Investigation and Monitoring of Antibody Aggregation in Biopharmaceutical Manufacturing

Mariana Medeira Silva Ressurreição

Thesis to obtain the Master of Science Degree in

Biological Engineering

Supervisor: Prof. José Cardoso Menezes,

Prof. Massimo Morbidelli

Examination Committee

Chairperson: Prof. Arsénio do Carmo Sales Mendes Fialho

Supervisor: Prof. José Monteiro Cardoso de Menezes

Member of the Committee: Prof. Duarte Miguel De França Teixeira dos Prazeres

November 2018

Preface

The work presented in this thesis was performed at the Institute of Chemistry and Bioengineering of ETH Zürich (Switzerland), during the period February-October 2018, under the supervision of MSc. Fabian Feidl and MSc. Ruben Wälchli. The thesis was co-supervised at Instituto Superior Técnico by Prof. José Cardoso Menezes.

Acknowledgements

First and foremost, I would like to thank Professor Morbidelli for the opportunity to work in his research group here at ETH Zürich.

I would also like to express my deepest gratitude to my supervisors, Ruben Wälchli and Fabian Feidl. Ruben for always being present and available for my never ending questions and for teaching more than I could ask for. Fabian for the enormous trust he put on me and his endless enthusiasm about the work, which helped me keep pushing forward.

I also want to thank my lab companion Francesca Fanizzi who has been, since the beginning, a constant in my life here, and all the other students who became important people in my life in the past 9 months, namely Caterina, Fabio, Matevz, Nicolo and Matteo.

I would also like to thank Sebastian Vogg for making the DSP side of the biolab a less boring place to work in and the whole Morbidelli group for how good it felt to belong here for these months.

Many thanks to my family of course, for their never ending support.

At last, I would like to express my regards to Prof. José Cardoso Menezes for supporting my work as my official supervisor at IST Lisbon and for presenting this opportunity to me.

Abstract

In monoclonal antibody (mAb) manufacturing, protein aggregates constitute one of the most relevant critical quality attributes, and their levels must be controlled thoroughly. This work focuses both on understanding the mechanism of aggregation and assessment of Raman spectroscopy as a tool for online aggregate quantification.

To that end, two mAbs were investigated. First, aggregation behaviour under pH-shift stress conditions was studied to learn what promotes aggregation during the low pH virus inactivation step in downstream processing (DSP). Further, a flow-cell was developed for integration in the chromatographic effluent of the final polishing step in DSP. Further, the ability to quantify antibody aggregates with Raman technology was tested.

Electrostatic repulsion between antibody molecules prevents mAb coagulation at pH values between 2.5 and 3.5 and low ionic strength. Formation of aggregates occurs during and after neutralization to higher pH and ionic strength values. Besides these two parameters, time of incubation and presence of additives also affect the process of aggregation at pH 3.5. The velocity at which the neutralization step is conducted does not influence aggregation.

Smoothing techniques for spectral data, such as Wavelet transform and Empirical Mode Decomposition, were investigated and compared to Savitzky-Golay, providing better results. Prediction of monomer concentration under process relevant ranges was achieved with good results. Although aggregates were not successfully quantified in process relevant ranges (<15%), information in increased ranges (2-60%) was predicted with relative root mean square error of 8%, laying the foundation towards online monitoring of aggregate concentration with Raman technology.

Keywords: monoclonal antibody; aggregation behaviour; Raman spectroscopy; chemometrics

Resumo

Na produção de anticorpos monoclonais (mAb), aggregados proteicos exigem um controlo rigoroso. Este trabalho foca-se em compreender os mecanismos de agregação de anticorpos, e avaliar a espectroscopia Raman como ferramenta de quantificação de agregados em tempo real num processo.

Dois mAbs foram estudados. O comportamento de agregação sob condições de *stress* impostas por mudanças de pH foi investigado para perceber o que promove agregação durante o passo de inactivação viral durante o *Downstrem Processing* (DSP). Uma célula de fluxo foi desenvolvida para integração no efluente cromatográfico do passo de purificação final do DSP. Finalmente, foi testado o potencial da espectroscopia Raman para quantificação de agregados de mAbs.

Entre pH 2.5 e 3.5 e baixa força iónica, repulsão electrostática entre as moléculas previne a coagulação dos mAb. Formação de agregados ocorre quando a solução é neutralizada para valores de pH e força iónica mais altos. Para além destes parâmetros, tempo de incubação e presença de aditivos também afetam o processo de agregação, quando a incubação é a pH 3.5.

Foram investigadas técnicas de suavização para dados espectrais e comparadas ao filtro *Savitzky-Golay*, apresentando melhores resultados. A previsão da concentração de monómero com um modelo de regressão apresentou bons resultados. A previsão da concentração de agregados numa gama relevante para o processo não foi possível. No entanto, numa gama mais alargada, o modelo de regressão conseguiu prever a mesma com um erro relativo de 8%, apresentando o primeiro alicerce para a possibilidade de monitorização *online* de agregados de mAbs com tecnologia Raman.

Palavras-chave: anticorpo monoclonal, comportamento de agregação, espectroscopia Raman, quimiometria

Contents

List of Figures

List of Tables

Nomenclature

1/k – Debye's length

A^- - Base compound

a – Ion activity

A, B – Temperature dependent coefficients

A_2 – Osmotic second virial coefficient

AH – Acid compound

AI/III – Amide I/III

ANS – 8-Anilino-1-naphthalenesulfonic acid

Arg – Arginine

Asp – Aspartic acid

BR – Background removal

C – Ion concentration

C – Ion dependent constant

CA – Cluster analysis

CCD – Charge coupled device

Coif – Coiflet wavelet function

CP – Centre point

C_{pol} – Concentration of macromolecules

CQA – Critical quality attributes

CV – Cross validation

Cys – Cysteine

d – Diameter of a hydrated ion

D – Diffusion coefficient

db – Daubechies wavelet function

DLS – Dynamic light scattering

DoE – Design of experiments

Dopt – D-optimal

DSP – Downstream processing

E – Electric filed

e – Electron charge

EMD – Empirical mode decomposition

F – Electrostatic force

FDA – Food and Drug Administration

fr – Frame length

FT – Fourier transform

FTIR – Fourier-transform infrared

FWHM – Full width at half maximum

Glu – Glutamic acid

H^+ - Proton

HCP – Host cell protein

His – Histidine

HPLC – High performance liquid chromatography

I – Ionic strength

IEX – ion-exchange chromatography

Ig – Immunoglobulin

IMF – Intrinsic mode function

IPA – Isopropanol

IR – Infrared

K – Optical constant, stoichiometric equilibrium constant

K_a – Thermodynamic equilibrium constant

LDA – Linear Discriminant Analysis

LHS – Latin Hypercube Sample

LOWESS - Locally weighted scatterplot smoothing

LV – Latent variables

Lys – Lysine

mAb – Monoclonal antibody

MALS – Multi-angle light scattering

M_w – Molecular weight

MWCO – Molecular weight cut-off

n – solvent refractive index

N_A – Avogadro's number

NIR – Near-infrared

ord – derivative order

P – PCA loadings

PAT – Process analytical technologies

PC – Principal component

PCA – Principal Component Analysis

pI – Isoelectric point

PLS – Partial Least Squares regression

pol – Polynomial order

Q – Charge

Q^2 – Coefficient of determination

r – Distance between to charges

relRMSECV – Relative root mean square error in CV

R_g – Radius of gyration

R_h – Hydrodynamic radius

RMSEP – Root mean square error in prediction

R_θ – Excess Rayleigh ratio

s – Seconds

SEC – Size-exclusion chromatography

SG – Savitzky-Golay

SLS – Static light scattering

SNR – Signal to noise ratio

SNV – Standard Normal Variate

Sym – Symmlet wavelet function

T – PCA scores

Tyr – Tyrosine

UV – Ultraviolet

VI – Viral inactivation

VWD – Van der Waals

w – Free energy

WT – Wavelet Transform

X – Data matrix

Y – Vector of dependent variables

z – Ionic valence

γ – Activity coefficient

ε – dielectric permittivity

η – Viscosity

θ – Scattering angle

λ – Wavelength of incident light in the medium

λ_0 – Vacuum wavelength

1 Introduction

1.1 Biopharmaceutical Manufacturing of Therapeutic Proteins

Due to their broad application range, therapeutic proteins are used worldwide in a variety of applications and are currently the fastest growing segment in biopharmaceutical industry. Increasing product demand for these proteins translates in the need of more efficient and economic processes while at the same time improvement in antibody quality [1].

1.1.1 Therapeutic Antibodies

Monoclonal antibodies (mAbs) are immunoglobulin (Ig) molecules that mimic the function of a natural Ig in the body, which is binding to pathogens or foreign molecules to neutralize and clear said entities. Igs are glycoprotein molecules produced by white blood cells. They act as a crucial part of the immune system response. MAbs function is either to bind to extracellular targets, cells or pathogens to neutralize them, block their function and remove them from circulation, or to mimic their activity. Thus, they are considered pharmaceutically active agents intended to modify a disease state by pharmacologically altering components of the disease [2].

As monoclonal antibodies, they can have monovalent affinity, binding always to the same antigen, recognized by the antibody. They are grouped into five classes, based on the sequence of their heavy chain constant regions: IgM, IgD, IgG, IgE and IgA. Of the five classes, IgG, or altered forms of IgG, are the most frequently used in immunotherapy, and will be the molecules studied in this work. Figure 1 presents the general structure of an IgG, which is composed of two heavy (H) and two light (L) chains, both comprising constant (C) and variable (V) regions. The Fc domain is composed only of heavy chains with constant regions, while the Fab domain of both heavy and light chains with constant and variable regions that give antigen specificity [3].

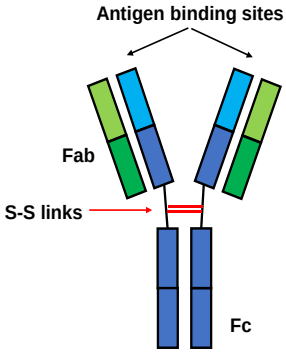

Figure 1 – General scheme of an Immunoglobulin G. Heave chains are represented in blue and light chains in green; dark and light represent the constant and variable chains, respectively. The hinge region is represented between the two domains, red being the disulphide bonds.

An IgG has a molecular weight of approximately 150 kDa and 4 subclasses described in human, mouse and rat: IgG1, IgG2, IgG3 and IgG4. They differ in the number of disulphide bonds and the length and flexibility of the hinge region, the flexible amino acid stretch in the central part of the heavy chains [4].

1.1.2 State of the Art on Biopharmaceutical Manufacturing Downstream Processing

Changes from batch to continuous processing have started to take place in the pharmaceutical industry both for small molecules and large biomolecules. The main motivators for this phenomenon are a decrease of production and operation costs as well as higher and constant product quality. Although the best manufacturing strategy changes for each scenario, the potential of continuous processing for the manufacturing of therapeutic proteins has been recognized. The large demand for these therapeutics

coupled with the large costs of Good Manufacturing Practices production spaces demand for more efficient technologies in the sector [5].

Downstream Processing

The harvest stream/cell-free supernatant obtained from cell culture in the bioreactor contains, besides the target compound (i.e. the monomeric antibody), a great variety of other substances. Those impurities include cell-culture media components, host cell proteins (HCPs), DNA, viruses as well as antibody aggregates. Since content of each impurity has to be within certain specificities (i.e. defined by regulatory authorities), it is necessary to resort to multiple purification steps based on a variety of separation principles such as affinity binding, charge or size to obtain the final product under specification [6]. This process is called downstream processing (DSP) of therapeutic antibodies and contains mainly preparative chromatography.

First step after cell culture is selective capture of the target protein, usually performed through Protein A affinity chromatography. The product binds specifically to the stationary phase and is later recovered by elution under different mobile phase conditions. The principle of Protein A chromatography lies in the specific interaction between the Fc region of antibodies and Protein A, a cell wall component of *Staphylococcus aureus*: a highly conserved histidyl residue in the centre of the protein A binding site of Igs pairs with a complementary histidyl residue on Protein A.

In the process, cell culture harvest is loaded onto the column at neutral pH, where binding between antibody molecules and protein A molecules immobilized on the surface of the column's stationary phase occurs. After a wash step at intermediate pH, designed to remove HCPs and other contaminants, product is eluted from the column by switching to an acidic mobile phase (i.e. pH between 3.0 and 4.1). At low pH, both amino acid residues involved in the specific antibody-resin interaction carry a positive charge, which disrupts the binding and releases the target from the stationary phase [6][5][7].

Viral inactivation (VI) aims at the inactivation of enveloped viruses by denaturation of their lipid envelope. Possible methods include exposure to low pH, detergents as well as incubation at elevated temperatures [8]. VI by low pH incubation is commonly implemented in antibody production after protein A capture, since the antibody elutes at a pH lower than 4 from the column, which is necessary for disruption of virus envelopes. Protein A product fractions are pooled, and solution pH might be further reduced by addition of an appropriate quantity of acid, followed by incubation (duration usually in the range between 30 and 120 minutes). At the end of the incubation, the antibody solution is neutralized by addition of base. Target pH depends on requirements of the subsequent purification step and is commonly mildly acidic (i.e. around pH 5) [9]. One important aspect in VI by low pH incubation is to ensure that pH is homogeneous across all fluid elements and that all of them experience sufficient residence time at low pH before neutralization [6].

Following VI, one or multiple polishing steps are performed to bring final product composition within specifications. This involves removal of residual process-related impurities (e.g. remaining HCPs, DNA, leached Protein A, etc.) as well as product-related contamination such as charge variants and antibody

aggregates. Aggregates are undesirable as they lack biological activity and can increase the risk of an immune response from the patient against the active protein monomer [10] [11].

For polishing, either ion-exchange or hydrophobic interaction chromatography are commonly used. Those chromatographic steps are conveniently operated in flow-through or frontal mode. For that purpose, the target compound should interact minimally with the stationary phase and pass through the column while impurities are retained by the stationary phase. High purity levels can be achieved by using two or more columns in series [5][7].

1.1.3 Antibody Aggregation in Downstream Processing

Most common and significant source of aggregate formation in DSP is the VI [12]. When incubated at low pH, proteins tend to denature, i.e. alter their conformation, which leads to exposure of hydrophobic residues previously buried within the protein's interior. If (partially) denatured protein molecules are unable or slow to refold back to their native conformation afterwards, hydrophobic side chains of different protein molecules may interact, leading to formation of dimers and higher order oligomers [13].

Even though it is well-known that acid-stress can cause protein aggregation, details of the process are still unclear to some extent. For a successful assembly of continuous DSP, profound knowledge about each unit operation is critical. Regarding low pH VI, impact of process parameters on antibody stability and aggregation behaviour merit further investigation for potential reduction of product loss.

1.2 Forces and Mechanisms Underlying Protein Aggregation

Protein aggregation designates the process by which proteins assemble into stable complexes composed of two or more monomeric units, with the biologically active form of the protein denoted as monomer. Within aggregates, protein molecules interact through strong, non-covalent forces, requiring some degree of conformational distortion or unfolding in order to expose amino acids residues necessary to form these strong contacts. Aggregation propensity comes firstly from the protein's amino acid sequence, but it is also dependent on the protein's environment (i.e. solution pH, ionic strength, protein concentration) [11].

1.2.1 Protein Structure and Molecular Conformation

Proteins consist of linear polymers of α-amino acids (called polypeptides) linked by the formation of amides between the α-carboxyl group of one amino acid and the α-amino groups of adjacent amino acid, as depicted in Figure 2. The sequence of the covalently bonded amino acids is known as a protein's primary structure. The secondary protein structure is essentially equal to the three-dimensional configuration of the polypeptide backbone. The two most abundant structural motifs are α-helices and β-sheets. The overall three-dimensional arrangement of the polypeptide chain(s), including the residues' side-chains, is referred to as tertiary structure. Further, the supra-molecular

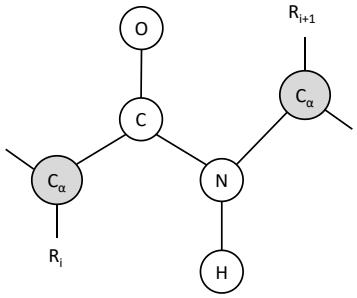

Figure 2 – Representation of the peptide bond between two amino acids (i and i+1). Each amino acid residue in a protein has a different side-chain attached to the α-carbon, represented by R.

association state of individual protein molecules is referred to as quaternary structure. The secondary, tertiary, and quaternary protein structure, which is most populated under physiological conditions and is responsible for biological activity is referred to as *native state* [14][13].

Naturally, proteins exist in aqueous environments and consequently the majority of non-polar side chains are assembled into hydrophobic cores on the inside of the protein structure, reducing the interfacial area with the surrounding solvent. These cores in the protein centre have high packing densities, suggesting a resemblance between the protein's interior and a crystal. As for charged groups, it is energetically unfavourable to remove them from water and so almost all of these are located at the protein's surface. They increase protein solubility and help direct the folding process of a polypeptide by staying in solution [15].

Under normal conditions, proteins in their native state are in equilibrium with a small amount of folding intermediates, which are further in equilibrium with the completely unfolded/denatured states. Folding intermediates generally present incorrect folding or a few domains in the unfolded state, and evidence suggests that they constitute precursors for the aggregation process, since these intermediates have typically more of their hydrophobic side-chains exposed and possess higher flexibility relative to the folded state [13] [16].

1.2.2 Interactions Relevant to Physical Protein Stability

Structure and stability of proteins depend on molecular scale interactions, namely: (i) van der Waals interactions between backbone and side chain atoms; (ii) electrostatic attractions between amino acids with opposite charges; (iii) electrostatic repulsions between amino acids with the same charge; (iv) hydrophobic attractions between nonpolar side chains; (v) loss or gain of chain entropy due to folding or unfolding; (vi) hydrogen bonding between backbone atoms [11].

Van der Waals Forces

Van der Waals (VDW) forces emerge from fluctuation in the electric dipole moments of molecules, which become correlated as the molecules come closer together, giving rise to an attractive force. There is always a VDW force between any two molecules or surfaces. It can be either attractive or repulsive, being that it is always an attractive force between similar molecules. Nevertheless VDW forces are relatively weak compared to other binding forces, such as ionic, H-bonding and hydrophobic that, when present, dominate the interaction [17].

Electrostatic Forces

Electrostatic force between two charged molecules or side chains is essentially given by Coulomb's law (equation (1.1)), representing the electric field E_1 of a charge Q_1 acting on a second charge, Q_2, at a given distance r, where ε is the dielectric permittivity of the medium.

$$F(r) = Q_2 E_1 = \frac{Q_1 Q_2}{4\pi\varepsilon_0 \varepsilon r^2} \tag{1.1}$$

From this relation, the free energy for the Coulomb interaction between Q_1 and Q_2 is obtained through equation (1.2), z standing for ionic valence, where the reference state of zero energy is represented by $r = \infty$.

$$w(r) = \int_{\infty}^{r} -F(r)dr \; = \frac{z_1 z_2 e^2}{4\pi\varepsilon_0 \varepsilon r} \tag{1.2}$$

The expression on the right is used for ionic interactions in aqueous solutions. Magnitude and sign of each ionic charge is given in terms of the elementary electron charge (e = 1.602 x 10^{-19} C) multiplied by z. For charges of equal sign, both *F(r)* and *w(r)* are positive and the force is repulsive, whereas for opposite charges both are negative, and the force is attractive. The force necessary to break an ionic bond is F(r)=w(r)/r, which is similar to the force needed to break a covalent bond. As such, it is quite a strong and of long-range interaction. However, in a solution containing dissolved salts and buffer species, the Coulomb force on a charge particle often involves many close and distant charges, both positive and negative, and so both attractive and repulsive. The actual net electrostatic interaction on a body is more complex and of shorter-range than the form of Coulomb's law suggests [18].

Electrostatic interactions of molecules in solution can be described at thermodynamic equilibrium by the second-order differential Poisson-Boltzmann equation. When solved, it gives the chemical potential, at any point in the gap between two surfaces. It neglects nonidealities such as ion-ion correlations or the effect of the size of the ions and assumes a continuum of charges in solution and on the surface [19]. The factor 1/k, known as the Debye length, expresses the characteristic decay length of the chemical potential and depends only on a solution's characteristics, namely on the temperature and electrolytes present [20]. The Debye length falls with increasing ionic strength and valence of the ions in solution, which can be interpreted as, for a given salt, the magnitude of k increases as salt concentration increases [11].

Only amino acids with acidic or basic side chains will be involved in electrostatic interactions and the majority of such residues are located at the protein's surface. Some of these are aspartic acid (Asp), glutamic acid (Glu), histidine (His), cysteine (Cys), tyrosine (Tyr), lysine (Lys) and arginine (Arg) [13]. Changing the pH changes the net charge and charge distribution on a protein, which consequently changes the inter- as well as intra-molecular repulsions and attractions [18].

Hydrophobic Effect

Generally, addition of a non-polar molecule or residue to water reduces the system's entropy, which is thermodynamically unfavourable. This decrease is commonly explained by the structuring of water molecules in the vicinity of hydrophobic entities to avoid interaction [13]. Intra-molecular hydrophobic interactions are often considered the driving force for protein folding. Non-polar residues are hidden in the inside of the proteins where they can efficiently avoid contact with water. However, when partially or fully unfolded, proteins expose their hydrophobic side-chains to the solvent [13]. As consequence, the hydrophobic effect can also lead to aggregation of denatured proteins, since association of hydrophobic residues from different molecules also decreases interfacial area with the surrounding water molecules [13][21].

Hydrogen Bonds

Hydrogen bonds occur when a hydrogen atom is shared between two electronegative atoms. In native proteins, although hydrogen bonds are abundant in the secondary structure, they are not a dominant factor in maintaining conformational integrity. The conformational stability of almost all globular proteins

is in the range of 5-20 kcal/mol, equivalent to 3 to 5 hydrogen bonds. Hypothetically considering a protein with 100 amino acids, and thus approximately 150 hydrogen bonds, the breakage of 3 among these most likely will not unfold the protein [15].

1.2.3 Critical Process Parameters in Low pH Viral Inactivation

The two key process parameters in the VI are undoubtedly duration and solution pH. Time is important because, on one hand, the process needs to be long enough to achieve a robust virus reduction, and, on the other, the longer the incubation at low pH, the higher the risk of product loss [9]. As for the pH of incubation, it must be below 4 to guarantee the destruction of the lipid envelopes of the viruses but the process value will depend on the aimed viral clearance. As such, viral inactivation needs to last long enough at a certain acidic pH to guarantee an effective reduction in viral activity.

Concerning the effect on the product, acidic pH can potentially induce partial or complete unfolding of the proteins. It will dictate the distribution of surface charges on proteins, affecting intra and intermolecular protein-protein interactions. At extreme values, proteins are heavily charged and a dense charge population on the proteins' surfaces can significantly increase repulsive interactions, leading to at least partial unfolding. Aggregation under these conditions will depend on the relative contribution of inter-molecular hydrophobic attraction and electrostatic repulsion [16].

Strongly related to pH is ionic strength of the protein solution. Most biological molecules have evolved to work under physiological conditions, including a certain constancy of pH and ionic strength. If the solution has an ionic strength different from the physiological state, then the protein's stability may change [22]. Ions in solution will interact with the proteins, leading to altered charge-charge interactions and possibly different behaviour of aggregation [16].

Furthermore, protein concentration may also influence protein behaviour towards aggregation in different ways: (i) decrease aggregation due to the macromolecular crowding effect, which favours the formation of compact conformations (i.e. limits (partial) denaturation); (ii) increase aggregation rates by increasing the frequency of molecular collisions; (iii) precipitation due to solubility limit of the protein [16].

1.2.4 Analytical Techniques for Characterization and Quantitation of Antibody Aggregates

Various methods can be used to obtain information about a protein's secondary, tertiary, and quaternary structure and it is often necessary to combine information from various techniques to validate conclusions. Some of the most common methods are optical spectroscopy, electrophoresis, and high-performance liquid chromatography (HPLC). Among these, spectroscopic and HPLC techniques were used in this work for the aggregation studies, in particular size-exclusion chromatography (SEC), light scattering and extrinsic fluorescence.

Size-exclusion Chromatography

Size-exclusion techniques are usually performed on stationary phases with suitable pore-size distributions allowing different analytes to be excluded to different extents from the pore volume of the packing [23]. As a results, SEC allows for separation of monomeric and aggregated antibody molecules on the basis of differences in hydrodynamic size. The higher order oligomers will be the first to elute as

they will not enter any of the pores, followed for lower order oligomers and finally by the monomer, that will be held for the longest time in the stationary phase. Coupled with an UV absorbance detector, it provides information regarding the concentration of the different species in the analysed sample.

Static Light Scattering

Static light scattering (SLS) provides a direct measure of the molecular weight (M_w) of biopolymers in solution through measurements of the scattered intensity at multiple angles (and analyte concentrations). Additionally, SLS gives information concerning molecular size in terms of the radius of gyration, R_g [24]. The R_g of a polymer chain is commonly defined as the root-mean-square distance of the segments of the molecule from its centre of mass. Use in combination with SEC, which serves as a pre-fractionation step, can increase the information content extracted from an SLS measurement since light scattering provides an averaged molecular weight of the molecules present inside the scattering volume. Being that the aim is to characterize the antibody aggregates, with this technique it is possible to analyse them separately from the monomers.

Analysis of light scattering data is commonly performed based on the Zimm equation (1.3) [24], where R_θ is the excess Rayleigh ratio (proportional to the intensity of light scattered by the investigated macromolecules/particles), c_{pol} the concentration of the macromolecules/particles, λ the wavelength of the incident light in the medium, θ the scattering angle and K the optical constant given by equation (1.4). In the latter, N_A is Avogadro's number, λ_0 the vacuum wavelength, n the solvent refractive index and $\frac{dn}{dc_{pol}}$ the refractive index increment in the polymer solution. A_2, the osmotic second virial coefficient, which provides a measurement of interactions between the solute molecules, can be determined from a batch-mode experiment [25].

$$\frac{Kc_{pol}}{R_\theta} = \left(\frac{1}{M_w} + 2A_2c_{pol} + \cdots\right)\left(1 + \frac{16\pi^2}{3\lambda^2}R_g^2 sin^2\left(\frac{\theta}{2}\right)\right) \tag{1.3}$$

$$K = \frac{4\pi^2 n^2}{\lambda_0^4 N_A}\left(\frac{\mathrm{d}n}{\mathrm{d}c_{pol}}\right)^2 \tag{1.4}$$

The solution eluting from the SEC column is usually very dilute. Therefore, the term $2A_2c_{pol}$ can be neglected without significant error, simplifying equation (1.3):

$$\frac{Kc_{pol}}{R_\theta} = \frac{1}{M_w}\left(1 + \frac{16\pi^2}{3\lambda^2}R_g^2 sin^2\left(\frac{\theta}{2}\right)\right) \tag{1.5}$$

A plot of $\frac{Kc_{pol}}{R_\theta}$ versus $sin^2\left(\frac{\theta}{2}\right)$, called Zimm plot, affords calculation of M_w from the vertical intercept and R_g^2 from the slope.

