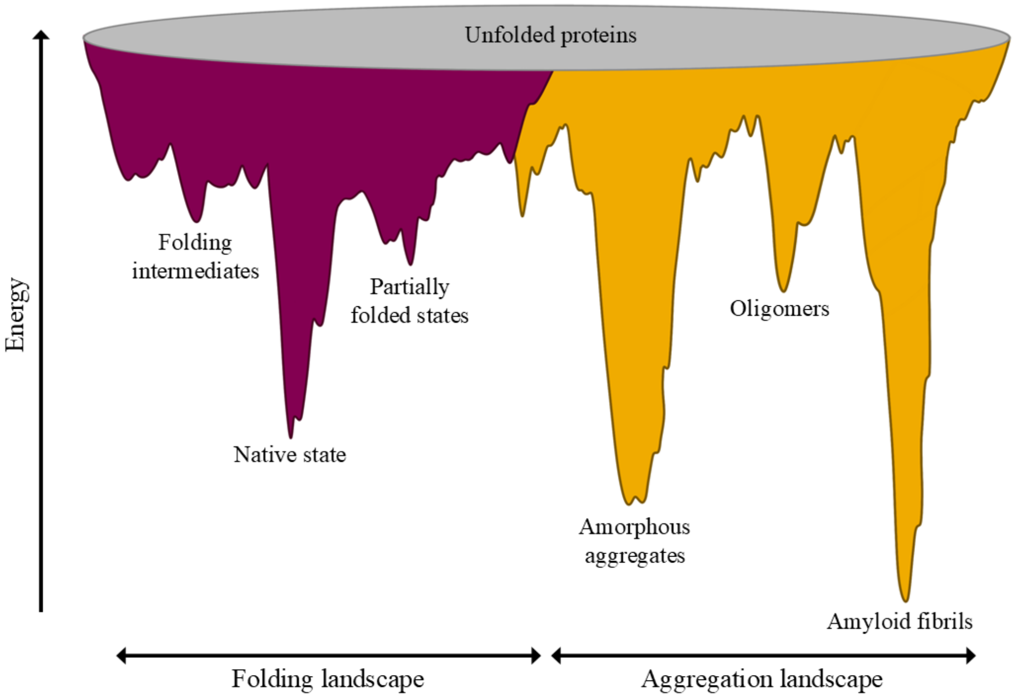


Figure 2.4: The folding funnel landscape scheme, together with the second aggregation funnel. The pink region to the left represents folding pathways that guide the protein toward its native structure, whereas the yellow region to the right illustrates pathways leading to the formation of amorphous aggregates or amyloid fibrils.

2.1.3.2 Immunogenicity

The formation of protein aggregates can be highly unfavorable, as the therapeutic effect of the drug can be negatively impacted. An adverse immune response may be triggered as the aggregates might be recognized as foreign or abnormal, resulting in the production of antibodies against the therapeutic protein, denoted anti-drug antibodies (ADA) [17], [24]. The binding of ADAs to therapeutic proteins can neutralize their activity, impede their delivery to target sites, or accelerate their clearance from the body. In more severe cases, ADA-induced immune responses can result in adverse events such as hypersensitivity, allergic reactions or immune-mediated diseases. Particles within the size range of 0.1-10 µm are identified as having significant immunogenic potential [27], although this potential can vary depending on the specific protein being analyzed [28], [29]. Aggregate size is therefore a critical factor influencing the likelihood of unwanted immune activation.

2.1.4 Proteins Investigated in This Work

Due to the propritery nature of the proteins explored in this work, specific details about their characteristics are kept confidential. Not disclosing information about their nature does not limit any discussions.

2.2 Inhaled Drug Technologies

Pulmonary delivery of drugs offers an alternative to parenteral delivery, allowing direct administration for treating lung diseases or systemic treatment for various other diseases. Continued development of formulations and inhalation devices has expanded its use to a wider range of treatments. In these products, the protein serves as the API, which is the primary substance responsible for eliciting the therapeutic effect. However, the manufacturing process must adress challenges such as protein stability and aggregation, which can compromise the drug effectiveness. To manage these risks, controlling SVP levels is essential. Quantification methodologies are implemented to monitor these SVPs, playing a key role in the safe delivery of the protein.

2.2.1 Deposition Mechanisms of Inhaled Drugs

The lungs enable a unique route for drug administration with their large surface area, high vascularization and highly permeable alveolar membrane [30], [31]. This route is advantageous over other administration routes as it provides rapid onset and targeted delivery while still being non-invasive and giving fewer side effects. The effectiveness of inhalation therapy depends not only on the drug's pharmacology but also on where and how much of it deposits in the respiratory tract. Aerosol deposition occurs through mechanisms such as inertial impaction, gravitational sedimentation and Brownian diffusion, with additional contributions from turbulence, electrostatic effects and interception [32]. Among the variables that influence the deposition pattern are particle size, lung anatomy and airflow dynamics. Larger particles, >5 µm, mainly settle in the upper and large airways, while particles between 2-5 µm tend to deposit in the central and small airways. Particles <2 µm reach the alveolus tissue [32], [33].

2.2.2 Pulmonary Drug Delivery Systems

Various delivery systems exist to effectively administer inhaled drugs to the human body. Pulmonary drug delivery (PDD) systems have been used for the treatment of diseases like asthma, respiratory tract infections, and chronic obstructive pulmonary diseases. PDD devices used today include soft mist inhalers (SMIs), metered dose inhalers (MDIs), dry powder inhalers (DPIs) and nebulizers [31]. While SMIs, MDIs and nebulizers deliver the API in a solution or suspension, DPIs contain the API in a solid state within a dry powder formulation, which is usually chemically more stable. Proper drug formulation and device selection are crucial for efficient lung delivery. Since drugs clear quickly via routes like phagocytosis by macrophages and dissolution, frequent dosing is needed, which can reduce patient compliance [30]. Prolonging particle retention in the lungs helps maintain therapeutic effects.

2.2.3 Spray Drying

Spray drying is a process in which a liquid solution or suspension is converted into a powdered solid [34], [35]. The liquid feed is pumped into a drying chamber through a nozzle, where it is dispersed into fine droplets that rapidly come into contact with a stream of hot drying gas. As the liquid evaporates, dry particles remain, which are then separated from the gas using a cyclone and/or a bag filter. The final powdered product is then collected in a vessel. The spray drying method is largely used in pharmaceutical formulations as it accurately controls parameters such as particle morphology and size distribution in the final product.

2.2.4 Excipients in Drug Formulation

For pulmonary delivery of powders, several particle characteristics must be met in order to ensure maximum efficiency and efficacy. The particle must be in an appropriate size range, which for delivery to the lungs means an aerodynamic diameter of 1-5 µm, while also having a low interparticle cohesion to ensure proper dispersion and deagglomeration [6]. The particles must also be physically and chemically stable enough to allow for long-term storage. Excipients, which are inert ingredients used in a drug formulation to enhance the properties of the API, are used to achieve these properties [36].

The choice and function of excipients are influenced by the dosage form and selection of a PDD device, as these factors impact formulation requirements. In MDIs, commonly used excipients include surfactants such as sorbitan trioleate, lecithin and oleic acid to aid in particle dispersion [7]. Stabilizers such as polyvinyl alcohol are used to enhance stability in biologics or water-soluble drugs. Taste-masking agents such as menthol may also be added to improve patient acceptability.

Nebulizer formulations include surfactants and preservatives to stabilize the drug and improve aerosol formation while buffers and salts adjust pH and osmolarity for lung comfort [7]. Some excipients like chitosan and cryoprotectants also help protect sensitive drugs and ensure consistent particle size.

DPIs commonly incorporate excipients that support powder flow, stability and aerosol performance [7], [37]. Fillers like lactose are frequently used as a carrier for low-dose drugs [7]. Amino acids such as L-leucine, glycine, L-alanine and trileucine are included to reduce particle adhesion and enhance dispersion by forming hydrophobic surfaces [7], [38]. They act as shell-forming excipients that reduce particle adhesion and improve dispersion by creating hydrophobic shells [31]. Additionally, sugars like mannitol and trehalose are used as alternatives to lactose and serve as stabilizers by forming a glassy matrix, but are often used in combination with other excipients due to their hygroscopic and cohesive nature [37], [39].

2.3 Sub-Visible Particle Quantification

SVP quantification is a critical aspect of ensuring the safety and efficacy of pharmaceutical products, with protein aggregates potentially comprising a portion of these particles. European and US regulatory compendial methods include the quantification of particles with an equivalent circular diameter (ECD) exceeding 10 µm and 25 µm [9], [40], [41]. The ECD is the diameter of a circle that has the same projected area as the particle. The techniques employed are LO and microscopic particle counting, capable of detecting particles within the 2-100 µm range, which defines SVPs [41]. SVPs refer to particles inherent to the solution, such as protein aggregates, but also comprises non-inherent particles that may have been introduced during manufacturing or handling, including silicon oil droplets, fibers, glass particles etc.

2.3.1 Light Obscuration

LO is a technique used to measure the size and concentration of particles in a liquid by detecting changes in light intensity as particles pass through a beam of light, with the resulting beam obstruction used to determine particle size, see Figure 2.5. The lower size limit for detection of particles using LO is approximately 1-2 µm, limited by the sensitivity of the photodiode in the detector [42].

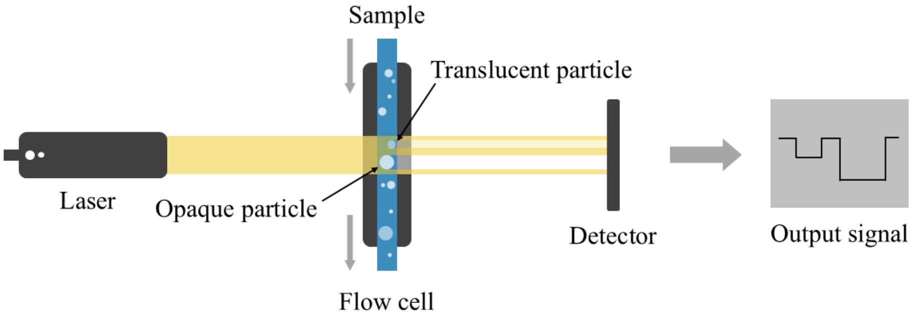


Figure 2.5: Schematic of the setup used in LO. The detector measures the changes in light intensity as particles pass through a beam of light, which is then translated into particle size.

While LO is a standard method and guidelines are provided for the quantification of SVPs using LO in USP <787> [8] and USP <788> [9], it is also known that LO can underestimate particle count for translucent particles smaller than <10 µm [40], [43], [44], as these particles might produce shadows that are insufficiently distinct for the detector to accurately register. Additionally, LO is unable to distinguish between different sub-populations of particles apart from size, such as differentiating between protein aggregates and other particulate matter.

2.3.2 Micro-Flow Imaging

MFI is a microscopy technology in which bright-field images are successively captured of a continuous sample stream that passes through a flow cell, positioned in the field of view of a camera. A schematic figure of the instrument principles and its configuration is shown in Figure 2.6. The captured images are processed using image analysis software that analyzes morphological characteristics and light intesity properties, which can ultimately be used to differentiate between populations [43], [44]. The MFI has a detection size range of 1-70 µm [45].

Compared to LO, MFI has been shown to have better sensitivity, particularly for detecting and characterizing small and translucent particles that may not cast sufficient shadows for LO detection [43], [44]. Additionally, MFI's capability to capture images of individual particles facilitates visual differentiation, enabling assessment of particle morphology and formulation composition, which is not possible using LO.

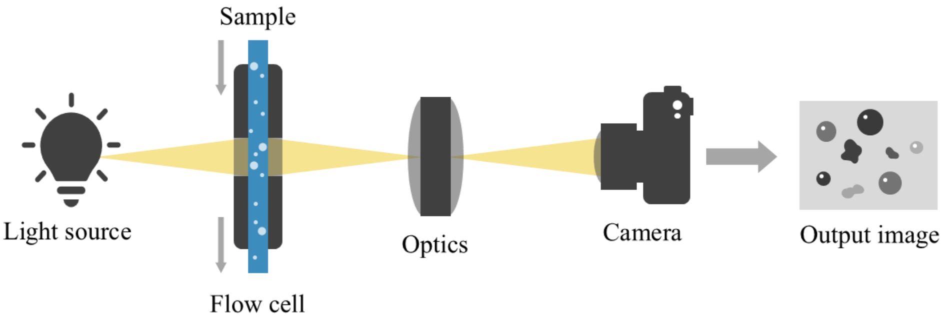


Figure 2.6: Schematic of the setup used in MFI, a bright-field microscopy technique that captures images of a continuous sample stream.

The MFI instrument has the capability to export individual grayscale images of each particle, which are obtained from a larger image of the entire flow cell. This process involves real-time particle detection using a binary masking technique to identify particles within the flow, although this binary data is not permanently stored. Particles are detected through their contrast against the background, which refers to the constant visual backdrop of the illuminated flow cell. The system operates by capturing multiple images per second, allowing each distinct particle to be extracted and analyzed from the continuous flow of the sample. This enables detailed examination of individual particle characteristics independent of the larger background.

Through the manufacturer's built-in software, the MFI outputs numerous morphological and intensity parameters. A complete list of these is presented in Table 2.1. In addition to these, particle position is also given in x- and y coordinates, as well as at what time of the run the particle is found, given in minutes and as a percentage of the total sample run duration.

Table 2.1: Table of the parameters being outputted for each detected particle by the built-in MFI software, along with their respective units and descriptions. Table adapted from [46].

Parameter	Unit	Description
Area	Pixels	The total number of pixels that represent the particle.
ECD	Micrometers	Equivalent circular diameter, the diameter of a circle with the same area as the given shape.
Perimeter	Pixels	The total number of pixels around the boundary of the object.
Maximum Feret Diameter	Micrometers	The longest distance between any two points on the object's boundary, independent of angular rotation.
Circularity	Dimensionless (0-1)	A measure of how close the object is to a perfect circle, calculated as $4\pi \cdot \text{Area}/\text{Perimeter}^2$.
Aspect Ratio	Dimensionless (0-1)	The ratio of the object's minor axis length over the major axis length, of a fitted ellipse with the same second moments as the object.
Intensity Minimum	Grayscale value (0-1023)	The lowest grayscale value of all pixels representing the particle.
Intensity Maximum	Grayscale value (0-1023)	The highest grayscale value of all pixels representing the particle.
Intensity Mean	Grayscale value (0-1023)	The average grayscale value of all pixels representing the particle.
Intensity Std	Grayscale value (0-1023)	The standard deviation of grayscale values of all pixels representing the particle.

2.3.2.1 Morphological Data

The morphological data the MFI Image Analysis software provides is based on how the software detects and defines particles. The algorithm used for this purpose is proprietary to Bio-Techne, the manufacturer of the MFI instrument, so the exact method of calculating the morphological data is not publicly disclosed. Attempting to calculate the data independenly is challenging due to the variable pixel-to-micrometer ratio, which is not constant. The relation between ECD squared and the area in pixels is close to linear, as shown by data provided by the manufacturer, see Figure 2.7, but not identically proportioned. Lacking a 1:1 relationship suggests that the apparent pixel size may vary with particle size. Work presented herein further shows that no linear relation exists also for particles $<10\,\mu\text{m}$. This is explored more in subsection 4.2.3.

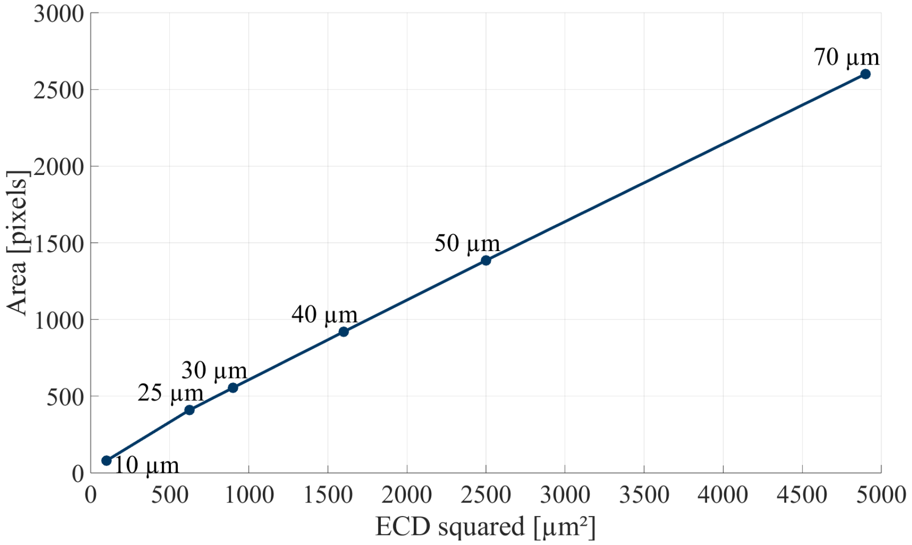


Figure 2.7: The relation between the size of a particle and the number of pixels it occupies in the MFI image analysis software. Image reworked from [47].

2.3.2.2 Light Intensity Data

The MFI Image Analysis software outputs the light intensity data in the form of a 10-bit, grayscale value, able to take on $2^{10} = 1024$ shades of gray, see Table 2.1. The value ranges from 0-1023, where 0 would be completely black and 1023 being completely white. In contrast, the traditional 8-bit format offers only $2^8 = 256$ shades of gray. The 10-bit format, therefore, provides four times as many shades, resulting in smoother images and more detailed gradients [48].

2.4 Spectroscopy

Additional analytical techniques to characterize proteins are briefly presented in this section, including fluorescence spectroscopy, spectrophotometry and X-ray powder diffraction.

2.4.1 Protein Analysis

Fluorescence spectroscopy is an analytical method used to gain insights into protein structure and function, as well as to quantify protein concentrations in biological samples. Fluorescence spectroscopy relies on the excitation of electrons when they absorb high-energy photons or other excited electrons. As the electrons return to the ground state, they emit lower-energy photons. This emitted light forms the basis of fluorescence detection, with the fluorescent molecules termed fluorophores. Fluorophores can either be extrinsic, such as radioactive probes and dyes, or intrinsic, like some amino acids, the most notable being tryptophan [49], [50], [51], [52]. Tryptophan has a maximum excitation wavelength of $\lambda_{ex} = 280\,\text{nm}$ and emits at longer wavelengths of about $\lambda_{em} = 350\,\text{nm}$. Measuring the fluorescence intensity enables estimation of protein concentration, given their proportional relationship. The protein structural state may also be probed by monitoring changes in fluorescence intensity, as buried residues in a hydrophobic core might exhibit higher quantum yield compared to exposed residues in a hydrophilic environment [52].

In addition to fluorescence-based techniques, UV-Visible (UV-Vis) spectrophotometry is a widely used method in molecular analysis. Unlike fluorescence spectroscopy, which measures emitted light, UV-Vis spectrophotometry is based on the absorption of light in the ultraviolet and visible range. It is commonly used for quantitative analysis, such as determining the concentration of proteins with the absorbance of light by aromatic amino acids at 280 nm being a standard approach [53], [54].

2.4.2 X-Ray Powder Diffraction

X-ray powder diffraction (XRPD) is a technique widely employed for determining phase composition and crystal structures [55]. A monochromatic X-ray beam is directed at the powdered sample, and the diffracted intensity is measured as a function of the scattering angle (2θ). The resulting diffraction pattern arises from constructive interference of the rays being scattered by the periodic arrangement of atoms in the crystal lattice. Analysis of the diffraction peaks, such as positions and intensities, provides information about the material.

XRPD is particularly advantageous when single crystals are unavailable, allowing for insight into bulk polycrystalline samples. The technique can determine polymorphism, detect amorphous content and identify complex phase mixtures [55]. Through further refinement, XRPD is also able to determine lattice parameters and phase identification through comparison with reference databases. Its broad applicability and non-destructive nature makes XRPD a widley used tool in fields such as materials science, solid-state chemistry and pharmaceutical development.

2.5 Particle Differentiation

Particle differentiation involves distinguishing between various sub-populations within a sample. By identifying different particle types, it provides insights into their characteristics and potential impacts on formulation performance. This section presents a brief overview of how a particle itself is identified, as well as different methods that can be used to differentiate between particles.

2.5.1 Particle Detection

In order to use data like morphological parameters to differentiate between particles, the particles themselves must first be identified. This involves analyzing images using binary identification techniques so that particle boundaries are clearly defined. Morphological parameters can then be extracted from the binary format.