Alternatively, M_w can be determined directly from absolute light scattering measurements, providing that the concentration of eluting protein is determined independently. Extrapolation of equation (1.3) to zero scattering angle gives [26]:

$$\frac{Kc_{pol}}{R_\theta} = \frac{1}{M_w} + 2A_2c_{pol} \tag{1.6}$$

Dynamic Light Scattering

In dynamic light scattering (DLS), temporal variations in scattered light intensity are measured at a fixed angle and represented by what is known as the intensity autocorrelation function. Intensity fluctuations will depend on the Brownian motion of the molecules/particles within the scattering volume. Larger molecules will diffuse more slowly and thus the scattering fluctuations they cause are correlated over longer time scales. Through the autocorrelation function, the temporal variation of the scattering intensity is quantified and the diffusion coefficient of the molecules (D) determined. Assuming sufficient dilution, size in terms of hydrodynamic radius (R_h) can be calculated based on the Stokes-Einstein equation, equation (1.7). R_h is defined as the radius of a spherical particle with the same diffusion coefficient as that of the protein molecule [26][27].

$$D = \frac{kT}{6\pi\eta R_h} \tag{1.7}$$

In theory, DLS will enable a mathematical separation of large and small molecules because they produce scattering fluctuations on different time scales. However, the resolution of the technique is quite low, and in practice, the difference in size must be considerably large for a possible distinction. Therefore, generally only the z-average hydrodynamic size, an average based on the scattering intensity, of molecules in the sample is considered.

Extrinsic Fluorescence Spectroscopy

Fluorescence is the emission of light form an illuminated sample, occurring from electronically excited states of, typically, aromatic molecules. It is a reliable tool in the study of proteins given its high sensitivity. Proteins may be studied through intrinsic or extrinsic fluorescence. One works with intrinsic fluorescence when the protein of interest contains one or more amino acids that are intrinsically fluorescent (phenylalanine, tyrosine or tryptophan). When the intensity is not great enough or the molecule does not have fluorescent amino acids, extrinsic fluorophores, such as 8-Anilino-1-naphthalenesulfonic acid (ANS), can be used to bind to the proteins. ANS is a commonly used non-covalent probe, which is very weakly fluorescent in water but fluoresces strongly when bound to non-polar entities. Upon unfolding, hydrophobic residues from the interior of a globular protein will be exposed. Consequently, when the protein is incubated with ANS, the latter will interact with said residues, which will increase its fluorescence, granting the possibility of studying the unfolding process of the protein [28][29].

1.2.5 Thermodynamics of Electrolyte Solutions & Acid-Base Equilibria

For a thorough study on antibody aggregation, control over solution pH and ionic strength in experiments is necessary. Through the estimation of ionic strength based on solution pH and buffer species' molarity, it is possible to exercise such control. However, due to the non-ideality of electrolyte solutions even at low concentrations [30], a significant error occurs when estimation is carried out with the species' concentration instead of activity. As ions in solution are generally shielded by counter ions of the opposite charge, they may act as less active than what is expected, according to their concentration. As such, activity expresses the actual influence of each ion and is given by equation (1.8), where, for an

ion *i*, *c* denotes its concentration, *γ* its activity coefficient and *a* its activity [22]. The activity coefficient expresses how far from ideality one solution is.

$$a_i = \gamma_i c_i \tag{1.8}$$

To circumvent this problem, activity models can be applied to calculate *γ* for every ion in solution, based on its ionic strength [31].

The Extended Debye-Hückel Activity Model

The effect of the concentration of ions is expressed by the ionic strength, that measures the electrical environment of the solution [32]. Ionic strength is given by equation (1.9), where *z* stands for the valence and *c* for the concentration of each ion *i*. The summation must include all ionic species present in solution.

$$I = \frac{1}{2}\sum_i z_i^2 c_i \tag{1.9}$$

Assuming that activity coefficients and equilibrium constants are independent of the medium composition at a certain ionic strength, activity coefficients can be calculated. The quotient of long-range Coulombic forces and short-range interactive forces between the central ion and ions with the opposite charge in its atmosphere defines the extended Debye-Hückel equation for the activity coefficient. Equation (1.10) represents the extended Debye-Hückel equation [31], being *A* and *B* temperature-dependent coefficients, 0.5085 $mol^{-0.5}(dm^3)^{0.5}$ and 3.281 $nm^{-1}mol^{-0.5}(dm^3)^{0.5}$ at 25°C, respectively, and *d* the diameter of the hydrated ion *i*, in nm. Nevertheless, the model is only valid for diluted solutions, and is only applicable in solutions with 100 mM or less of ionic strength.

$$-log\gamma_i = \frac{Az_i^2 I^{0.5}}{1 + BdI^{0.5}} \tag{1.10}$$

Control of Solution pH and Ionic Strength with Sodium Citrate Buffer and Sodium Chloride

Upon defining the solution's conditions, namely the total buffer molarity, its volume, pH and ionic strength, one can determine the right concentrations of each compound to be added. This is accomplished by combining the model presented above with acid-base equilibrium theory. To every acid-base reaction, exemplified in equation (1.11), where AH, H^+ and A^- represent, respectively, the acid, the proton and the base, an equilibrium constant is associated. One refers to it as a thermodynamic equilibrium constant (K) when dealing with the activity of the species responsible for the reaction, and values can be obtained in literature [32] [33]. When dealing with the concentration of the species, a stoichiometric equilibrium constant (K_a) is used. Thus, to obtain the concentrations necessary of each buffer compound, the latter constant needs to be determined.

By applying the extended Debye-Hückel equation, one obtains the *γ* of each buffer species in a solution of fixed ionic strength. After this, equation (1.12) can be applied and K_a determined [35].

$$HA + H_2O \leftrightharpoons A^- + H_3O^+ \tag{1.11}$$

$$K_a = K\frac{\gamma_{AH}}{\gamma_{H^+}\gamma_{A^-}} \tag{1.12}$$

As a poly-protic buffer [22], citrate has three ionisable groups that can generate protons, each producing one equilibrium condition for the possible acid-base equilibria. With citric acid, $C_6H_8O_7$, and the three species deriving from it (i.e. $C_6H_7O_7^-$, $C_6H_6O_7^{2-}$ and $C_6H_5O_7^{3-}$), these equilibrium conditions can be used to express their concentrations as a function of the concentration of just one of the species, as depicted in equations (1.13), (1.14) and (1.15), the value of each equilibrium constant already know.

$$[C_6H_7O_7^-] = K_{a,1}^{\text{Citr}} \cdot \frac{[C_6H_8O_7]}{[H_3O^+]} \tag{1.13}$$

$$[C_6H_6O_7^{2-}] = K_{a,2}^{\text{Citr}} \cdot K_{a,1}^{\text{Citr}} \cdot \frac{[C_6H_8O_7]}{[H_3O^+]^2} \tag{1.14}$$

$$[C_6H_5O_7^{3-}] = K_{a,3}^{\text{Citr}} \cdot K_{a,2}^{\text{Citr}} \cdot K_{a,1}^{\text{Citr}} \cdot \frac{[C_6H_8O_7]}{[H_3O^+]^3} \tag{1.15}$$

However, in the present case, the buffer species (citrate) is not present alone in solution. Counter-ions from the basic buffer component (i.e. sodium) and salt (i.e. sodium chloride) that is used to adjust the solution's ionic strength have to be taken into account as well. Thus, for determination of the concentration of all those species, additional equations have to be considered.

On one hand, one assumes that these acid-base reactions are the only reactions taking place in the solution, which leads to the mass balance condition in equation (1.16). The subscript 0 stands for the concentration of each buffer species before the acid-base equilibrium is established. On the other, the bulk solution has to be electro-neutral, meaning that the total concentration of cationic species must be equal to the concentration of anionic species, as is depicted in equation (1.17).

$$\begin{aligned}[C_6H_8O_7]_0 + [C_6H_7O_7^-]_0 + [C_6H_6O_7^{2-}]_0 + [C_6H_5O_7^{3-}]_0 \\ = [C_6H_8O_7] + [C_6H_7O_7^-] + [C_6H_6O_7^{2-}] + [C_6H_5O_7^{3-}]\end{aligned} \tag{1.16}$$

$$[Na^+] + [H^+] = [Cl^-] + [C_6H_7O_7^-] + 2[C_6H_6O_7^{2-}] + 3[C_6H_5O_7^{3-}] \tag{1.17}$$

1.3 Monitoring and Control of Protein Aggregation in Downstream Processing

MAbs, as complex proteins, consist of a distribution of many chemical species and slight process changes can affect the product quality profile [36]. To ensure consistency of product quality and reduce batch-to-batch variability, Process Analytical Technology (PAT) tools come as of great interest [37]. PAT is described by the Food and Drug Administration (FDA) as being part of process design and intended to contribute to process control. That is, to be taken actively into account for process decisions [38]. Additionally, current trends towards continuous integrated manufacturing may require better process control to ensure steady state over large periods of time. This sort of control can be simplified by the implementation of PAT tools as real-time monitoring tools of CQAs [39]. It would allow the simplification of root cause analysis and improvement of process understanding. Ideally, conjugation of the latter with real time monitoring capabilities could ensure that the CQA profile of the final product lies within acceptable quality limits based solely on real time measurements [36]. A considerable challenge in the implementation of PAT tools is related to the flexibility of use in various unit operations [40] and standardized communication between them.

1.3.1 Application of Spectroscopy as a Process Analytical Technology Tool

Spectroscopy is a strong tool for process monitoring as measurements can often be performed in-line [41]. This is pertinent especially in chromatography, as processes are very non-linear and have quite sharp concentration variations – the CQAs of chromatography effluents, such as the mass fraction of impurities, are quickly changing [42]. Mathematical tools for multivariate data analysis are then applied to extract information from spectroscopic measurements [43].

Spectroscopic methods have been widely used as tools for the structural analysis of proteins in DSP [36]. With UV/vis spectroscopy or Fourier-Transform Infrared (FTIR), the assessment of the protein's sequence and total concentration is possible through the primary structure of the proteins, by measuring the absorption of aromatic amino acids or polypeptide backbone [44]. The secondary structure is usually measured by vibrational spectroscopy, such as Infrared and Raman spectroscopy, which allow the measuring of the vibrational modes of the backbone of polypeptides [44][45][46]. The tertiary structure is accessible over the fluorescence of aromatic amino acids such as tryptophan, and structural changes affecting the local environment of tryptophan residues can be detected by fluorescence spectroscopy [28]. At last, the quaternary structure, may be assessed from protein's size by quasi-elastic light scattering methods (*e.g.* SLS and DLS) [44].

Raman Spectroscopy

Raman spectroscopy is a vibrational spectroscopic technique based on the inelastic scattering of photons, also known as Raman effect [47]. It measures changes in the scattered light frequencies between molecule's ground and excited vibrational states when they are interacting with a beam of light, resulting in a Raman spectrum containing bands correspondent to molecular configuration [48]. When a sample is irradiated by monochromatic light, the molecules absorb photons and transit to virtual states. Then relaxation occurs almost instantaneously as the molecules emit the photons they have exchanged energy with. Most of the light is scattered by the sample at the same wavelength as that of the incoming radiation. This occurs when the photons relax back to the ground state, and it is known as Rayleigh scattering. However, a small portion of the light is scattered at a wavelength shifted from the original laser wavelength. Such an event happens in two different situations: (i) when the energy of the emitted photon is weaker than that of the absorbed photon, and the molecule transits to a still excited state instead of returning to its ground state; (ii) when the incoming photon gains energy from a molecule that is already in an excited vibrational state. Relaxation to the first vibration level, situation (i), results in Stokes Raman shift. Its scattered light is of lower energy (longer wavelength) than that of the incoming light. When the process starts from the excited vibrational level, situation (ii), relaxation to the ground level can occur, and the scatter is of higher energy than that of the incoming laser. The process is known as anti-Stokes Raman. About one photon out of a million exhibits the Stokes Raman scattering and the probability is even smaller when considering anti-Stokes Raman, since the population of molecules in the excited vibrational state is always smaller than the population in ground state, according to Boltzmann distribution law [49].

Vibrations known to be strong in infrared (IR), also a vibrational spectroscopy, such as hydroxyl or amine stretching vibrations or vibration of carboxyl groups, are typically weak in a Raman spectrum. Whereas

non-polar functional group vibrations produce very strong Raman bands, for example stretching vibrations of carbon double and triple bonds, as well as symmetric vibrations of aromatic groups. As such, Raman can provide important information about the structure of molecules, allowing the determination of the molecular identity of samples [49].

It offers several advantages when compared to other analytical techniques: (i) no two different molecules present the exact same spectrum; (ii) it is non-destructive; (iii) its operational wavelength is independent of vibrational modes and can be used in ranges from UV to near-infrared (NIR); (iv) water, the natural medium for proteins, is a weak Raman scatterer, and does not interfere significantly; (v) does not require large samples or extensive preparation and the spectra can be obtained in many physical states. Nevertheless, Raman is a very weak physical phenomenon, as very few photons in the millions that interact with the molecules will present Stokes Raman scattering. Additionally, in the visible region, fluorescence emitted from the sample has a quantum yield much higher than that of Raman scattering, overwhelming its signal [48].

The polypeptide backbone and side chain groups of proteins are the main contributors to Raman spectra. Denaturation of native proteins followed by aggregation often involves changes in the secondary structure [48]. The Raman bands related to peptide linkage are designated as amide bands, being Amide I (AI) and Amide III (AIII) the predominantly sensitive bands to the secondary structure. [50]. As for signals obtained from side chain groups, these derive from localized vibrational modes of specific amino acids. The marker bands of aromatics are quite strong and recognizable in visible Raman spectra but frequently overlap with amide bands. The side chain modes are usually analyzed with UV Resonance Raman, which provides better resolution than standard Raman technique [48].

The AI band is present in the wavenumber region between 1610 and 1700 cm^{-1}, and is strongly dependent on the hydrogen bonds accepted from the carbonyl group and donated to the amine group. AI's intrinsic wavenumber depends on the peptide's environment: it decreases with increasing strength of the hydrogen bonding between the solvent and CO and NH groups, and with increasing electrostatic environments. As for the AIII band, it normally appears in the region between 1200 and 1340 cm^{-1}, the exact wavenumber will also depend on the environment and the structure of the molecule [45]. In the case of monoclonal antibodies, which are mainly β-sheet proteins, the AI band would be approximately at 1670 cm^{-1} while the AIII exhibits, for the majority of mAbs, a band at approximately 1245 cm^{-1} [48]. Considering aromatic side chain signals, phenylalanine, tyrosine, tryptophan and cysteine have been identified in Raman spectra [51] [52] [53].

1.3.2 Multivariate Analysis Approaches for Raman Spectroscopic Data

The identification and quantification of biological components by spectroscopic methods is usually difficult given the sample's diversified nature. To be able to gain meaningful information, the acquired data needs to be processed and analysed [43]. Spectroscopic data consists of thousands of variables (wavenumbers) and measurements (observations). To make use of all the information of complex spectra, univariate analysis, which deals with only one variable at a time, is not applicable. Thus, multivariate analysis, data analytical methods that deal with more than one variable, is necessary. The

application of multivariate statistical methods to chemistry and biology is known as Chemometrics [54]. The main aim of such techniques is to understand the relationship between the variables, which allows for pattern recognition in the data that could be modelled and consequently used to predict newly acquired data of similar nature [43].

Multivariate methods can be divided into two groups: unsupervised or supervised. Unsupervised methods, *e.g.* PCA, have no guidance available, and are useful to find hidden relations in unlabelled data. Supervised methods, e.g. *PLS*, label classes to be discriminated and consist of two phases: the training phase, a passive modelling stage that uses a training data set (previously labelled) to find patterns in the data; and a prediction phase, an active stage, where new data is validated using the model built in the first phase [43]. Methods such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Cluster Analysis (CA) and Partial Least Squares regression (PLS) have already been employed in vibrational spectroscopy data analysis. They enable the analysis of ample spectral distribution and are able to discriminate between spectra of different samples that show only minor changes [55]–[57]. All spectral data analysis conducted in this work was performed with PCA and PLS. Prior to the implementation of such methods, pre-treatment of the raw data is usually warranted in order to eliminate unwanted signal, such as fluorescence, detector noise or laser power fluctuations, and enhance discriminating features from sample components [43].

1.3.2.1 Pre-processing

Even though various techniques have been studied and are available, a standardized pre-processing procedure is difficult to create. Generally, pre-processing steps will depend on the problem, on empirical observations and prior experience on the matter. Nevertheless, a few steps are deemed advisable and others even indispensable [43], [58], [59].

Spectral axis alignment and cosmic ray removal are two basic steps always required before any sort of data treatment. The former, to guarantee that spectra of the same material measured on different days/months have the same Raman shift axis; the latter, to correct the spectra for Cosmic ray events, which are generated due to high-energy particles that pass through the charge couples device (CCD) and are interpreted as signal [60], [61]. Following those corrections, background correction, smoothing of the signal, normalization of the spectra and removal of outliers are usually performed on the data [34].

From intrinsic fluorescence from biomolecules to ambient light and detector thermal noise, multiple factors can contribute to the background. Background removal eliminates the effect of broad bands or low frequency components present [43]. Various methods can be used for this purpose, namely, first and second order differentiation, polynomial fitting, or frequency domain filtering [59], [62]. Raman spectra can also be affected by detector noise and intensity fluctuations of the radiation source, conferring a low signal to noise ratio (SNR) to the signal due to these high frequency components [43]. SNR can be improved by the application of smoothing filters, such as Savitzky-Golay (SG) or Locally Weighted Scatterplot Smoothing (LOWESS) [63], [64]. Fluctuations from the environment also adds variability to the spectra that is not related to chemical or structural differences in the samples. Normalization techniques can be used to overcome such differences [43]. Finally, due to instrumental artifacts or unwanted variations in the samples, among other problems, some spectra may diverge

significantly from what is to be expected, and these measurements are known as outliers. Detection of such outliers is usually conducted using analysis methods such as Principal Component Analysis (PCA), discussed below [43].

Alternative Techniques for Smoothing

The SG filter is generally applied for smoothing given its easy applicability and improved algorithm. Nevertheless, it is not optimal as the algorithm tends to smooth every sharp peak, including narrow features potentially characteristic of a sample [65]. Additionally, as a least-square fitting method, it is suitable to be applied to data that can be expressed analytically, *e.g.* an exponential decay. It also requires knowledge based on experimental conditions. The limitation of such a technique comes when experimental results cannot be approximated to a reasonable expression [66]. This is quite common in spectroscopic data. Other, more sophisticated smoothing techniques were found, whose application might not lead to such consequences, such as the Fourier Transform (FT), the Wavelet Transform (WT) and the Empirical Mode Decomposition (EMD).

When a Fourier transform is used, all the information contained in the data is represented in the Fourier domain by the amplitudes of the sine and cosine functions [66]. Knowing random noise is located in high frequencies, high frequency components can be set to zero in the Fourier domain. Once the inverse Fourier transform is applied to the modified data, the noise will be left out [67]. However, it is necessary to have a criterion to separate the noise from the signal. A limitation of such a procedure is that not only noise may be filtered out. Quite an extensive number of filtering techniques can be applied for the smoothing of data [66] [68].

Concerning WT, wavelets are functions localized in both time and frequency domains and have been applied for a number of procedures, namely denoising, spike removal and background removal [69]–[71]. It decomposes the signal into a series of detail and approximation coefficients. It enables the separation of noise, made of high frequency components from Raman information, of moderate frequency [72]. Similarly to FT, the basis function applied in the transformation is made sines and cosines, able to effectively extract both time and frequency information. Given the fact that one is working with spectra, the data will be discrete and the time is in fact represented by the wavenumbers. There are different families of wavelet functions, the most common being the Daubechies, Symmlet and Coiflet [69]. WT can decompose the signal into detail and approximation coefficients multiple times,

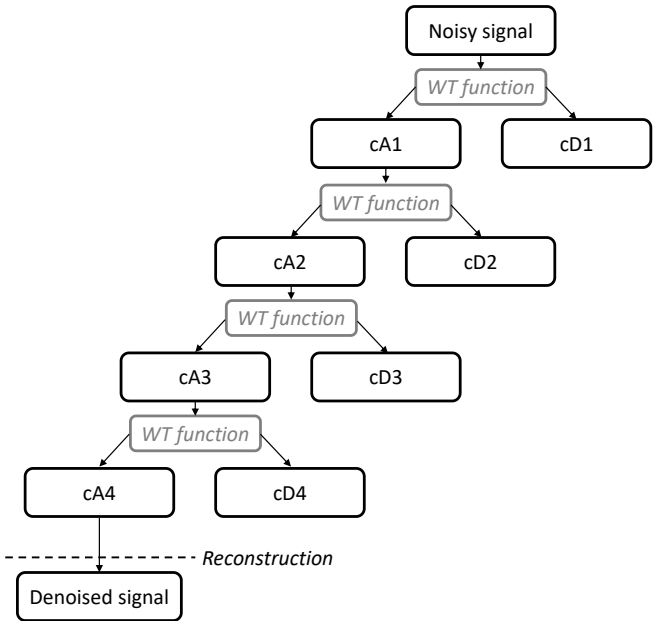

Figure 3 - Scheme of a signal decomposition by WT. cA represents the approximation coefficients and cD represents the detail coefficients.

each constituting a level of decomposition, as shown in an example in Figure 3. Each decomposition will capture contributions from different frequencies. In each level, the detail coefficients retain the information of a certain frequency and the approximation coefficient represents what is left from the signal. Thus, the smooth component is captured by the approximation coefficients. When decomposition has reached a level where no more noise is being filtered out, the resulting approximation coefficients are transformed back to the time domain. This is carried by the inverse wavelet transform. Once again, the possible problem with such a technique is knowing until what level of decomposition should go and what base wavelet function to apply.

At last, EMD is an adaptive nonlinear time-frequency data analysis technique. The goal of EMD is to decompose the signal into a sum of intrinsic mode functions (IMFs), functions with an equal number of extremes (maxima and minima), zero crossings and symmetric with respect to zero. By the end of the decomposition of the signal, all that is left are the obtained IMFs and a monotonic function (the residue) from which no IMF can be extracted. The original spectrum is given by the sum of both components [73]. High frequency noise can be eliminated by assuming that noise has a zero mean, while the signal has a nonzero mean. Thus, a statistical test, *e.g.* t-test, can be used to identify IMFs with zero mean, using a certain significance level. The test is applied to the first IMF and if the mean is zero, the test is then applied to the sum of the first and second IMFs. It carries on while the result is mean zero and stops when the mean in nonzero, excluding the last IMF added to the test. However, if, by change, some signal also presents zero mean and is included in these IMFs, deformations in the denoised signal can occur. Consequently, usually only up to 5 IMFs are considered to model the noise, in a situation where more the 4 are necessary. Having identified the IMFs representing only noise, these are subtracted from the total sum of IMFs to obtain the denoised signal. Additionally, background removal with EMD has also been suggested, stating that the sum of the final residue and the last IMF can be used to model background of a Raman spectrum. The residue contains the general tendency of the spectrum and the last IMF represents the major fluctuations around the residue [47].

1.3.2.2 Principal Component Analysis

PCA is a common tool for data reduction, simplification and outlier detection [36]. Some combinations of variables (wavenumbers) in a given set are highly correlated with each other and, if one can capture these correlations, the data set can be compacted using fewer variables. These new variables are linear combinations of the original ones. Vibrational spectra contain important information about the sample in different wavenumbers (bands) and some of these have the same source of variation, resulting in high correlations between few variables and allowing for dimensionality reduction [43]. PCA reduces a set of variables into a smaller set of orthogonal (and thus completely independent) principal components (PCs) in the direction of maximal variation, reducing the dimensionality and retaining the most significant information for further analysis. It decomposes a data matrix X with m object rows and n variable columns according to equation (1.18), in a structured part S and a noise part E.

$$X = S + E \tag{1.18}$$

The m objects are the different observations (spectra) and the n variables the wavenumbers of each object. PCA transforms the “variable space” into a more relevant “PC coordinate space”. The first PC

(PC1) represents the direction of the greatest variance in the data, PC2 illustrates the largest residual variance along a direction orthogonal to PC1 and so forth. Although the number of PCs can reach the number of variables or one less than the number of objects, the relevant information (spectral variance) is usually explained by the first few dominant PCs [43]. The structured part (S) obtained from PCA can be represented by equation (1.19).

$$S = TP^T \tag{1.19}$$

Matrix *T* represents the score values of each component and *P* the loading values. The goal is to explain as much as possible of X with the minimal number of PCs. Loading are coefficients representing each PC that indicate the contributions of each variable in the original coordinate space. The values of each object in the new coordinate system are the scores of the object, which corresponds to a point in the new PC-coordinate system. As an unsupervised method, PCA is capable of identifying some important structural information in the data but does not have a good discriminatory power [36][43][74].

1.3.2.3 Partial Least Squares Regression

A different analysis can be performed with multivariate regression methods such as PLS. Regression is used to quantify particular dependent variables – the expected outcome of an experiment. Two matrices are used, one with dependent variables Y and the other the already introduced matrix X of spectral data. From these two matrices, a regression model is built in the training phase and is used to predict Y values from new measurements of X [43]. In PLS, the original variables and measurements are projected into a lower dimensionality space where the latent variables are not correlated with each other, as explained for PCA. The covariance between predictor and response variables is subjected to Singular Value Decomposition to obtain the latent variables, which are further used to predict the Y [75]. Additionally, PLS considers the possibility that the dependent variables which are being estimated from the same pool of independent variables are correlated with each other, being able to handle one or more dependent variables [43]. Spectroscopic data has a great number of dependent variables, highly correlated with each other. PLS is one of the most widely used analytical techniques in vibrational spectroscopy data analysis to estimate and quantify components in a given samples [75], [76], [77].

1.4 Aim of the Work

The aim of this work was to gain a deeper understanding of antibody aggregation in downstream processing and assess the ability of online monitoring of mAb aggregates with Raman spectroscopy. Additionally, the development and optimization of a flow-cell was carried out to aid in Raman spectra acquisition.

Firstly, two monoclonal antibodies, mAb-1 and mAb-2, were studied under conditions that mimic a low pH VI step. Identification of key parameters for protein stability during the process was conducted by analysis of antibody aggregation behaviour. MAbs were subjected to low pH incubation and pH-shift experiments, similar to what occurs in the VI. A hypothesis for the aggregation mechanism under these conditions was formulated based on the results.

Secondly, the potential of a flow-cell for the robust acquisition of spectra was tested and optimization of the device to be implemented in a chromatographic effluent was carried out.

Further, Raman spectra of mAb-1 and mAb-2 solutions were acquired and analysed with multivariate data analysis. Resorting to DoE, multiple samples of decorrelated monomer and aggregate concentration were measured. Different pre-treatment techniques were tested on the spectroscopic data. With a PLS regression model, the ability of Raman to identify and quantify aggregated and monomeric antibody was determined.

2 Materials and Methods

2.1 Materials

MAb-1 was expressed by Chinese hamster ovary cells inside a perfusion bioreactor. Cell-free harvest was purified by protein A affinity chromatography. The eluate was immediately neutralized using 1 M Tris-HCl pH 8.0 buffer solution and stored in the fridge prior to use. mAb-2 was provided as neutralized protein A eluate.

The chemicals used for the formulation of buffer solutions were sodium phosphate (Sigma-Adrich, USA), L-arginine (Sigma, USA), sodium azide (Sigma-Adrich, USA), citric acid (Fluka Analytical, Switzerland), trisodium citrate (Sigma-Aldrich, USA), sodium chloride (Sigma-Aldrich, USA), sodium hydroxide (Sigma-Aldrich, USA), isopropanol (Sigma-Aldrich, USA), D-Sorbitol (Sigma, USA) and 8-Anilino-1-naphthalenesulfonic acid (ACROS ORGANICS, Belgium).

2.2 Sample Preparation for Aggregation Studies

2.2.1 Protein Stock Solution in Ultrapure Water

High concentration protein stock solutions in ultrapure water were prepared in order to have full control over sample composition for aggregation studies.

Up-Concentration and Buffer Exchange

Protein A capture eluates were up-concentrated to approximately 100 g/L by reducing solution volume using Amicon Ultra centrifugal filter units equipped with a 10 kDa molecular weight cut-off regenerated cellulose membrane (Merck Millipore, USA). Centrifugation was conducted at 3200xg and 4 °C. The resulting solution was dialyzed against ultrapure water using Slide-A-Lyzer dialysis cassettes with 7 kDa MWCO (Thermo Scientific, USA). Volume ratio of protein solution to surrounding reservoir was approximately 1:500. The Millipore water was exchanged 5 times over the course of 24 hours and the entire procedure was performed at 4°C under mild agitation. The dyalized solution was filtered into a 2 mL plastic tube (Sarstedt AG & Co., Germany) using a Millex-GV syringe filter unit with 0.22 μm pore size (Merck Millipore, USA) and stored in the fridge prior to further use.

Analysis of Protein Stock Solution

The prepared solutions were analysed by SEC-MALS. An aliquot of stock solution was diluted 10x using 10 mM sodium phosphate pH 7.0. 5 μL of diluted solution were injected into a Superdex 200 column using 100 mM sodium phosphate pH 7.0, 200 mM L-arginine and 1 g/L sodium azide as eluent. The injected sample was eluted at a flow rate of 0.5 mL/min. The procedure was always conducted in duplicates. Analysis of the protein stock solutions was repeated on a weekly basis to assess stability under those solution conditions.

2.2.2 Stock Solutions of Buffer Components

To rapidly produce samples with different solution conditions, concentrated stock solutions of buffer components were prepared. Those included 200 mM citric acid, 200 mM trisodium citrate, 1 M sodium chloride and 100 mM sodium hydroxide. Appropriate mass of substance was dissolved in ultrapure water and total solution volume was adjusted using a volumetric flask. All stock solutions were filtered

using 0.45 µm cut-off Durapore membrane filters (Merck Millipore, USA) and stored in the fridge prior to use.

2.2.3 Stock Solutions of Extrinsic Fluorescence Studies

A 2.5 mM ANS stock solution was prepared in two steps. First, an appropriate mass of ANS (ACROS ORGANICS, Belgium) was dissolved in dimethyl sulfoxide (Fluka Analytical, Switzerland) to achieve a concentration of 25 mM. This solution was diluted 10x with pure DMSO, aliquoted into 2 mL plastic tubes and stored frozen at -5 °C to ensure negligible degradation. Aliquots defrosted and diluted 10x with ultrapure water prior to use. Defrosted and diluted ANS solutions were stored in the fridge and used for no longer than five days. Further, containers were always wrapped in aluminium foil to minimize risk of photo-bleaching.

A 2 M D-Sorbitol stock solution was prepared by dissolving an adequate amount of D-Sorbitol (Sigma, USA) in ultrapure water and adjusting solution volume using a volumetric flask.

2.3 Characterization and Quantification of Protein Aggregation

2.3.1 Aggregation and Unfolding Studies at Low pH

1 g/L protein solutions were prepared by mixing appropriate volumes of protein, citric acid, trisodium citrate, and sodium chloride stock solutions as well as ultrapure water to achieve desired molarity, pH, and ionic strength. The effect of the pH during the incubation was analysed by measuring solutions in pH 2.50, 3.00 and 3.50, while maintaining every other parameter constant, i.e. protein concentration, buffer molarity and ionic strength. The effect of the ionic strength was evaluated by fixing the pH at 3.5 and varying the ionic strength of the solution between 25, 50 and 100 mM.

Dynamic Light Scattering Measurements

Average hydrodynamic size of the protein during incubation at low pH was measured over the course of one hour by means of *in-situ* dynamic light scattering using a Malvern ZetaSizer Nano ZS instrument (Malvern, UK) equipped with a 633 nm He-Ne and detection at 173° scattering angle. The attenuator was set to the value of 10 and measurement position to 4.2 cm. Immediately after preparation, solutions were filtered directly into a ZEN2112 quartz cuvette with 3 mm path-length (Malvern Panalytical, USA) through 0.22 µm cut-off syringe filter. All measurements were conducted at 25 °C. The entire procedure was conducted three times for every investigated condition. Solution pH of the samples was always measured for verification using a Jenco 6230N pH meter (JENCO, USA).

Extrinsic Fluorescence Measurements

Extrinsic protein fluorescence was measured on a PerkinElmer EnSpire 2300 Multilabel Reader instrument (PerkinElmer, USA). Fluorescence emission was recorded at 490 nm upon excitation at 403 nm. Sample preparation was identical to the one for DLS measurements. Only differences were that fluorescence measurements were conducted at 0.25 g/L protein concentration, 25 µM ANS was added through the aforementioned procedure, and in some cases, excipients were added using appropriate stock solutions. The molar mAb:ANS ratio was 1:10. To avoid lengthy exposure of ANS to light as well as exposure of protein to low pH before start of the measurements, those two components were always added at the end of the preparation of a sample. Immediately after mixing all required stock

solutions with ultrapure water, 250 µL of sample were transferred into a well of a Greiner Bio-One µClear non-binding black 96 well plate (Greiner Bio-One, Germany) and measurement was started. Each condition was measured at least two times and remaining sample volume was used to validate solution pH as described above.

2.3.2 Aggregation Studies for Neutralized Solutions after Temporary Exposure to Low pH

Antibody aggregation after incubation under acidic conditions for a defined period of time followed by neutralization was systematically studied as well. Sample preparation for low pH incubation was identical to that reported above. All experiments were conducted using 1.5 mL plastic reaction tubes without agitation. 500 µL of acidic solution with 2 g/L protein concentration were produced by mixing appropriate volumes of component stock solutions and ultrapure water. Solutions were left at room temperature for a predetermined amount of time before pH was raised by adding an equal volume of neutralization solution containing appropriate quantities of sodium hydroxide and sodium chloride to reach target solution pH and ionic strength. Those quantities were calculated based on the theory presented in section 1.2.5. Neutralization solution was added using an adjustable micropipette and the resulting solution was mixed immediately by pipetting up and down.

Additionally, the effect of velocity of neutralization on protein aggregation behaviour under otherwise identical conditions was investigated. To that end, the neutralization solution was added using a NE-300 Just Infusion syringe pump (New Era Pump Systems Inc., USA), while the protein solution was stirred by a magnetic stirring bar inside a glass vial. Volumetric flow rate at the syringe pump was adjusted to ensure complete neutralization after a predetermined amount of time.

After completed neutralization, monomer content, average aggregate mass as well as average hydrodynamic size of the protein were followed as function of time by SEC-MALS and DLS as detailed below. Further, solution pH was measured to ensure proper neutralization.

Dynamic Light Scattering Measurements

Immediately after neutralization, 100 µL of solution were filtered through a 0.22 µm cut-off syringe filter unit into a quartz cuvette. Fluctuations in scattered light intensity were measured in periodic intervals over the course of 10 h. Instrument and measurement protocol were identical to those reported above for the low pH incubation studies.

Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering

Neutralized solutions were also analysed by SEC-MALS using a Superdex 200 10/300 GL size-exclusion column (GE Healthcare, USA) connected to an Agilent 1200 series HPLC system (Agilent Technologies, USA), consisting of quaternary pump, auto-sampler and UV absorbance detector. 50 µL of sample were injected into the column and eluted for 60 minutes at a constant flow rate of 0.5 mL/min using 100 mM sodium phosphate pH 7.0, 200 mM L-arginine and 1 g/L sodium azide as mobile phase. UV absorbance of the effluent was recorded at 280 nm wavelength. Injections were made initially after neutralization and repeated once per hour, for 10 hours. Further, static light scattering (SLS) from the liquid leaving the column was measured using a DAWN HELEOS-II multi-angle light scattering device (Wyatt Technology, USA) covering detection angles from 14.5 to 163.3°. The instrument is equipped

with a 663.5 nm GaAs laser. Agilent ChemStation (Agilent Technologies, USA) and Wyatt Astra V (Wyatt Technology, USA) software was used for analysis of UV absorbance and light scattering data, respectively.

2.4 Raman Spectra Measurement and Analysis

2.4.1 Data Acquisition

Raman spectra of protein aqueous solutions were acquired using a RamanRxn2 Multi-channel Raman Analyzer (Kaiser Optical Systems, Inc., USA). The spectrometer is equipped with a Charge Coupled Device (CCD) detector, a HPG-785 grating and an excitation laser at 785 nm wavelength and 1 cm^{-1} of spectral resolution. Each spectrum was measured from 100 to 3425 cm^{-1}. The laser power employed for the measurements was 400 W. All data was collected in iC Raman 4.1.917.0 SP2 software (Mettler Toledo & Kaiser Optical Systems, Inc., USA) and processed and analysed using Matlab 8.6.0. software.

Experiments were conducted offline and in batch mode with a flow-cell, described in Chapter 3.2.1.

Batch Mode Measurements with the Flow-cell

Samples were injected in the flow-cell where a 3 mm diameter channel was in the direct path of the laser source and detector, with a void volume dependent on the length of the channel. Inlet and outlet channels allowed for the insertion and removal of the samples, as well as avoidance of air bubbles. Before initiating the measurements, the time of spectrum acquisition needed to be defined. This depends on the saturation of the CCD detector. At saturation, pixels lose their ability to accommodate additional charge, and valuable information can be lost. In the iC Raman software, information about detector saturation is provided by the Pixel Fill, defined as the largest value, in counts, at any wavenumber in the collected spectrum. By adjusting the exposure time of the laser, one could know how saturated the detector was by the percentage of the Pixel Fill achieved after one measurement. A spectrum should display between 40 and 80% Pixel Fill. Below 40%, not enough information is being acquired, and above 80% the detector begins to saturate, and information can be lost. As such, for each experiment, the exposure time was set with the most signal intense sample and maintained for all the samples. After being measured, the sample would be removed from the flow-cell and the cannel cleaned with ultrapure water and dried with air before the next sample was injected and measured.

To increase the signal to noise ratio (SNR) more than one spectrum of each sample was acquired and during data analysis the collected spectra were average.

2.4.2 Data Analysis

The general data analysis path, consisting of pre-treatment and the construction of predictive model, was developed within the group for the application of Raman spectroscopy in upstream and downstream processing.

Pre-treatment

Besides Cosmic spike removal and intensity correction, which were performed automatically by the iC Raman software, all other pre-treatment processes were handled in Matlab.

Background correction and smoothing of the spectra was performed by applying a Savitzky-Golay (SG) smoothing and differentiation filter. The SG filter is a moving window based local polynomial fitting procedure, which has as smoothing input parameters the size of the moving window (frame length, *fr*) and the polynomial order (*pol*). Additionally, for the differentiation of the Raman signal and consequently elimination of background, a derivative order parameter (*ord*) was also an input [43]. Thus, for smoothing, all spectra were fitted to a second order polynomial (*pol*=2) and the frame length adopted was selected according to the best result obtained through hyper-parameter tuning procedure; for baseline correction the first derivative was applied (*ord*=1). Furthermore, normalization by the Standard Normal Variate (SNV) technique was also applied to the spectra to guarantee that the intensity of the Raman bands was as similar as possible across all spectra recorded. Finally, column-wise mean centring was performed to the normalized data.

Outlier detection was executed with PCA using Matlab's *pca* function, providing the principal component's (PC) loading, scores, and explained variance of each PC.

As an alternative for SG, different techniques were compared specifically for the smoothing of the data, namely Fourier Transform (FT), the Wavelet Transform (WT) and the Empirical Mode Decomposition (EMD). As most of them, including SG, present input parameters whose values will affect the treatment of the data, tuning of these parameters was performed. After obtaining the best input parameter value for each technique, these were implemented and compared.

Hyper-parameter Tuning of Pre-treatment Techniques

Table 1 shows the techniques for which parameter tuning was necessary and the respective parameters tested. For SG smoothing, the frame length range value was decided based on process experience. A window of size below 17 is too short for polynomial fitting and too large a window might lead to information loss. The smoothing was carried out with the standard SG filter, from Matlab. For WT denoising, the Matlab functions *wavedec* and *idwt* were applied for decomposition of the data and transformation back to time domain, respectively. As inputs, for the former, one needs to provide the data set, the level of decomposition and the base function that is to be applied. For the latter, the approximation coefficients after the final decomposition, and again the base function. Both the level of decomposition and the base wavelet function were tested as hyper-parameters. Concerning the FT method, the Matlab *fft* and *ifft* functions were applied to execute with the smoothing of the data. The filtering of the data in the Fourier domain was simply performed by setting to zero high frequency components. The level of decomposition range for WT smoothing and the filter range for FT smoothing were arbitrary. Also for WT, the chosen base functions were taken from literature as functions declared well suited to be applied in Raman spectra, *db* representing the Daubechies, *Coif* the Coifflet family and *Sym* the Symmlet family [70]–[72].

Table 1 - Techniques for which hyper-parameter tuning was performed, the respective hyper-parameters and the tested values.

Technique	Hyper-parameter	Tuning
SG	Frame-length (fr)	17 – 51 (odd numbers)
WT	Decomposition level	1 – 10
	Base Function	db4, db5, db6, db7, db8, db9, db10, db11, db12, Coif4, Coif5, Sym8
FT	Filter	10 – 1000 (even numbers)

PLS Regression Model

An independent PLS regression model was built for each data set. 80% of the data set was used for the training and cross validation of the model and the remaining 20% were used for its validation. For a greater confidence in the root mean square error in prediction (RMSEP), a rotation procedure was implemented. In each rotation, 5 in total, different measurements were assigned to the external predicted set. Additionally, the measurements assigned for the prediction set were evenly distributed between the ranges of the to-be-predicted variables. Resorting to this method, from each model receives five distinct results, one for each rotation.

For the calibration, to assess the robustness and prediction strength of the model, cross-validation (CV) was performed on the training set by application of the *k-fold* CV method [43]. The training data was split in two subsets as well. With a 5 fold partition (k=5) of the data in a random order, a calibration subset containing 4 of the defined groups and a prediction subset containing the 5th were used for the cross-validation of the data.

Before calibration, the averaged spectra were pre-treated with the technique selected after hyper-parameter tuning for smoothing, then SNV normalization and mean centering. After pre-treatment, the extremes of the spectra, below 450 and above 3100 cm^{-1}, were cut, as it was known that no information would come from these regions. Moreover, two short regions known to be affected by the laser according to the Raman device manufacturer were also cut: 1820 – 1880 and 2530 – 2590 cm^{-1}.

The optimal number of latent variables (LV) was determined by the lowest relative root mean square error in CV (relRMSECV). RMSE is given by equation (2.1), where N is the total amount of samples, $y_{ref,I}$ is the reference value given by an analytical measurement and $y_{pred,I}$ is the value predicted from the Raman spectrum of the samples using the regression model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(y_{ref,i} - y_{pred,i}\right)^2}{N}} \tag{2.1}$$

The relative root mean square error is obtained by dividing the RMSE by the range of the calibration or prediction set, if one is defining the error in CV or external prediction, respectively.

The coefficient of determination, R^2 for the training phase or Q^2 for the validation phase, reveals how much of the variance of the Y (monomer or aggregate concentration) is explained by the independent variable (the spectra). It is determined by the quotient of explained variance in the model (SSE) by total variance in the data set (SST). In a perfect correlation between observed and predicted values, it equals 1. SSE and SST are given by equations (2.2) and (2.3), respectively.

$$SSE = \sum_{i=1}^{N} (y_{ref,i} - y_{pred,i})^2 \tag{2.2}$$

$$SST = \sum_{i=1}^{N} (y_{ref,i} - \overline{y_{ref}})^2 \tag{2.3}$$

3 Results and Discussion

3.1 Characterization of Antibody Aggregation

3.1.1 Aggregation Studies at Low pH

As pH and ionic strength (I) are assumed to be the major variables affecting protein aggregation, their effect on mAb self-association during low pH incubation was assessed by measuring the average hydrodynamic size of mAb-1 and mAb-2 *in-situ* by dynamic light scattering over the course of 1 hour. Table 2 contains the set of performed experiments.

Table 2 - Investigated conditions for studying antibody aggregation under acidic conditions. 50 mM sodium citrate was used as buffer system and sodium chloride was added in appropriate quantities to set the ionic strength to a value of 50 mM. Antibody concentration was 1 g/L for all conditions. Solution pH was always within ± 0.1 unit from target.

pH [-]	I [mM]
2.5	50
3.0	25
	50
	100
3.5	50

Impact of Solution pH

Figure 4 shows the evolution of average the hydrodynamic radius ($<R_h>_z$) for both antibodies at 50 mM ionic strength as a function of time and solution pH. Over the course of 1 hour, hydrodynamic size of mAb-1 and mAb-2 remains constant within statistical uncertainty, indicating that no significant aggregation is occurring under acidic conditions at low ionic strength.

Caution has to be taken when comparing absolute values of measured $<R_h>_z$ between mAbs and conditions. Dynamic light scattering determines a species' mobility based on the recorded scattered light intensity autocorrelation. Size is then inferred by applying the Stokes-Einstein equation (equation (1.7)). Only under conditions of negligible interaction between particles, this conversion is completely correct. Any non-zero particle-particle interaction will distort the calculated value for $<R_h>_z$.

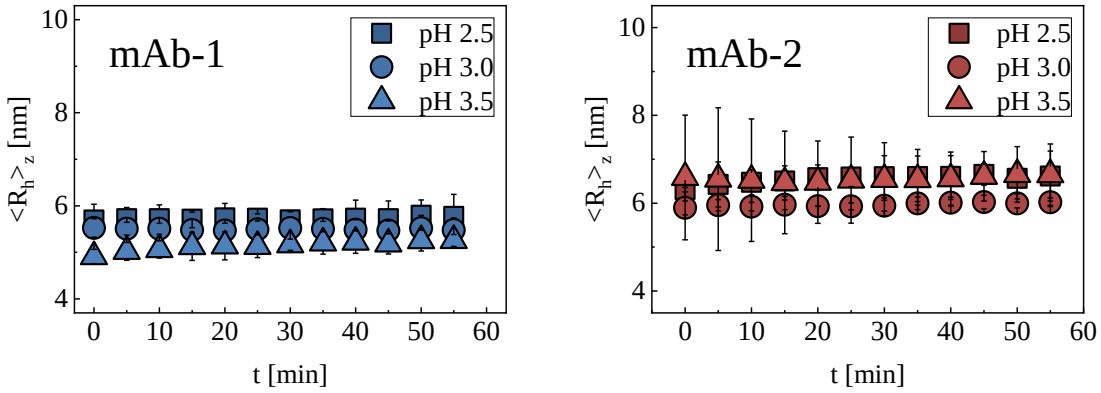

Figure 4 – Time evolution of the average hydrodynamic radius of 1 g/L mAb-1 (left) and mAb-2 (right) solutions in 50 mM sodium citrate at pH 2.5 (square), 3.0 (circle) and 3.5 (triangle) with total ionic strength fixed at 50 mM. Error bars represent 90% confidence intervals for the mean, which was calculated by averaging results from three independent measurements.

The unexpected outcome of undetectable aggregate formation at low pH could be explained by the dominant contribution of repulsive electrostatic interactions over any other type of attractive intermolecular interaction, preventing mAb molecules from forming oligomers. Under acidic conditions, far from the antibodies' isoelectric point (pI), polypeptide chains are highly protonated and will strongly repel each other by charge-charge interactions.

Impact of Ionic Strength

Another important factor that might help explain these results is the low ionic strength of 0.05 M that will render electrostatic interactions long-ranged. As ions contribute to the screening of surface charges, increase or decrease of the ionic strength will have an effect on the electrostatic forces that are assumed to be the dominant contribution to the overall protein-protein interaction. Following this line of thought, the impact of the ionic strength on the proteins was tested for 1 g/L of protein in 50 mM sodium citrate at pH 3.0.

Figure 5 shows the evolution of the $\langle R_h \rangle_z$ of mAb-1 and mAb-2 as a function of time for 25, 50 and 100 mM ionic strength at pH 3.0, constant buffer molarity and protein concentration. Again, average hydrodynamic size does not change through the 1 hour of analysis for all three conditions.

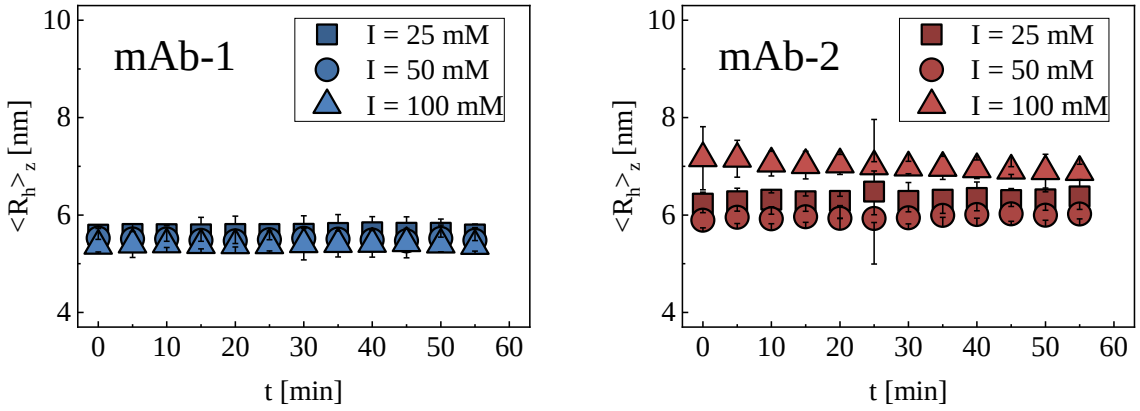

Figure 5 - Time evolution of the average hydrodynamic radius of 1 g/L mAb-1 (left) and mAb-2 (right) solutions in 50 mM sodium citrate at pH 3.0 with total ionic strength varying between 25 (square), 50 (circle) and 100 (triangle) mM. Error bars represent 90% confidence intervals for the mean, which was calculated by averaging results from three independent measurements.

Even 100 mM of ionic strength appear to be not sufficiently high to promote aggregation, and so the electrostatic repulsion between proteins still dominates. VI is usually conducted between two chromatographic purification steps, i.e. capture and polishing. Often, the first polishing step is performed with ion-exchange chromatography (IEX), a process in which the feed is loaded onto the column under low ionic strength conditions [12]. Additionally, elution from the protein A column is usually performed at moderate buffer molarities [78]. Therefore, higher values of ionic strength are usually not encountered in mAb downstream processing and thus were not tested.

3.1.2 Protein Unfolding Studies at Low pH

Even if aggregation does not occur under low pH conditions, proteins are expected to undergo conformational changes when incubated in acidic environments (i.e. unfold at least partially) [79][80],

which is a possible intermediate step within the process of aggregation [81][82]. Structural information about proteins can be obtained through extrinsic fluorescence spectroscopy, indicating the extent of exposure of hydrophobic residues to the solvent. This can give useful information regarding the effect of incubation under acidic conditions on antibody structure and how it might contribute to aggregate formation during and after VI.

Figure 6 shows the time evolution of the fluorescence intensity of 0.25 g/L mAb-1 and mAb-2 solutions at pH 2.5, 3.0, and 3.5 with 50 mM buffer molarity and 50 mM ionic strength. For mAb-1, measured fluorescence intensity remains virtually constant over the course of one hour, when the protein is incubated at pH 2.5 and 3.0. On the other hand, when mAb-1 is exposed to pH 3.5, fluorescence intensity increases gradually and almost reaches the level observed for pH 2.5 and 3.0 after one hour. These measurements indicate that surface hydrophobicity of mAb-1 reaches its new steady-state value (compare with control measurement at pH 5.0) almost instantaneously upon exposure to pH 2.5 and 3.0, whereas it takes some time for it to reach similar levels at pH 3.5. Since changes in surface hydrophobicity are commonly associated with conformational rearrangements, it can be concluded that a new conformational equilibrium is reached very fast when pH is lowered to 2.5 or 3.0 for mAb-1. In contrast, conformational changes occur much slower at pH 3.5, which would account for the increase of fluorescence intensity over time, as more hydrophobic binding sites are gradually made available for ANS molecules.

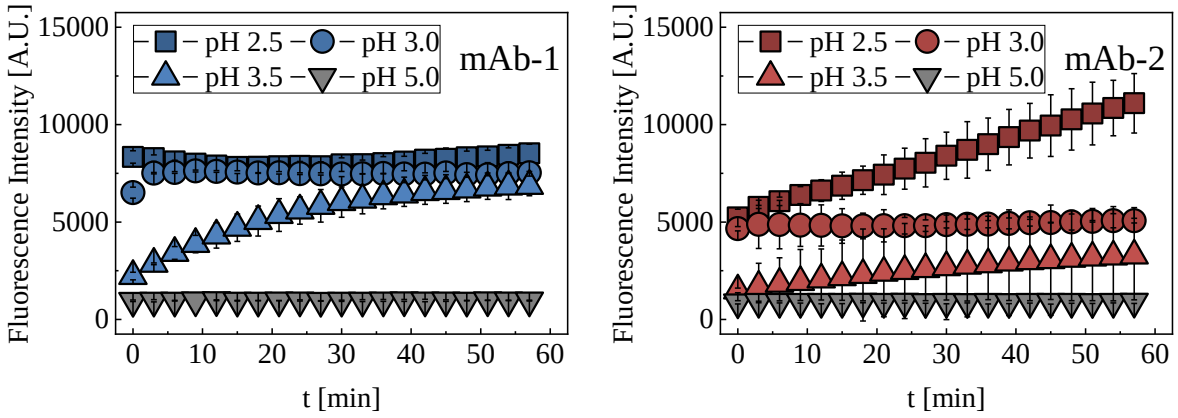

Figure 6 - Time evolution of fluorescence intensity of 0.25 g/L mAb-1 (left) and mAb-2 (right) solutions in 50 mM sodium citrate with I equal to 50 mM. ANS was added to the solutions in ten-fold molar excess with respect to the protein. Error bars: 90% confidence intervals for the mean.

Figure 6 also contains the results for mAb-2 under otherwise identical conditions. At pH 2.5, fluorescence intensity increases substantially as function of time over the course of one hour. When mAb-2 is incubated at pH 3.0, fluorescence intensity remains almost constant at a level slightly lower than that observed for pH 2.5 immediately at the beginning. Further increasing solution pH to 3.5 again changes both qualitative and quantitative behaviour of fluorescence intensity as function of time. It starts from a comparatively low level and gradually increases over the course of one hour but never reaching the value observed for pH 3.0. Along the line of argument presented above, conformational dynamics of mAb-2 in response to exposure to an acidic pH of either 2.5 or 3.5 seem to happen on time-scales comparable to those relevant for viral inactivation in therapeutic antibody manufacturing. On the other

hand, conformational steady state seems to be reached almost instantaneously at the intermediate pH of 3.0.