2.5.1.1 Object Detection Techniques

Numerous techniques exist today for detecting objects using image analysis. Thresholding is an image processing technique that converts a grayscale image into a binary image by assigning pixels as either foreground or background based on whether their intensity is above or below a certain value, threshold T [56]. In the context of MFI, the background refers to the illuminated flow cell against which particles are deteced. There are different methods for determining this threshold, with the simplest being the use of a single, global threshold. This method would be fast and require minimal computational power, but would fail with variable backgrounds. Otsu's method is another way of determining the threshold, where each possible threshold is iterated over and a global optimal threshold is calculated by minimizing intra-class variance [57]. Otsu's method is efficient when the image has a clear bimodal histogram, but would struggle otherwise. Adaptive thresholding calculates a different threshold for each region of the image, based in its local neighbor [58]. This could possibly create artifacts in smooth regions and requires more computational power. It would however be able to handle varying lighting conditions and complex backgrounds.

Edge detection can also be used to identify particles. The traditional Canny algorithm is a widely used edge detection method that consists of smoothing the image, calculating gradients, suppressing non-maximal values, and applying a double threshold to connect edge points [59]. Although effective, it faces challenges such as noise sensitivity and the use of fixed thresholds, which can lead to false edge detection and reduced adaptability.

Another way of detecting objects is to look at textures instead. While thresholding only involves grayscale histograms computed from the images, texture analysis also considers the spatial interaction between pixels. One approach is the use of Haralick texture features [60], [61]. The features are calculated from a gray-level co-occurrence matrix, which captures how often pairs of gray levels occur at specific distances and orientations. This provides a way to numerically describe the texture in an image. Haralick features are however not ideal for small objects, especially if they lack distinct textures, and also requires a high image resolution.

Machine learning, particularly supervised learning, can be applied to particle detection by training a model on labeled data to distinguish particles from the background [62], [63]. In supervised learning, the algorithm learns from example inputs paired with correct outputs, in this case images where particles have been manually marked. Features such as shape, intensity, or texture are used to build the model. However, for this approach to work well, there must be a clear visual difference that a human could identify, as the model relies on patterns detectable in the labeled training data. This also applies for the texture features.

2.5.1.2 Refinement of Binary Mask

Additional refinements on the binary equivalence can be made to further enhance and clean the binary mask. A list of such operations used is presented in Table 2.2.

Table 2.2: Morphological operations and their effects on binary patterns. Each transformation shows the input on the left and the result on the right.

Operation	Description
Bridge	Bridges unconnected pixels, a 0-valued pixel is set to 1 if it has two unconnected non-zero neighbors. $\begin{matrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{matrix} \rightarrow \begin{matrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{matrix}$
Clean	A 1-valued pixel is set to 0 if it is surrounded by 0-valued pixels. $\begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix} \rightarrow \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix}$
Fill	Fills isolated interior pixels, meaning 0-valued pixels that are surrounded by 1-valued pixels. $\begin{matrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{matrix} \rightarrow \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}$
Spur	Removes small extrusions, meaning 1-valued pixels that extend from objects with minimal connection to other 1-valued pixels. $\begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{matrix} \rightarrow \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{matrix}$

2.5.2 Principal Component Analysis

Principal component analysis (PCA) is a multivariate statistical method used to analyze datasets where observations are characterized by mutiple interrelated quantitative variables [64], [65]. The technique aims to reduce data dimensionality by transforming the original correlated variables into a smaller set of uncorrelated variables, called principal components, that retain most of the original variance. The first principal component captures the greatest amount of variation in the data, followed by the second component, which accounts for the next largest source of variation, and so on [66]. The principal components are all orthogonal to each other, guaranteeing that they are uncorrelated to each other. Variation between data points is shown graphically in a scores plot, while variables are shown in a loadings plot.

Data is collected into a matrix $\boldsymbol{X}$ of size $I \times J$ with rows usually representing observations and columns representing variables [64]. Individual variables are denoted $\boldsymbol{x}$, of which linear combinations can be written as $\boldsymbol{t} = w_1 \times \boldsymbol{x}_1 + \cdots + w_j \times \boldsymbol{x}_j$. The vector $\boldsymbol{t}$ is now in the same I-dimensional space as the original data, where the normalized weights w_1,...w_j are chosen to maximize the variance of $\boldsymbol{t}$. The problem can therefore be summarized as

$$\arg\max_{\|w\|=1} \operatorname{var}(t). \tag{2.2}$$

These weights are the principal components, which can be used to create visualizations that reveal patterns in the data. Plotting the scores of the principal components against each other can reveal clusters and outliers among observations, making it easier to interpret groupings or trends. A loadings plot can show how variables contribute to each component, revealing correlated variables and dominant features.

PCA generates statistical metrics that help assess the reliability and relevance of the model. Two indicators are R^2X, which shows the proportion of variation in the dataset explained by the model, and Q^2, which reflects the model's predictive performance [66]. The Q^2 value is obtained through cross-validation by omitting portions of the data, predicting them using the model, and comparing predicted versus actual values. Together, R^2X and Q^2 help evaluate the model's robustness. Both values should be high (>0.5, ideally >0.9), with low difference between them (<0.3 difference) indicating low overfitting and a robust, low-noise model, see Figure 2.8.

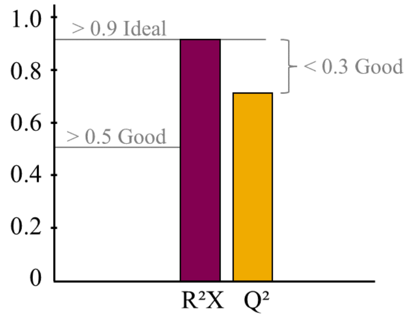


Figure 2.8: Graphical representation of preferable relations between R^2X and Q^2 values.

3

Methods

This chapter presents materials used in this work, and sample preparation. A more detailed explanation on how the MFI instrument is operated is also given, as well as the method used to analyze particle images.

3.1 Materials

Equipment used to prepare samples include pipettes, together with 1-10 mL, and 200-1000 µL and 20-200 µL pipette tips, for handling liquid materials. An analytical balance was used for weighing dry materials, alongside a thermohygrometer. Samples were prepared in 10 mL vials, and 2 mL glass vials. A magnetic stirrer with magnetic stir bars, and a roller mixer were used for proper dissolution. A thermomixer was used for simultaneous heating and mixing.

Substances used include Thermo Scientific™ COUNT-CAL™ Count Precision Standards, of sizes 2 µm, 5 µm, 10 µm and 25 µm, and ethylene tetrafluoroethylene (ETFE) particles from NIST [67]. ETFE particles are chemically inert polymers meant to mimic proteinaceous particles. Milli-Q® Ultrapure water was used as a diluent.

The primary protein analyzed will hereinafter be referred to as Protein X, while Protein Y and Protein Z were used for comparison of formulations. Excipients will be denoted A, B, C and D. See Table 3.1 for a complete list. Drug substance (DS) contains the API in the form of a liquid. A drug product (DP, active or placebo) on the other hand, refers to spray-dried material in capsules containing the API and the excipients in different ratios. Different ratios of the same protein-excipient combinations are denoted L (low), M (medium) and H (high) corresponding to relative protein concentration within each group.

Table 3.1: Table listing all analyzed samples, organized by category and available formulations.

Category	Sample	Available in
Placebo	AB Placebo	-
	ABC Placebo	-
	ABCD Placebo	-
Protein X	DP-X-AB	L, M, H
	DP-X-ABC	L, M, H
	DP-X-B	L, M, H
	DS-X	-
Protein Y	DP-Y-ABCD	L, M, H
Protein Z	DP-Z-ABCD	L, M, H

3.2 Sample Preparation

To prepare the samples that were measured in the MFI instrument, dry material such as drug product or excipients, was weighed and deposited into 10 mL vials. This was done in a controlled environment with a low relative humidity at room temperature monitored by a thermohygrometer, to reduce water uptake during sample preparation. Water was added as a diluent to the dry powder. The vials were left on a roller mixer for at least 60 minutes under gentle stirring, until the sample was completely dissolved. In cases where samples underwent thermal stress, a 1 mL aliquot was taken of the sample and deposited into a 2 mL glass vial, and heated while shaking to ensure even heating.

Samples containing standard size beads and ETFE particles were by directly pipetting and diluting the substance with water in the 96-well plate used for the MFI instrument.

3.3 Micro-Flow Imaging

Further details on the MFI method, and particle detection interface are presented in this section.

3.3.1 Micro-Flow Imaging Method

The samples, including liquids or solids dissolved in diluent, were analyzed using an MFI5200, manufactured by ProteinSimple, Bio-Techne, equipped with a Bot1 to enable automation of sample analysis. In accordance to the internal method previously developed,1 mL of sample material was deposited into a well of a 96-well plate, and several samples could be measured in parallel. Molecular biology grade water is used for flushing the flow cell, with low concentrated detergent to ensure adequate cleaning. Conductive pipette tips are used, both to dispense sample and detergent but also to flush water.

Analyzing the batch of samples can be managed by creating custom run sequences. The run sequence used for this project was based on a previously developed method, see section 1.2. The method is proved to be robust, ensuring sufficient flow cell cleaning and conditioning, and correct homogeneity and loading of the sample.

3.3.2 Particle Detection

The MFI Image Analysis software takes pictures of the flow cell while the sample is being passed. An example of a flow cell image is shown in Figure 3.1. Individual particles in these larger images are then detected and extracted as separate images.

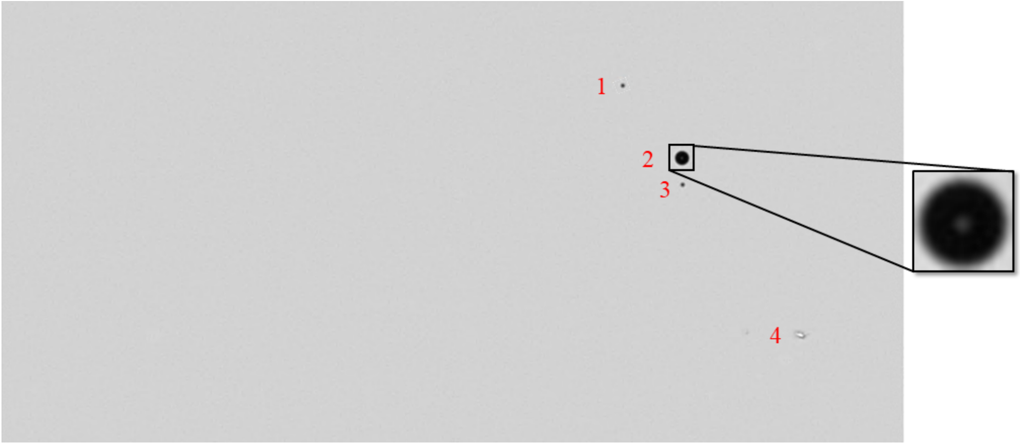


Figure 3.1: An image taken of the flow cell, where four particles can be seen against the background. The inset provides a magnified view of one of these particles.

A quick view of the detected particles is available in the software, as showcased in Figure 3.2, where each image shows one individual particle.

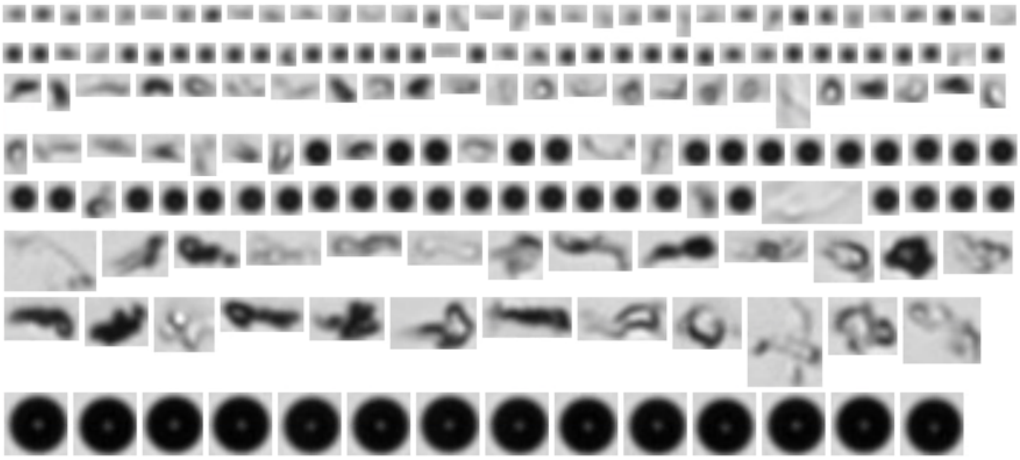


Figure 3.2: An MFI Image Analysis output example, showing particles ordered by ECD. Each image contains one particle.

3.4 Image Analysis

Although the MFI Image Analysis software provides data parameters like morphological data and light intensity data, inconsistencies in these results prompted the need for custom analysis. This is explained more in subsection 4.2.1. The individual particle images obtained from the MFI were therefore exported to be analyzed independently. It is noted that although the MFI Image Analysis software seems to handle the images using 10-bit format, exporting the images transforms them into 8-bit format. The number of grayscale shades therefore decreases from 1024 to 256, but this was deemed sufficient and that it would not have affected any results to a greater extent. Image analysis was performed using MATLAB.

Particle detection was performed empirically and by trying different methods and threshold levels. Smaller sets of particles were created, for which images were chosen to span the biggest visual variance between the particles. Intermediary images were created to test the different effects of the methods. These include

- Original: The original exported image of the particle on the background, in 8-bit format.
- Bright binary: Binary image with corresponding pixels brighter than a certain threshold.
- Dark binary: Binary image with corresponding pixels darker than a certain threshold.
- Combined binary: The combined image with the bright and dark images. In some cases, refinement was applied to the combined binary, mentioned in subsubsection 2.5.1.2.
- Masked: Final image with the combined binary image masked over the original image, so that the original particle can be seen with the background removed.

Some methods did not require the bright and dark binary, instead going directly to a final binary version and then the masked image. Morphological data would be determined from the obtained binary image using a pixel-to-micrometer ratio. Light intensity data could either be determined directly from the original image without any masking, or on the masked image.

3.5 Principal Component Analysis

PCA was performed using SIMCA® to reduce the dimensionality of the datasets and to visualize sample distribution. The analysis produced scores and loadings plots to assess grouping patterns and variable relationships, as well as R^2X and Q^2 values. A 95% confidence interval is included in the scores plot.

3.6 Fluorescence Spectroscopy

For fluorescence testing, a multi-mode microplate reader was used, together with 0.22 µm filters and a 6-well plate. Fluorescence spectroscopy was performed for additional protein particle quantification. A placebo sample was prepared as control, and a DS sample of protein X was prepared as well. Samples were passed through the filter using a syringe in order to catch larger insoluble particles. Diluent was then flushed through several times to ensure proper dissolvement of smaller, soluble particles. The filter, containing the insoluble particles from the solution, was placed in a well plate, in turn placed in the microplate reader. An internal method, leveraging the intrinsic fluorescence of the tryptophan residues of the protein, was used to evaluate the fluorescence emission from the particles caught on the filter and estimate protein content.

3.7 UV-Visible Spectrophotometry

UV-Vis spectrophotometry was performed on samples where the protein concentration of the dry powder was unknown. The sample was dissolved to a known concentration, and then dispensed into a cuvette. A spectrophotometer was then used to perform the measurements to reveal the protein concentration. It is noted that this method was only used to determine protein concentration of samples which required a concentration re-test.

3.8 X-Ray Powder Diffraction

For XRPD analysis, the sample was evenly distributed on a sample holder. The sample was then analyzed using an X-ray diffractometer to generate diffraction patterns. These patterns were analyzed to determine the material's structure by comparison with internal reference data.

4

Results and Discussions

Results and corresponding discussions are presented in this chapter. Future outlooks are also presented.

4.1 Correlation Among Particle Attributes

This section presents initial methods of sub-population differentiation, and how these methods showed correlation between parameters.

4.1.1 Initial Validation with Standard Particles

An initial study was conducted to assess the capabilites of MFI technology for sub-population differentation. This work investigated standard size beads and ETFE particles. The standard size beads served as a model due to their appearance as distinct, dark and spherical particles, allowing for straightforward analysis of morphological parameters. In contrast, ETFE particles offered a bigger variation, mimicking the optical characteristics of proteins, thereby presenting a test case for the instrument's capability.

PCA was conducted on standard size beads and ETFE particles based on the data given by the MFI instrument directly. A scores plot for samples containing the standard size beads, ETFE particles, and a mix of all of these can be seen in Figure 4.1. Although the standard size beads fall in respective clusters, there is significant overlap between the ETFE particles and the smaller size beads. The distinct clustering observed in the figure is primarily due to strong influence of size-related variables, which contribute significantly to the principal components. The clusters holding the standard size beads are all positioned within the 95% confidence interval, indicating homogeneity within the samples. In contrast, the ETFE particles display a broader spread, with several data points extending beyond the confidence boundary. This indicates greater internal variability and deviation from the model's structure, reflecting heterogeneity or noise within the data points. Additionally, this suggests that the ETFE particles, intended to mimic proteinaceous particles, may not exhibit clearly defined characteristics. This might be extended to protein particles possibly also not having clear distinguishing features that can be used for sub-population differentiation.

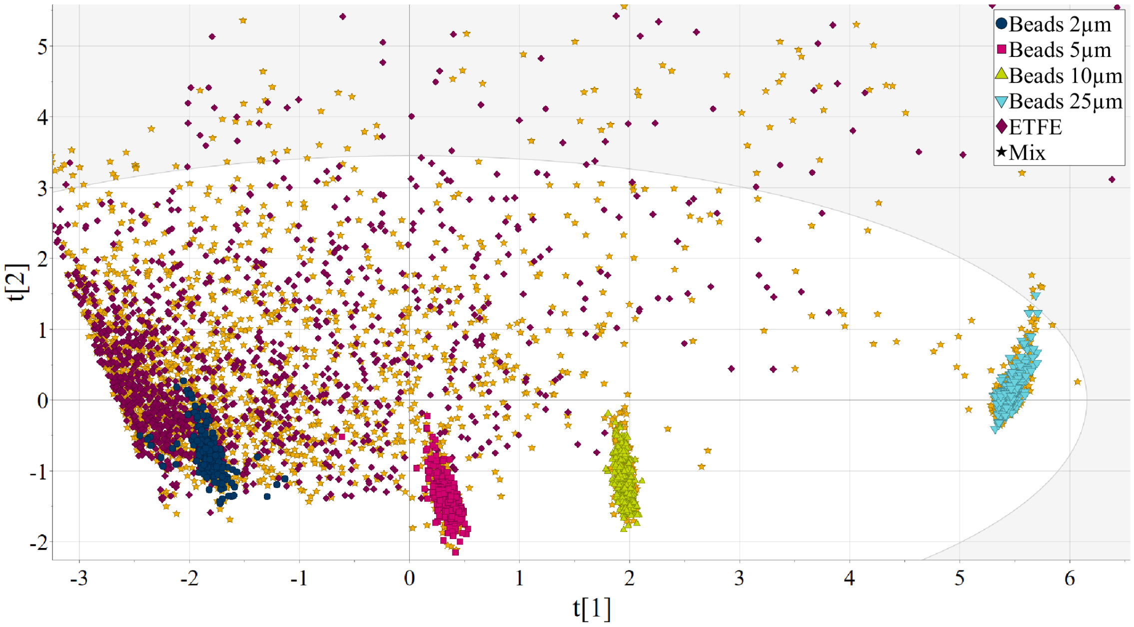


Figure 4.1: PCA scores plot showing the distribution of standard size beads and ETFE particle samples in the space defined by the first two principal components. The points are colored according to sample, each point representing an individual particle. The ellipse around the origin shows the 95% confidence region.

Figure 4.2 presents the cumulative R^2X and Q^2 values of the model covering the standard size beads and ETFE particles. The model exhibits fairly high R^2X values from the first principal component, indicating that a significant portion of the variance is explained. Values surpassing 0.9 around the third and fourth components further demonstrate the model's explanatory capabilites. Additionally, the model also has a high Q^2 value, showing that the model has a strong predictive ability across the components. The small difference between the R^2X and Q^2 values further suggests low overfitting and robustness of the model.

Although some fluctuations in the Q^2 values are observed, it is not necessarily indicative of model instability. Such variations are common, often resulting from sampling variability, and typically do not detract from the overall reliability of the model unless there is a significant drop. While high R^2X and Q^2 values show the proficiency and predictive capabilites of the models, they do not inherently reveal the existence of distinct sub-populations. They primarily indicate that most data points fit the model well, but don't highlight any underlying structure or differences between data sub-populations. Additional analysis would be needed to identify if these exist in the dataset.

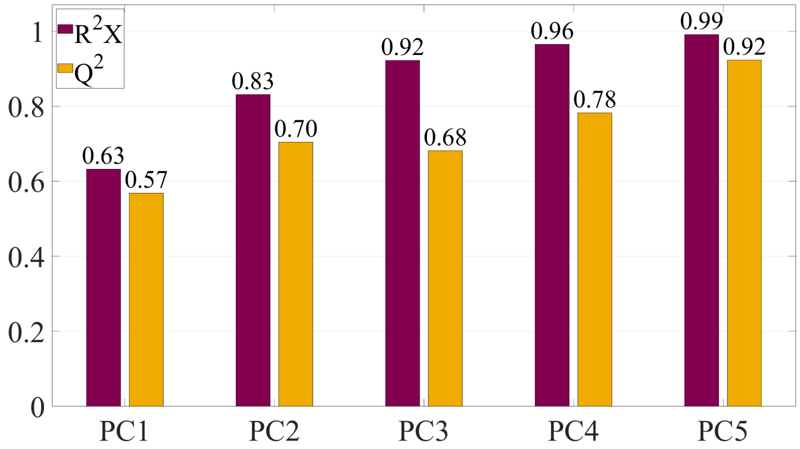


Figure 4.2: Cumulative R^2X and Q^2 values for model of samples containing standard size beads and ETFE particles.

While ETFE particles were used to mimic protein particles, no clear differentiation or trends were observed, potentially due to the simplicity of the tested system. Further work was peformed on the target samples (placebo and DP samples).

4.1.2 Challenges with Drug Product Analysis

Consequently, PCA was similarly conducted on actual placebo and DP samples (Protein X, excipients AB) to assess whether the varying formulations would exhibit any clustering, see section A.1. However, the scores plot showed significant overlap without distinct clustering, calling for the use of other approaches instead.

As the PCA scores plots did not inherently reveal any clear differentiation in both models, see Figure 4.1 and Figure A.1, the next approach was to use scatter plots and distributions to see if any relation could be seen. Loadings plots are used in PCA to identify which variables are best suited for plotting against each other by selecting those located on opposite sides of the graph. This process highlights contrasting characteristics and can reveal meaningful patterns. The loadings plot provides insight into how variables contribute to and relate within the principal components. The further from the origin, the stronger the contribution to the principal components. Parameters closer together exhibit strong correlation to each other.

It can be seen from Figure A.3 that size-related parameters like maximum feret diameter, perimeter, area and ECD are all closely correlated, as well as aspect ratio with circularity, which is expected. It is also revealed that the size-related parameters correlate negatively to intensity minimum and intensity mean. Since the main aim is to obtain a deeper understanding of the nature of each detected particle beyond its diameter, which can already be determined by LO techniques, other parameters such as intensity minimum/intensity mean and aspect ratio/circlarity were chosen to be studied. Intensity maximum and intensity standard deviation against intensity minimum/intensity mean was also a possibility, focusing only on the intensity data instead of morphological data.

Investigating these plots and distributions was proven more difficult than first expected - the graphs showed unexpected features, such as the clear cut-off for the intensity maximum showed in Figure 4.3. Due to the natural variability in size, shape and optical properties, particles are expected to exhibit a range of intensity maxima, resulting in an even distribution rather than uniformity. This prompted the need to look more closely at the images themselves and how the built-in software in the MFI instrument was handling the particles, indicating potential issues with the software's particle detection algorithm.

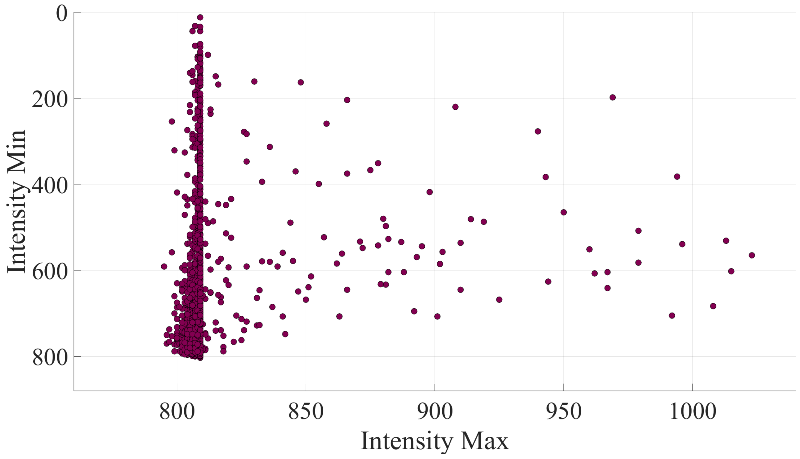


Figure 4.3: Scatter plots of intensity maximum against intensity minimum using the data from the built-in algorithm.

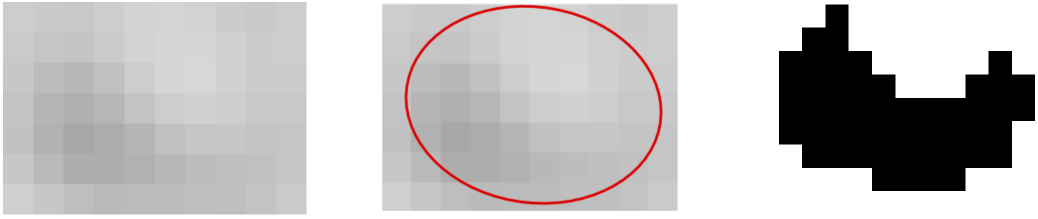
4.2 Optimized Particle Detection

The inadequacy of the vendor's software of detecting particles is presented and discussed in this section. A custom-built algorithm used to combat this issue is presented as well.

4.2.1 Detection Mechanisms

As the particle detection algorithm in the built-in software seemed not to be accurate, as previously mentioned, the binary view in the software was explored. A more detailed view of one example particle is shown in Figure 4.4, where the original image is shown together with a presumed approximate edge and the same particle in binary mode. The image of the binary mode is taken from a live run of a sample, as the binary images are not stored. It is noticed that the bright region in the upper right corner is not correctly detected by the software as being part of the particle. This also results in the particle having a low aspect ratio of 0.53, while it is distinctly fairly rounded.

This proved that the particle detection algorithm itself is inadequate, meaning that all the data parameters provided by the software might also be inaccurate. While the algorithm is proprietary to Bio-Techne, as mentioned in subsubsection 2.3.2.1, all parameters are seemingly calculated only on the masked binary equivalence on the detected particle, light intensity parameters included. This means that, since the particle detection is inaccurate, so might all data parameters also be.



(a) Image of a particle measured in MFI.

(b) The particle's presumed approximate edge.

(c) Binary image of the particle.

Figure 4.4: A detected particle using the MFI, along with an approximate edge determined visually, and the equivalent binary image.

4.2.2 Custom Algorithm

A custom particle detection algorithm was constructed for the individual particle images. While numerous methods were tested, see section A.2, the final algorithm uses dual intensity thresholds - one for darker particles and one for brighter particles, in relation to a fixed background intensity, combined with refinements on the binary mask. The MFI flow cell background intensity was determined to be a normal distribution centered around 210 with low variance, see Figure 4.5 for an example.

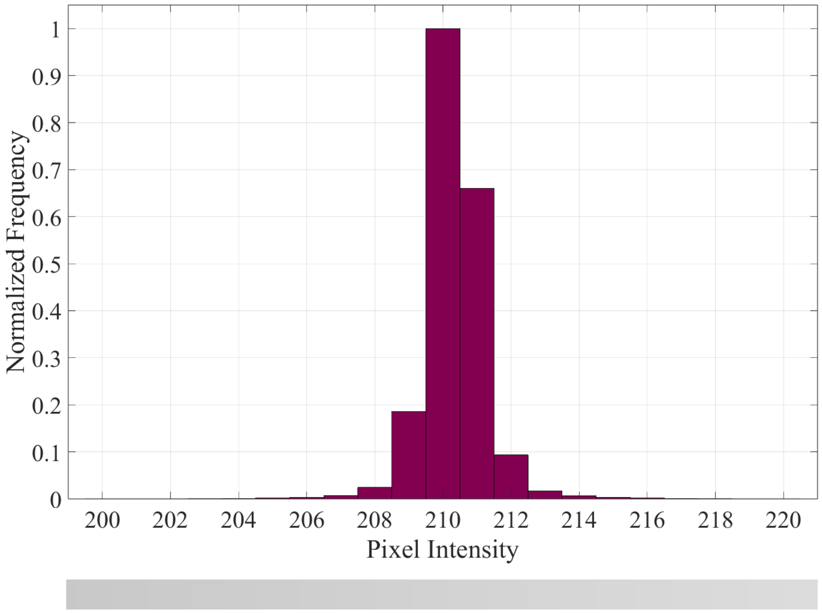


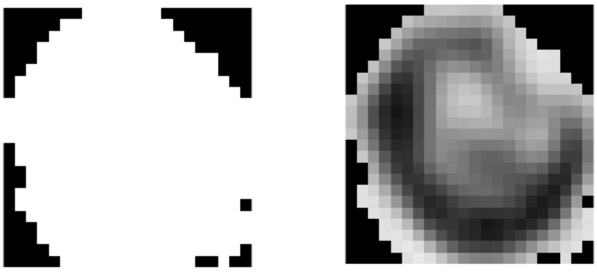
Figure 4.5: Normalized histogram for the pixel intensity distribution of the MFI flow cell background, with a corresponding gradient between intensities 200-220 is below.

As the instrument requires an illumination adjustment for each sample, this implies potential differences in background intensity for each sample. Consequently, backgrounds from several runs were tested, as presented in Figure A.4. The figures show only slight variations in the background pixels. Due to how the images of the particles were taken however, the background intensity was set to span between 200-220, since some smaller particles occupied the entire image, leaving no visible background. Every pixel outside the background intensity span is automatically considered to be part of the particle. Pixels values below 200 are considered part of the particle in the dark binary intermediary, while the limit is above 220 for the bright binary intermediary. The final Matlab script is presented in section A.5.

The final algorithm exhibited certain deficiencies, explored in more detail here. Results of a representative particle, using the different intermediary images separating the bright and dark pixels and subsequently masking the original image, are shown in Figure 4.6 and Figure 4.7. Both figures are analyzed using the same algorithm, but the different appearances of the particles highlight varying levels of success with the new, custom algorithm. Figure 4.6 clearly shows that, in this specific case, using only the dark binary would result in a more accurate detection algorithm. The brigher pixels show the presence of a halo around the particle, indicating that some diffraction and phase contrast effects are occuring. The halo only appears for particles that are relatively large and dark, probably due to their greater ability to refract and scatter light, leading to optical interference at their edges.

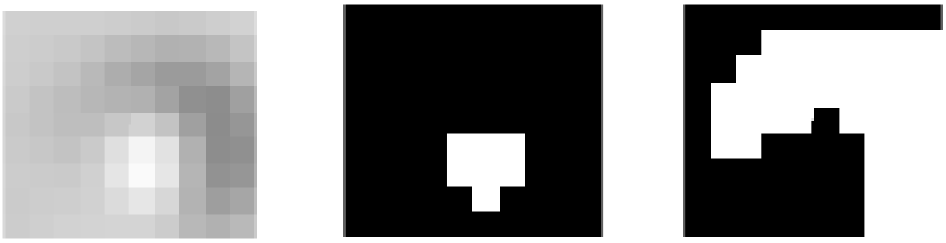
While Figure 4.6 shows that using only the dark binary detection is superior to using both bright and dark, Figure 4.7 shows why the brighter particles cannot be omitted. The original image shows a particle with a distinct brighter region that would be lost if only the dark binary was used, much like Figure 4.4. Here, the final masked image shows the detected particle to include both darker and brighter pixels, resulting in a more accurate particle shape. The extent of the shape accuracy may however be debated. Conclusively, both darker and brighter pixel thresholds have been used in all further analysis presented in upcoming sections.

(a) Original image. **(b)** Bright binary. **(c)** Dark binary.

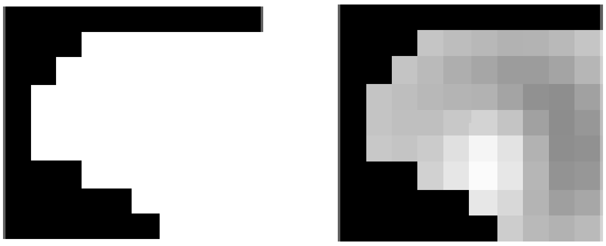


(d) Combined binary. **(e)** Masked image.

Figure 4.6: Example of an original image of a relatively large and dark particle, along with the intermediary images, and the final masked image.



(a) Original image. **(b)** Bright binary. **(c)** Dark binary.



(d) Combined binary. **(e)** Masked image.

Figure 4.7: Example of an original image of a particle with a bright region, along with the intermediary images, and the final masked image.

4.2.3 Intensity and Morphological Analysis

The new, custom particle detection algorithm recalculated the light intensity data parameters by looking at pixel values. Morphological data would theoretically also be able to be recalculated. Parameters measured in pixels (area, perimeter, circular, aspect ratio) could be obtained directly from the binary images by counting pixels, whereas parameters measured in micrometers (ECD, maximum feret diameter) could be calculated using a conversion factor between pixel size and micrometers. The proprietary algorithm from the vendor also encompasses this conversion, see Figure 2.7. Similar attempts were made to determine this using experimentally obtained data.

Table 4.1 shows the morphological parameters ECD and area of four different particles, each taken from measurements of the standard beads. The particles presented all have an aspect ratio of 1.00 to ensure proper conversion between circular area in micrometers and pixels. As the table shows, the calculated pixel size is not constant, meaning there is no set pixel size. Similar calculations using perimeter values instead of area were also performed, with comparable results. Figure 4.8 shows that the relation between ECD and pixel area is not strictly linear.

Table 4.1: Measured values for ECD in micrometers and area in pixels of four different particles using MFI, along with the calculated pixel size, based on the experimentally obtained data.

Particle	ECD [µm]	Area [pixels]	Pixel size [µm²/pixel]
1	2.125	28	0.127
2	4.875	56	0.333
3	9.875	118	0.649
4	24.375	418	1.116

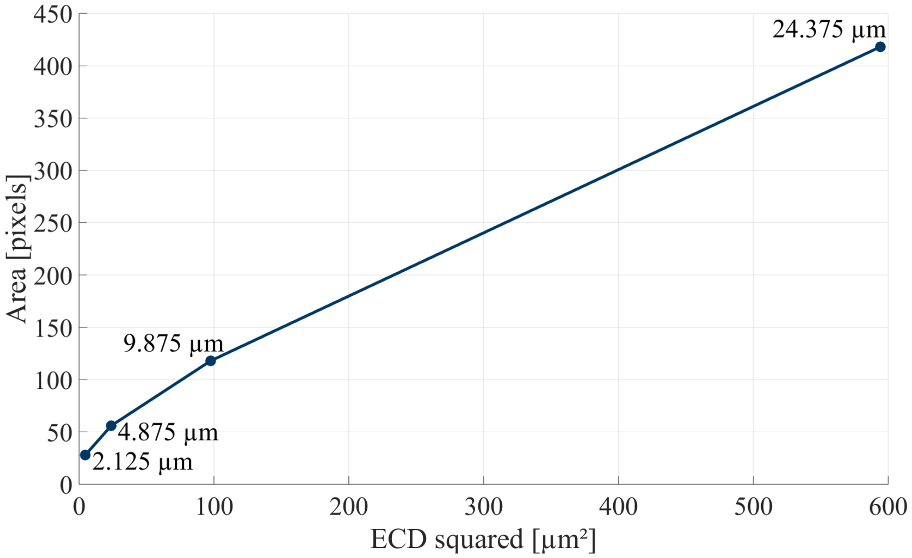


Figure 4.8: Graph showing the non-linear relation between ECD squared and the area in pixels, based on data presented in Table 4.1.

Although pixel sizes could be calculated, as in the example above, the non-linear relationship between ECD and pixel size makes these calculations more complicated. This specific case was simplified by the particles all having an aspect ratio of 1.00, enabling calculations based on the assumption of circular-shaped particles. For particles with different aspect ratios other than 1.00, using the same method to determine pixel size would not be applicable. While interpolating pixel size data based on circular particles and applying these to non-circular particles to determine their morphological data is possible, this method was proved unreliable due to aforementioned factors. Conesequently, this approach was ultimately abandoned. An additional argument to this is that the custom particle detection algorithm in itself was not very reliable either, as presented earlier.

4.3 Intensity Analysis and Comparability

Instead of attempting improvement on calculations of morphological data, focus shifted to light intensity data instead. The inaccurate detection system of the built-in software also gave inaccurate intensity data, prompting the need to recalculate these. Intensity data such as intensity minimum and maximum were calculated from the original image (examples of which are shown in Figure 4.6a and Figure 4.7a), while intensity standard deviation and mean were calculated on the masked binary equivalence (examples shown in Figure 4.6e and Figure 4.7e). This was to ensure that the full range of intensity was covered for the intensity minimum and maximum, while the standard deviation and mean focused on the particle, excluding any background. Particles that only have pixels darker than the background will have an intensity maximum at background level. Particles that are completely brighter than the background will have an intensity minimum at background level. Note that the darker the pixel, the lower the intensity value, and the brighter the pixel, the higher the intensity value. While this approach of using light intensity data focuses less on individual particles compared to morphological data, looking at distributions instead would still give insight into sample compositions. Such distributions are presented and discussed in subsection 4.3.1.

The custom algorithm's success at reading intenstity levels is showcased in Figure 4.9. Whereas the aforementioned cut-off can be seen in Figure 4.3, the new figure shows the expected, more even distribution. It is noted that the data from the built-in algorithm uses 10-bit format, while the custom algorithm that was based on exported uses 8-bit format, hence the different scales on the x-axis. The formats are, however, proportional.

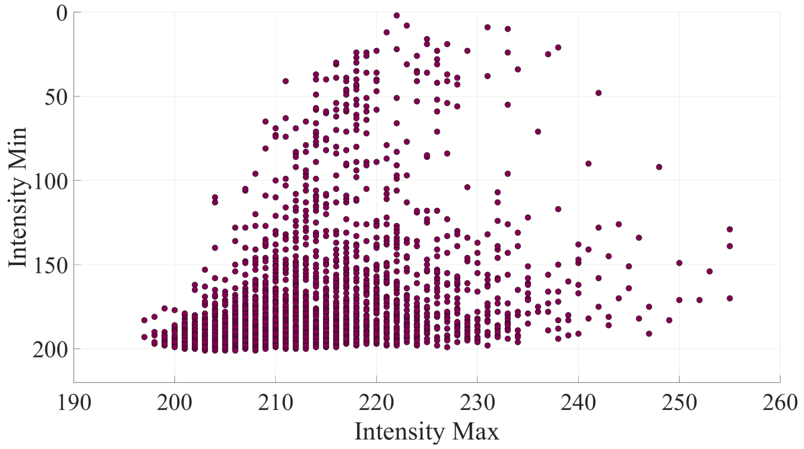


Figure 4.9: Scatter plots of intensity maximum against intensity minimum using the data from the custom algorithm.

4.3.1 Evaluation of Protein X with Excipients AB

Light intensity data from DP particle images are plotted in Figure 4.10, corresponding to an intensity distribution for DP-X-AB samples. Note that the roughness of the curve is variability due to comparatively low particle counts; increasing particle count increases smoothness. The distribution peaks at around 205, which is slightly lower than the background that was previously calculated to be at 210. As mentioned previously, this is due to some of the smaller particles spanning the entire image, so no pure background is left exposed. As the intensity maxima peak at background intensity, this means that background level is the most common occurence, while everything to the sides of the peak is occuring less frequently. See Figure A.11 for all plotted distributions of maximum intensity for the different sample series. Figure A.12 show the intensity minimum distributions, peaking at around 195. The intensity maximum and intensity minimum distributions imply that most of the pixels across all images would fall in the narrow span between 195-205. This is validated by the fact that the majority of the particles are small, and with the resolution of the MFI can only be seen as a faint shadow.