These results led to further studies on antibody unfolding in the presence of D-sorbitol, a polyol sugar. Osmolytes such as polyols promote preferential hydration of protein molecules due to unfavourable interactions between with the protein surface. Preferential hydration relates to the exclusion of a species from the protein domain promoting an excess of water molecules present [83]. This hydration is likely to favour more compact molecular conformations [84], possibly reducing its degree of unfolding upon exposure to stress conditions.

Experiments were conducted with 0.25 g/L mAb-1 in 50 mM sodium citrate pH 3.5 at 50 mM ionic strength in presence of 250 and 500 mM of D-sorbitol to understand the possible impact of the additive on conformational changes. Figure 7 shows the time evolution of fluorescence intensity also including that in absence of the polyol. While fluorescence keeps increasing over time in a similar manner for all three cases, a lower value is observed at a given time point for higher D-sorbitol concentration, indicating that unfolding is inhibited to a certain extent.

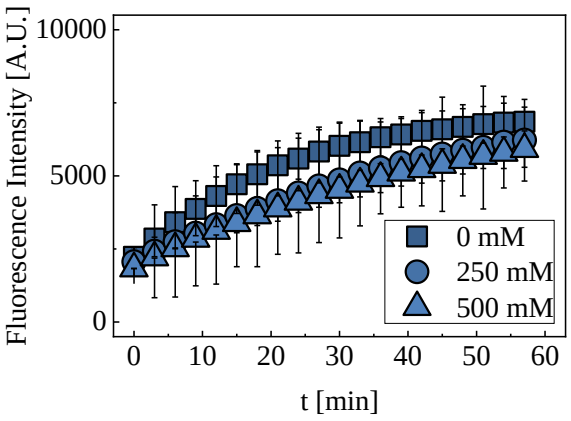

Figure 7 - Time evolution of fluorescence intensity of 0.25 g/L mAb-1 solutions in 50 mM sodium citrate and pH 3.5 and 50 mM I, in the presence of 0, 250 and 500 mM of D-sorbitol. ANS was added to the solutions in ten-fold molar excess with respect to the protein. Error bars represent 90% confidence intervals for the mean, which was calculated by averaging results from three independent measurements.

3.1.3 Aggregation Studies Following Neutralization after Temporary Exposure to Low pH

From the results presented above, the suggested explanation is that exposure to acidic conditions may contribute to the process of aggregation by exposing part of a mAb's hydrophobic residues upon denaturation. Virus inactivation is, however, composed of two steps: incubation at low pH and neutralization to neutral or mildly acidic pH values. This pH-shift upon neutralization will reduce the electrostatic repulsion found dominant during low pH incubation. Therefore, the effect of pH-shift stress on the investigated mAbs was studied systematically to understand how it might lead to formation of mAb aggregates during the VI step.

3.1.3.1 Impact of pH and Incubation Time

Table 3 contains the set of experiments performed to evaluate aggregation after neutralization, resulting in a combination of different pH values and incubation periods. Antibody solutions were analysed by SEC coupled to MALS as well as batch DLS over the course of many hours after neutralization. The final pH and ionic strength were always 5.0 and 100 mM, respectively.

Table 3 - Conditions investigated for the study of antibody aggregation induced by pH-shift stress. The antibodies were incubation in 50 mM sodium citrate and 50 mM ionic strength, and, at the end of low pH incubation, solutions were always neutralized to pH 5.0 and 100 mM ionic strength. Solution pH was within ± 0.1 of target for all conditions.

pH [-]	Duration [min]
2.5	20
	40
	60
3.0	20
	40
	60
3.5	20
	40
	60

Figure 8 shows examples of SEC chromatograms obtained for neutralized solutions of either mAb over time. Prior to neutralization, antibodies were incubated for 40 minutes in 50 mM sodium citrate pH 3.0 and 50 mM ionic strength. Qualitative shape of chromatograms is similar with appearance of earlier eluting species over time indicating aggregate formation. Relative amounts of mAb monomer and aggregates were then determined based on areas under the curve.

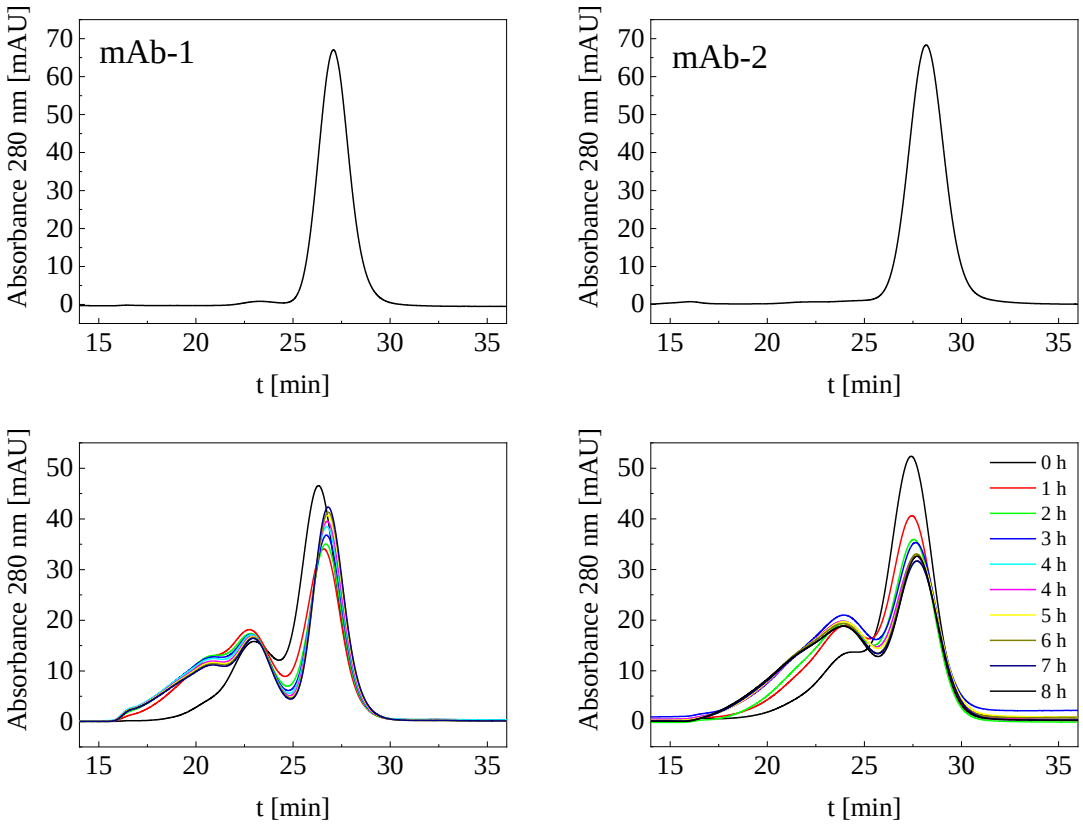

Figure 8 – SEC chromatograms of MAb-1 (left) and Ab-2 (right) solutions. On the top: chromatograms of unstressed antibody, in 10 mM Sodium Phosphate pH 7.0; On the bottom: chromatograms measured after neutralization to pH 5.0 and 100 mM ionic strength. Prior to neutralization, the proteins were incubated for 40 minutes at pH 3.0 and 50 mM ionic strength. Once every hour, 50 µL of neutralized solution were injected into the column.

In the case of mAb-1, monomer content dropped within the first two hours after neutralization, followed by a slight increase. This suggest that monomer loss does not stop shortly after neutralization, but monomers keep getting incorporated into aggregates over time, while dissociation of monomers from aggregates becomes relevant after longer times. Further, analysis of the MALS data gives insight about the identity of the peaks in the chromatograms: (i) It confirms that the main peak in the chromatograms eluting after approximately 25 minutes represents the monomeric protein; (ii) the measured molecular weight of the peak eluting next to the monomer's corresponds to mAb-1 dimers; (iii) that after the first hour, the average molecular weight at the respective peak maxima does not represent integer multiples of the monomer's molecular weight anymore, which is explained by co-elution of different oligomeric species due to lack of chromatographic resolution.

For mAb-2, a more significant decrease in monomer peak area (and concomitant increase in aggregate peak area) is observed over time after neutralization. Contrary to mAb-1, area of the monomer peak decreases monotonically as function of time.

Figure 9 sums up the results for mAb-1 and all conditions presented in *Table 3*. In addition to fraction of monomeric antibody, it presents average hydrodynamic radius of neutralized mAb-1 solutions as measured by DLS.

Monomer content drops significantly within the first hour after neutralization for all investigated pHs and incubation times. Further, mild increase in monomeric fraction is detected after the third hour. As for the $<R_h>_z$, it steeply increases in the first two hours, after which it slightly decreases for the remaining time of analysis. In accordance with this observation, are the results obtained by MALS (Appendix 2), that indicate an increase in the average molecular weight of protein aggregates in the first hours until reaching a maximum. Important to mention, mass recovery in chromatographic analysis was constant and always above 90%, meaning that there was no significant protein loss either by adsorption of the protein to the stationary phase or by formation of aggregates too large to pass through the chromatographic bed. Regarding the effect of pH, the measured monomer fraction after neutralization is smaller at every time point for pH 2.5 and 3.0 compared to pH 3.5. At pH 3.5, fraction of residual monomer never is below 70%, whereas at pH 2.5 and 3.0 it decreases to approximately 50%. The contrast between pH 3.5 and the other two conditions is also represented in the DLS results.

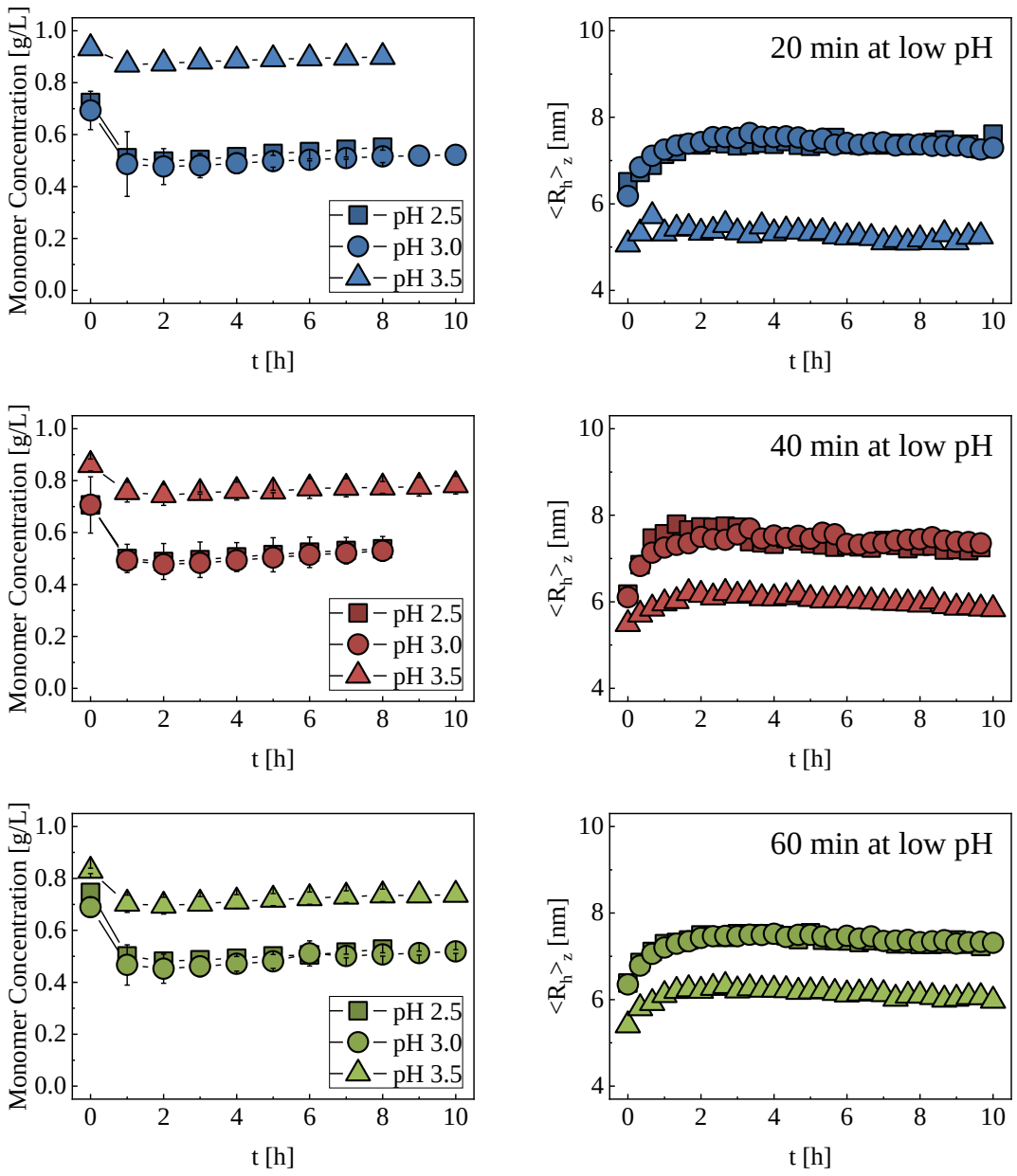

Figure 9: Results obtained for 1 g/L mAb-1 solutions after neutralization to pH 5.0 and 100 mM ionic strength. Prior to neutralization, the protein was incubated at pH 2.5 (squares), 3.0 (circles) and 3.5 (triangles). Each row represents a different incubation time: upper row – 20 minutes; middle row – 40 minutes; lower row – 60 minutes. The left side column shows the monomer content as a function of time through the analysis by SEC; the right side column shows the time evolution of the average hydrodynamic radius measured by DLS. Error bars represent 90% confidence intervals for the mean, which was calculated by averaging results from three independent measurements.

For mAb-2, results are presented in Figure 10. The monomer content decreases monotonically as a function of time after neutralization, while the $\langle R_h \rangle_z$ increases until reaching a plateau. Also for mAb-2, chromatographic mass recovery was above 90% in all experiments. Overall, aggregation behaviour is similar to that of mAb-1. However, a few differences must be acknowledged: (i) the extent of aggregation after neutralization when mAb-2 is incubated at pH 3.5 is very small, while in the case of mAb-1 there is an evident decrease in monomer content and increase in average hydrodynamic size; (ii) monomer loss in the first hours happens at a slower pace for mAb-2 than for mAb-1; (iii) while for mAb-1 a small

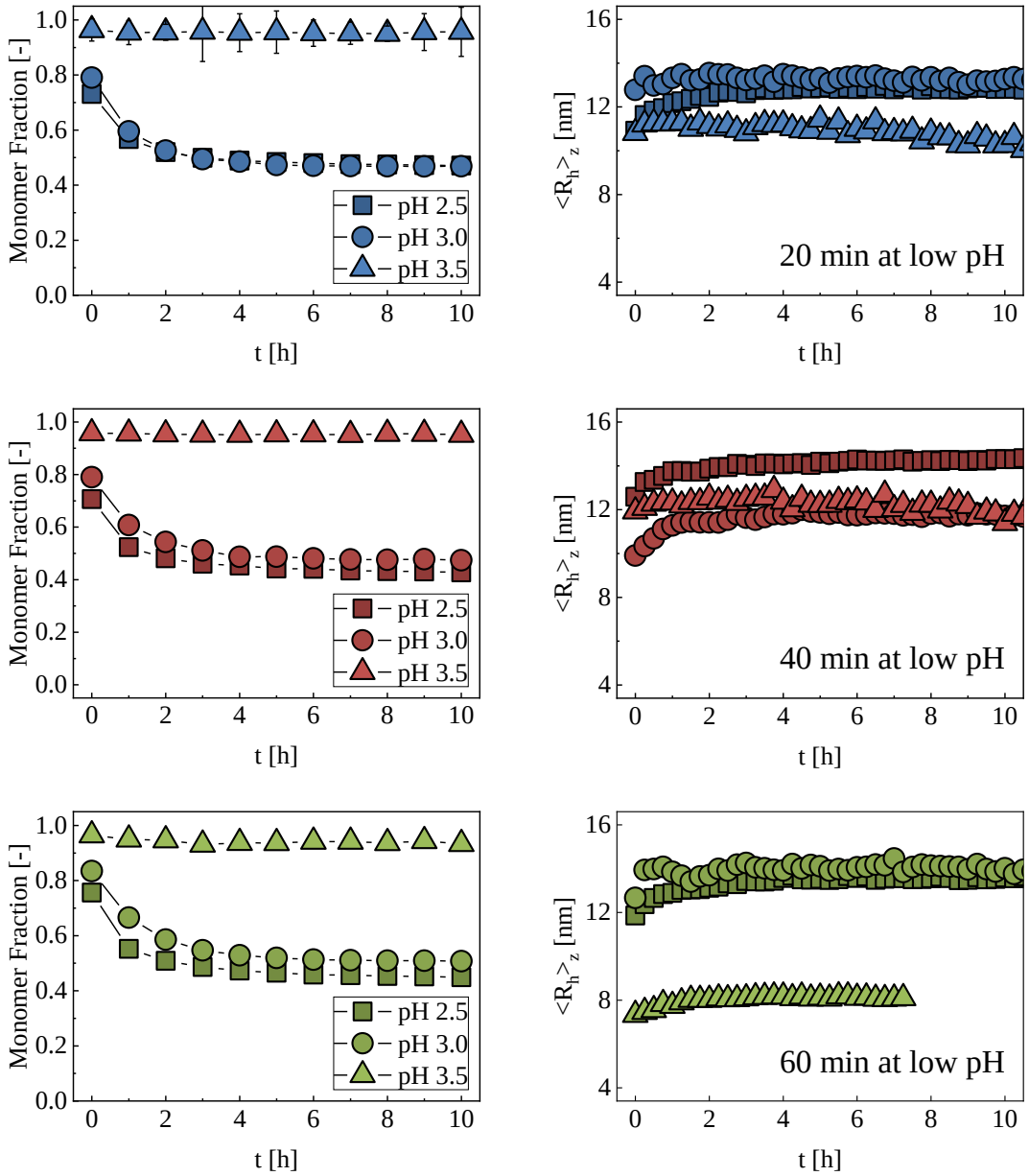

Figure 10: Results obtained for 1 g/L mAb-2 solutions after neutralization to pH 5.0 and 100 mM ionic strength. Prior to neutralization, the protein was incubated at pH 2.5 (squares), 3.0 (circles) and 3.5 (triangles). Each row represents a different incubation time: upper row – 20 minutes; middle row – 40 minutes; lower row – 60 minutes. The left side column shows the monomer content as a function of time through the analysis by SEC; the right side column shows the time evolution of the average hydrodynamic radius measured by DLS. Error bars represent 90% confidence intervals for the mean, which was calculated by averaging results from three independent measurements.

increase in monomer content could be observed in the last hours, for mAb-2 monomer content is decreasing throughout the analysis.

Based on ANS fluorescence results, surface hydrophobicity of mAb-2 increases even during incubation at pH 3.5, which implies some degree of conformational changes (Figure 6). The fact that no significant monomer loss is observed after neutralization might indicate that these changes are not sufficient to lead to subsequent aggregation. The evolution of $\langle R_h \rangle_z$ after neutralization is very similar for pH 2.5 and 3.0. Further, there is no clear dependence on incubation time. On the other hand, peculiar behaviour is

observed for pH 3.5. Based on the limited extent of aggregation determined by SEC, one would expect negligible increase of $<R_h>_z$ over time. For 60 minutes incubation that is actually the case. However, for 20 and 40 minutes of incubation, much higher values of $<R_h>_z$ after neutralization are observed (comparable to those for pH 2.5 and 3.0). This result might be explained by the presence of a small number of large impurities (*e.g*. aggregates formed by residual HCPs). Such particles would strongly scatter light and their presence would interfere with the obtained average value for R_h.

Concerning the impact of low pH incubation time, in the case of mAb-1, for both pH 2.5 and 3.0 no noteworthy difference in the evolution of monomer fraction and average hydrodynamic radius can be detected between incubation for 20, 40 or 60 minutes. This is in accordance with the results obtained from ANS fluorescence. Following the hypothesis that a new conformational equilibrium is reached within a relatively short amount of time, the fraction of partially unfolded mAb after 60 minutes should not be higher than after 20 or 40 min. Indeed, a dependence of the dynamics after neutralization on the duration of the previous low pH incubation is observed at pH 3.5 in line with this reasoning and ANS fluorescence results.

In the case of mAb-2, incubation time has limited impact on aggregation behaviour after neutralization for all conditions. Only for pH 3.5, the extent of monomer loss at a given time after neutralization has slight dependence on incubation time at low pH, but due to the very small amount of aggregation observed for that pH, differences are not statistically significant.

Overall, substantial evidence has been collected suggesting that mAb aggregation does not occur during the low pH hold but only after sharply raising pH. Furthermore, after raising solution pH back towards neutral/mildly acidic values, mAbs rapidly form aggregates and subsequently the aggregation rate decreases over the course of several hours. This can be explained by strong electrostatic repulsion between mAb molecules at low pH, which prevents self-association. Nevertheless, mAb molecules (partially) denature/unfold under acidic conditions. When pH and ionic strength are subsequently increased, repulsion is reduced and denatured mAb molecules can start to aggregate.

Further, results indicate that duration of the hold at low pH can have an impact on the extent of aggregation that is observed after neutralization. This can be the case for low pH conditions that lead to progression of partial denaturation of mAb molecules over time. Nevertheless, it is important to point out that a slower progression of partial unfolding at low pH is not directly associated with a dependence of mAb aggregation after neutralization on low pH incubation time. Such a case is mAb-2 incubated at pH 2.5. ANS fluorescence keeps increasing over the course of 1 hour, but the amount of aggregate formation after neutralization is independent of incubation time at pH 2.5. Compared to pH 3.0, additional hydrophobic binding sites for ANS on the mAb surface might become gradually available over time at pH 2.5. However, those sites seem to be unimportant in the aggregation process. Either, because they are buried very fast within the mAb's interior when pH is raised or because they are sterically not available for formation of inter-molecular contacts.

3.1.3.2 Impact of Neutralization Velocity

In the experiments described above, the neutralizing solution was added all at once as soon as the desired incubation time at low pH had passed. However, the velocity at which the neutralization is conducted might be an important process parameter, which might influence the aggregation process triggered by the temporal acid exposure. A slower increase of pH and ionic strength could affect the rate of aggregation and refolding events in a different way.

For this study, all other parameters were kept identical to those of the previous experiments. Neutralization velocities of 0.5 and 1 mL/h were investigated for both mAbs after 40 minutes incubation in 50 mM sodium citrate pH 3.0 and 50 mM ionic strength. The 500 µL of neutralization solution were added continuously under mild stirring to neutralize the mAb solutions to pH 5.0 and 100 mM ionic strength.

Figure 11 and Figure 12 show the time course of monomer content and average hydrodynamic radius for mAb-1 and mAb-2, respectively. Similar results were obtained at all three neutralization velocities for both mAbs, which indicates that the velocity of neutralization does not have a significant impact on the aggregation process.

Two possible explanations can be put forward for that observation. Either unfolding events that are triggered by acid exposure are irreversible simply by increasing pH. Recently, some evidence for this hypothesis has been put forward in literature [85]. The other explanation would be that the pH at which the mAb has significantly folded back is higher than the pH at which the energy barrier for self-association becomes comparable to the thermal energy of the mAb molecules. Further experiments would have to be conducted to reach a definitive conclusion.

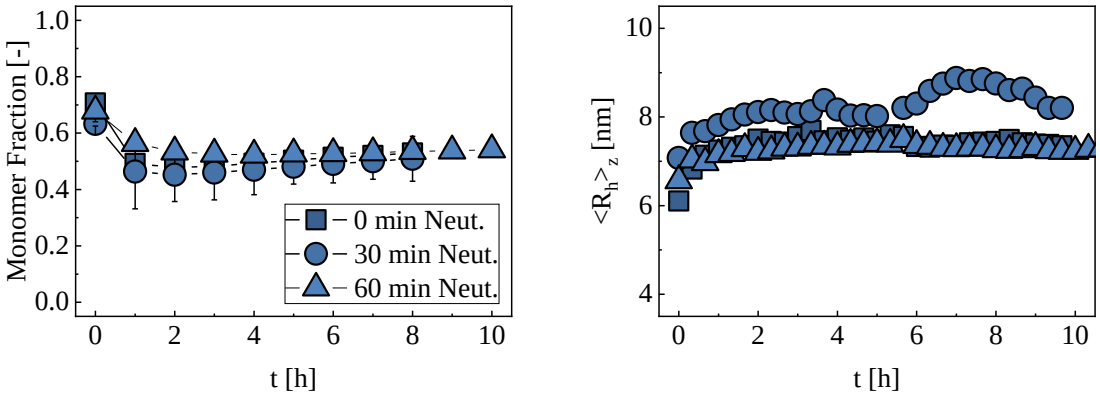

Figure 11 - Results obtained for 1 g/L mAb-1 solutions after neutralization to pH 5.0 and 100 mM ionic strength. Comparison between the effect of instantaneous (squares) and slow (circles and triangles) addition of the neutralization solution. Prior to neutralization, the protein was incubated for 40 minutes in 50 mM sodium citrate pH 3.0 and 50 mM ionic strength. The left side shows the monomer content as a function of time through the analysis by SEC; the right side shows the time evolution of the average hydrodynamic radius measured by DLS.

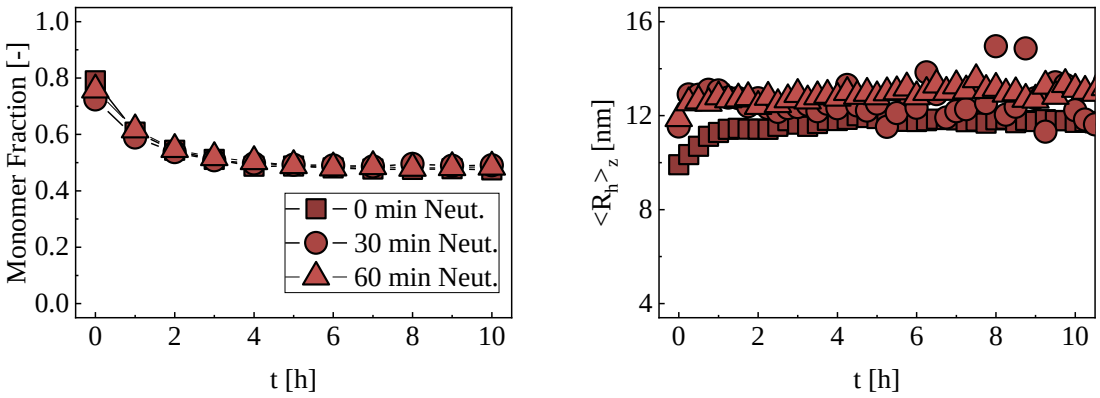

Figure 12 - Results obtained for 1 g/L mAb-2 solutions after neutralization to pH 5.0 and 100 mM ionic strength. Comparison between the effect of instantaneous (squares) and slow (circles) addition of the neutralization solution. Prior to neutralization, the protein was incubated for 40 minutes in 50 mM sodium citrate pH 3.0 and 50 mM ionic strength. The left side shows the monomer content as a function of time through the analysis by SEC; the right side shows the time evolution of the average hydrodynamic radius measured by DLS.

3.1.3.3 Impact of Additives

Based on the results presented in the Chapter 3.1.2, the effect of D-sorbitol on mAb aggregation after a pH-shift was investigated as well.

For this study, both mAbs were incubated in low pH with 250 and 500 mM of D-Sorbitol. The effect was tested under different conditions for the two mAbs given the results obtained above regarding the aggregation behaviour of mAb-1 and mAb-2. For the former, experiments were conducted in 50 mM sodium citrate pH 3.5. Under this condition, aggregation takes place, but not to an extreme extent. Further, it represents a realistic condition for actual manufacturing. For mAb-2 however, there would be no use in testing the effect of D-sorbitol in that case, as no significant aggregation occurs. Therefore, experiments were only performed at pH 3.0. Figure 13 and Figure 14 present the time evolution of the monomer content and average hydrodynamic radius upon neutralization for mAb-1 and mAb-2, respectively.

In the case of mAb-1, presence of D-sorbitol does have an impact on aggregate formation after neutralization. At a given point in time after neutralization, residual monomer fraction increases as function of D-sorbitol concentration. In the presence of 500 mM D-sorbitol the sharp initial decrease in monomer content immediately after neutralization does not occur. Further, values of $\langle R_h \rangle_z$ are lower for 500 mM D-sorbitol compared to the other two conditions. These results are consistent with ANS fluorescence in the presence of D-sorbitol (see Figure 7). If denaturation is inhibited in part by the polyol, then the degree of mAb aggregation after neutralization is bound to change with lesser aggregates being formed.

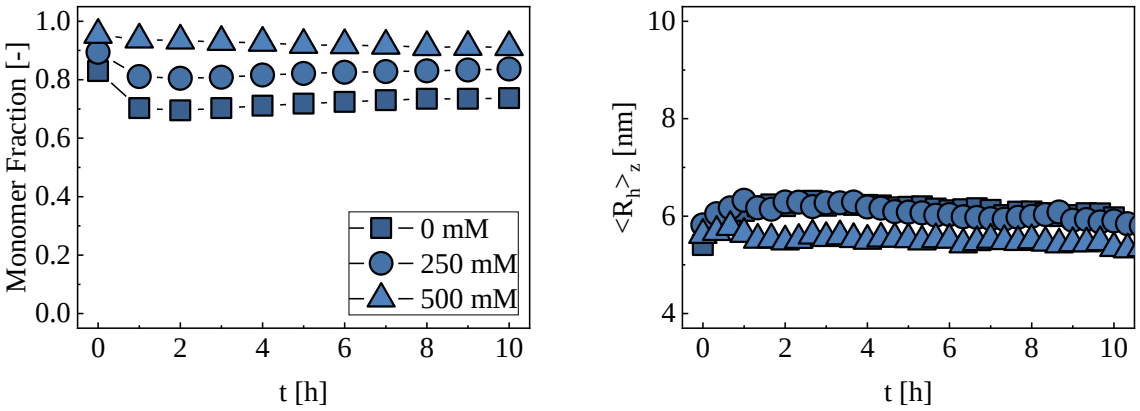

Figure 13 - Results obtained for 1 g/L mAb-1 solutions after neutralization to pH 5.0 and 100 mM ionic strength. Comparison between the effect of 0, 250 and 500 mM of D-sorbitol present during low pH incubation. The protein was incubated for 60 minutes in 50 mM sodium citrate pH 3.5 and 50 mM ionic strength. The left side shows the monomer content as a function of time through the analysis by SEC; the right side shows the time evolution of the average hydrodynamic radius measured by DLS.

Concerning mAb-2, presence of D-sorbitol did not show an impact on the aggregation process under the investigated conditions. Monomer content and average hydrodynamic radius follow the same path after neutralization for all three experiments. Under more harsh conditions, presence of moderate concentrations of D-sorbitol seems to be insufficient to reduce the amount of mAb denaturation and aggregate formation.

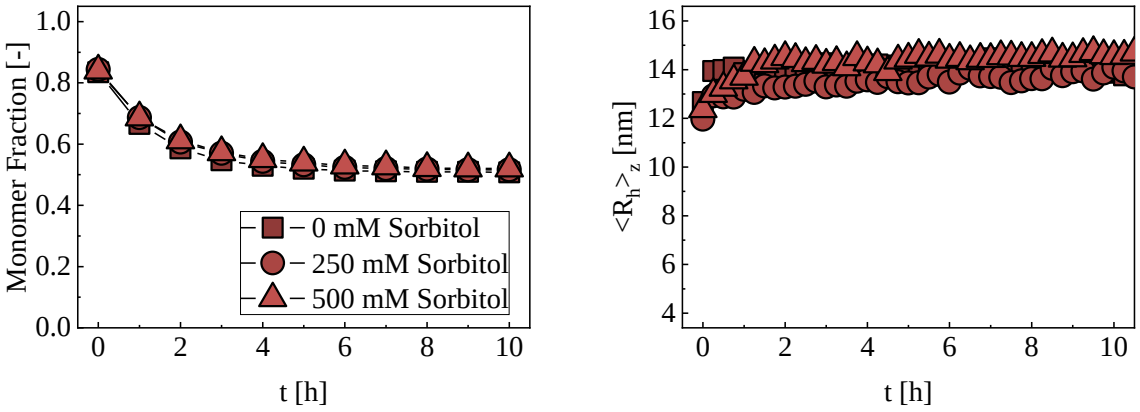

Figure 14 - Results obtained for 1 g/L mAb-2 solutions after neutralization to pH 5.0 and 100 mM ionic strength. Comparison between the effect of 0, 250 and 500 mM of D-sorbitol present during low pH incubation. The protein was incubated for 60 minutes in 50 mM sodium citrate pH 3.0 and 50 mM ionic strength. The left side shows the monomer content as a function of time through the analysis by SEC; the right side shows the time evolution of the average hydrodynamic radius measured by DLS.

3.2 Implementation of Raman Spectroscopy for Monitoring and Control

One of the most important critical quality attributes (CQA) in antibody manufacturing are aggregate forms of the product. Under the scope of continuous processing, it's monitoring and control in real time is of extreme importance. As shown in Figure 15, aggregate monitoring should start in the VI step, where the aggregates are supposedly formed. This would feed information to the following polishing step, as feedforward control, where the monomer is separated from the aggregates. During the polishing step, one would monitor the elution from the column to control the level of the aggregates. The purpose of

such control would be to obtain the highest possible yield without compromising purity of the final product, guaranteeing the CQA profile was within regulatory specifications.

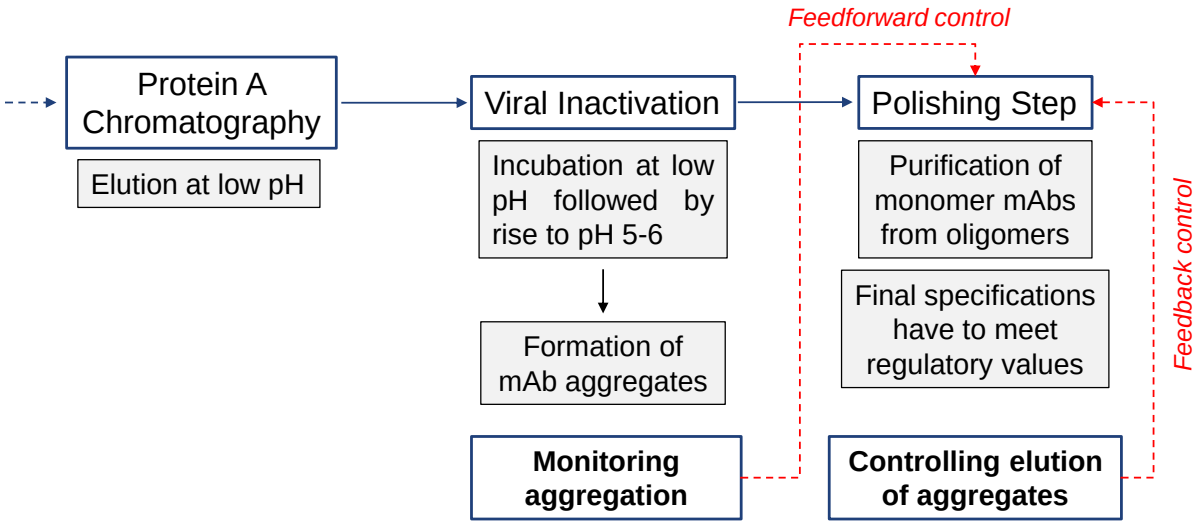

Figure 15 – Scheme for the implementation of monitoring and control of aggregates in a DSP of antibody manufacturing. Following elution from protein A, during the VI step aggregation of the product is monitored and this information is used for the feeding of the column in the polishing step. During the latter step, control of aggregate levels occurs during elution.

Given the results presented above for mAb-1 and mAb-2, it can be concluded that monitoring the VI step for these two antibodies does not provide any direct knowledge concerning aggregate formation, as this only occurs after neutralization.

Concerning the control of aggregate levels during the polishing step, a flow-cell was built to adjust to the process set-up, presented in Figure 16. It has been successfully implemented in capture Simulated Moving Bed (SMB) for online monitoring of protein concentration (Feidl et al., in submission). The present goal is to use the device for aggregate monitoring. It circumvents some of DSP challenges, such as reduced volume streams and short processing times. It would enable the quantification of eluted product from a chromatographic effluent and therefore online control. While optimizing this device and before its integration in a bioprocess, the flow-cell was used for offline batch mode measurements of antibody solutions, for the assessment of Raman as a spectroscopic tool capable of detecting aggregates.

3.2.1 Flow-cell Development and Optimization

The flow-cell set-up is composed of several separate units that, when assembled together, allow for the solution inside to be measured without any interference from visible light. It is composed of two pieces in each side, that hold a reflector and a non-contact probe, and a middle piece, where the sample flows through. The path for the sample is a 3 mm diameter channel, containing an inlet and an outlet. The Raman signal emitted from the samples needs to return to the probe, where the detector is located. For increased signal intensity, on the opposite side of the channel a mirror reflects back to the detector the signal that goes in its direction.

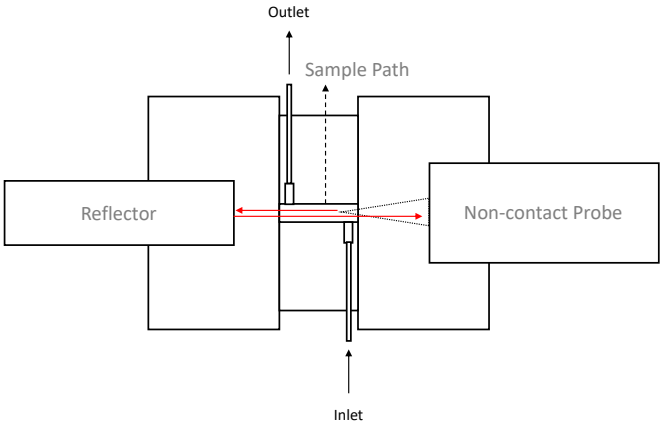

Figure 16 - Scheme of the original flow-cell set-up, composed of three main blocks: reflector holder, non-contact probe holder and channel middle piece. The sample's path is aligned with the probe and the reflector and is connected to the outside by two capilaries, an inlet and an outlet.

As a device to be implemented in a continuous process, the flow-cell presents several advantages: (i) minimized dead volume; (ii) increased signal enhancement due to the presence of the reflector; (iii) optimized flow path, which makes it suitable for chromatographic applications; (iv) use of non-contact probes, allowing a single-use approach; (v) modular design, enabling the adjustment of both the reflector and the non-contact optic to the necessary distances; (vi) grants the possibility of introducing additional optical supplements, such as filters.

The flow-cell was firstly tested with isopropanol (IPA), which is known for being a Raman active chemical. This presented the opportunity to define a measurement procedure to later be applied in protein experiments. Aqueous solutions with IPA content ranging from 0 to 100% were produced and measured in a random order in batch mode. Triplicate measurements were performed for three of the samples. Before initiating the measurements, the non-contact probe and the reflector were fixed. Then, after the insertion of the first sample, the acquisition time of each spectrum was defined according to its Pixel Fill (Chapter 2.4.1). Between each sample, the channel was cleaned with ultrapure water and dried.

Figure 17 shows the raw spectra of the measured samples and a score plot of the two first principal components obtained after a Principal Component Analysis (PCA) performed on the acquired data. Prior to the PCA, the data was pre-treated according to the methods described in Chapter 2.4.2. In theory, higher peaks in the spectra would be associated with higher IPA content samples. One can observe the increase in height of each sample, which goes in accord with that statement. As for the score plot, two observation can be made: (i) it is possible to identify a very defined trend of increased IPA concentration; (ii) the triplicate measures have the same scores for all three samples, being almost indistinguishable.

From an experimental point of view, one can say the procedure seems to produce accurate results, and thus it was adopted for the protein experiments.

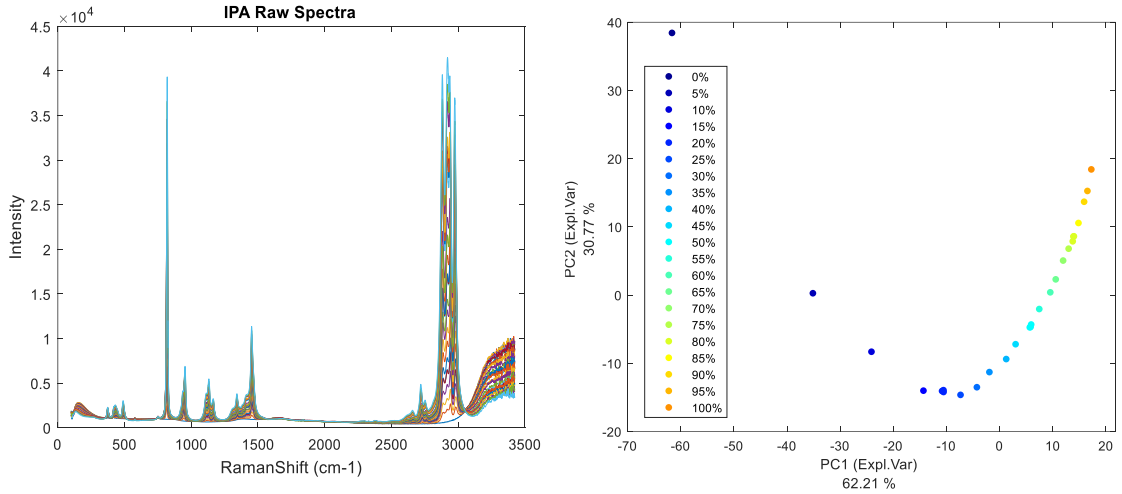

Figure 17 - On the left: Raman spectra of IPA samples ranging from 0 to 100% content in ultra-pure water, measured in a random order with the flow-cell, 10 spectra per sample were measured with an acquisition time of 2 seconds; on the right: score plot of PC1 and PC2 of the data obtained from a PCA model.

Nevertheless, protein solutions do not have such a high Raman signal and it was necessary to increase the signal's intensity as much as possible to guarantee that molecular information was captured at its maximum. Figure 18 shows a single spectrum of mAb-1 in an aqueous solution measured in the flow-cell with a 10 mm channel. If compared with a water spectrum (Appendix 4) and knowing where biological information would be expected (Chapter 1.3.1), it is safe to say that protein bands are not as visible in the spectrum as it would be desirable. Higher signal intensity would probably be advantageous. However, the high intensity band above 3100 cm^{-1} might constitute a limitation for increased signal intensity, given the possibility saturation of the CCD detector.

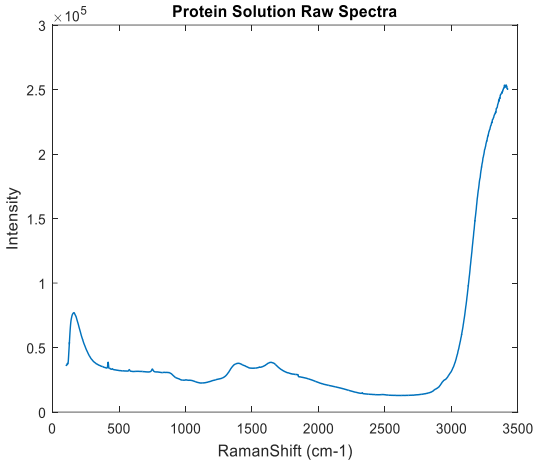

Figure 18 - Raman spectrum of mAb-1 in ultra-pure water, obtained with an acquisition time of 30 seconds.

Concerning the first observation, as instrument wise there was not the possibility at the moment to increase laser power, the only option for increasing intensity was by optimization of the device. As already explained, Raman is a weak phenomenon and the number of particles presenting Raman scattering behaviour instead of Rayleigh's is small. One way to increase the probability of Raman scattering is by increasing the chances of the laser hitting protein molecules. As higher protein concentration is not a general solution in a real bioprocess, a possibility was increasing the sample's volume inside the flow cell. This was accomplished by expanding the path's length inside it. With this purpose, two longer pathways were tested, 20 and 30 mm.

As for the second observation, it is known that no biological information is captured above 3100 cm^{-1} [48][51][52][53]. Thus, one can assume that the high intensity band present in that region of the spectrum is not related with the sample itself, and so it was regarded as background noise. One of the implications of this high intensity band is the fast saturation of the detector. A spectrum could not be acquired for a longer period of time given the extremely high intensity of this band. Longer exposure length, the detector saturated earlier than what would be desirable, and less sample-related information would be obtained. According to the manufacturer, above 80% of detector saturation information might be lost. To extend the acquisition time of each spectrum, a notch filter was tested on the flow-cell between the non-contact probe and the sample.

Channel Length Tests

Figure 20 shows normalized spectra of 70% IPA and antibody solutions, measured with all three channels. Being IPA such a Raman active compound, the exposure length necessary to reach 50% of detector saturation with the 10 mm path length was 15 seconds (s). As for the antibody solution, an exposure length of 60 s was necessary. Changing the channel of the flow-cell, both IPA and antibody were measured with the same exposure time as before. By observing the plots, one can see that the background in the spectra measured in the 10 mm channel is different from the background of spectra measured with both 20 and 30 mm channel. Additionally, the spectra corresponding to the 20 mm channel measurements are not visible, because they are superimposed by the 30 mm spectra. This indicates that changing between these 2 lengths does not produce any visible differences in the spectra. While an IPA spectrum has very distinct peaks and can almost hide background contribution, an antibody solution does not. A few bands in the 10 mm spectrum stand out below wavenumber 1000 cm^{-1}, which seem to be repressed with the longer channels. Although biological information could be present in that region, these bands can also be identified in a water spectrum, which led to the conclusion that they were not significant. Regardless, a conclusion about the effect on mAb is very difficult as no distinct peaks can be observed.

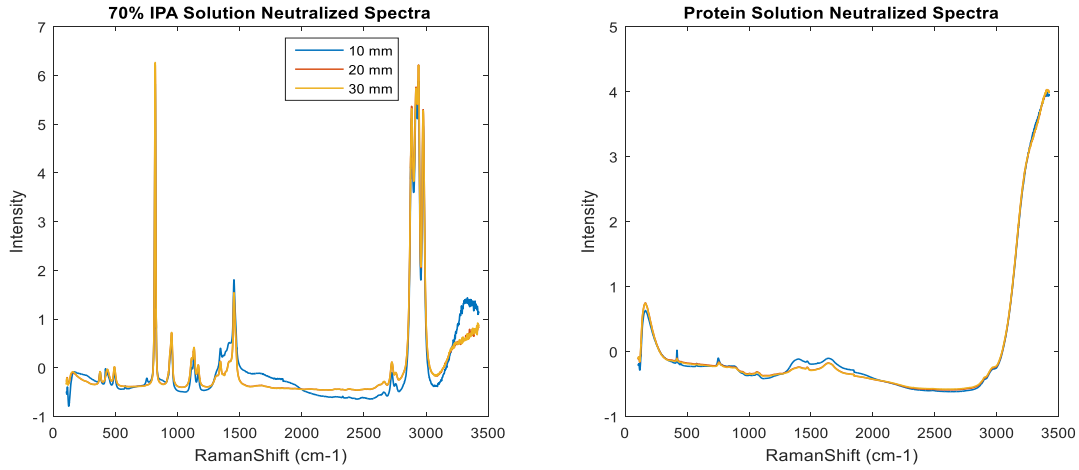

Figure 20 - On the left: normalized Raman spectra of 70% IPA in ultrapure water, 10 spectra per sample were measured and averaged, with an acquisition time of 15 seconds. On the right: normalized Raman spectra of mAb-2 in ultrapure wate, 10 spectra per sample were measured and averaged, with an acquisition time of 30 seconds. Comparison between spectra measured in all three channel lengths, 10, 20 and 30 mm.

Table 4 presents the degree of saturation reached for these measurements, and additional ones. In the presence of both IPA and antibody, for 15 and 60 s, respectively, for the 20 mm channel, the detector's saturation reaches above 80%, which is the recommended superior limit. As it would be expected, the higher the path length, the greater is the signal's intensity, which can be beneficial for the acquisition of chemical information. However, when a shorter period of acquisition is tested, one can see that the intensity signal difference between 20 and 30 mm channel long is not significant. This might be related with the fact that, for too long a channel, the signal disperses too much. Additionally, one might have to consider the void volume that is being introduced once the flow-cell is integrated in a chromatographic step. A compromise between both effects was met by choosing the 20 mm long channel to proceed with the experiments.

Table 4 - Saturation level of Raman CCD detector upon acquisition of IPA or antibody spectra, in various flow-cell set-ups and acquisitions times.

	IPA		Antibody	
Exposure Time (s)	2	15	30	60
10 mm	~50%	~50%	~40	~50%
20 mm	~50%	>80%	~70	>80%
30 mm	~50%	>80%	~70	>80%

Notch Filter Implementation

Notch filters are band-rejection filters, designed to transmit most wavelengths with little intensity loss, while attenuating the light of a specific wavelength range [86]. A scheme of the blocking process within the flow-cell is exemplified in Figure 21. The rationale behind implementing such a filter is allowing for a longer exposure time for the acquisition of a spectrum without saturating the detector. Available notch filters exist for specific wavelengths at which transmission is almost zero, and a blocking region from that centre wavelength with increasing transmission. The properties regarding the chosen filter are presented in Table 5. A centre wavelength of 1064 nm corresponds to the Raman wavenumber 3340 cm^{-1}. A blocking region with a full width at half maximum (FWHM) of 44 nm represents a blocking range at 50% transmission of 1038 – 1074 nm, *i.e.* 3100 – 3425 cm^{-1}.

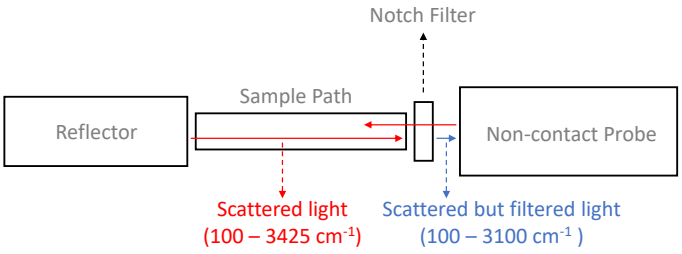

Figure 21 - Scheme of flow-cell's inside with a specific Notch filter implemented. The Raman signal emitted from the sample is scattered and before reaching the detector goes through a band rejection filter which blocks a certain wavenumber region from reaching the detector.

Table 5 - Notch filter properties of centre wavelength and blocking range of transmission.

Centre Wavelength	FWHM of Blocking Region
1064 ± 2 nm	44 nm

Spectra with and without the filter were acquired for both IPA and antibody solutions. Figure 22 shows the resulting averaged spectra after acquisition of 5 spectra in each case. The averaging of the acquired

spectra is performed to decrease the signal to noise ratio (SNR). By comparing the IPA spectra, although the decrease in intensity after 3100 cm^{-1} can be observed, it does not affect in any way the signal's intensity, as IPA bands are significantly stronger. However, for the antibody solution, the decrease of the aimed band is considerable. Concerning the saturation of the detector, while for the measurement with filter a saturation of 70% was reached under 30 s, the implementation of the filter allowed an increase of 15 s in exposure time to reach the same saturation. Nevertheless, there does seem to be a drawback, as the background changes in the presence of the filter. It manifests, on one hand, a decrease in intensity and, on the other, the appearance of bands not related with chemical information below 1000 cm^{-1}. Thus, with only these experiments it is not possible to understand if, for the measurement of antibody and its aggregates, the implementation of a notch filter would be beneficial. A more extensive experiment was conducted and is explained below.

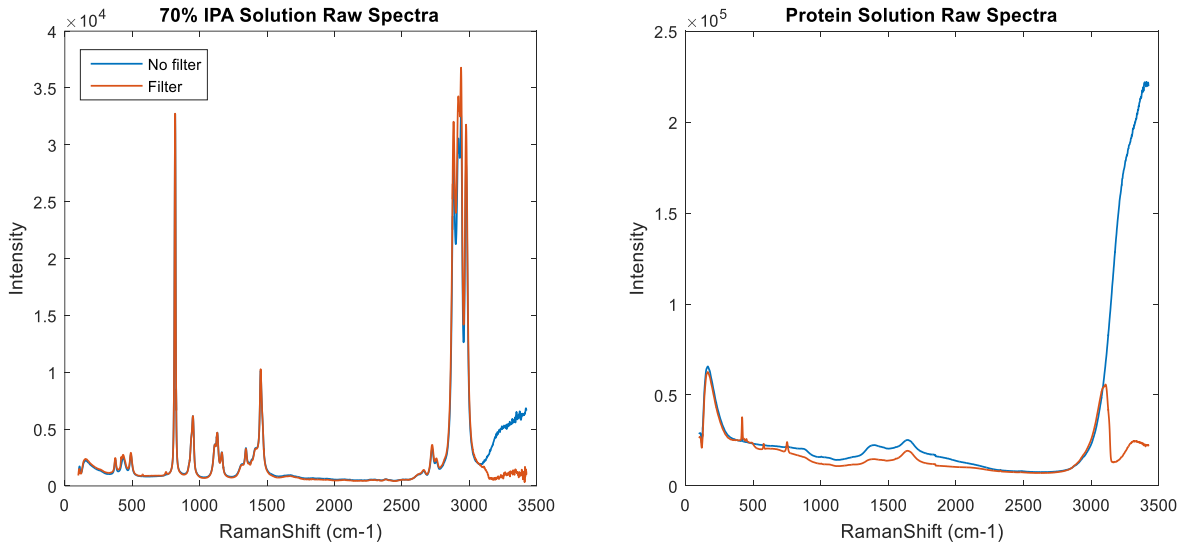

Figure 22 - On the left: raw Raman spectra of 70% IPA in ultrapure water, 10 spectra per sample were measured and averaged, with an acquisition time of 15 seconds. On the right: raw Raman spectra of Ab-2 in ultrapure water, 10 spectra per sample were measured and averaged, with an acquisition time of 30 seconds. Comparison between spectra measured with and without filter implemented on the flow-cell set-up.

3.2.2 Data Acquisition

To determine if Raman was able to detect mAb aggregates or even monomeric protein, samples of known monomer and aggregate concentration of both mAb-1 and mAb-2 were measured in batch mode with the 20 mm long flow-cell. Multiple samples were generated by mixture of appropriate volumes of original protein stock solutions of different total antibody concentration and aggregate content. In a first experiment, both mAbs were tested within a large range of antibody aggregate content. Then, the experiment was repeated for mAb-2 with a set of samples representing a narrower and more realistic range of aggregate content. For this last experiment, every sample was measured both with and without the notch filter.