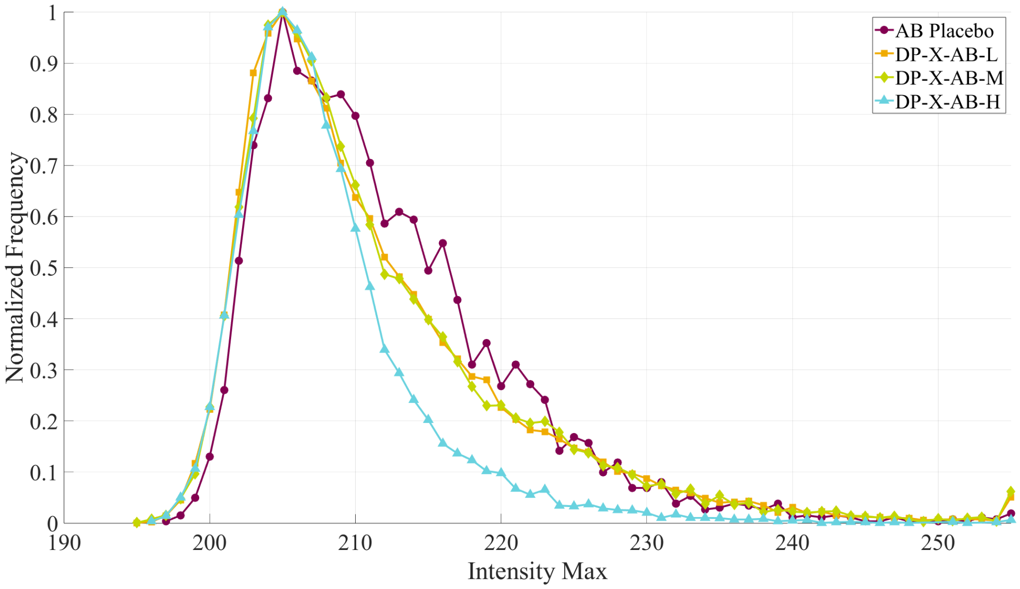
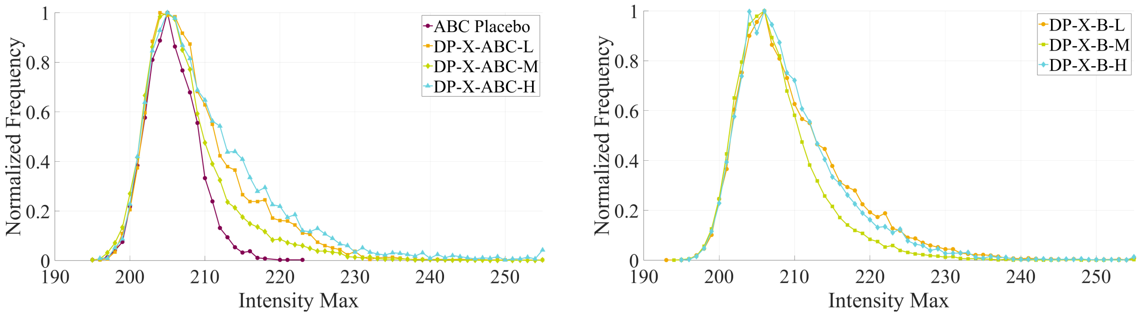


Figure 4.10: Distibution of intensity maximum against normalized frequency, for DP-X-AB samples along with AB Placebo.

While the peaks of the intensity maximum distributions all fall in the same background range for all samples, significant changes can be observed in the distibution profiles toward higher intensity values. The intensity distribution in Figure 4.10 show a clear separation between the different samples. The frequency at higher intensities is lower for the highest strength sample (DP-X-AB-H; highest protein/excipient ratio), while the profile is very similar for both medium (DP-X-AB-M) and low (DP-X-AB-L) strengths. Furthermore, the placebo sample has an even higher ratio of brighter particles compared to the DP samples. The graph seems to indicate that particle brightness is correlated to the concentration of excipients, and that protein particles might be darker than the excipients. This is further supported by Figure A.11d, as the DS sample contains only protein and shows an even darker distribution than the high strength (DP-X-AB-H) sample.

4.3.2 Analysis of Alternative Formulations and Proteins

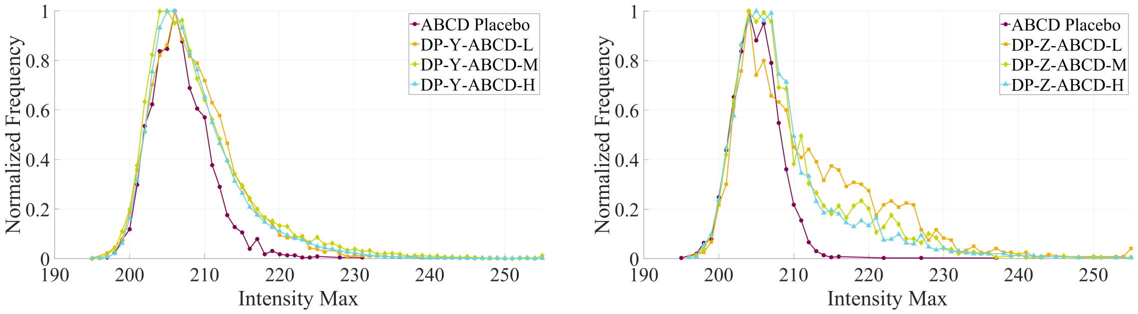
To further support the findings, other formulations containing the same protein but with distinct excipient compositions were analyzed. Figure 4.11 shows the intensity maximum distributions for DP-X-ABC and DP-X-B samples. It is noted that a placebo sample for the DP-X-B series is missing due to unavailability of the material, so no direct comparisons of placebos can be made for this case. Figure 4.11a shows, in contrast to Figure 4.10, that the ABC Placebo sample now seemed to contain darker particles than the DP samples. Additionally, the DP distributions no longer align in order of protein concentration in both Figure 4.11a and Figure 4.11b, suggesting that other factors may be influencing the observed results. This discrepancy highlights that more complex relationships and interactions are likely at play, potentially involving not only the inherent properties of the particles but also the interactions amongst various components within the different formulations.



(a) X-ABC. **(b)** X-B.

Figure 4.11: Distributions plots showing intensity maximum for particles in DP-X-ABC and DP-X-B samples.

Other proteins were also analyzed, and the distributions are shown in Figure 4.12. Again, the placebos, now with ABCD excipients, showed the lowest ratio of bright particles.



(a) Y. **(b)** Z.

Figure 4.12: Distributions plots showing intensity maximum for particles in DP-Y-ABCD and DP-Z-ABCD samples.

Analysis of intensity distributions revealed some patterns and discrepancies. If the comparability of the distributions is focused only on the placebos, the reason behind the differences falls on the excipients. The AB Placebo exhibits clear differences from the ABC(D) Placebos, meaning that the presence of the additional excipients might change the behavior when the individual excipients are in solution. Further examination of the placebos and excipients may reveal their impact on the formulation's characteristics.

4.4 Advanced Characterization Techniques

This section presents a deeper analysis of the particles, and how they are characterized. First, an overview of attempts at protein analysis is given, followed by characterization of excipients.

4.4.1 Protein Analysis Using Fluorescence and Spectrophotometry

Deeper analysis of the samples was needed in order to establish any valid relations in the intensity distributions plots. Spectrophotometry and fluorescence spectroscopy was utilized in an attempt to further characterize the SVPs observed in the DP samples. This setup aimed to determine the concentration of protein particles, offering insights into the proportion of SVPs that were actual protein as opposed to excipients or other particles. Samples were filtered, and the insoluble particles collected on the filter. Initial protein concentration and fluorescence tests successfully detected protein content among the insoluble particles on the filter, as evidenced by clear fluorescent emission from the DP sample. The emission spectrum obtained matched the expected pattern for tryptophan emission. However, inconsistencies in measurements and sample preparation were detected, with high background noise from the soluble protein even after several filter washing cycles. The fluorescence strategy was thus ultimately abandoned.

4.4.2 Excipient Characterization through MFI and XRPD

As previously mentioned, it was noted, that the AB and ABC(D) Placebos showed differences in their respective intensity maximum distributions. Excipient A or B seemed to be impacted by the presence of other excipients (C, D). In order to understand which excipient was critical, further characterization was made by analyzing SVP images from individual excipients (A, B) in solution. It was determined that excipient B was not generating any SVPs, as measurements of samples containing non-spray-dried excipient B in water showed a close to zero particle count. Excipient A was therefore the component that generated SVPs, and that appeared to be precipitated in the AB Placebo, but with a different particle shape compared to when in the ABC(D) Placebo. Excipient A can be precipitated in different types of crystals. Two of the main ones were available (A_i, A_j) and were tested in the MFI upon reconstitution in water.

Looking at Figure 4.13, similarities between A_i and the ABC Placebo, as well as between A_j and the AB Placebo can be appreciated. The A_i sample and ABC Placebo contains almost exclusively small, rounded, black particles, while the A_j sample and AB Placebo contains both the small, black particles but this time less rounded, as well as brighter, also more oval particles. This was further confirmed by the intensity distributions, presented in Figure 4.14, that showed A_i containing darker particles, and A_j containing brighter particles.

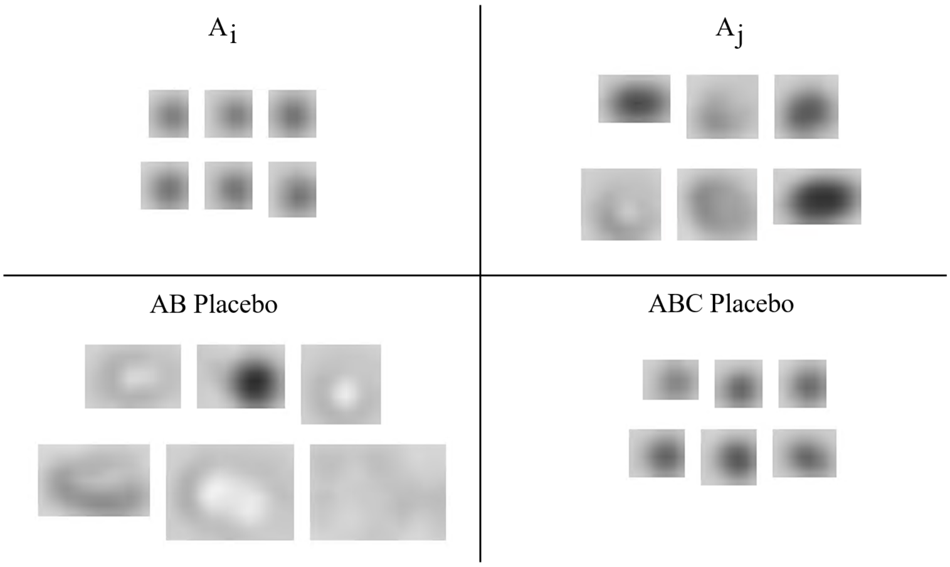
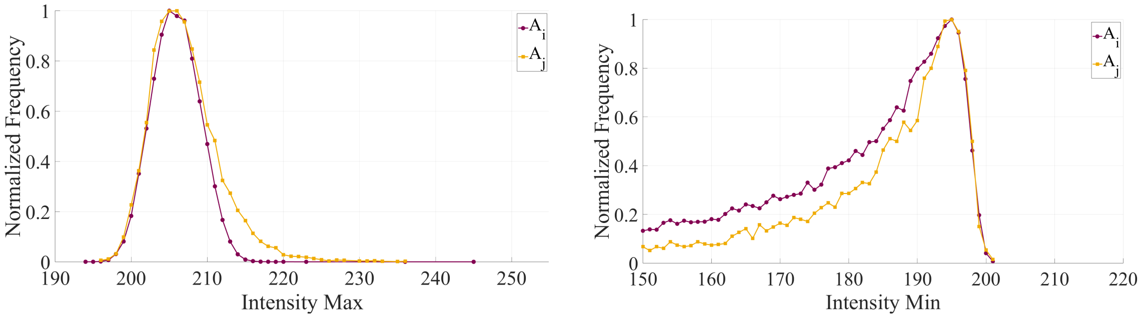


Figure 4.13: Particle images for the A$_i$ and A$_j$ excipients, and AB and ABC Placebos. ABCD Placebo showed a similar appearance to the ABC Placebo, not shown here for simplicity.



(a) Intensity maximum.

(b) Intensity minimum.

Figure 4.14: Intensity maximum and intensity minimum distributions for the A$_i$ and A$_j$ excipients.

Although the same type of crystal (A$_i$) is expected to be formed in both types of placebo (AB, ABC(D)), the findings proposed the possibility of the different placebos containing different crystal types of excipient A. This would explain why different formulations may affect which crystal type appears. In order to confirm that A$_j$ crystals were formed upon reconstitution of the AB Placebo, visible crystals were required for XRPD analysis. To that end, precipitation of the AB Placebo was induced by exposing the sample to thermal stress. XRPD analysis was then performed on the AB Placebo precipitates, which confirmed that the AB Placebo contains A$_j$ crystals. The ABC(D) Placebo is considered to be more stable due to a higher proportion of solutes in the solution, therefore not leading to the same excipient crystal type.

The verification of the different types of crystals supports the above mentioned discussion on the intensity profiles on placebo and DP samples. While more work might be required to further understand the distributions in the ABC and ABCD Placebos and DP samples, the trends for Protein X in AB formulation is understood. The brighter the particles, the lower the protein ratio. The potential to use MFI to identify sub-populations has been shown.

4.5 Revised Correlation of Attributes

Seeing a possibility to differentiate sub-populations using the new light intensity data distributions, PCA was revisited. A new PCA model was constructed using the light intensity data obtained from the custom algorithm, but still using the morphological data provided by the built-in software, as attempts to recalculate these was abandoned.

The scores plot, Figure 4.15 of the new model resembles that of the old model, indicating that the new light intensity data did not provide any clear differentiation. Slight differences can be seen in the loadings plot, however, indicating that the relations between the different parameters have changed. The same significant overlap is still present in the new model. Additionally, the R^2X and Q^2 values of the new model, Figure 4.16, do not show any significant difference in the model's robustness compared to the old one.

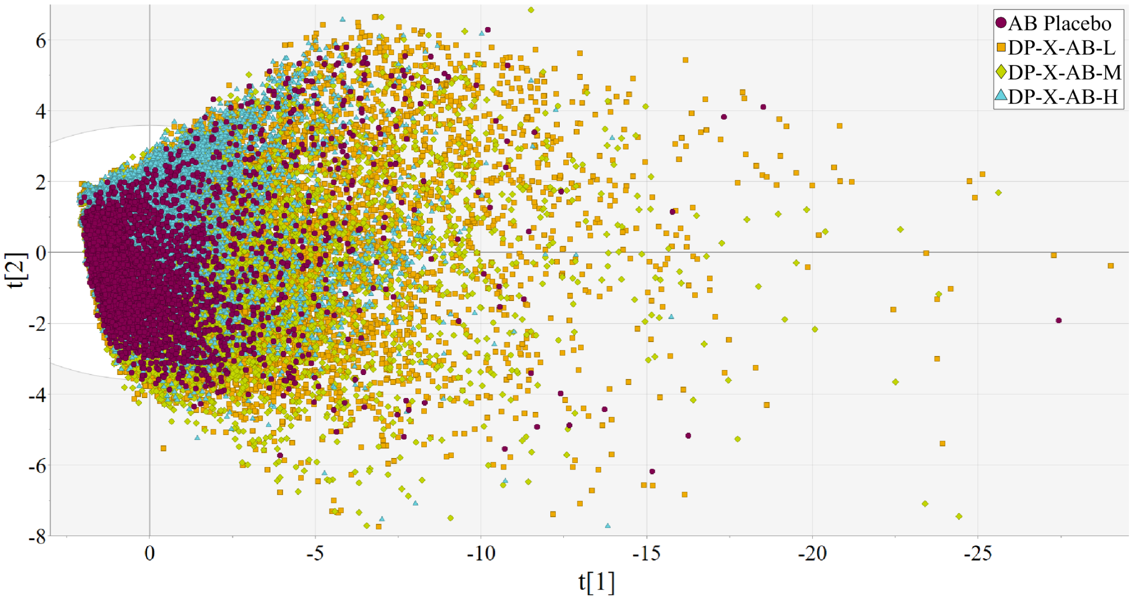


Figure 4.15: PCA scores plot of the model on the data obtained using the custom algorithm, showing the distribution of DP-X-AB samples along with AB Placebo.

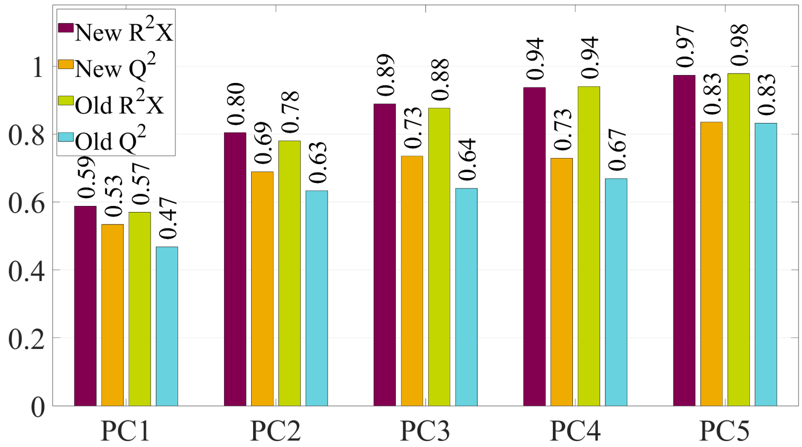


Figure 4.16: A comparison of the cumulative R^2X and Q^2 values for the new model with samples containing DP-X-AB samples and AB Placebo, against the old model.

A slight difference is observed in the new loadings plot, Figure 4.17 compared to that of the old model, Figure A.3. The new model reveals that intensity maximum is closely related to the particle size. While this relationship may not have been initially anticipated, an inspection of individual particle images confirms it - only larger particles exhibit brighter pixel regions, as the camera resolution of the flow cell limits this possibility for smaller particles. The new intensity data also prompted new scatter plots to evaluate relations between parameters. Plots of selected morphological and intensity data against each other is shown in section A.4. Note that these use new intensity data from the custom algorithm, but old morphological data from the vendor's software.

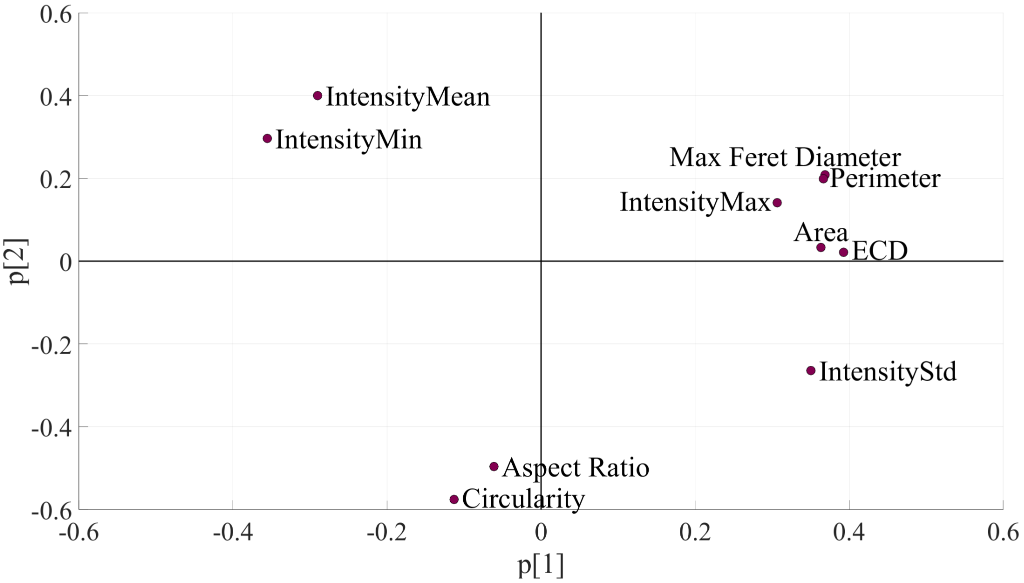


Figure 4.17: PCA loadings plot of the new model with DP-X-AB and AB Placebo samples, indicating how each variable contributes to the first two principal components.

4.6 Future Outlooks

Multiple runs on the same preparation were conducted to obtain more data points, and these showed no significant differences. The obtained data from the MFI instrument could be strengthened by preparing more replicates for each formulation, and subsequently analyzing the images. By doing so, variability within and between the samples can be better assessed. This was already explored and confirmed during previous work. Additionally, analyzing multiple replicates helps ensure that any observed trends are representative of the formulation rather than artifacts of a single sample. However, it is noted that each run detected at least several thousands, upwards of hundreds of thousands of particles, mitigating this risk.

The project was severly limited by the resolution of the camera belonging to the MFI instrument. While machine learning could be a powerful tool, the resolution of the particle images was deemed too low, representing a limitation to explore machine learning and object classification methods. The same goes for Haralick texture features. These methods would work on the larger particles, but the majority of the problems encountered were with the smaller particles, rendering the implementation of machine learning techniques futile. More advanced image analysis and refinements could possibly make machine learning successful.

While PCA should in theory be a straight-forward approach for sub-population differentiation, further exploration of this could enable more precise identification and separation. The main problems lie with the particle detection algorithm - the built-in software was inadequate, and a non-linear relation between pixel size and micrometer size rendered own calculations difficult. Calculating the morphological data is possible if this conversion factor is established, but then again the low resolution of the camera prevents any deeper analysis.