Broad-range Aggregate Content Experiment

Within the original protein stocks, solutions of high and low concentration were produced with both high and low content of aggregates. Table 6 shows the values of each stock for both mAb-1 and mAb-2 in

the first experiment conducted. High aggregate concentration was met by inducing pH-shift stress on the molecules. The mAbs were subjected for 1 hour to 50 mM sodium citrate in pH 2.5 and I=50 mM, after which the samples were neutralized to pH 5.0 and I=100 mM. To make sure the buffer molarity was identical in all four solutions, the stocks of low aggregate concentration were prepared by diluting the antibodies to the conditions of neutralization of the high content ones. For reference values, all samples were analysed by SEC.

Table 6 - Total antibody concentration and aggregate content in the four original stocks produced for mAb-1 and mAb-2.

	mAb-1				**mAb-2**			
Stock	1	2	3	4	1	2	3	4
Total Protein Concentration [g/L]	0.92	0.95	3.88	3.87	0.94	0.86	3.84	3.67
Aggregate Content [%]	2	43	2	68	2	49	2	58

A design of experiments (DoE) based on a D-optimal (Dopt) design was conducted to enable a complete decorrelation between protein concentration and aggregate content. Although the aim was the prediction of monomer and aggregate concentration, at the time these experiments were performed the focus was still on the percentage of aggregates in the samples. Matlab functions *cordtexch* and *daugment* were used for this purpose. The former resorts to a coordinate-exchange algorithm to generate a Dopt design, in the present case applying a linear regression model with constant, linear and interaction terms. The latter uses a coordinate exchange algorithm to add runs to an existing experimental design. A quadratic model was used for this. This combination allowed for the generation of a statistically significant number of samples for a first experiment. As can be observed from Figure 23 and Figure 24, no significant correlation between protein concentration and aggregate content exists between samples, and a wide distribution of aggregate concentration was met.

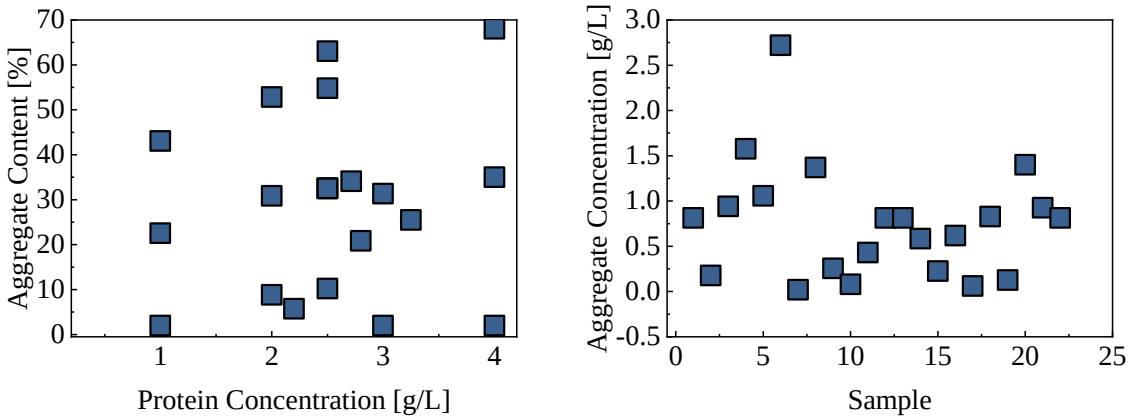

Figure 23 – D-opt design result for mAb-1. On the left, the relation between aggregate content and protein concentrations for each sample. On the right, the distribution of mAb aggregate concentration per sample.

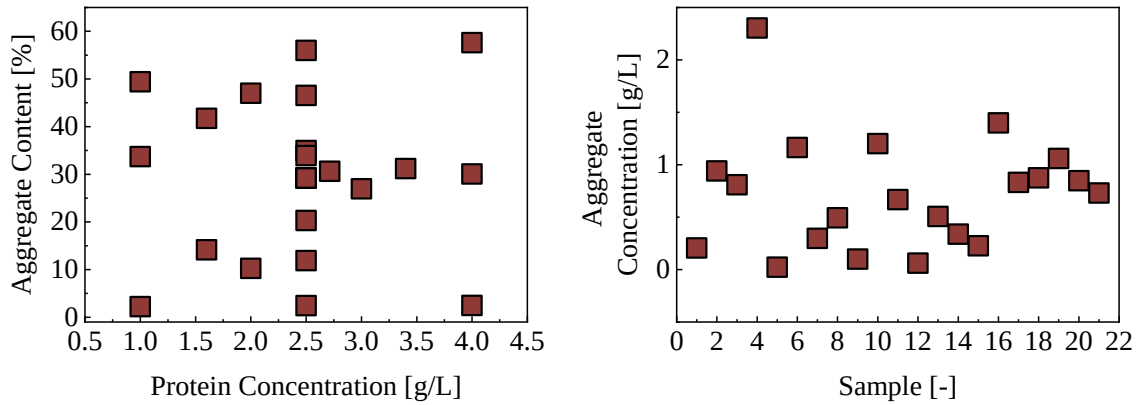

Figure 24 - D-opt design result for mAb-2. On the left, the relation between aggregate content and protein concentrations for each sample. On the right, the distribution of mAb aggregate concentration per sample.

For each experiment, the reflector and the non-contact probe were adjusted in such a way that the detector would receive as much intensity as possible, without saturating. Given the fact that the experiments were performed in different days slight changes in the environment might have occurred, such as the position or direction of the laser's cable, which is known to have an influence. Having this in mind, different acquisition times were necessary to achieve between 65 and 75% of the detector's saturation. For mAb-1 measurements, the acquisition of each spectra took 30 seconds, for mAb-2 measurements, 36 seconds. To increase SNR, 10 spectra of each sample were measured and then averaged into 1, for further analysis. Figure 25 shows the resulting average spectra for both mAbs. Very low intensity spectra in mAb-1 plot can be observed, which is associated with the presence of air bubbles inside the flow-cell, that consequently affect the measurement. In mAb-2 plot not so many significant differences can be detected between the spectra in this sense. Nevertheless, with PCA, one can confirm the presence of outliers.

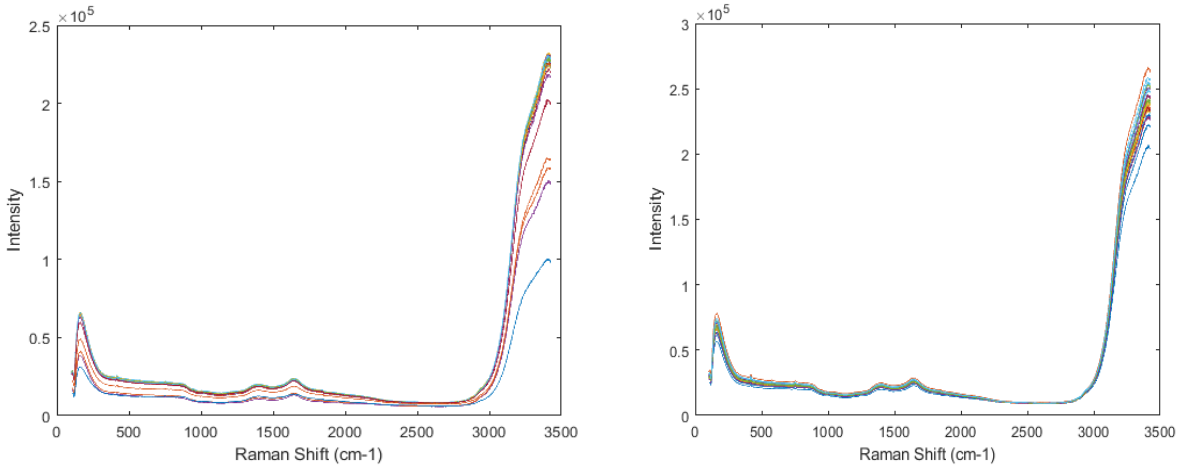

Figure 25 – Raman spectra of mAb-1 (left) and mAb-2 (right) long-range experiments. Each spectra represents the average of 10 different spectra acquired for each sample, with an acquisition time of 30 and 36 seconds for mAb-1 and mAb-2, respectively.

Narrow-range Aggregate Content Experiment

With mAb-2, a new experiment was performed in order to understand if Raman had enough resolution to obtain information under a more realistic range of aggregate content solutions, to identify and quantify antibody aggregates. This was important given the usually low aggregate concentration that is met in the elution of the polishing step, where aggregates levels should be controlled. Given the results obtained for the aggregate behaviour study of mAb-2 (Chapter 0), it was not possible to achieve levels around 10 - 20% of aggregates with the protocol established. Therefore, a different protocol was used, to stress protein with a pH-shift from 3.2 to 5.8, using 0.1 N hydrochloric acid for the incubation at low pH, and 0.5 M Tris buffer for neutralization. This resulted in the aggregated levels indicated in Table 7. Only 3 original stocks were produced for these experiments, as it was the only way of ensuring that no correlation could exist between the buffer molarity and either monomer or aggregate concentration. An additional 4th stock of high aggregate and low monomer concentration would require a dilution of the original protein stock before stressing the antibody, which would cause disparities in the buffer system.

Table 7 - Total antibody concentration and aggregate content in the four original stocks produced for mAb-2.

	mAb-2		
Stock	1	2	3
Total Protein Concentration [g/L]	4.03	4.54	1.75
Aggregate Content [%]	18	2	3

For decorrelation of monomer and aggregate concentration, the DoE was based on Latin Hypercube Sample (LHS) design with Matlab function *lhsdesign*. This method was chosen over Dopt because of its more centric distribution of aggregate content (Figure 26), as Dopt provides more solutions for the extreme values.

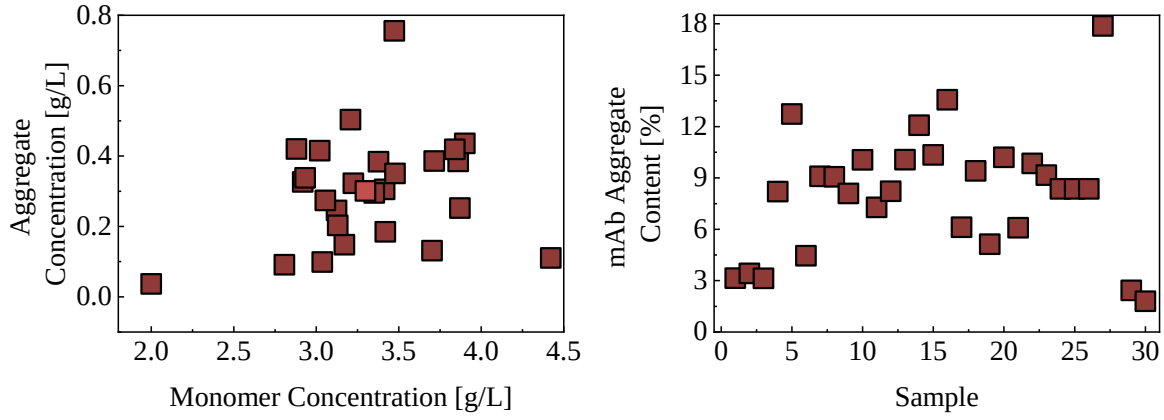

Figure 26 - LHS design result for mAb-2. On the left, the relation between aggregate and monomer concentrations for each sample. In light blue is represented the CP. On the right, the distribution of mAb aggregates per sample.

The acquisition time of each spectra was 30 seconds for the experiment without filter, and 45 seconds for the experiment with filter. This means the presence of the filter increased by 1/3 the acquisition time of a spectrum. Once again, multiple spectra were taken for each sample, although this time only 5

spectra, and averaged for an increased SNR, before further analysis. The number of spectra acquired per sample was reduced because it was proven that the difference in effect was not significant and 5 spectra were enough for the purpose of increasing SNR. No great differences between the spectra in each experiment, shown in Figure 27, can be observed.

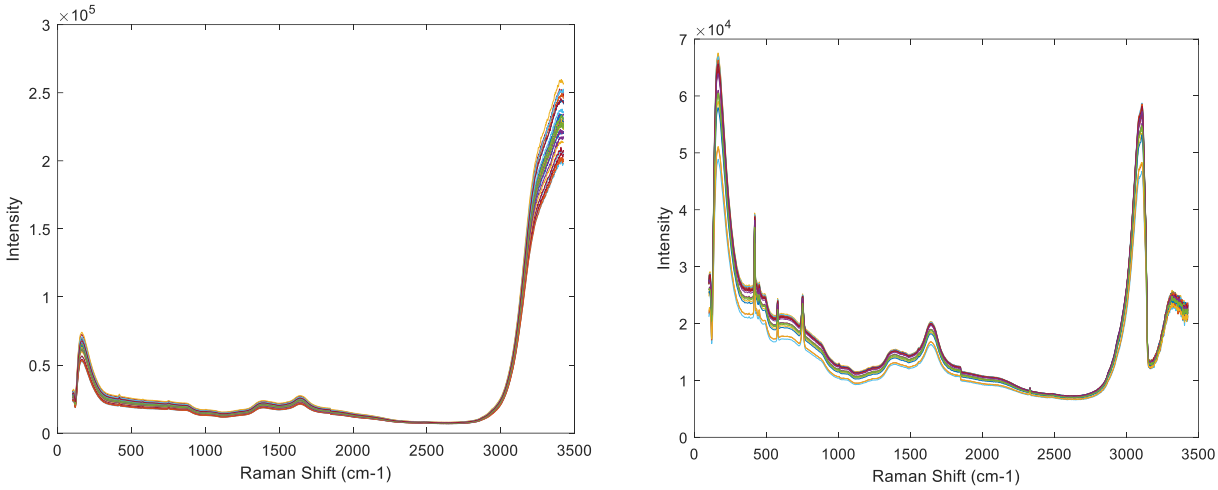

Figure 27 – Raman spectra of mAb-2 without the Notch filter (left) and with the Notch filter (right) short-range experiments. Each spectra represents the average of 5 different spectra acquired for each sample, with an acquisition time of 30 and 45 seconds without filter and with filter, respectively.

3.2.3 Data Analysis

Data analysis of the acquired spectra was performed with PLS. In a first instance, PCA was used for the detection of outliers (Appendix 5). After their identification, these samples were removed from the data set. To make sure the best results were obtained once calibrating the regression model, different techniques were tested for smoothing, namely, SG with and without the derivative function, Wavelet transform (WT), Fourier transform (FT) and Empirical Mode Decomposition (EMD) with and without background removal (BR). As most of them have parameters whose choice of value will impact the result, a tuning of these parameters was performed for such techniques.

3.2.3.1 Hyper-parameter Tuning

As the contributions of aggregate and monomer mAb variance in the spectra are expected to be different, the best parameter values in each smoothing technique were found for both cases separately. Then, a comparison between the best result of each technique was made to determine which granted the best prediction of each variable. This classification was based on the root mean square error in prediction (RMSEP). Only one data set was subjected to these tunings as it was expected that, whichever technique is determined to be the most suited, the selection would be transversal to all data sets. As such, mAb-2's broad-range of aggregate content experiment was chosen.

Figure 28 shows the resulting RMSEP as a function of *fr*, for the SG filter with and without the application of the derivative. It is instantly obvious the difference in results, as the minimum RMSEP is quite distinct in both cases. It is also to be noted that there almost seems to be two minima, especially when the

derivative is applied. Analysing both plots, one can conclude that the SG filter without the application of the derivative produces, for this range, better results.

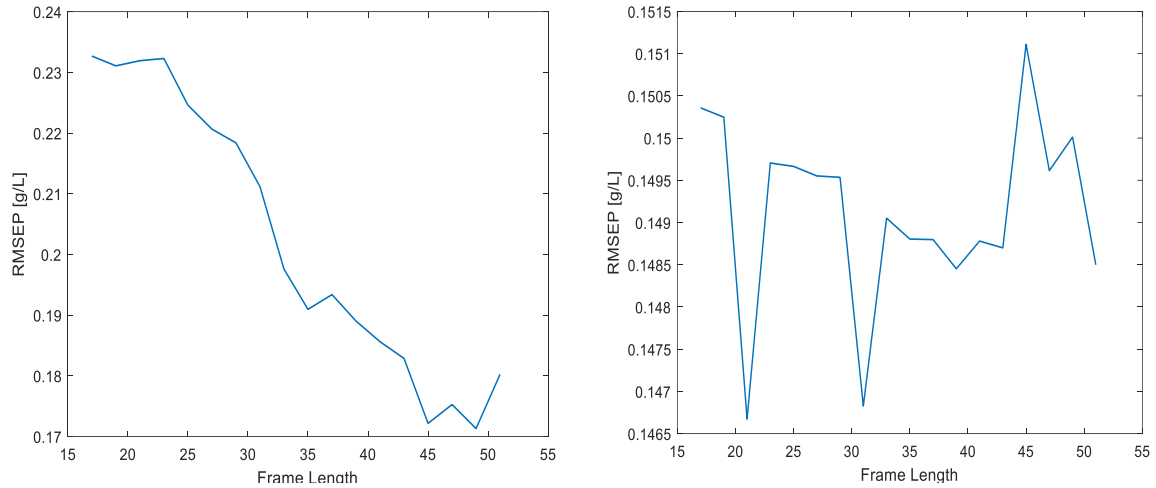

Figure 28 - RMSEP (g/L) result from PLS for the prediction of mAb-2 aggregate concentration as a function of the frame length in the SG filter. On the left, using the derivative function of the filter (ord=1); on the right, using only the smoothing option of the function (ord=0).

Figure 29 shows the resulting RMSEP as a function of the FT filter. On a first instance, one can say this method does not appear to be very stable, which presents a risk in applying it to other data sets just based on this result.

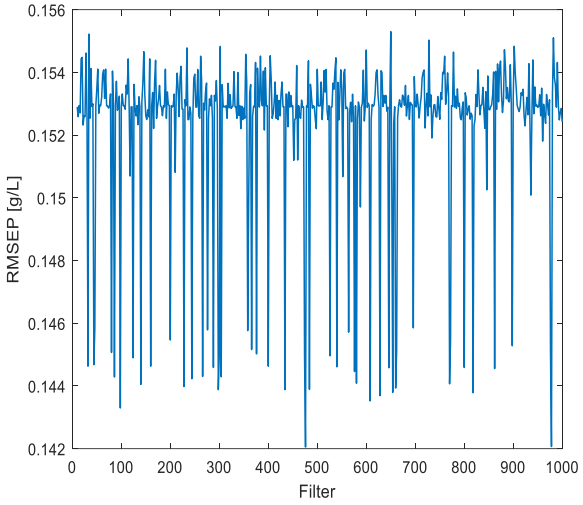

Figure 29 - RMSEP (g/L) result from PLS for the prediction of mAb-2 aggregate concentration as a function of the filter after smoothing the data with FT.

Finally, Figure 30 shows the resulting RMSEP as a function of the decomposition level for each base function tested. Most base functions appear to present their best result when decomposition is performed to the 4th level. Even so, there are some exceptions. The general trend seems to be that, beyond the 4th or 5th level of decomposition, important spectral information is being removed from the spectra and not only noise, leading to higher RMSEP values.

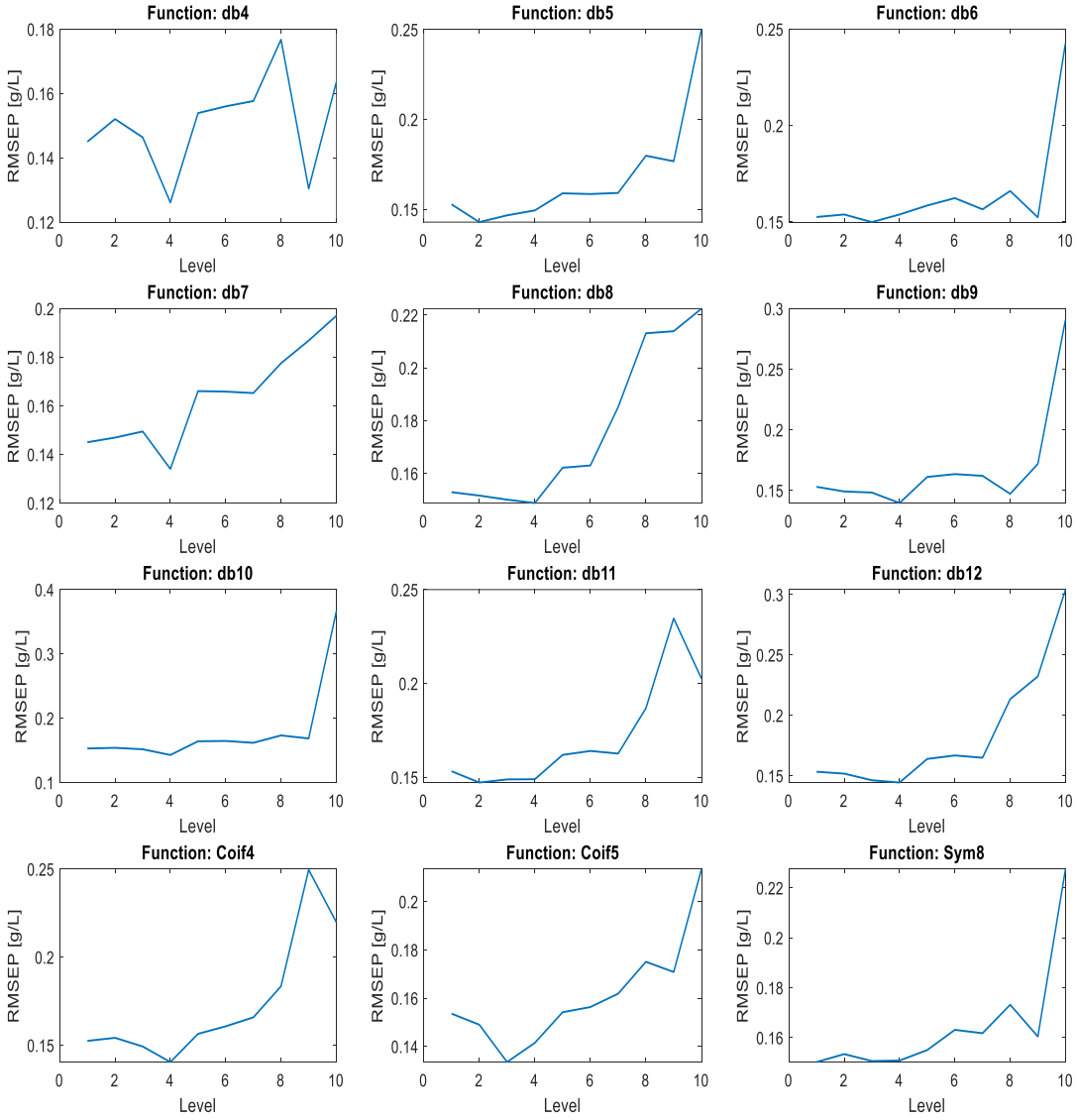

Figure 30 - RMSEP (g/L) result from PLS modelling for the prediction of mAb-2 aggregate concentration as a function of the level of decomposition after smoothing the data with WT for each base functions studied.

Considering the best result for each of these techniques and, additionally, the result obtained for EMD both with and without BR, Figure 31 provides the answer to which technique is the most suitable to be used in the pre-treatment of the data. The same procedure was carried out for the prediction of mAb-2 monomer concentration in the same experiment (plots in Appendix 6). The result is also presented in Figure 31. The RMSEP values are discriminated in Table 8.

Table 8 - Minimum RMSEP values for aggregate and monomer prediction of mAb-2 long-range aggregate content experiment by each of the tested techniques.

Technique	RMSEP of aggregate concentration [g/L]	RMSEP of monomer concentration [g/L]
SG	0.171	0.216
SG w/o derivative	0.147	0.129
FT	0.142	0.125
WT	0.140	0.106
EMD	0.563	0.892
EMD w/o BR	0.121	0.147

Although the difference is quite small, when predicting mAb-2 aggregate concentration, the most appropriate technique for smoothing the spectra is EMD without BR, with a RMSEP of 0.1208 g/L. As for the prediction of the monomer concentration, the most suitable technique is WT, presenting a RMSEP of 0.1058 g/L, with Coif4 as a base function and level 4 of decomposition.

As both selected techniques are only smoothing techniques without any spectral changes concerning the baseline, no problem was seen in applying them to each of the data sets. Thus, adopting these two techniques for each of the cases, mAb-1 and mAb-2 aggregate and monomer concentration were predicted for each of the experiments.

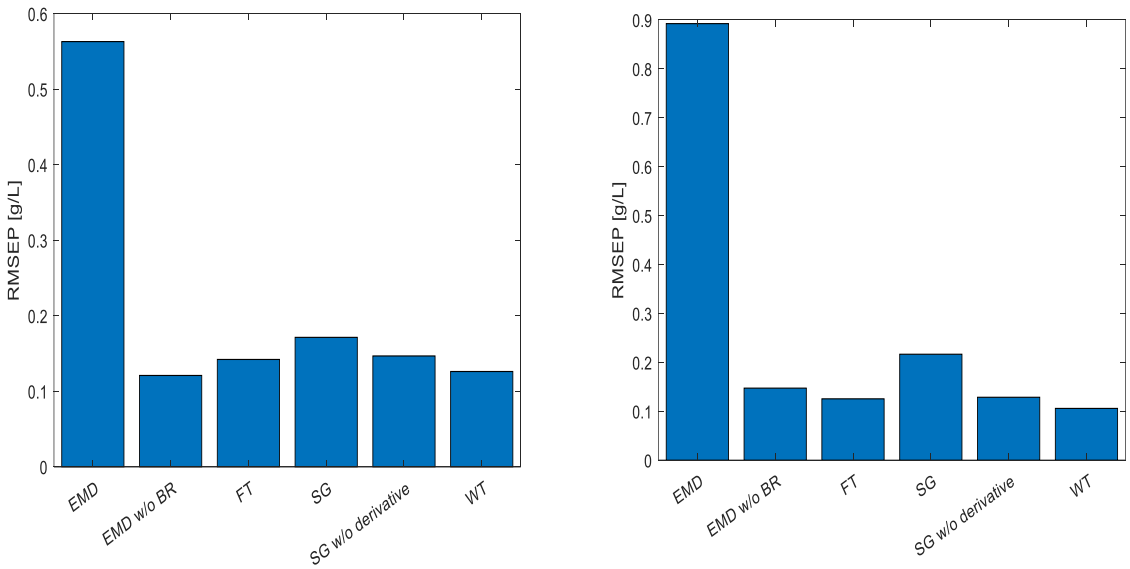

Figure 31 - RMSEP (g/L) results from PLS for each tested technique for mAb-2 broad-range experiment, using, as a Y for the PLS modelling, the aggregate concentration on the left plot and the monomer concentration on the right plot.

3.2.3.2 PLS model

The procedure for the regression model implemented in Matlab had the following line:

1. Pre-treatment of the averaged raw spectra;
2. Wavelength cut;

3. Cross-validation and PLS;
4. External prediction.

Broad-range Aggregate Content Experiment

Table 9 presents the average result obtained from the 5 rotations for the prediction of mAb-1 and mAb-2 monomer and aggregate concentrations. Original results from each rotation can be found in Appendix 7.

Concerning the results obtained for mAb-1, the regression model for the prediction of monomer concentration presented a relative root mean square error in prediction (relRMSEP) of 35%, which translates in a RMSEP of 0.942 g/L in a range between 0.536 and 3.81 g/L. As for the regression model calibrated for the prediction of aggregate concentration, a relRMSEP of 21% was obtained, reflecting a RMSEP of 0.471 g/L in a concentration range between 0.0145 and 2.63 g/L. It may be concluded that the prediction of either monomer or aggregate mAb-1 was not very successful. A prediction model with error above 20% is not suitable for monitoring of either monomer or aggregates in DSP.

Regarding the results from mAb-2, prediction of monomer concentration presented a relRMSEP of 3.5%, expressing a RMSEP of 0.106 g/L in a concentration range between 0.437 and 3.76 g/L. For aggregate prediction, the regression model presented a relRMSEP of 7.51%, with a RMSEP of 0.121 g/L in a range between 0.0189 and 2.13 g/L. One can thus say that the PLS model predictions were both quite successful for mAb-2 and both variables were decently predicted. Additionally, a Q^2 above 90% in both cases translates the ability of the models to successfully explain the variance in the spectral data.

This presents a first proof of concept that Raman spectroscopy can be applied for aggregate prediction. Comparing both antibodies, mAb-2 clearly presents more promising results. Nevertheless, it must be taken into account the size of the data set used for each. Due to the removal of outliers, mAb-1's data set lost a considerable number of samples. In other words, it lost valuable information that could have improved the model.

Table 9 - PLS prediction results of aggregate and monomer concentration for both mAb-1 and mAb-2.

	Predicted Variable	**N° of Samples**	**Calibration Range [g/L]**	**Prediction Range [g/L]**	**LV**	**Q^2**	**RMSEP [g/L]**	**rel RMSEP**
mAb-1	Monomer concentration	18	0.536-3.81	0.536-3.81	4	-0.573	0.942	0.345
	Aggregate concentration		0.0145-2.63	0.0145-2.634	6	0.471	0.424	0.205
mAb-2	Monomer concentration	24	0.437-3.76	0.437-3.76	14	0.964	0.106	0.0350
	Aggregate concentration		0.0189-2.13	0.0189-2.13	14	0.913	0.121	0.0751

Figure 32 shows observed *vs* predicted plots for mAb-1 aggregate and monomer concentration, for both the cross-validation and validation sets. The observable distance to the diagonal shows how the model lacks the capability of prediction. Additionally, the lack of robustness of the models is also evident by how for the predictions in CV of the same value are between rotation. Figure 33 presents the observed *vs* predicted plots for mAb-2. A significant difference can be acknowledged in the quality of prediction. An important observation to be made is that in the lowest range of aggregate concentration the distance to the centre increases. This might be related to Raman's limited detection.

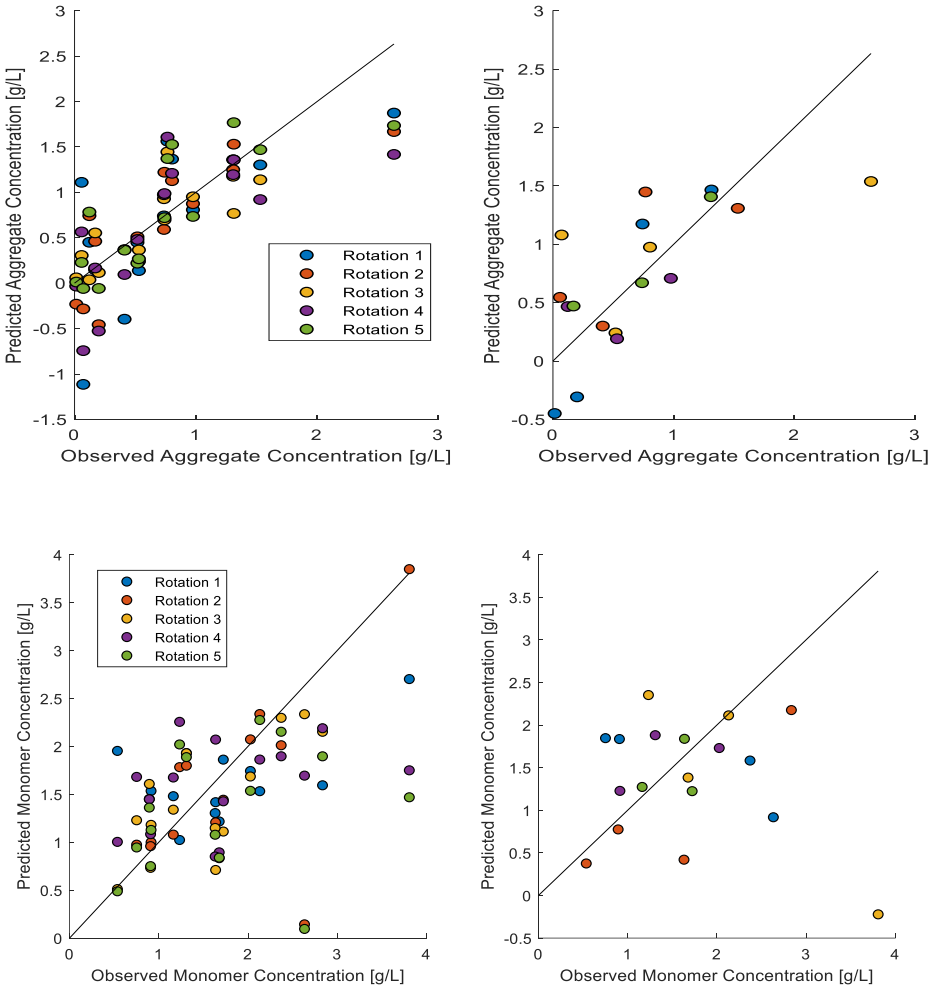

Figure 32 – Observed versus predicted plots of mAb-1 PLS model for the prediction of aggregate (upper plots) and monomer (lower plots) concentration in the samples. Each colour is associated with rotation. On the left, the results obtained for cross-validation using a 5-fold method; on the right the external prediction results.

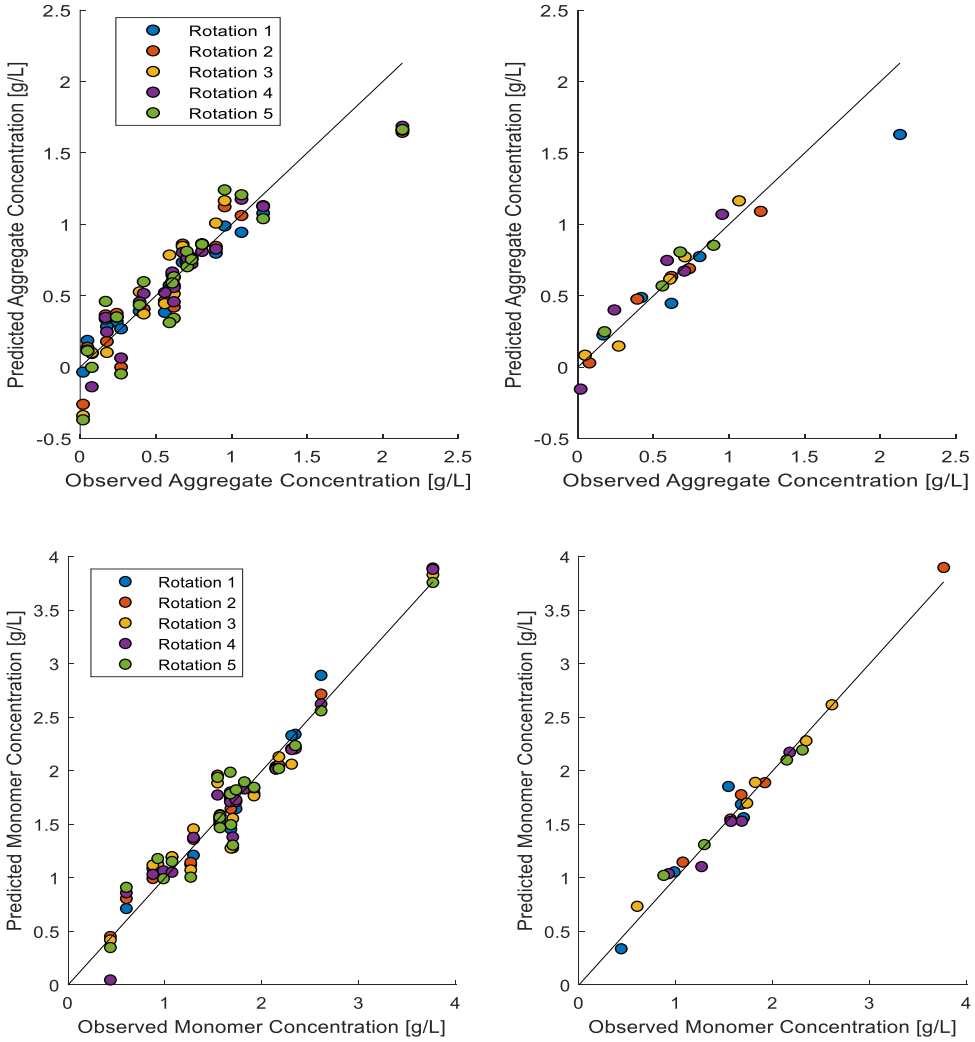

Figure 33 – Observed versus predicted plots of mAb-2 PLS model for the prediction of aggregate (upper plots) and monomer (lower plots) concentration in the samples from the broad-range experiment. Each colour is associated with rotation. On the left, the results obtained for cross-validation using a 5-fold method; on the right the external prediction results.

Narrow-range Aggregate Content Experiments

Table 10 presents the PLS results for the prediction of monomer and aggregate concentration of mAb-2 in a short-range of aggregate content, measured both with and without the presence of a Notch filter.

For the measurements without the Notch filter, the PLS model once again predicts monomer concentration quite accurately, with a relRMSEP of 6.1%, translated in a RMSEP of 0.146 g/L in a range between 1.69 and 4.45 g/L. However, for the prediction of aggregate concentration, the model presents a relRMSEP of 25%, reflecting a RMSEP of 0.113 g/L in a range between 0.0527 and 0.581 g/L.

The experiment performed in the presence of the Notch filter produced similar results. The regression model successfully predicted mAb-2 monomer concentration with a reRMSEP of 6.7%, translated in a RMSEP of 0.15 g/L in a range between 1.69 and 4.45 g/L. When calibrated for mAb-2 aggregate concentration, the predictions presented a relRMSEP of 27%, which represents a RMSEP of 0.116 g/L in a concentration range between 0.0527 and 0.581 g/L.

It appears that, under this range, Raman is not suitable for the quantification of aggregates. Nevertheless, the prediction capability of mAb-2 monomer is quite good.

Comparing the results between both experiments, there does not seem to be a significant difference in the predictions. The acquisition time that was gained by the presence of the filter does not seem to have sorted the desired effect on captured information.

Table 10 - PLS prediction results of aggregate and monomer concentration for both mAb-2 for samples measured both with and without the notch filter.

	Predicted Variable	**N° of samples**	**Calibration Range [g/L]**	**Prediction Range [g/L]**	**LV**	**Q^2**	**RMSEP [g/L]**	**relRMSEP**
mAb-2 without filter	Monomer concentration	28	1.69-4.45	1.69-4.45	12	0.875	0.1462	0.0605
	Aggregate concentration		0.0527-0.581	0.0527-0.581	2	-0.112	0.113	0.254
mAb-2 with filter	Monomer concentration	26	1.69-4.45	1.69-4.45	6	0.883	0.153	0.0674
	Aggregate concentration		0.0527-0.581	0.0527-0.581	1	-0.298	0.116	0.270

Figure 34 shows the observed *vs* predicted plots of the models built from mAb-2 samples without the notch filter. The disperse results of the plots related to the aggregate prediction corroborate the observations made above. Once again, the lack of predicting capability by the model, is evident by the fact that in the CV plot, for the same observed value, in each rotation a different value is predicted. The compactness presented for the monomer prediction plots are in line with the results shown in Table 10. The same analysis can be made about the resulting observed *vs* predicted plots for the measurement of mAb-2 samples with the notch filter, shown in Figure 35.

As a conclusion, considering all the results shown above, one can say that under the studied conditions, Raman spectroscopy is not able to detect mAb-1 and mAb-2 aggregates to a level warranted for its monitoring. This might be explained by different reasons, such as the lack of spectroscopic resolution, not enough experimental data points to train a regression model well enough for the prediction of aggregates or lack of data analysis tools applied. Regarding the first point, resolution could be improved by increased laser power. Experimentally, further experiments must be conducted for the acquisition of more spectral measurements and consequently bigger data sets for the calibration of the models. At last, improvements in data analysis can also be tested with the application of different tools, namely wavelength selection tools or PLS2 [87] [88].

Nevertheless, it has also been shown that monomer antibody of mAb-2 is detected by Raman spectroscopy quite well and can be applied in DSP monitoring. This is important in the sense that monomer monitoring is also a relevant aspect in the scope of continuous processing. Online monitoring

of antibody concentration is necessary during the capture Protein A chromatography step and could also provide important information in the final polishing step.

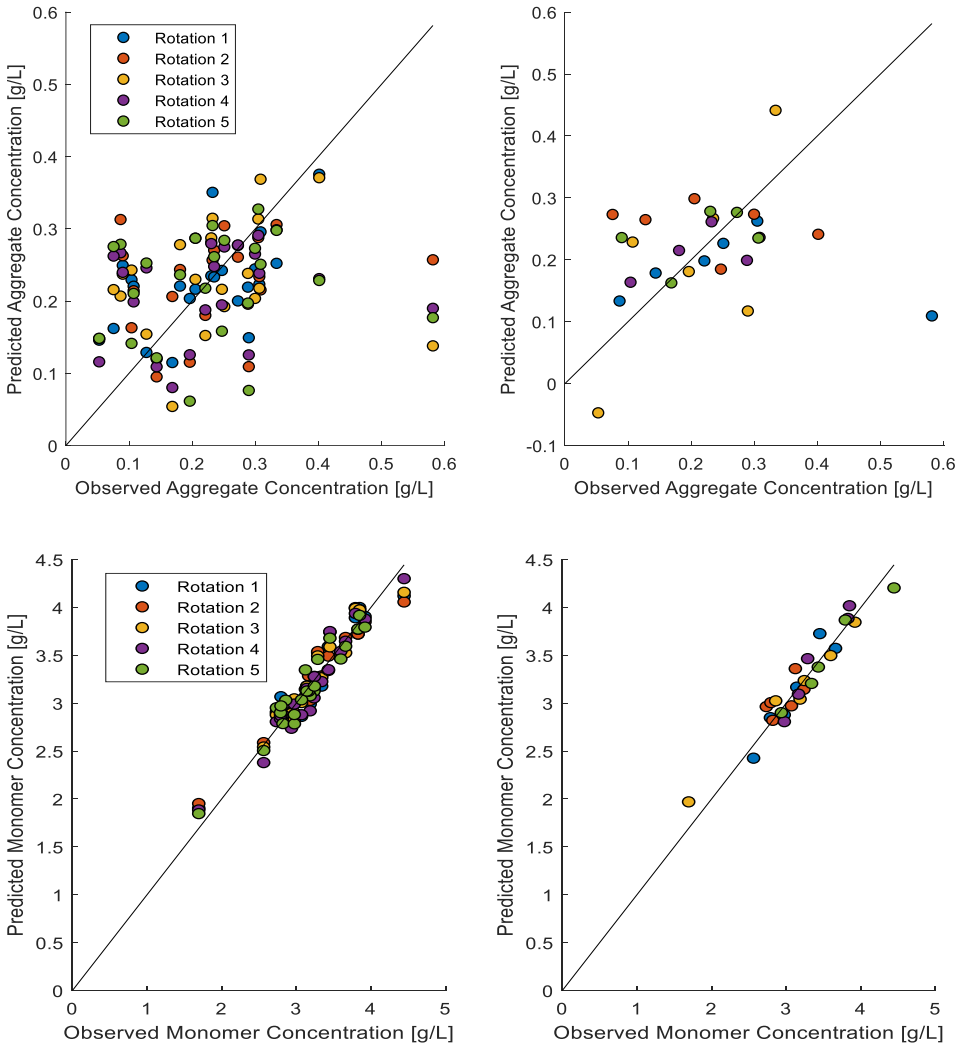

Figure 34 – Observed versus predicted plots of mAb-2 PLS regression model for the prediction of aggregate (upper plots) and monomer (lower plots) concentration in the samples, for the experiment of narrow-range of aggregate content samples and without the filter. Each colour is associated with a rotation. On the left, the results obtained for cross-validation using a 5-fold method; on the right the external prediction results.

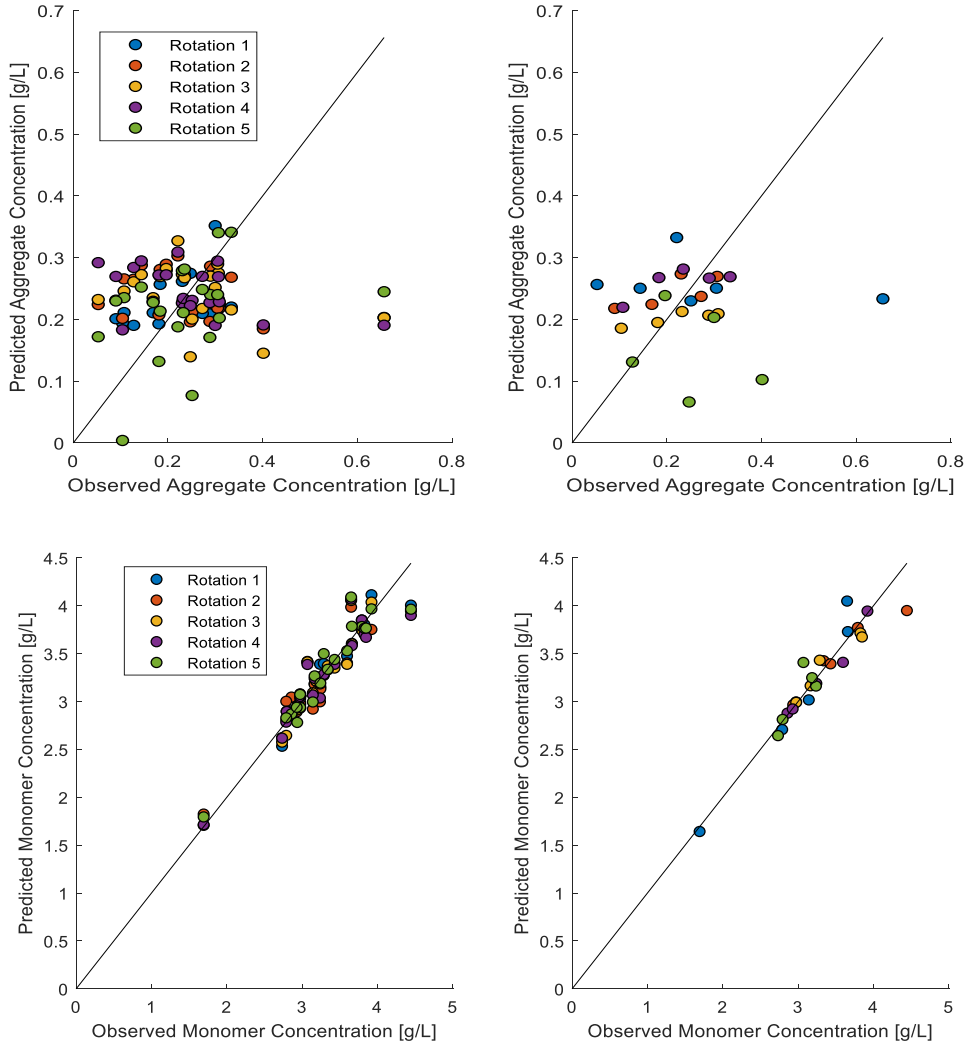

Figure 35 – Observed versus predicted plots of mAb-2 PLS model for the prediction of aggregate (upper plots) and monomer (lower plots) concentration in the samples, for the experiment of narrow-range of aggregate content samples and in the presence of the filter. Each colour is associated with rotation. On the left, the results obtained for cross-validation using a 5-fold method; on the right the external prediction results.

4 Final Remarks

Considering the aggregation behaviour of mAb-1 and mAb-2, it has been concluded that acidic conditions at low ionic strength do not immediately lead to formation of aggregates. Instead, neutralization of the solutions to mildly acidic pH values and higher ionic strength will initiate aggregation. For both mAbs, aggregation after neutralization starts with rapid oligomerization. For mAb-1, slight recovery of monomer takes place at later times, while for mAb-2 the protein monomer keeps aggregating at a decreasing pace. The most influential parameter for protein aggregation is the pH value at which the protein is incubated prior to neutralization. It supposedly determines the degree of denaturation the mAbs suffer and consequently how hydrophobic the protein surface has become right before neutralization. Also of importance, but only when moderate unfolding takes place, is the incubation time. Under less harsh conditions (i.e. at pH 3.5), the longer the protein is incubated, the higher is the loss of monomeric mAb after neutralization. Addition of D-sorbitol during low pH incubation was tested for alleviation of aggregate formation. Promising results were obtained for mAb-1. Finally, it was studied, whether the rate of neutralization has an impact on the aggregation process. It was found that the velocity at which the neutralization solution is added does not affect the aggregation outcome for both mAbs in the range from 0 to 60 minutes of neutralization.

Based on the experimental observations, the following was concluded: at low pH and ionic strength, proteins partially unfold and expose part of their hydrophobic residues (confirmed by extrinsic fluorescence spectroscopy). Depending on the pH value, a new conformational steady-state is reached at different velocities. At the same time, electrostatic repulsion dominates the interaction of mAb molecules and prevents their self-association. When pH and ionic strength are subsequently increased, mAb surface charges are screened and inter-molecular repulsion decreases. This allows exposed hydrophobic patches of different mAb molecules to interact, leading to aggregation. For mAb-1, fewer hydrophobic patches are exposed at pH 3.5 in presence of D-sorbitol, which decreases the amount of aggregates formed after neutralization.

In future studies, the influence of other important parameters should be tested: protein concentration, ionic strength after neutralization as well as other additives besides D-sorbitol. It was mentioned in Chapter 1.2.3 that the concentration can affect protein conformation and aggregation in distinct ways. This should be investigated as higher mAb concentrations might be encountered in manufacturing [5]. In this work, ionic strength is believed to have a very significant influence in mAb aggregation. However, it is still to be tested if its effect after in aggregation after neutralization will shift if one increases the final ionic strength. The extended Debye-Hückel model was successfully implemented, allowing for simultaneous control of pH and ionic strength with low error (i.e. ±0.1 pH unit). Nevertheless, as mentioned in Chapter 1.2.5, the model is not applicable for calculation of ion activity coefficients for ionic strengths higher than 0.1 M. Thus, use of a more sophisticated model is necessary to carry out this study, *e.g.* the Pitzer ion activity model [31]. Besides other sugars and polyols, amino acids, water-soluble polymers, or surfactants might be investigated as co-solutes to reduce the amount of aggregate formation. Some of them might behave as chemical chaperones that promote protein folding and thereby inhibit aggregation [16].

Concerning the optimization of the flow cell, enhancement of the signal was possible, as well as shortening the exposure time of acquisition. A higher signal intensity was reached by changing the length of the sample's channel from 10 to 20 mm, at the same time decreasing background contribution. Reduced exposure time was accomplished by the implementation of a Notch filter that successfully blocks an unwanted region of the spectrum, avoiding the saturation of the detector for a longer time. For the measurements of mAb-2, the Notch filter granted the possibility to increase by 1/3 the acquisition time of each spectrum. Although the implementation showed no favourable effect on the prediction results, it could be very beneficial for other applications, in which the ratio between unwanted regions and information regions is further increased.

Concerning the different smoothing techniques tested, one can say that, excluding Empirical Mode Decomposition (EMD) with the subtraction of background, all of them gave similar results. Nevertheless, better alternatives to the Savitzky-Golay filter were found and applied. For the prediction of aggregate concentration, EMD without background removal was applied, and for the prediction of monomer concentration, the Wavelet transform to the 4^{th} level of decomposition with the base function Coif4.

Regarding the ability to predict antibody aggregates with Raman spectroscopy, it was concluded that it was possible under a broad-range of aggregate content. With a RMSEP of 0.12 g/L in a range between 0.019 and 2.13 g/L, one can say the regression model was successful in predicting mAb-2 aggregates' concentration. Nevertheless, when shortening the aggregate content range of the samples, a RMSEP of 0.11 g/L in a range between 0.053 and 0.58 g/L is not acceptable. As for the prediction of mAb-2 monomer concentration, the regression models were successful in every experiment, with a relRMSEP always below 10% and a coefficient of determination always above 85%. The bad prediction results obtained for mAb-1 may be related to the short number of samples used in the data set for the training of the model.

As a conclusion, this work provides the proof of concept that Raman spectroscopy is able to access antibody aggregate information. However, further intensive optimisations need to be exercised to allow the application of this technique in more realistic aggregate ranges, common in the DSP of antibody manufacturing. Improvement might be achieved both experimentally and computationally wise to grant better results. Considering data acquisition, the Raman device that is used here constitutes its simplest version, when compared to techniques such as Raman Optical Activity or UV Resonance Raman [89]. The device is a commercial available process analyser optimized for upstream purposes. The low resolution might be what is inhibiting the prediction of such low aggregate values. A further possible improvement is to increase the laser power of the device, allowing for a stronger light hitting the sample and igniting a bigger response. As for improvements in the data analysis field, wavelength selection techniques could be applied. Apart from the extremes of the spectra that were cut for the analysis, also specific ranges of wavelengths in the intermediate region may not have meaningful information. One way to increase the robustness of the regression model might be to eliminate these sections of the spectra, possibly responsible for leading to wrong model correlations [87]. Additionally, PLS2 can also be tested, by having both monomer and aggregate concentration as variables Y, as it is known to better explain a higher degree in the data set variance [88].

5. Bibliography

[1] P. Gronemeyer, R. Ditz, and J. Strube, “Trends in Upstream and Downstream Process Development for Antibody Manufacturing,” *Bioengineering*, vol. 1, no. 4, pp. 188–212, 2014.

[2] R. S. William and M. S. Lila, “Chapter 1: Introduction to biologics and monoclonal antibodies,” in *Therapeutic Antibody Engineering*, Elsevier, 2012, pp. 1–14.

[3] L. M. Weiner, R. Surana, and S. Wang, “Monoclonal antibodies: versatile platforms for cancer immunotherapy,” *Nat. Rev. Immunol.*, no. 10, pp. 317–327, 2010.

[4] ThermoFisher Scientific, “Immunoglobulin IgG Class.” [Online]. Available: https://www.thermofisher.com/us/en/home/life-science/antibodies/antibodies-learning-center/antibodies-resource-library/antibody-methods/immunoglobulin-igg-class.html. [Accessed: 23-Jun-2018].

[5] D. J. Karst, F. Steinebach, and M. Morbidelli, “Continuous integrated manufacturing of therapeutic proteins,” *Curr. Opin. Biotechnol.*, vol. 53, pp. 76–84, 2018.