5

Conclusion

This project aimed to enhance MFI technology and exploring its possible usage for differentiation between proteinaceous and non-proteinaceous sub-populations. Different drug product formulations and excipients were analyzed to establish any potential correlation between the data obtained from the MFI and particle characteristics.

A first approach using principal component analysis showed no differentiation between sub-populations. The models created were robust and the parameters were shown to have certain correlations, indicating a possibility of success using this approach. Nevertheless, the models were unable to capture any underlying structure or differences between sub-populations. While exploring a second approach - using intensity data distributions, significant problems with the MFI software and particle detection algorithm were encountered. This prompted the need for a custom particle detection algorithm, basing binary equivalences of the particle images on pixel-wise intensity levels. The algorithm showed success at calculating intensity data, but struggled with obtaining morphological data as the relation between pixel size and micrometer length is non-linear.

Using data obtained from the custom algorithm to construct intensity distribution plots revealed differences between various formulations, specifically between different placebo formulations. Drug product samples also showed variation, although not in order of protein concentration, possibly indicating the presence of more complex relations and interactions. Deeper analysis of the excipients used in the placebos found the reason to be one excipient, which revealed the role that this excipient might have played. The different crystals have distinct characteristics, which can be seen in the images taken by the MFI instrument, which affected the intensity distributions. Advanced characterization of the particles confirmed the findings of the existence of the different excipient crystals in the formulations.

Future work on a more advanced particle detection algorithm would enable more detailed findings. Obtaining accurate morphological data could possibly enable principal component analysis to be used to distinguish sub-populations, not relying only on light intensity data anymore.

Bibliography

[1] G. Mezo, "Peptide and protein based pharmaceuticals," in *Amino Acids, Peptides and Proteins*, E. Farkas and M. Ryadnov, Eds., vol. 38, The Royal Society of Chemistry, Nov. 2013, p. 0, ISBN: 978-1-84973-585-8. DOI: `10.1039/9781849737081-00203`.

[2] C. J. Roberts, "Protein aggregation and its impact on product quality," *Current Opinion in Biotechnology*, vol. 30, pp. 211–217, 2014, ISSN: 0958-1669. DOI: `10.1016/j.copbio.2014.08.001`.

[3] M. Rahban, F. Ahmad, M. A. Piatyszek, T. Haertlé, L. Saso, and A. A. Saboury, *Stabilization challenges and aggregation in protein-based therapeutics in the pharmaceutical industry*, Dec. 2023. DOI: `10.1039/d3ra06476j`.

[4] F. Depreter, G. Pilcer, and K. Amighi, "Inhaled proteins: Challenges and perspectives," *International Journal of Pharmaceutics*, vol. 447, no. 1, pp. 251–280, 2013, ISSN: 0378-5173. DOI: `10.1016/j.ijpharm.2013.02.031`.

[5] T. Martonen and Y. Yang, "Deposition mechanics of pharmaceutical particles in human airways," *LUNG BIOLOGY IN HEALTH AND DISEASE*, vol. 94, pp. 3–27, 1996.

[6] W. H. Finlay, *The mechanics of inhaled pharmaceutical aerosols: an introduction.* Academic press, 2001.

[7] C. Yousry, M. Goyal, and V. Gupta, "Excipients for novel inhaled dosage forms: An overview," *AAPS PharmSciTech*, vol. 25, pp. 1–10, 2 Feb. 2024, ISSN: 15309932. DOI: `10.1208/S12249-024-02741-W/TABLES/2`. [Online]. Available: `https://link.springer.com/article/10.1208/s12249-024-02741-w`.

[8] U. S. P. Convention, "United states pharmacopeia (usp) <787>: Subvisible particulate matter in therapeutic protein injections,"

[9] U. S. P. Convention, "United states pharmacopeia (usp) <788> particulate matter in injections,"

[10] K. L. Ewalt, P. Schimmel, and J. M. Manning, "Protein," *AccessScience*, 2021. DOI: `10.1036/1097-8542.550200`.

[11] S. B. Ebrahimi and D. Samanta, "Engineering protein-based therapeutics through structural and chemical design," *Nature Communications 2023 14:1*, vol. 14, no. 1, pp. 1–11, Apr. 2023, ISSN: 2041-1723. DOI: `10.1038/s41467-023-38039-x`.

[12] S. Bondos and K. Matthews, "Protein folding," *AccessScience*, 2021. DOI: `10.1036/1097-8542.801070`.

[13] K. Ahlgren, *Sweet Stability, An Investigation of Disaccharides as Protein Stabilisers and Amyloid Fibril Inhibitors.* Gothenburg, 2024.

[14] A. V. Finkelstein, N. S. Bogatyreva, D. N. Ivankov, and S. O. Garbuzynskiy, "Protein folding problem: enigma, paradox, solution," *Biophysical Reviews 2022 14:6*, vol. 14, no. 6, pp. 1255–1272, Oct. 2022, ISSN: 1867-2469. DOI: `10.1007/S12551-022-01000-1`.

[15] P. Salahuddin, "Protein Folding, misfolding, aggregation and amyloid formation: Mechanisms of Ab oligomer mediated toxicities," *Journal of Biochemistry and Molecular Biology Research*, vol. 1, no. 2, pp. 36–45, 2015.

[16] M. C. Manning, D. K. Chou, B. M. Murphy, R. W. Payne, and D. S. Katayama, "Stability of protein pharmaceuticals: An update," *Pharmaceutical Research*, vol. 27, no. 4, pp. 544–575, Apr. 2010, ISSN: 07248741. DOI: `10.1007/S11095-009-0045-6`.

[17] N. B. Pham and W. S. Meng, "Protein aggregation and immunogenicity of biotherapeutics," *International Journal of Pharmaceutics*, vol. 585, p. 119 523, Jul. 2020, ISSN: 0378-5173. DOI: `10.1016/J.IJPHARM.2020.119523`.

[18] O. Jonsson, A. Lundell, J. Rosell, S. You, K. Ahlgren, and J. Swenson, "Comparison of Sucrose and Trehalose for Protein Stabilization Using Differential Scanning Calorimetry," *Journal of Physical Chemistry B*, vol. 128, no. 20, pp. 4922–4930, May 2024, ISSN: 15205207. DOI: `10.1021/ACS.JPCB.4C00022`.

[19] O. N. Ógáin, J. Li, L. Tajber, O. I. Corrigan, and A. M. Healy, "Particle engineering of materials for oral inhalation by dry powder inhalers. I—Particles of sugar excipients (trehalose and raffinose) for protein delivery," *International Journal of Pharmaceutics*, vol. 405, no. 1-2, pp. 23–35, Feb. 2011, ISSN: 0378-5173. DOI: `10.1016/J.IJPHARM.2010.11.039`.

[20] M. F. Sun et al., "Effects of polyol excipient stability during storage and use on the quality of biopharmaceutical formulations," *Journal of Pharmaceutical Analysis*, vol. 12, no. 5, pp. 774–782, Oct. 2022, ISSN: 2095-1779. DOI: `10.1016/J.JPHA.2022.03.003`.

[21] J. Bergfreund, S. Siegenthaler, V. Lutz-Bueno, P. Bertsch, and P. Fischer, "Surfactant Adsorption to Different Fluid Interfaces," *Langmuir*, vol. 37, no. 22, pp. 6722–6727, Jun. 2021, ISSN: 15205827. DOI: `10.1021/ACS.LANGMUIR.1C00668`.

[22] W. Wang, S. Nema, and D. Teagarden, "Protein aggregation—Pathways and influencing factors," *International Journal of Pharmaceutics*, vol. 390, no. 2, pp. 89–99, May 2010, ISSN: 0378-5173. DOI: `10.1016/J.IJPHARM.2010.02.025`.

[23] K. Ahlgren, F. Havemeister, J. Andersson, E. K. Esbjörner, and J. Swenson, "The inhibition of fibril formation of lysozyme by sucrose and trehalose," *RSC Advances*, vol. 14, no. 17, pp. 11 921–11 931, Apr. 2024, ISSN: 2046-2069. DOI: `10.1039/D4RA01171F`.

[24] W. Wang, S. K. Singh, N. Li, M. R. Toler, K. R. King, and S. Nema, "Immunogenicity of protein aggregates—Concerns and realities," *International Journal of Pharmaceutics*, vol. 431, no. 1-2, pp. 1–11, Jul. 2012, ISSN: 0378-5173. DOI: `10.1016/J.IJPHARM.2012.04.040`.

[25] P. Mishra and S. K. Jha, "The native state conformational heterogeneity in the energy landscape of protein folding," *Biophysical Chemistry*, vol. 283, p. 106 761, Apr. 2022, ISSN: 0301-4622. DOI: `10.1016/J.BPC.2022.106761`.

[26] K. Teilum, J. G. Olsen, and B. B. Kragelund, "Functional aspects of protein flexibility," *Cellular and Molecular Life Sciences 2009 66:14*, vol. 66, pp. 2231–2247, 14 Mar. 2009, ISSN: 1420-9071. DOI: `10.1007/S00018-009-0014-6`.

[27] U. Food and D. A. (FDA), "Guidance for industry: Immunogenicity assessment for therapeutic protein products," 2014. [Online]. Available: `https://www.fda.gov/media/85017/download`.

[28] G. Kijanka et al., "Submicron Size Particles of a Murine Monoclonal Antibody Are More Immunogenic Than Soluble Oligomers or Micron Size Particles Upon Subcutaneous Administration in Mice," *Journal of Pharmaceutical Sciences*, vol. 107, no. 11, pp. 2847–2859, Nov. 2018, ISSN: 0022-3549. DOI: `10.1016/J.XPHS.2018.06.029`.

[29] M. Ahmadi et al., "Small amounts of sub-visible aggregates enhance the immunogenic potential of monoclonal antibody therapeutics," *Pharmaceutical Research*, vol. 32, no. 4, pp. 1383–1394, Oct. 2015, ISSN: 1573904X. DOI: `10.1007/S11095-014-1541-X/FIGURES/4`.

[30] Y. Guo, H. Bera, C. Shi, L. Zhang, D. Cun, and M. Yang, "Pharmaceutical strategies to extend pulmonary exposure of inhaled medicines," *Acta Pharmaceutica Sinica B*, vol. 11, no. 8, pp. 2565–2584, Aug. 2021, ISSN: 2211-3835. DOI: `10.1016/J.APSB.2021.05.015`.

[31] N. Alhajj, N. J. O'Reilly, and H. Cathcart, "Leucine as an excipient in spray dried powder for inhalation," *Drug Discovery Today*, vol. 26, no. 10, pp. 2384–2396, Oct. 2021, ISSN: 1359-6446. DOI: `10.1016/J.DRUDIS.2021.04.009`.

[32] C. Darquenne, "Deposition Mechanisms," *Journal of Aerosol Medicine and Pulmonary Drug Delivery*, vol. 33, no. 4, pp. 181–185, Jul. 2020, ISSN: 19412703. DOI: `10.1089/JAMP.2020.29029.CD`.

[33] A. J. Hickey, *Inhalation Aerosols*, 2nd Edition, C. Lenfant, Ed. Boca Raton: CRC Press, Oct. 2006, vol. 221, ISBN: 9780429131851. DOI: `10.3109/9781420019537`.

[34] M. Davis and G. Walker, "Recent strategies in spray drying for the enhanced bioavailability of poorly water-soluble drugs," *Journal of Controlled Release*, vol. 269, pp. 110–127, Jan. 2018, ISSN: 0168-3659. DOI: `10.1016/J.JCONREL.2017.11.005`.

[35] R. Vehring, W. R. Foss, and D. Lechuga-Ballesteros, "Particle formation in spray drying," *Journal of Aerosol Science*, vol. 38, no. 7, pp. 728–746, Jul. 2007, ISSN: 0021-8502. DOI: `10.1016/J.JAEROSCI.2007.04.005`.

[36] G. Pilcer and K. Amighi, "Formulation strategy and use of excipients in pulmonary drug delivery," *International Journal of Pharmaceutics*, vol. 392, no. 1-2, pp. 1–19, Jun. 2010, ISSN: 0378-5173. DOI: `10.1016/J.IJPHARM.2010.03.017`.

[37] D. Zillen, M. Beugeling, W. L. Hinrichs, H. W. Frijlink, and F. Grasmeijer, "Natural and bioinspired excipients for dry powder inhalation formulations," *Current Opinion in Colloid & Interface Science*, vol. 56, p. 101 497, Dec. 2021, ISSN: 1359-0294. DOI: `10.1016/J.COCIS.2021.101497`.

[38] T. Sou, L. M. Kaminskas, T. H. Nguyen, R. Carlberg, M. P. McIntosh, and D. A. Morton, "The effect of amino acid excipients on morphology and solid-state properties of multi-component spray-dried formulations for pulmonary delivery of biomacromolecules," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 83, pp. 234–243, 2 Feb. 2013, ISSN: 0939-6411. DOI: `10.1016/J.EJPB.2012.10.015`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0939641112003414`.

[39] A. Nieto-Orellana et al., "Targeted PEG-poly(glutamic acid) complexes for inhalation protein delivery to the lung," *Journal of Controlled Release*, vol. 316, pp. 250–262, Dec. 2019, ISSN: 0168-3659. DOI: `10.1016/J.JCONREL.2019.10.012`.

[40] M. Kiyoshi et al., "Collaborative study for analysis of subvisible particles using flow imaging and light obscuration: Experiences in japanese biopharmaceutical consortium," *Journal of Pharmaceutical Sciences*, vol. 108, pp. 832–841, 2 Feb. 2019, ISSN: 0022-3549. DOI: `10.1016/J.XPHS.2018.08.006`.

[41] V. Corvari et al., "Subvisible (2–100 µm) particle analysis during biotherapeutic drug product development: Part 2, experience with the application of subvisible particle analysis," *Biologicals*, vol. 43, pp. 457–473, 6 Nov. 2015, ISSN: 1045-1056. DOI: `10.1016/J.BIOLOGICALS.2015.07.011`.

[42] R. Strehl et al., "Discrimination between silicone oil droplets and protein aggregates in biopharmaceuticals: A novel multiparametric image filter for sub-visible particles in microflow imaging analysis," *Pharmaceutical Research*, vol. 29, no. 2, pp. 594–602, Feb. 2012, ISSN: 07248741. DOI: `10.1007/s11095-011-0590-7`.

[43] D. K. Sharma, D. King, P. Oma, and C. Merchant, "Micro-flow imaging: Flow microscopy applied to sub-visible particulate analysis in protein formulations," *AAPS Journal*, vol. 12, no. 3, pp. 455–464, Sep. 2010, ISSN: 15507416. DOI: `10.1208/s12248-010-9205-1`.

[44] I. Fawaz, S. Schaz, A. Boehrer, P. Garidel, and M. Blech, "Micro-flow imaging multi-instrument evaluation for sub-visible particle detection," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 185, pp. 55–70, 2023, ISSN: 0939-6411. DOI: `10.1016/j.ejpb.2023.01.017`.

[45] ProteinSimple, *MFI 5000 Series*, 2018. [Online]. Available: `https://www.bio-techne.com/p/micro-flow-imaging-mfi/mfi-5000-series_4000-011-001#product-documents`.

[46] Bio-Techne, *Image analysis guide for particle analysis*, 2019.

[47] ProteinSimple, *Micro-Flow Imaging™ (MFI) IQ/OQ Validation*, Ottawa, Ontario, Sep. 2017.

[48] J. Vanne, M. Viitanen, A. Koivula, and T. D. Hamalainen, "Comparative study of 8 and 10-bit HEVC encoders," *2014 IEEE Visual Communications and Image Processing Conference, VCIP 2014*, pp. 542–545, Feb. 2015. DOI: `10.1109/VCIP.2014.7051626`.

[49] I. F. Domingos, L. B. Carvalho, J. L. Capelo, C. Lodeiro, and H. M. Santos, "Dtt protein equalization and tryptophan protein quantification as a powerful tool in analytical proteomics.," *Journal of Integrated OMICS*, vol. 14, pp. 7–11, 1 Apr. 2024, ISSN: 2182-0287. DOI: `10.5584/JIOMICS.V14I1.229`.

[50] A. B. Ghisaidoobe and S. J. Chung, “Intrinsic tryptophan fluorescence in the detection and analysis of proteins: A focus on förster resonance energy transfer techniques,” *International Journal of Molecular Sciences 2014, Vol. 15, Pages 22518-22538*, vol. 15, pp. 22 518–22 538, 12 Dec. 2014, ISSN: 1422-0067. DOI: `10.3390/IJMS151222518`.

[51] N. Hellmann and D. Schneider, “Hands on: Using tryptophan fluorescence spectroscopy to study protein structure,” *Methods in Molecular Biology*, vol. 1958, pp. 379–401, 2019, ISSN: 19406029. DOI: `10.1007/978-1-4939-9161-7_20/FIGURES/5`.

[52] G. Misra, “Fluorescence spectroscopy,” *Data Processing Handbook for Complex Biological Data Sources*, pp. 31–37, Jan. 2019. DOI: `10.1016/B978-0-12-816548-5.00003-4`.

[53] K. Reinmuth-Selzle et al., “Determination of the protein content of complex samples by aromatic amino acid analysis, liquid chromatography-uv absorbance, and colorimetry,” *Analytical and Bioanalytical Chemistry*, vol. 414, pp. 4457–4470, 15 Jun. 2022, ISSN: 16182650. DOI: `10.1007/S00216-022-03910-1/FIGURES/6`.

[54] C. LibreTexts, *Spectrophotometry.* [Online]. Available: `https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Kinetics/02%3A_Reaction_Rates/2.01%3A_Experimental_Determination_of_Kinetics/2.1.05%3A_Spectrophotometry`.

[55] M. Ermrich and D. Opper, “Xrd for the analyst: Getting acquainted with the principles,” 2011.

[56] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, “Cascade object detection with deformable part models,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2241–2248, 2010, ISSN: 10636919. DOI: `10.1109/CVPR.2010.5539906`.

[57] S. L. Bangare, A. Dubal, P. S. Bangare, and S. Patil, “Reviewing Otsu’s method for image thresholding,” *International Journal of Applied Engineering Research*, vol. 10, pp. 21 777–21 783, 2015. [Online]. Available: `https://www.researchgate.net/profile/Sunil-Bangare-2/publication/282282124_Reviewing_Otsu’s_Method_For_Image_Thresholding/links/57c6e3e408ae9d64047e90dc/Reviewing-Otsus-Method-For-Image-Thresholding.pdf`.

[58] X. Shi, Z. Li, and H. Yu, “Adaptive threshold cascade faster rcnn for domain adaptive object detection,” *Multimedia Tools and Applications*, vol. 80, pp. 25 291–25 308, 16 Jul. 2021, ISSN: 15737721. DOI: `10.1007/S11042-021-10917-W/METRICS`. [Online]. Available: `https://link.springer.com/article/10.1007/s11042-021-10917-w`.

[59] L. Xuan and Z. Hong, “An improved canny edge detection algorithm,” *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, vol. 2017-November, pp. 275–278, Jul. 2017, ISSN: 23270594. DOI: `10.1109/ICSESS.2017.8342913`.

[60] R. M. Haralick, I. Dinstein, and K. Shanmugam, “Textural Features for Image Classification,” *IEEE Transactions on Systems, Man and Cybernetics*,

vol. SMC-3, no. 6, pp. 610–621, 1973, ISSN: 21682909. DOI: `10.1109/TSMC.1973.4309314`.

[61] T. Löfstedt, P. Brynolfsson, T. Asklund, T. Nyholm, and A. Garpebring, "Gray-level invariant Haralick texture features," *PLOS ONE*, vol. 14, no. 2, e0212110, Feb. 2019, ISSN: 1932-6203. DOI: `10.1371/JOURNAL.PONE.0212110`.

[62] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, pp. 3212–3232, 11 Nov. 2019, ISSN: 21622388. DOI: `10.1109/TNNLS.2018.2876865`.

[63] E. Sangineto, M. Nabi, D. Culibrk, and N. Sebe, "Self paced deep learning for weakly supervised object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, pp. 712–725, 3 Mar. 2019, ISSN: 19393539. DOI: `10.1109/TPAMI.2018.2804907`.

[64] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, pp. 2812–2831, 9 Apr. 2014, ISSN: 17599679. DOI: `10.1039/C3AY41907J`.

[65] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, pp. 433–459, 4 Jul. 2010, ISSN: 19395108. DOI: `10.1002/wics.101`.

[66] J. Fick, P. L. Andersson, M. Johansson, and M. Tysklind, "Selection of antibiotics: A chemometric approach," pp. 143–150, 2004.

[67] D. Ripple et al., "Nist special publication 260-193: Reference material 8634 ethylene tetrafluoroethylene for particle size distribution and morphology," DOI: `10.6028/NIST.SP.260-193`.

A

Appendix 1

A.1 Principal Component Analysis

This section presents supplementary figures regarding PCA. Figures below show scores, R^2X and Q^2 values, and loadings for the model based on the data obtained from the built-in software, i.e., before any custom particle detection was applied.

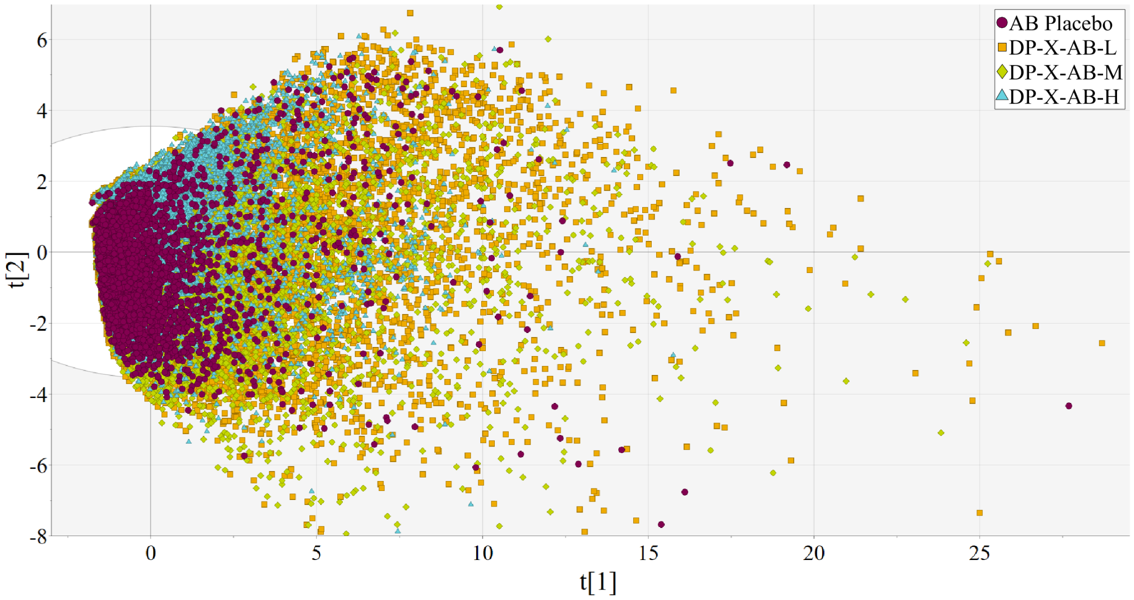


Figure A.1: PCA scores plot of the model on the data from the built-in software, showing the distribution of DP-X-AB samples along with AB Placebo.

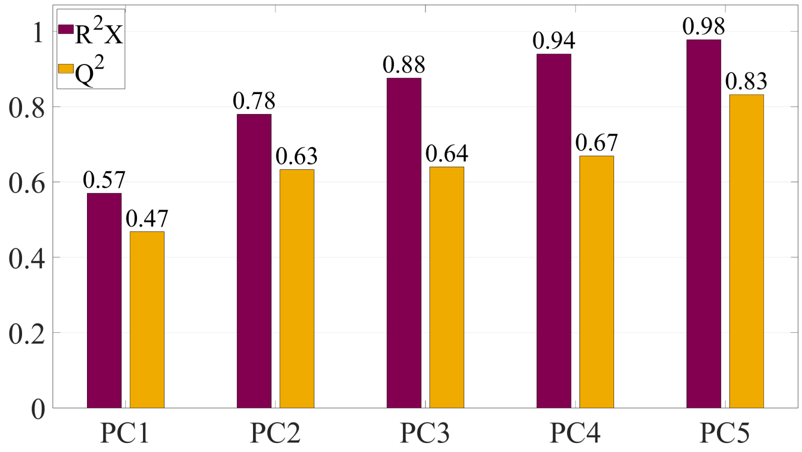


Figure A.2: Cumulative R^{2}X and Q^2 values for model of samples containing DP-X-AB samples and AB Placebo, using the old model.

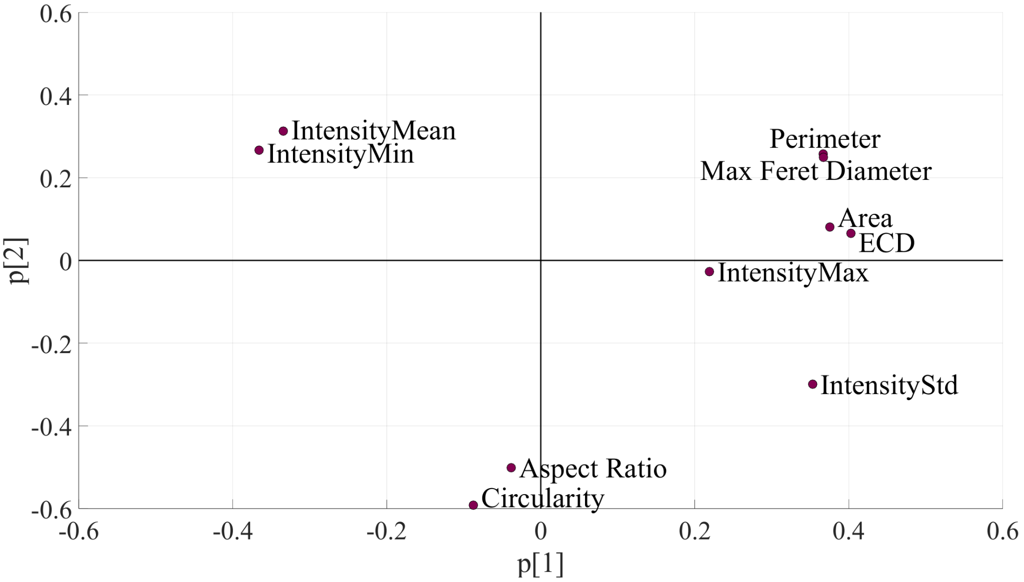


Figure A.3: PCA loadings plot of the old model with DP-X-AB and AB Placebo samples, indicating how each variable contributes to the first two principal components.

A.2 Particle Detection

The background intensity had to be determined in order to do further pixel analysis of the particle images. In each MFI run, the background intensity was adjusted for each sample. To ensure that the images could be run through the same algorithm, histograms of the background pixels were created, shown in Figure A.4. A majority of the backgrounds are centered around 210, while the rest are centered around 211. Only slight background variations is observed. This confirmed that custom background thresholds was not needed for each individual sample.

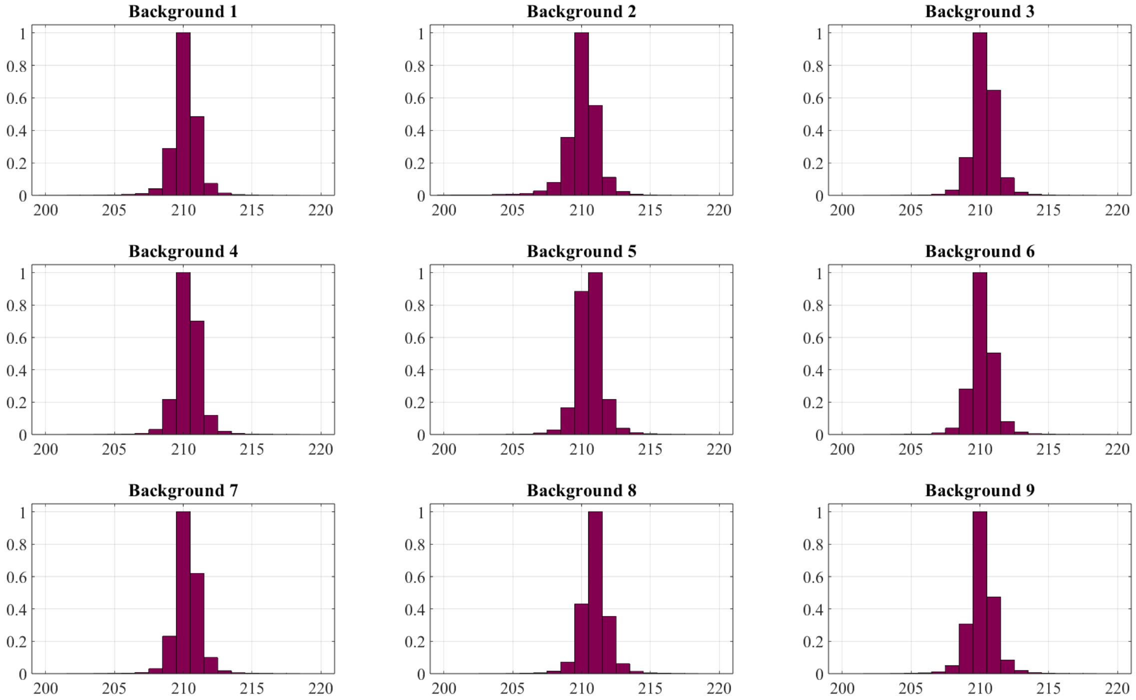


Figure A.4: Normalized histograms for flow cell backgrounds of 9 different samples.

Particle detection on the images provided by the MFI was attempted using different methods, including global thresholds, adaptive threshold, Otsu's threshold and edge detection. Test sets were created in which particles were chosen to span the biggest variety among the particles, taking both bright and dark, as well as particles of varying size. This section shows the results of the different methods, on a set of 20 particles, together with a brief presentation of successes/failures of the methods. All images are created using MATLAB, utilizing different built-in functions that comes with the software.

Figure A.5 shows a fairly successful particle detection, with most particles being clearly discerned, However, there are some isolated pixels adding noise such as in particle 11 and particle 12, and larger areas are not filled in, as in particle 1 and particle 20.

Figure A.6 uses the same threshold levels as on Figure A.5, but with binary masks refinement. This shows a clear improvement in noise levels, as well as correctly filled in particles. The algorithm still struggles with low-contrast particles, such as particle 7, 8 and 12 however.

Figure A.7 has a narrower range for background. Although Figure A.4 shows that these levels should theoretically be fine for the background, the algorithm fails with some basic particle shapes, as in particle 3 and particle 15.

Figure A.8 shows great particle detection, with the basic shapes being mostly accurate. There is, however, significant noise all throughout the images, as the adaptive thresholds is also applied to the background.

Figure A.9 shows overall great success with a few particles being problematic, such as particle 3 and particle 17. The outline is mostly correct, although some bigger parts of the outline is missing, so that not even binary mask refinement could correct it.

Figure A.10 focuses on edges instead, which shows success where there is a clear difference in intensity, such as for particle 2 and particle 17. The detection is however significantly worse for lower-contrast particles, such as particle 7 and particle 8. It would, however, be difficult to continue working with the detected edges, even using binary mask refinement as most of the particles are not completely circled so there is no clear interior.

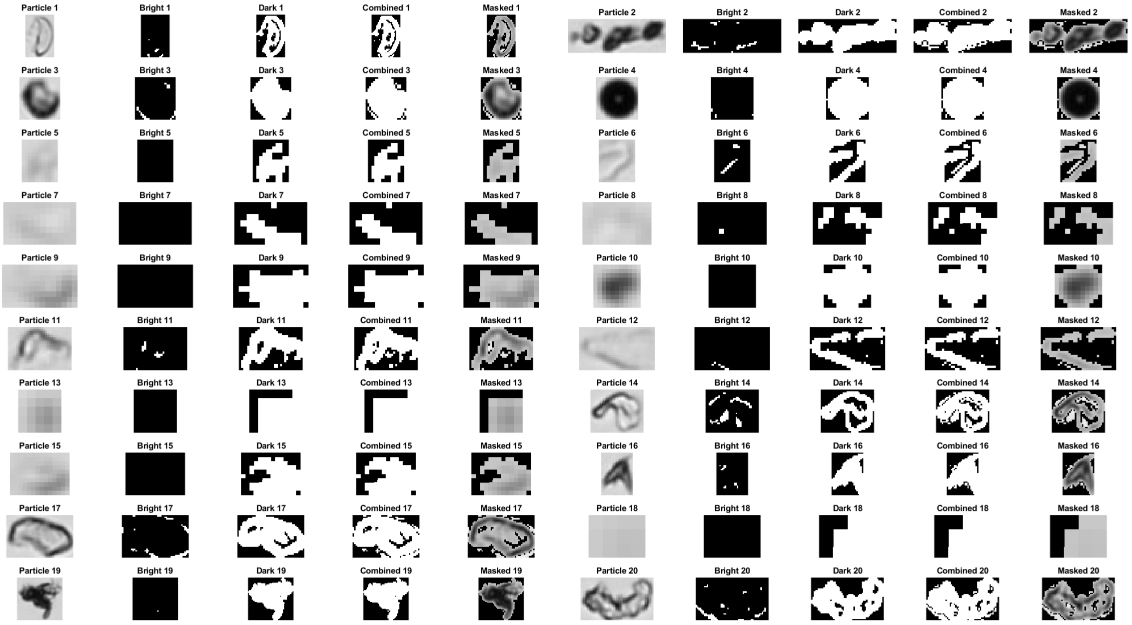


Figure A.5: Particle detection using a dark threshold < 200 and bright threshold > 220, combining these together using no refinement.

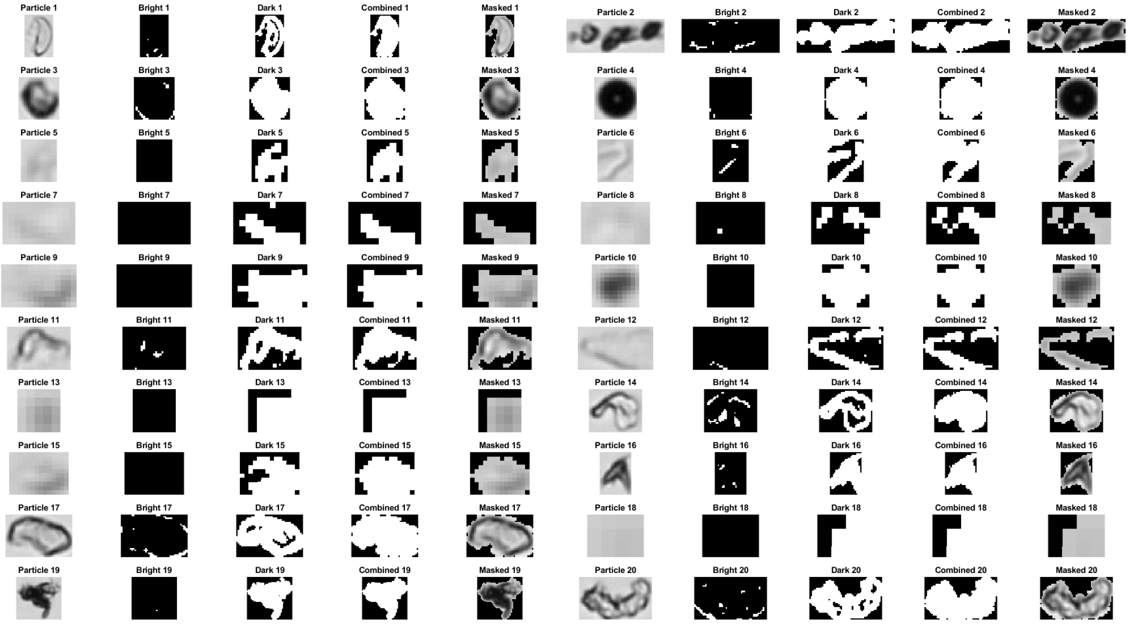


Figure A.6: Particle detection using a dark threshold < 200 and bright threshold > 220, combining these together with refinement using filling holes, bridging and cleaning.

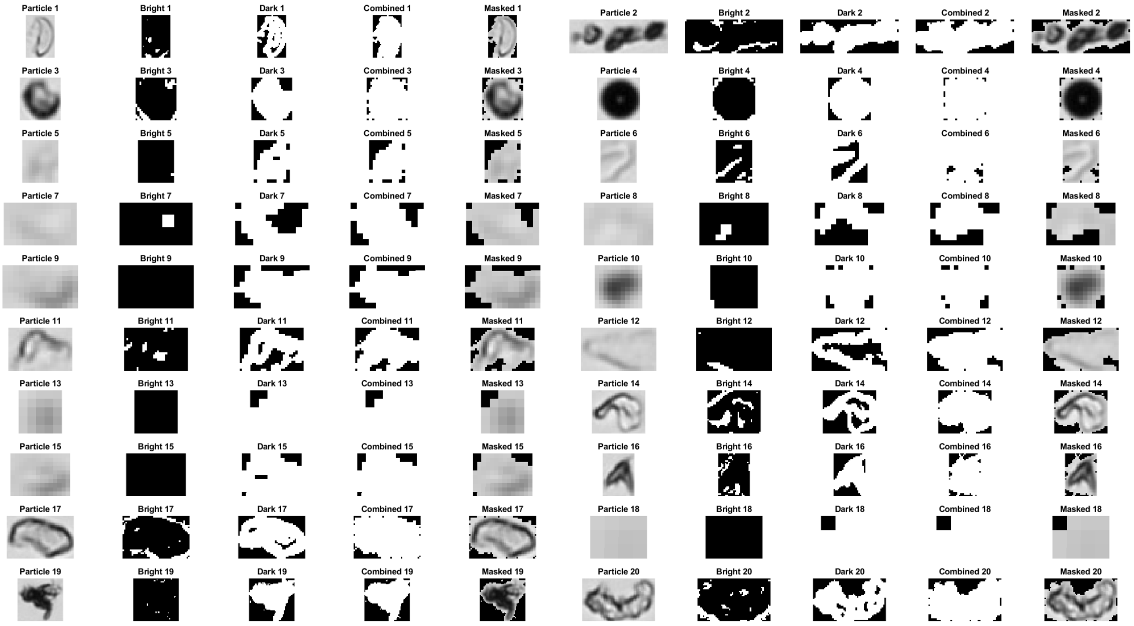


Figure A.7: Particle detection using a dark threshold < 205 and bright threshold > 215, combining these together with refinement using filling holes, bridging and cleaning.

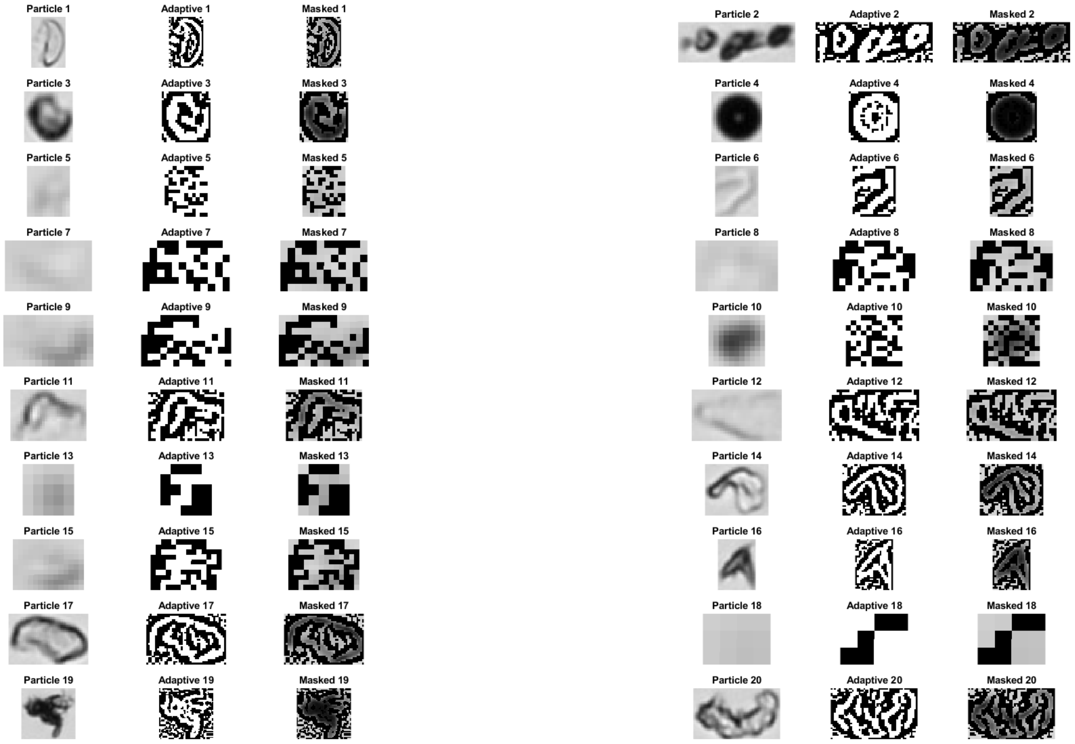


Figure A.8: Particle detection using an adaptive threshold with sensitivity 0.6.

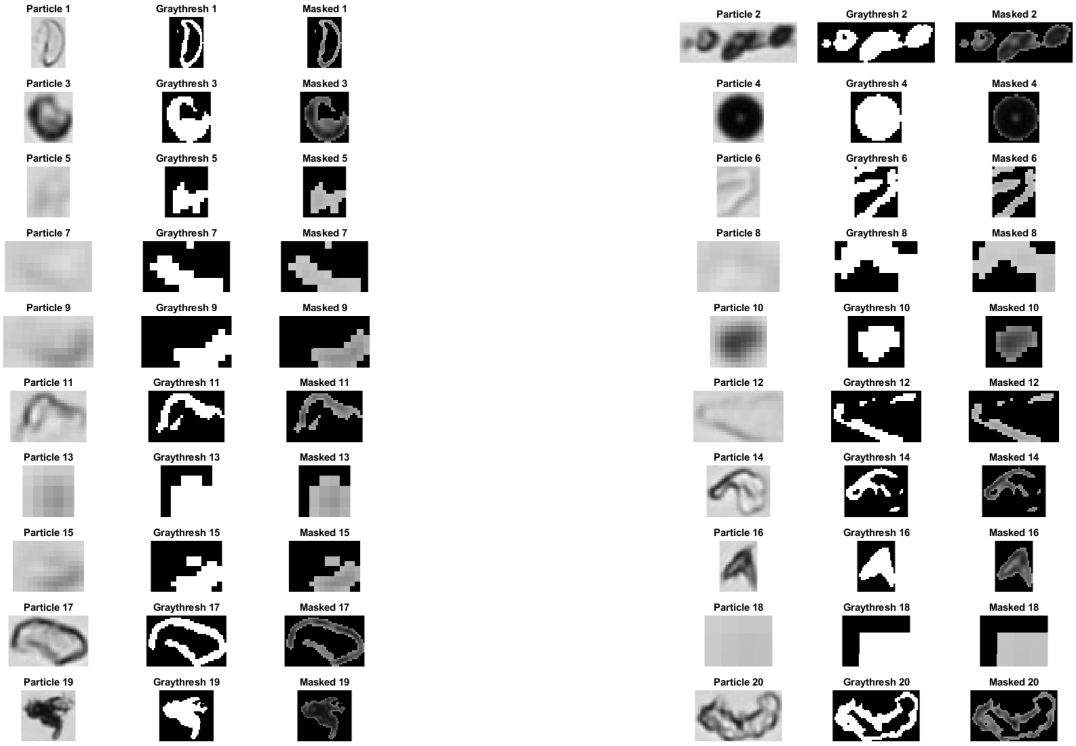


Figure A.9: Particle detection using Otsu's threshold, and masking over the original image.

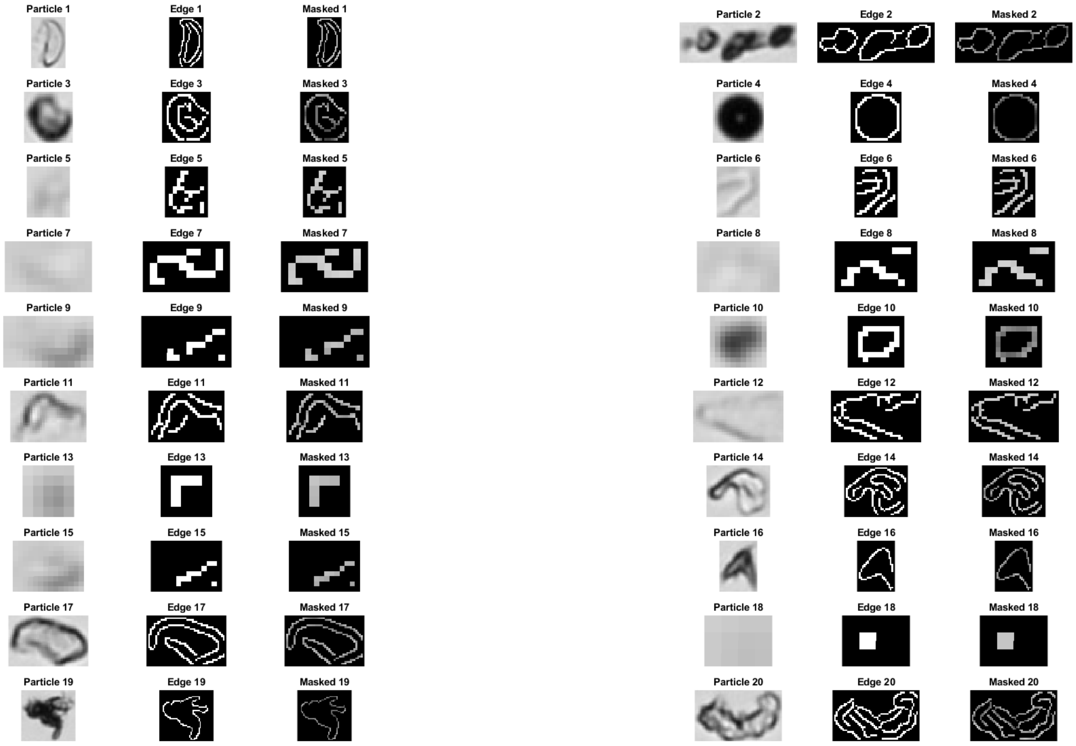
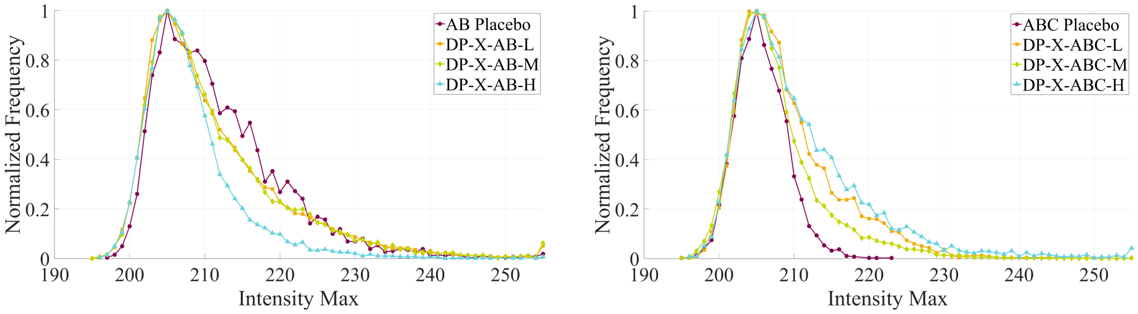


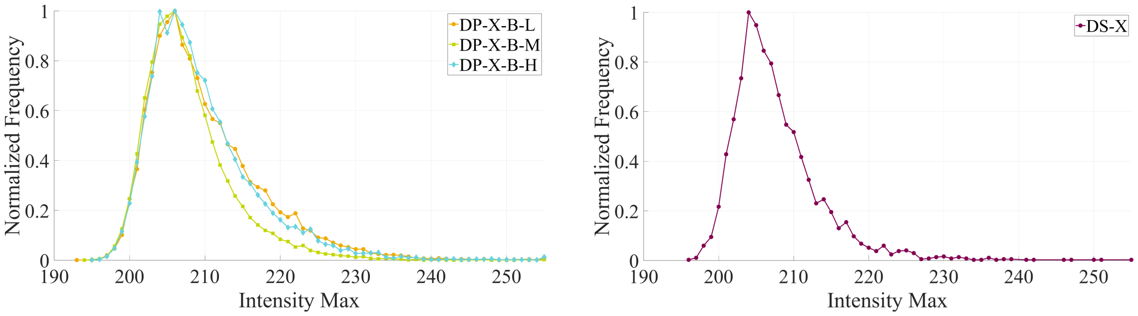
Figure A.10: Particle detection based on edge detection, using the Canny method.

A.3 Intensity Distributions

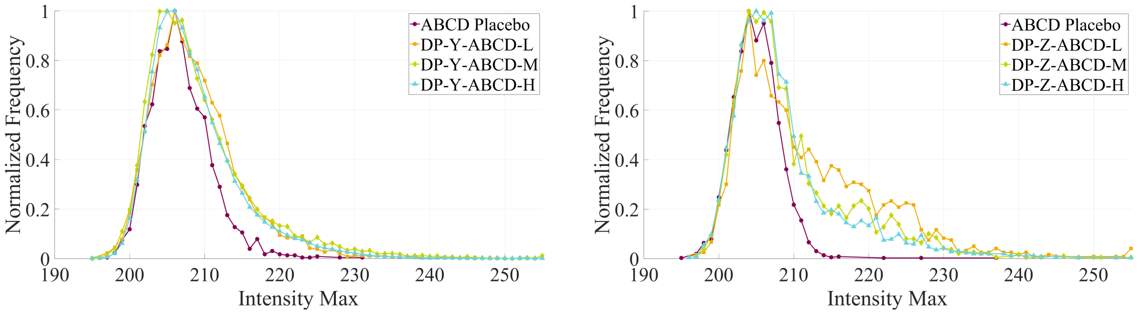
The distributions of intensity maximum and mininum for all analyzed sample series are collected in this section for easier comparisons. The distributions for intensity mean and standard deviation are not included due to the high variability in the graphs, which made it difficult to interpret any meaningful data.



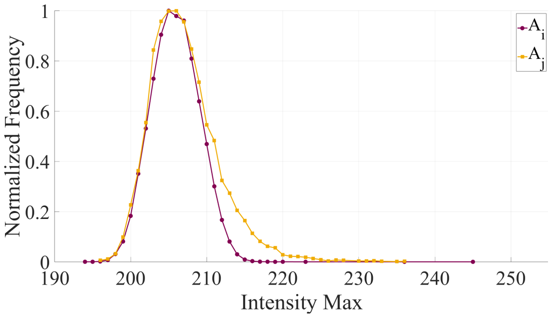
(a) X-AB. **(b)** X-ABC.



(c) X-B. **(d)** DS-X.

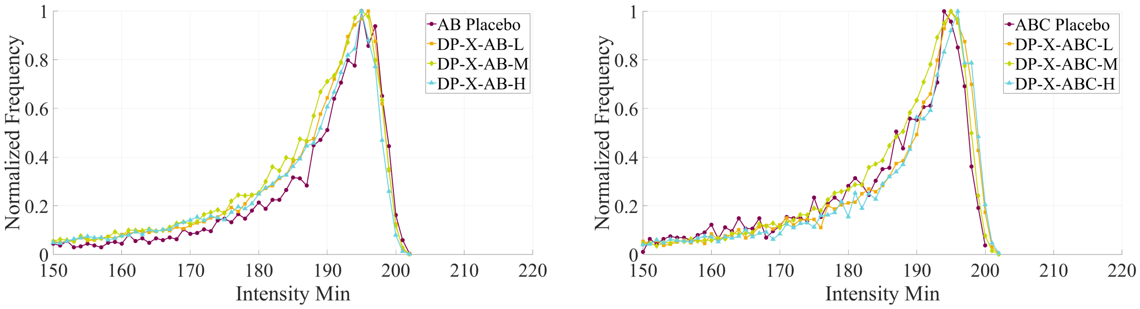


(e) Y. **(f)** Z.

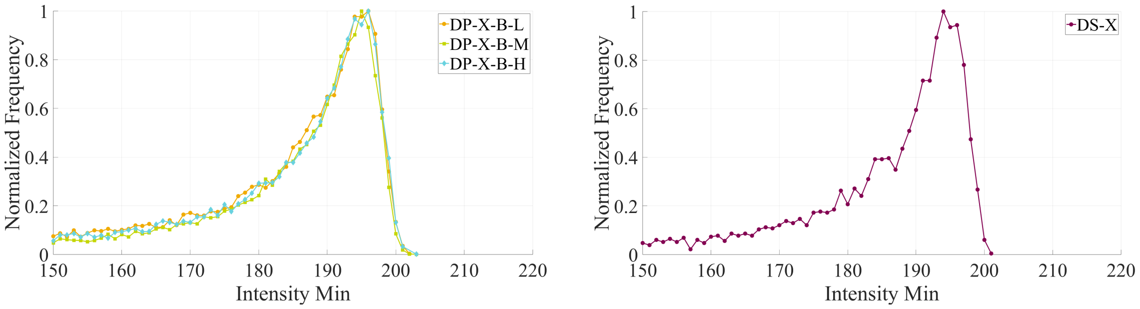


(g) A_i, A_j.

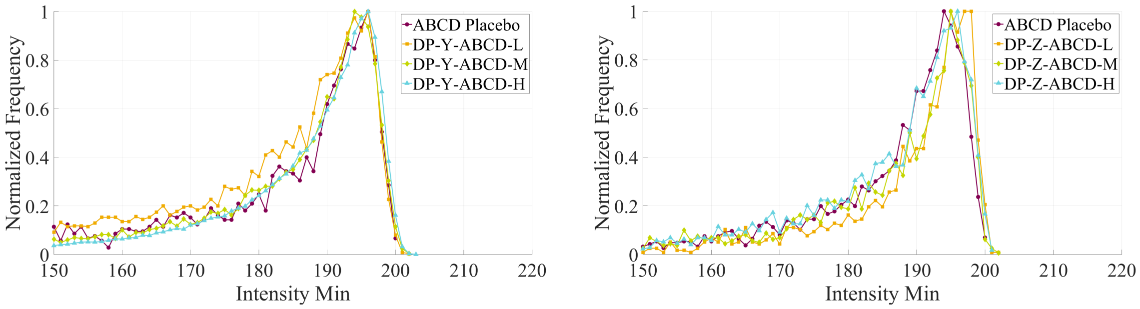
Figure A.11: Distribution plots showing intensity maximum of all particle images, for seven sample categories.



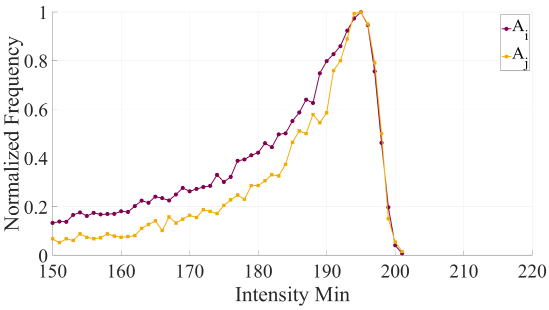
(a) X-AB. **(b)** X-ABC.



(c) X-B. **(d)** DS-X.



(e) Y. **(f)** Z.



(g) A_i, A_j.

Figure A.12: Distribution plots showing intensity minimum of all particle images, for seven sample categories.

A.4 Scatter Plots of Morphological and Intensity Features

This section provides plots of selected morphological and intenstity parameters, using AB Placebo and DP samples of Protein X. It is noted that the four different samples showed significant differences in particle count, affecting the characteristics of the scatter plots.

It is difficult to discern any correlation between the parameters looking at these figures. All comparisons are also made more difficult by the varying particle counts between the samples.

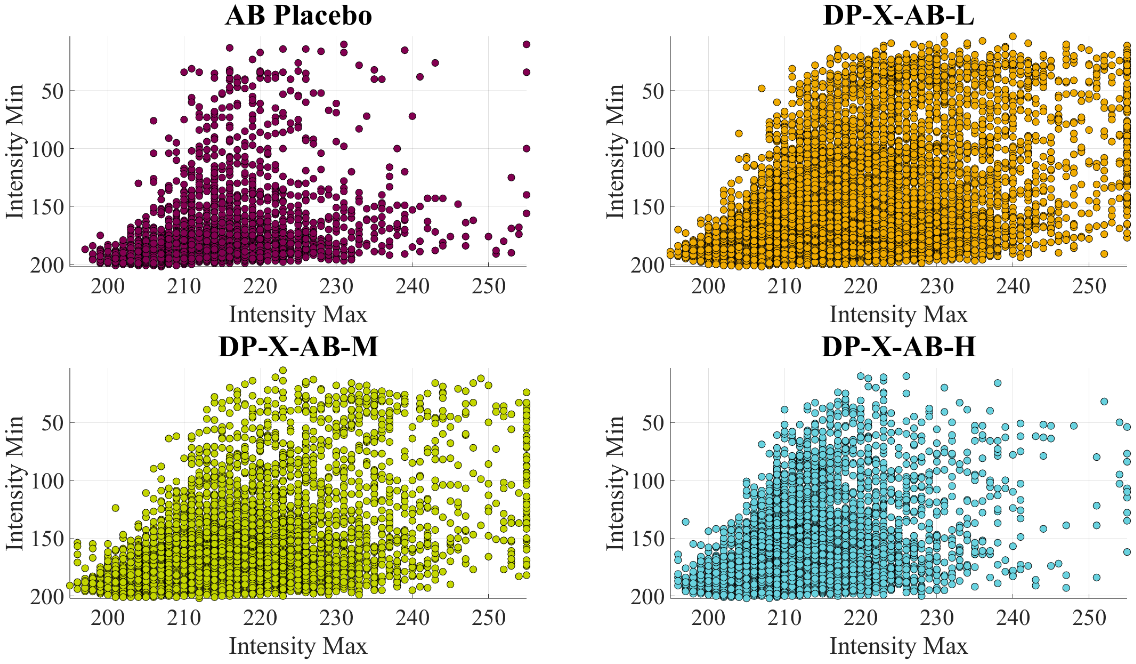


Figure A.13: Scatter plot of DP-X-AB and AB Placebo samples, of intensity maximum against intensity minimum.

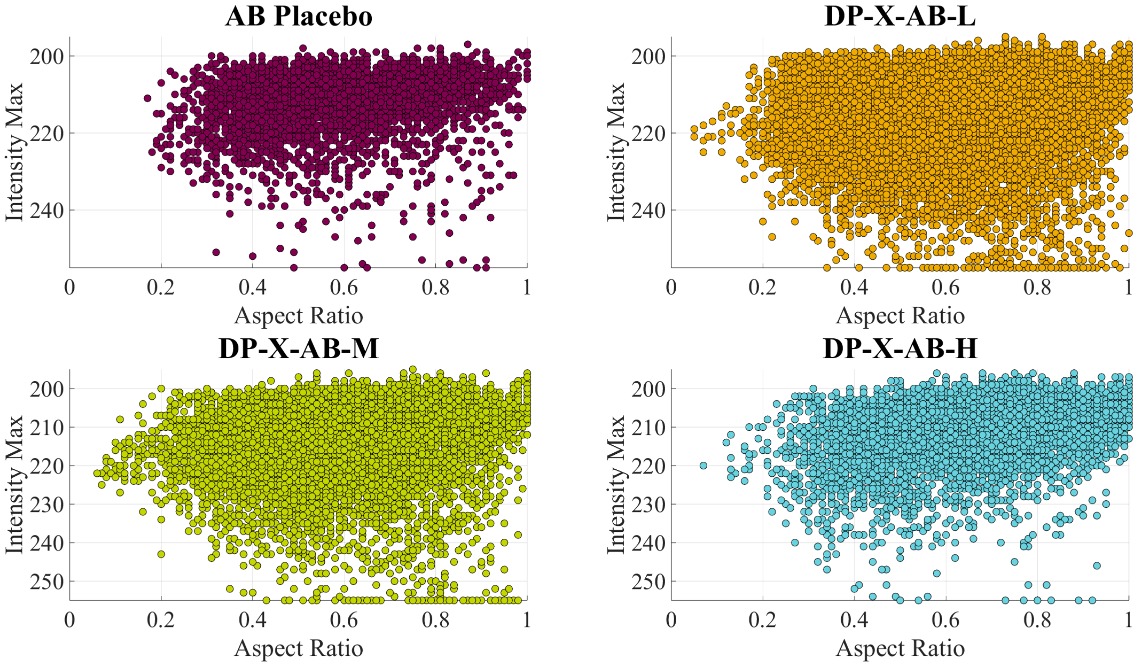


Figure A.14: Scatter plot of DP-X-AB and AB Placebo samples, of aspect ratio against intensity maximum.

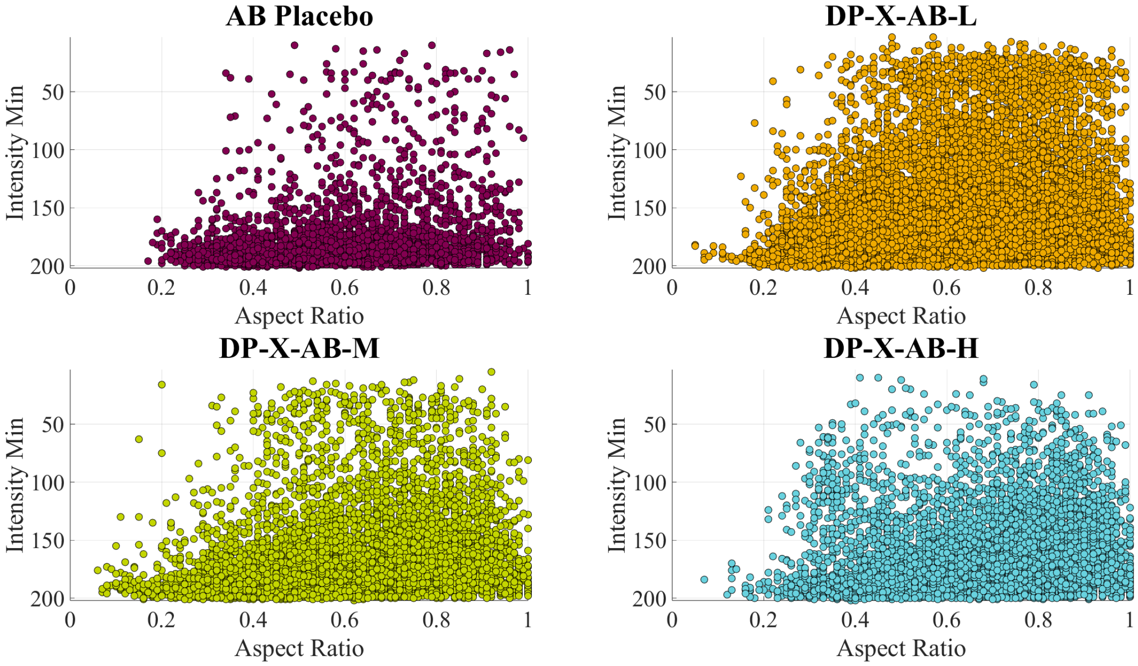


Figure A.15: Scatter plot of DP-X-AB and AB Placebo samples, of aspect ratio against intensity minimum.

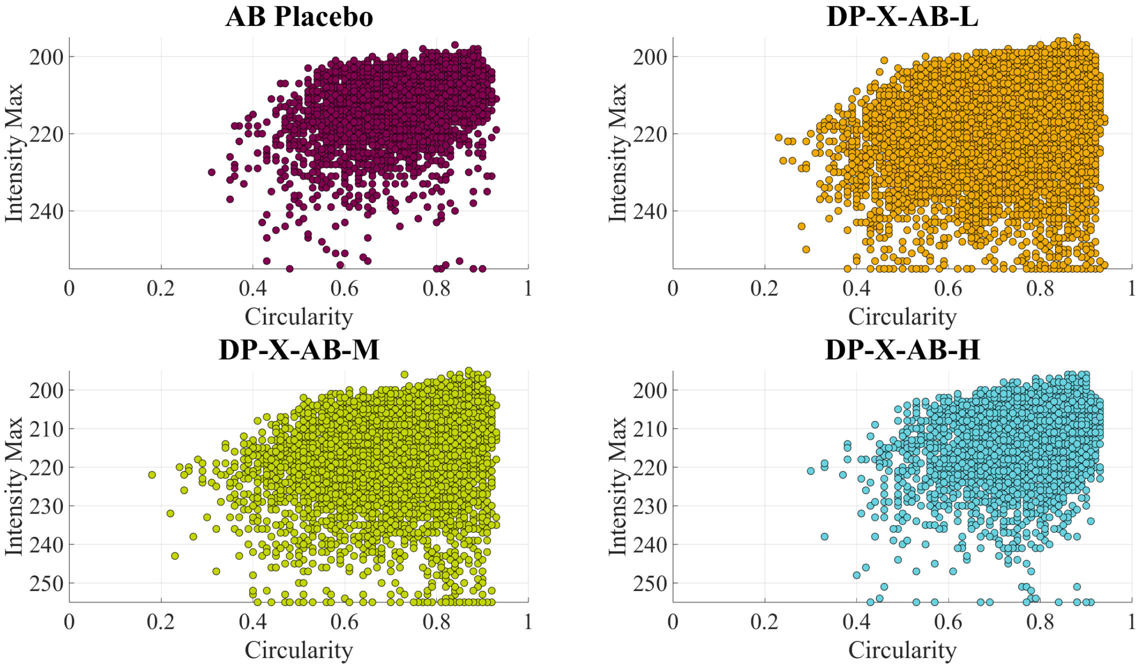


Figure A.16: Scatter plot of DP-X-AB and AB Placebo samples, of circularity against intensity maximum.

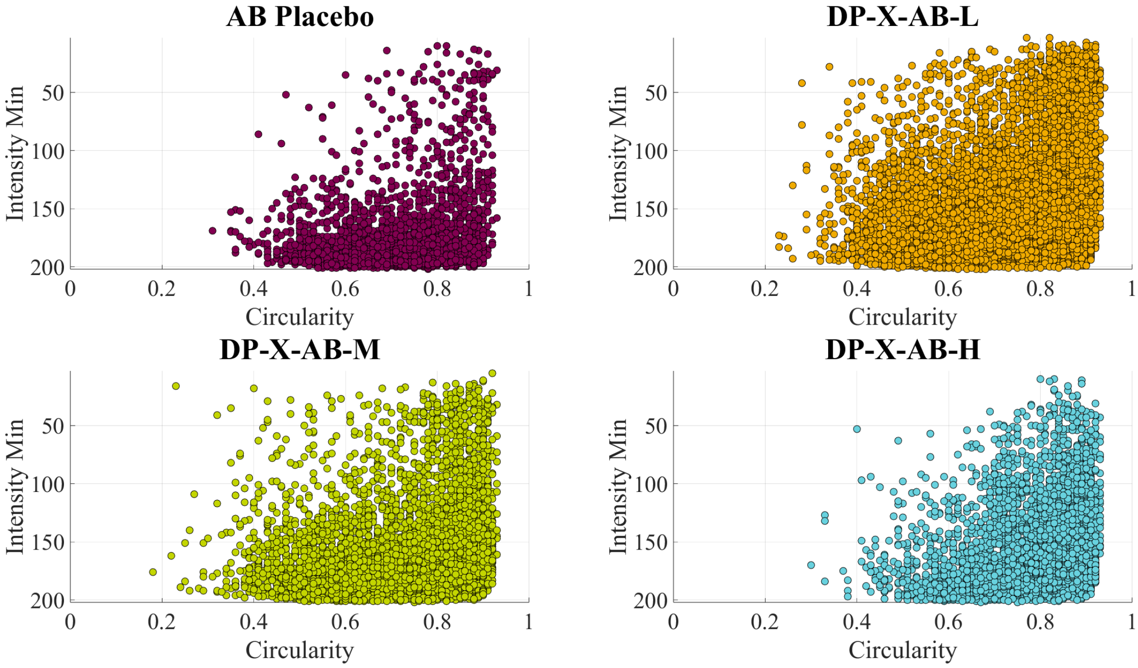


Figure A.17: Scatter plot of DP-X-AB and AB Placebo samples, of circularity against intensity minimum.

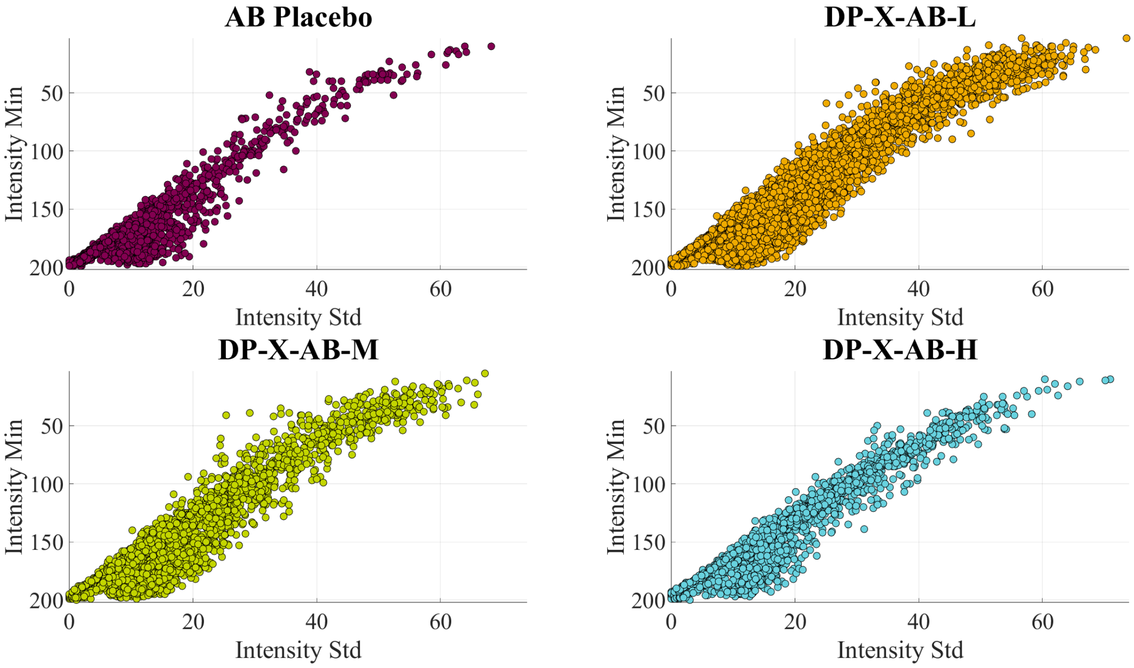


Figure A.18: Scatter plot of DP-X-AB and AB Placebo samples, of intensity standard deviation against intensity minimum.

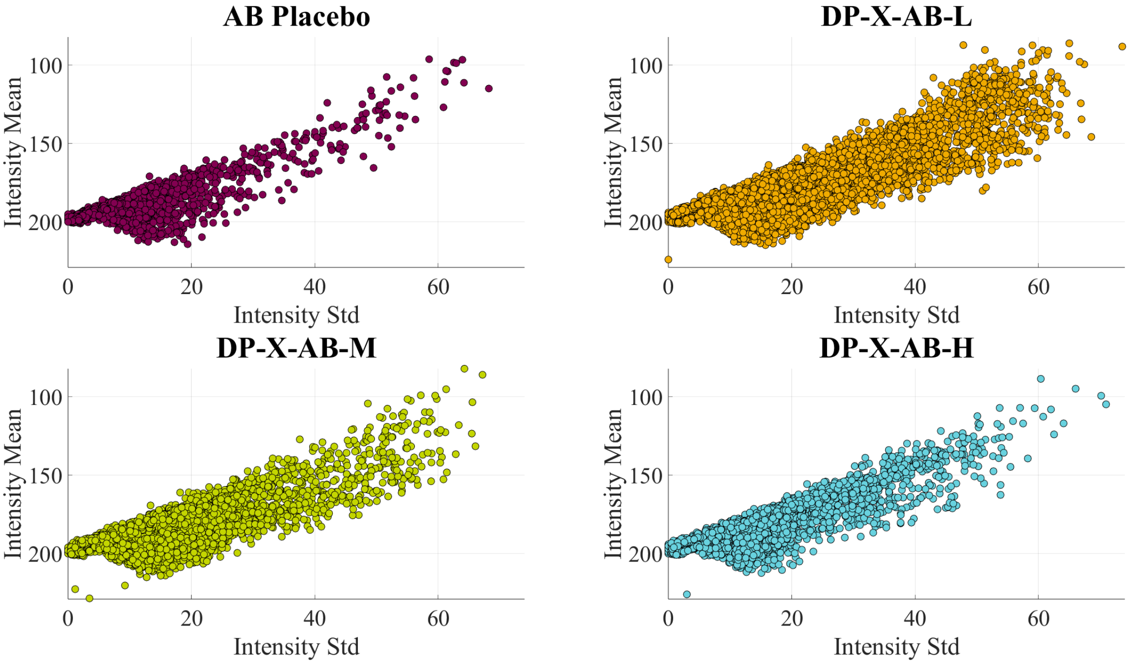


Figure A.19: Scatter plot of DP-X-AB and AB Placebo samples, of intensity standard deviation against intensity mean.

A.5 MATLAB Code

This section provides the MATLAB script used for particle detection and intensity calculation purposes.

```matlab
%% Folder and path setup
imageFolder = 'FOLDER PATH';

% Get all .png files in the folder
imageFiles = dir(fullfile(imageFolder, '*.png'));

% Displays the number of .png files in the folder
disp(['Number of files in this folder: ', num2str(length(imageFiles))]);

%% Particle detection and masking, intensity calculations and Excel export
clc;
tic;

% Prompt for how many images to display
disp('How many images do you want to display?');
disp('"none" - Show no images in folder');
disp('"all" - Show all images in folder');
disp('Or, enter a number to show that many images.');
userInput = input('Your choice: ', 's');
fprintf('----------------------------------------\n');

% Validate user input
if strcmpi(userInput, 'none')
    numToDisplay = 0;
elseif strcmpi(userInput, 'all')
    numToDisplay = length(imageFiles);
else
    numToDisplay = str2double(userInput);
    if isnan(numToDisplay) || numToDisplay < 1
        disp('Invalid input. Defaulting to showing no images.');
        numToDisplay = 0;
    else
        numToDisplay = min(numToDisplay, length(imageFiles));
    end
end

% Prompt for which display mode to use
if numToDisplay > 0
    disp('How do you want to display the images?');
    disp('"s" - Separate. Each image set in its own figure.');
    disp('"c" - Combined. All images in a single figure (subplots).');
    displayMode = input('Your choice: ', 's');

    % Validate user input
    if ~strcmpi(displayMode, 's') && ~strcmpi(displayMode, 'c')
        disp('Invalid input. Defaulting to "combined".');
        displayMode = 'c';
    end
end
fprintf('----------------------------------------\n');

% If using combined mode, determine subplot layout
if strcmpi(displayMode, 'c') && numToDisplay > 0
    rows = ceil(sqrt(5 * numToDisplay));
    cols = ceil((5 * numToDisplay) / rows);
    figure;
end

% Initialize arrays to store intensity statistics
numImages = length(imageFiles);
maxIntensities = zeros(1, numImages);
minIntensities = zeros(1, numImages);
meanIntensities = zeros(1, numImages);
stdIntensities = zeros(1, numImages);
imageNames = strings(1, numImages);

% Prompt for intensity display options
disp('Enter the display option for intensity statistics:');
disp('"all" - Show all statistics');
disp('"max" - Show only max intensity');
disp('"min" - Show only min intensity');
disp('"mean" - Show only mean intensity');
disp('"std" - Show only standard deviation');
disp('"none" - Do not display anything');
disp('"overall" - Show overall mean stats');
userInputStats = input('Your choice: ', 's');

% Validate user input
validStatsOptions = {'all', 'max', 'min', 'mean', 'std', 'none', 'overall'};
if ~ismember(userInputStats, validStatsOptions)
    disp('Invalid input. Defaulting to "none".');
    userInputStats = 'none';
end
fprintf('----------------------------------------\n');
```

```
% Loop through all images
for i = 1:numImages
    % Image reading
    imagePath = fullfile(imageFolder, imageFiles(i).name);
    img = imread(imagePath);

    % Convert the image to grayscale
    % Note that image appears to be in grayscale but is actually stored as RGB
    if size(img, 3) == 3
        img = rgb2gray(img);
    end

    % Single thresholds
    % Otsu's threshold
    %threshold = graythresh(img); % Computes a global, single threshold for the whole image

    % Adaptive threshold
    %threshold = adaptthresh(img, 0.7); % Computes a locally adaptive threshold for each neighborhood

    % Edge detection
    %binaryCombined = edge(img, 'Canny');

    % Bright and dark thresholds
    T_global = 210;
    brightThreshold = T_global + 10;
    darkThreshold = T_global - 10;

    % Detect bright and dark particles
    binaryBright = img > brightThreshold;
    binaryDark = img < darkThreshold;

    % Combine both detections
    binaryCombined = binaryBright | binaryDark;

    % Detection clean-up
    binaryCombined = imfill(binaryCombined, 'holes');
    binaryCombined = bwmorph(binaryCombined, 'bridge');
    binaryCombined = bwmorph(binaryCombined, 'clean');
    binaryCombined = imfill(binaryCombined, 'holes');
    binaryCombined = bwmorph(binaryCombined, 'spur');

    % Create masked image
    maskedImage = uint8(img + 1); % Avoid zero intensity background
    maskedImage(~binaryCombined) = 0; % Set background pixels to black

    % Display images if needed
    if i <= numToDisplay
        if strcmpi(displayMode, 's')
            figure;
            subplot(2,3,1), imshow(img), title('Original Image');
            subplot(2,3,2), imshow(binaryBright), title('Bright Particles');
            subplot(2,3,3), imshow(binaryDark), title('Dark Particles');
            subplot(2,3,4), imshow(binaryCombined), title('Final Combined');
            subplot(2,3,5), imshow(maskedImage), title('Masked Image');
            pause(0.5);
        else
            % Combined display mode
            subplot(rows, cols, 5*i-4), imshow(img), title(imageFiles(i).name, 'Interpreter', 'none');
            subplot(rows, cols, 5*i-3), imshow(binaryBright), title(sprintf('Bright %d', i));
            subplot(rows, cols, 5*i-2), imshow(binaryDark), title(sprintf('Dark %d', i));
            subplot(rows, cols, 5*i-1), imshow(binaryCombined), title(sprintf('Combined %d', i));
            subplot(rows, cols, 5*i), imshow(maskedImage), title(sprintf('Masked %d', i));
        end
    end

    % Get min and max intensity from the original image
    maxIntensities(i) = max(img(:));
    minIntensities(i) = min(img(:));

    % Extract intensity values from masked image (excluding background)
    validPixels = maskedImage(maskedImage > 0); % Take all pixels above level 0 (background)

    if ~isempty(validPixels)
        meanIntensities(i) = mean(validPixels);
        stdIntensities(i) = std(double(validPixels));
    else
        meanIntensities(i) = NaN;
        stdIntensities(i) = NaN;
    end

    % Display processing progress
    fprintf(repmat('\b', 1, length(sprintf('Processing images: %d/%d', i-1, numImages))));
    fprintf('Processing images: %d/%d', i, numImages);

    % Get particle ID from file-name
    fileName = erase(imageFiles(i).name, {'p', '.png'});
    imageNames(i) = fileName;

    % Display intensity statistics based on user choice
    if ~strcmpi(userInputStats, 'none')
        if strcmpi(userInputStats, 'all')
            fprintf('Image: %s\n', fileName);
            fprintf('Intensity max: %d\n', maxIntensities(i));
            fprintf('Intensity min: %d\n', minIntensities(i));
```

```
            fprintf('Mean intensity: %.2f\n', meanIntensities(i));
            fprintf('Standard deviation: %.2f\n', stdIntensities(i));
            fprintf('----------------------------------------\n');
        elseif strcmpi(userInputStats, 'max')
            fprintf('Image: %s - Intensity max: %d\n', fileName, maxIntensities(i));
        elseif strcmpi(userInputStats, 'min')
            fprintf('Image: %s - Intensity min: %d\n', fileName, minIntensities(i));
        elseif strcmpi(userInputStats, 'mean')
            fprintf('Image: %s - Mean intensity: %.2f\n', fileName, meanIntensities(i));
        elseif strcmpi(userInputStats, 'std')
            fprintf('Image: %s - Standard deviation: %.2f\n', fileName, stdIntensities(i));
        end
    end
end

% Compute overall statistics
overallMeanMax = mean(maxIntensities, 'omitnan');
overallMeanMin = mean(minIntensities, 'omitnan');
overallMeanMean = mean(meanIntensities, 'omitnan');
overallMeanStd = mean(stdIntensities, 'omitnan');

% Display overall statistics if user prompt 'Overall'
if strcmpi(userInputStats, 'overall')
    fprintf('\nOverall Mean Intensity Statistics Across All Images:\n');
    fprintf('Mean of Maximum Intensities: %.2f\n', overallMeanMax);
    fprintf('Mean of Minimum Intensities: %.2f\n', overallMeanMin);
    fprintf('Mean of Mean Intensities: %.2f\n', overallMeanMean);
    fprintf('Mean of Standard Deviations: %.2f\n', overallMeanStd);
end

elapsedTime = toc;
fprintf('Processing completed in %.2f seconds.\n', elapsedTime);

% Prompt for Excel exporting wishes
disp('Do you want to export and open the data in Excel?');
excelChoice = input('Note that this will export all intensities regardless of what is displayed. (y/n): ', 's↩
     ');

% Validate user input
if ~strcmpi(excelChoice, 'y') && ~strcmpi(excelChoice, 'n')
    disp('Invalid input. Defaulting to "n" (no export).');
    excelChoice = 'n';
end

if strcmpi(excelChoice, 'y')
    T = table(imageNames', ...
    round(meanIntensities', 2), ...
    round(stdIntensities', 2), ...
    round(minIntensities', 2), ...
    round(maxIntensities', 2), ...
    'VariableNames', {'Particle Id', 'Intensity mean', 'Intensity Std', 'Intensity Min', 'Intensity Max'});

    tempExcelFile = [tempname, '.xlsx'];
    writetable(T, tempExcelFile);
    winopen(tempExcelFile);
end
```

DEPARTMENT OF PHYSICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se