[6] M. Vázquez-Rey and D. A. Lang, “Aggregates in monoclonal antibody manufacturing processes,” *Biotechnol. Bioeng.*, vol. 108, no. 7, pp. 1494–1508, 2011.

[7] C. L. Gaughan, “The present state of the art in expression, production and characterization of monoclonal antibodies,” *Mol. Divers.*, vol. 20, no. 1, pp. 255–270, 2016.

[8] G. Miesegaes, S. Lute, H. Aranha, and K. Brorson, “Chapter 38: Virus Retentive Filters,” in *Downstream Intrustrial Biotechnology, Recovery and Purification*, M. C. FLICKINGER, Ed. Wiley, 2013, pp. 641–652.

[9] S. Klutz, M. Lobedann, C. Bramsiepe, and G. Schembecker, “Continuous viral inactivation at low pH value in antibody manufacturing,” *Chem. Eng. Process. Process Intensif.*, vol. 102, pp. 88–101, 2016.

[10] M. L. Chiu and G. L. Gilliland, “Engineering Antibody Therapeutics,” *Curr. Opin. Struct. Biol.*, vol. 38, pp. 163–173, 2016.

[11] C. J. Roberts, “Therapeutic Protein Aggregation: Mechanisms, Design, and Control,” *Trends Biotechnol.*, pp. 372–380, 2004.

[12] B. Somasundaram, K. Pleitt, E. Shave, K. Baker, and L. H. L. Lua, “Progression of continuous downstream processing of monoclonal antibodies: current trends and challenges,” *Biotechnol. Bioeng.*, pp. 0–2, 2018.

[13] T. Chen, “Formulation concerns of protein drugs,” *Drug Dev. Ind. Pharm.*, vol. 18, no. 11–12, pp. 1311–1354, 1992.

[14] H. F. Gilbert, “Peptide Bonds, Disulfide Bonds and Properties of Small Peptides,” *Encycl. Life Sci.*, pp. 1–8, 2010.

[15] W. Saenger, *Principles of Protein Structure*. 1984.

[16] W. Wang, S. Nema, and D. Teagarden, “Protein aggregation-Pathways and influencing factors,” *Int. J. Pharm.*, vol. 390, no. 2, pp. 89–99, 2010.

[17] D. Leckband and J. Israelachvili, “Introduction: overview of forces in biology,” in *Intermolecular Forces in Biology*, Cambridge University Press, 2001, pp. 108–155.

[18] J. N. Israelachvili, “Strong Intermolecular Forces: Covalent and Coulomb Interactions,” in *Intermolecular and Surface Forces: Third Edition*, 2011, pp. 53–68.

[19] P. M. Biesheuvel, “Simplifications of the Poisson–Boltzmann Equation for the Electrostatic Interaction of Close Hydrophilic Surfaces in Water,” *J. Colloid Interface Sci.*, vol. 238, no. 2, pp. 362–370, Jun. 2001.

[20] D. Leckband and J. N. Israelachvili, “Introduction: Overview of Forces in Biology,” in *Intermolecular Forces in Biology*, 2001, pp. 108–155.

[21] K. A. Dill, “Dominant Forces in Protein Folding,” *Biochemistry*, vol. 29, no. 31, pp. 7133–7155, 1990.

[22] R. J. Beynon and J. S. Easterby, *Buffer Solutions*. Oxford University Press, 1996.

[23] M. C. Flickinger, Ed., “Scaledown of Biopharmaceutical Purification Operations,” in *Downstream Intrustrial Biotechnology, Recovery and Purification*, 2013, pp. 129–146.

[24] M. P. Tarazona and E. Saiz, “Combination of SEC/MALS experimental procedures and theoretical analysis for studying the solution properties of macromolecules,” *J. Biochem. Biophys. Methods*, vol. 56, no. 1–3, pp. 95–116, 2003.

[25] J. R. Alford, B. S. Kendrick, J. F. Carpenter, and T. W. Randolph, “Measurement of the Second Osmotic Virial Coefficient for Protein Solutions Exhibiting Monomer-Dimer Equilibrium,” *Anal Biochem.*, pp. 128–133, 2008.

[26] P. C. Hiemenz and R. Rajagopalan, “Chapter 5: Static and Dynamic Light Scattering and Other Radiation Scattering,” in *Principles of Colloid and Surface Chemistry*, Marcel Dekker, Inc, 1997, pp. 193–247.

[27] U. Nobbmann *et al.*, “Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies,” *Biotechnol. Genet. Eng. Rev.*, vol. 24, no. 1, pp. 117–128, 2007.

[28] P. McPhie and J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. 2000.

[29] D. M. Togashi, A. G. Ryder, and D. O’Shaughnessy, “Monitoring local unfolding of bovine serum albumin during denaturation using steady-state and time-resolved fluorescence spectroscopy,” *J. Fluoresc.*, vol. 20, no. 2, pp. 441–452, 2010.

[30] C. G. McCarty and E. Vitz, “pH Paradoxes: Demonstrating That It Is Not True That pH ≡ -log[H+],” *J. Chem. Educ.*, vol. 83, no. 5, p. 752, 2006.

[31] M. Luckas and J. Krissmann, “Aktivitätskoeffizientenmodelle für Elektrolytlösungen,” in *Thermodynamik der Elektrolytlösungen*, Springer, 2001, pp. 135–161.

[32] C. J. RITSEMA, “Estimation of activity coefficients of individual ions in solutions with ionic strengths up to 0.3 mol dm−3,” *J. Soil Sci.*, vol. 44, no. 2, pp. 307–315, 1993.

[33] A. Ruvarac, “Determination of the thermodynamic equilibrium constants of ion-exchange processes,” *Mater. Chem. Phys.*, vol. 35, no. 3–4, pp. 247–249, Oct. 1993.

[34] R. G. Bates and G. D. Pinching, “Resolution of the Dissociation Constants of Citric Acid at 0 to 50°, and Determination of Certain Related Thermodynamic Functions,” *J. Am. Chem. Soc.*, vol. 71, no. 4, pp. 1274–1283, Apr. 1949.

[35] F. Crea, C. De Stefano, F. J. Millero, and V. K. Sharma, “Dissociation Constants for Citric Acid in NaCl and KCl Solutions and their Mixtures at 25 ◦ C,” vol. 33, no. 11, 2004.

[36] M. Rüdt, T. Briskot, and J. Hubbuch, “Advances in downstream processing of biologics--Spectroscopy: An emerging process analytical technology,” *J. Chromatogr. A*, vol. 1490, pp. 2–9, 2017.

[37] S. Flatman, I. Alam, J. Gerard, and N. Mussa, “Process analytics for purification of monoclonal antibodies,” *J. Chromatogr. B*, vol. 848, no. 1, pp. 79–87, Mar. 2007.

[38] US FDA, “Guidance for Industry PAT—A framework for innovative pharmaceutical development, manufacturing, and quality assurance,” 2004.

[39] K. B. Konstantinov and C. L. Cooney, “White paper on continuous bioprocessing May 20-21, 2014 continuous manufacturing symposium,” *J. Pharm. Sci.*, vol. 104, no. 3, pp. 813–820, 2015.

[40] A. A. Shukla and U. Gottschalk, “Single-use disposable technologies for biopharmaceutical manufacturing,” *Trends Biotechnol.*, vol. 31, no. 3, pp. 147–154, Mar. 2013.

[41] K. A. Bakeev, *Process analytical technology : spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries*. Wiley, 2010.

[42] G. Carta and A. Jungbauer, *Protein chromatography : process development and scale-up*. Wiley-VCH, 2010.

[43] R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, “Review of multidimensional data processing approaches for Raman and infrared spectroscopy,” *EPJ Tech. Instrum.*, vol. 2, no. 1, p. 8, 2015.

[44] W. Jiskoot and D. J. A. Crommelin, *Methods for Structural Analysis of Protein Pharmaceuticals*. American Association of Pharmaceutical Scientists, 2055.

[45] R. Schweitzer-Stenner, “Advances in vibrational spectroscopy as a sensitive probe of peptide

and protein structure," *Vib. Spectrosc.*, vol. 42, no. 1, pp. 98–117, Oct. 2006.

[46] A. Barth, "Infrared Spectroscopy of Proteins," *Biochim Biophys Acta*, 2007.

[47] F. León-Bejarano, M. Ramírez-Elías, M. O. Mendez, G. Dorantes-Méndez, M. D. C. Rodríguez-Aranda, and A. Alba, "Denoising of Raman spectroscopy for biological samples based on empirical mode decomposition," *Int. J. Mod. Phys. C*, vol. 28, no. 9, 2017.

[48] Z.-Q. Wen, "Raman Spectroscopy of Potein Pharmaceuticals," *Wiley Intersci.*, 2006.

[49] M. G. Friberger, "Introduction To Raman Spectroscopy," *Thermo Fischer Sci.*, vol. 1, 2010.

[50] B. S. Krimm and J. Bandekart, "Vibrational Spectroscopy and Conformation of Peptide, Polypeptides and Proteins," in *Advances in protein chemistry*, 1986, pp. 181–364.

[51] M. Pezolet, "Laser raman investigation of the conformation of human immunoglobulin G," *Biochim. Biophys. acta. Protein Struct.*, vol. 453, no. 2, pp. 502–512, 1976.

[52] P. C. Painter and J. L. Koenig, "Raman Spectroscopic Study of Structure of Antibodies," *Biopolymers*, vol. 14, no. 3, pp. 457–468, 1975.

[53] R. Gömez De La Cuesta, R. Goodacre, and L. Ashton, "Monitoring antibody aggregation in early drug development using raman spectroscopy and perturbation-correlation moving windows," *Anal. Chem.*, vol. 86, no. 22, pp. 11133–11140, 2014.

[54] S. N. Deming, "Chemometrics: An overview," *Clin. Chem.*, vol. 32, no. 9, pp. 1702–1706, 1986.

[55] M. O'Connell, A. Ryder, M. Leger, and H. T., "Qualitative analysis using Raman spectroscopy and chemometrics: a comprehensive model system for narcotics analysis," *Appl Spectrosc.*, vol. 64, 2010.

[56] H. RB1, D. E, M. M, and F. Vogt, "Fourier transform infrared spectroscopy and improved principal component regression (PCR) for quantification of solid analytes in microalgae and bacteria," *Appl Spectrosc*, vol. 65, no. 4, 2011.

[57] M. Ritz, L. Vaculíková, and E. Plevová, "Application of infrared spectroscopy and chemometric methods for the identification of selected minerals," *Acta Geodyn. Geomater.*, vol. 8, no. 1, pp. 47–58, 2011.

[58] P. Lasch, "Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging," *Chemom. Intell Lab.*, vol. 117, 2012.

[59] T. Bocklitz, A. Walter, K. Hartmann, P. Rösch, and J. Popp, "How to pre-process Raman spectra for reliable and stable models?," *Anal Chim Acta.*, vol. 704, pp. 47–56, 2011.

[60] S. Li and L. Dai, "An improved algorithm to remove cosmic spikes in Raman spectra for online monitoring.," *Appl Spectrosc.*, vol. 65, 2011.

[61] J. Burger and P. Gelad, "Hyperspectral NIR image regression part I: calibration and correction,"

J Chemom., vol. 19, 2005.

[62] G. Schulze, A. Jirasek, M. Yu, A. Lim, R. Turner, and M. Blades, “Investigation of selected baseline removal techniques as candidates for automated implementation,” *Appl Spectrosc.*, vol. 59, 2005.

[63] B. Bussian and W. Härdle, “Robust Smoothing Applied to White Noise and Single Outlier Contaminated Raman Spectra,” *Appl Spectrosc.*, 1984.

[64] A. Savitzky and M. J. Golay, “Smoothing and Differentiation of Data by Simplified Least Squares Procedures,” *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.

[65] M. Clupek, P. Matejka, and K. Volka, “Noise reduction in Raman spectra: Finite impulse response filtration versus Savitzky–Golay smoothing,” *J. Raman Spectrosc.*, vol. 38, pp. 1174–1179, 2007.

[66] E. L. Kosarev and E. Pantos, “Optimal smoothing of ‘noisy’ data by fast Fourier transform,” *J. Phys. E.*, vol. 16, no. 6, pp. 537–543, Jun. 1983.

[67] R. Horstman *et al.*, “Deconvolution, Derivation, and Smoothing of Spectra Using Fourier Transforms,” *J. Test. Eval.*, vol. 12, no. 2, p. 78, 1984.

[68] J. K. Kauppinen, D. J. Moffatt, H. H. Mantsch, and D. G. Cameron, “Smoothing of spectral data in the Fourier domain.,” *Appl. Opt.*, vol. 21, no. 10, pp. 1866–72, 1982.

[69] P. M. Ramos and I. Ruisánchez, “Noise and background removal in Raman spectra of ancient pigments using wavelet transform,” *J. Raman Spectrosc.*, vol. 36, no. 9, pp. 848–856, Sep. 2005.

[70] F. Ehrentreich and L. Summchen, “Spike removal and denoising of Raman spectra by wavelet transform methods,” *Anal. Chem.*, vol. 73, no. 17, pp. 4364–4373, 2001.

[71] Y. Hu, T. Jiang, A. Shen, W. Li, X. Wang, and J. Hu, “A background elimination method based on wavelet transform for Raman spectra,” *Chemom. Intell. Lab. Syst.*, vol. 85, no. 1, pp. 94–101, 2007.

[72] H. Chen, W. Xu, N. Broderick, and J. Han, “An adaptive denoising method for Raman spectroscopy based on lifting wavelet transform,” *J. Raman Spectrosc.*, vol. 49, no. 9, pp. 1529–1539, 2018.

[73] N. E. Huang *et al.*, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[74] P. Geladi, “Chemometrics in spectroscopy. Part 1. Classical chemometrics,” *Spectrochim. Acta - Part B At. Spectrosc.*, vol. 58, no. 5, pp. 767–782, 2003.

[75] A. Berger, T. Koo, I. Itzkan, G. Horowitz, and M. Feld, “Multicomponent blood analysis by near-infrared Raman spectroscopy,” *Appl Opt.*, 1999.

[76] H. Lui, J. Zhao, D. McLean, and H. Zeng, “Real-time Raman spectroscopy for in vivo skin cancer diagnosis,” *Cancer Res.*, vol. 72, no. 10, 2012.

[77] G. McLaughlin, K. Doty, and I. Lednev, “Discrimination of human and animal blood traces via Raman spectroscopy,” *Forensic Sci. Int.*, 2014.

[78] T. M. Pabst, J. Thai, and A. K. Hunter, “Evaluation of recent Protein A stationary phase innovations for capture of biotherapeutics,” *J. Chromatogr. A*, vol. 1554, pp. 45–60, 2018.

[79] S. Wang, R. Ionescu, N. Peekhaus, J. Leung, S. Ha, and J. Vlasak, “Separation of post-translational modifications in monoclonal antibodies by exploiting subtle conformational changes under mildly acidic conditions,” *J. Chromatogr. A*, vol. 1217, no. 42, pp. 6496–6502, Oct. 2010.

[80] D. Ejima *et al.*, “Effects of acid exposure on the conformation, stability, and aggregation of monoclonal antibodies,” *Proteins Struct. Funct. Bioinforma.*, vol. 66, no. 4, pp. 954–962, Dec. 2006.

[81] P. Arosio, S. Rima, M. Lattuada, and M. Morbidelli, “Population balance modeling of antibodies aggregation kinetics,” *J. Phys. Chem. B*, vol. 116, no. 24, pp. 7066–7075, 2012.

[82] P. Arosio, G. Barolo, T. Müller-Späth, H. Wu, and M. Morbidelli, “Aggregation stability of a monoclonal antibody during downstream processing,” *Pharm. Res.*, vol. 28, no. 8, pp. 1884–1894, 2011.

[83] S. N. Timasheff, “Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components,” *Proc. Natl. Acad. Sci.*, vol. 99, no. 15, pp. 9721–9726, 2002.

[84] C. Ebel, H. Eisenberg, and R. Ghirlando, “Probing Protein-Sugar Interactions,” vol. 78, no. January, pp. 385–393, 2000.

[85] J. B. Rowe *et al.*, “Submicron Aggregation of Chemically Denatured Monoclonal Antibody,” *Mol. Pharm.*, vol. 15, pp. 4710–4721, 2018.

[86] Thorlabs, “Notch Filters.” [Online]. Available: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=3880&gclid=EAIaIQobChMIr57Gs5mc3gIVFOh3Ch3xOAj_EAAYASAAEgI0xfD_BwE. [Accessed: 23-Aug-2018].

[87] C. M. Andersen and R. Bro, “Variable selection in regression-a tutorial,” *J. Chemom.*, vol. 24, no. 11–12, pp. 728–737, Nov. 2010.

[88] V. A. Lozano, J. M. Camiña, M. S. Boeris, and E. J. Marchevsky, “Simultaneous determination of sorbic and benzoic acids in commercial juices using the PLS-2 multivariate calibration method and validation by high performance liquid chromatography,” *Talanta*, vol. 73, no. 2, pp. 282–286, Sep. 2007.

[89] Z. Wen, “Raman spectroscopy of protein pharmaceuticals,” *J. Pharm. Sci.*, vol. 96, no. 11, pp.

2861–2878, Nov. 2007.

Appendix 1: Control dynamic light scattering measurements of mAb-1 and mAb-2 in 25 mM sodium citrate pH 5.0 and I=100 mM.

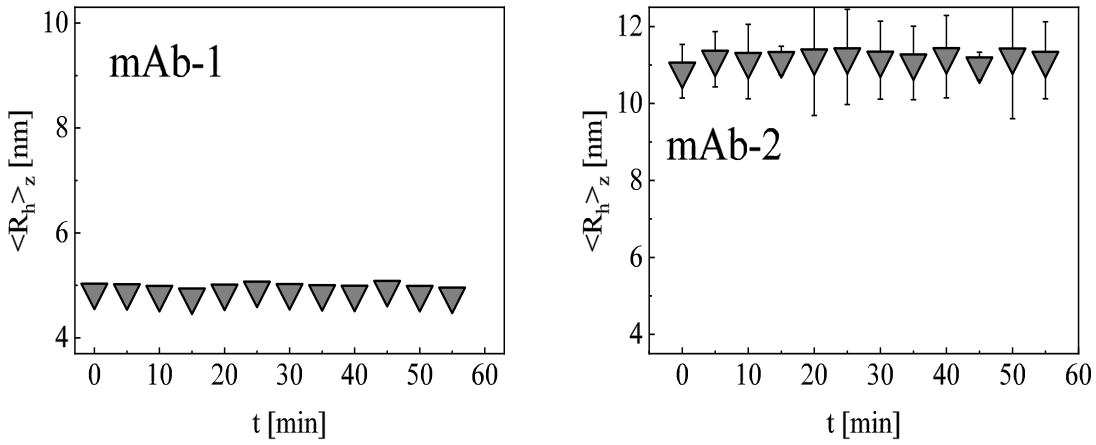

Figure 36 - Time evolution of the average hydrodynamic radius of 1 g/L mAb1 (left) and mAb-2 (right) solutions in 25 mM sodium citrate at pH 5.0 and 100 mM ionic strength.

Appendix 2: Averaged molecular weight of mAb-1 and mAb-2 aggregate peaks

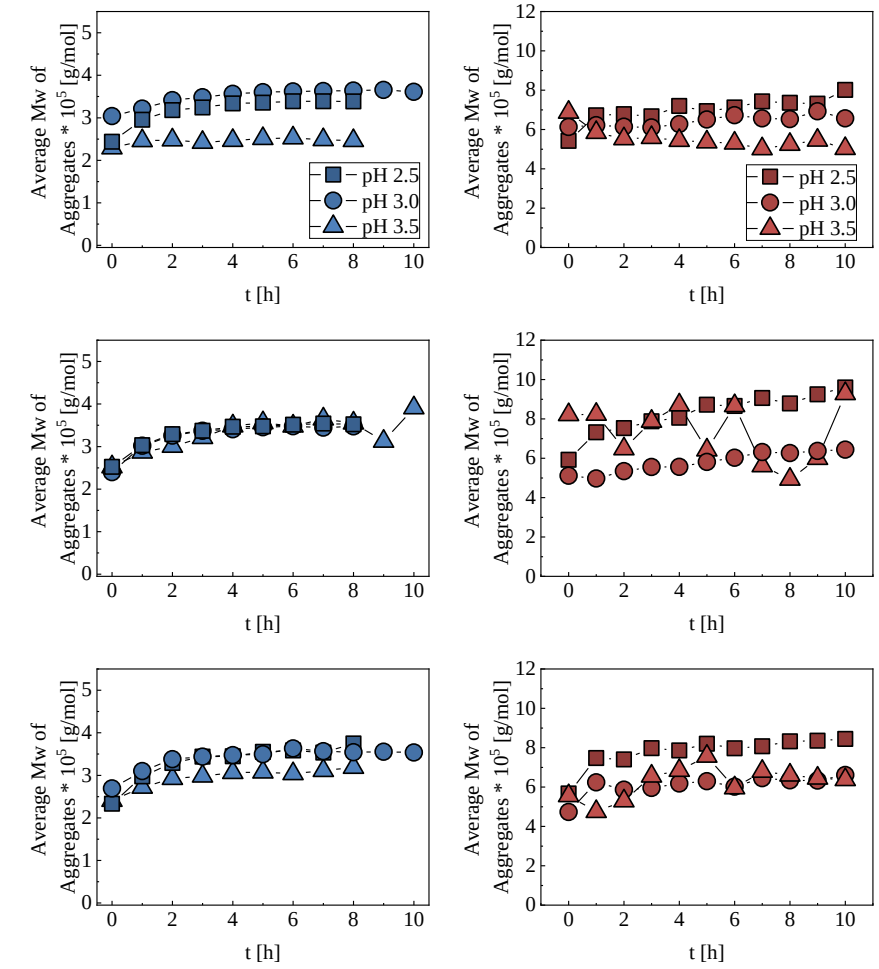

Figure 37 – Average molecular weight of aggregate peak determined by static light scattering as a function of time after neutralization of mAb-1 (left) and mAb-2 (right) solutions to pH 5.0 and ionic strength 100 mM.

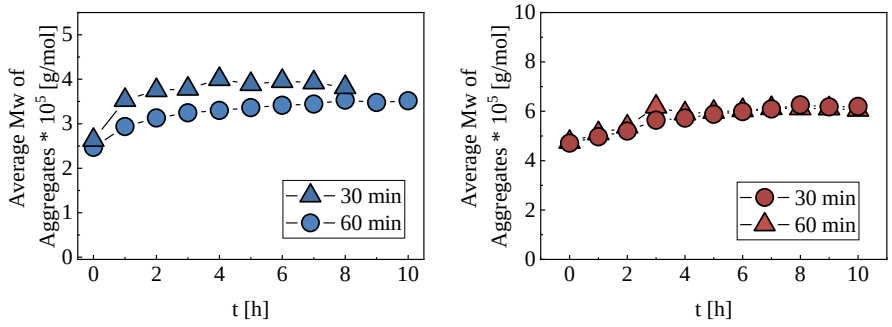

Figure 38 - Average molecular weight of aggregate peak determined by static light scattering as a function of time after neutralization of mAb-1 (left) and mAb-2 (right) solutions to pH 5.0 and ionic strength 100 mM for slow neutralization studies.

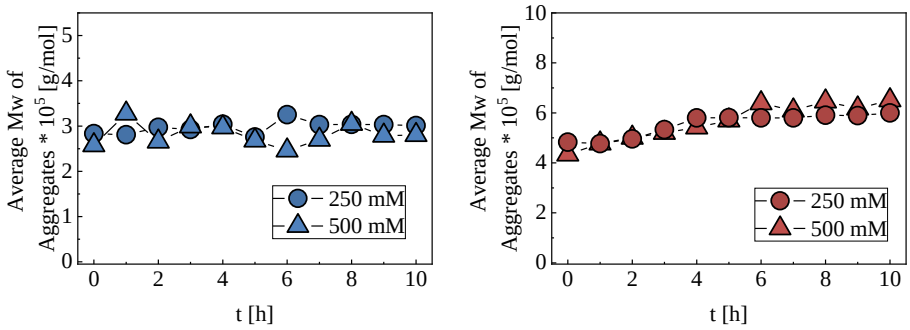

Figure 39 - Average molecular weight of aggregate peak determined by static light scattering as a function of time after neutralization of mAb-1 (left) and mAb-2 (right) solutions to pH 5.0 and ionic strength 100 mM for studies in the presence of D-sorbitol.

Appendix 3: Mass recovery plots of mAb-1 and mAb-2 from every size-exclusion analysis.

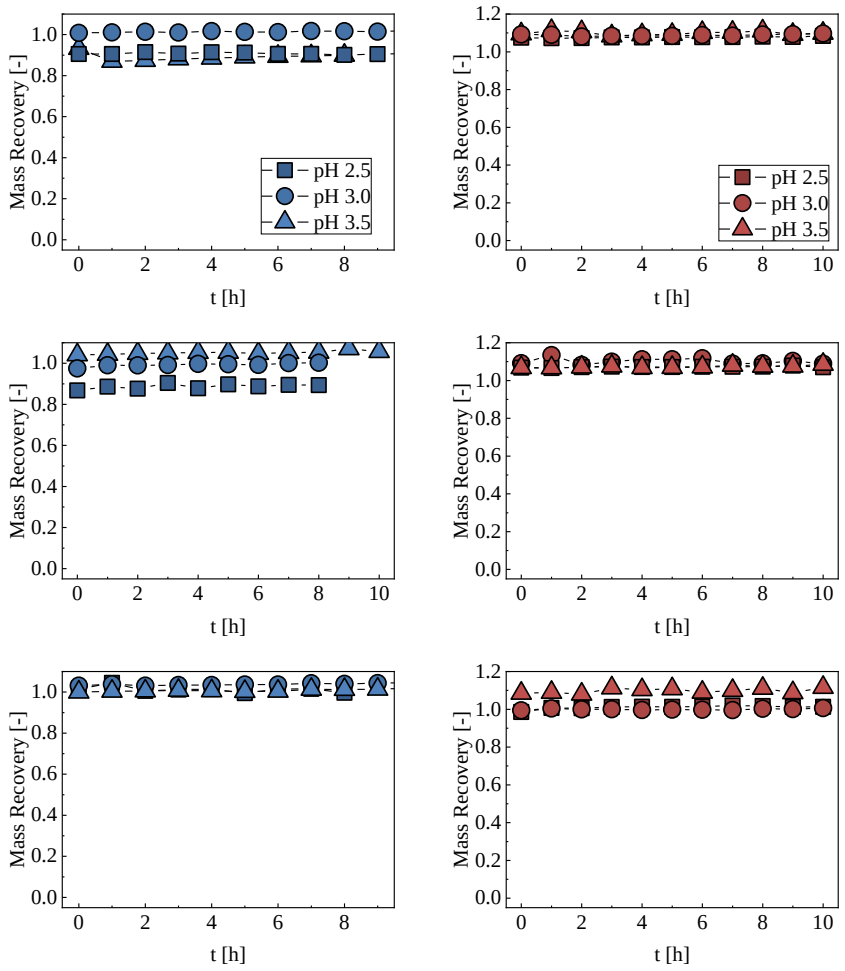

Figure 40 –Mass recovery in SEC analysis as a function of time after neutralization of mAb-1 (left) and mAb-2 (right) solution to pH 5.0 and 100 mM ionic strength.

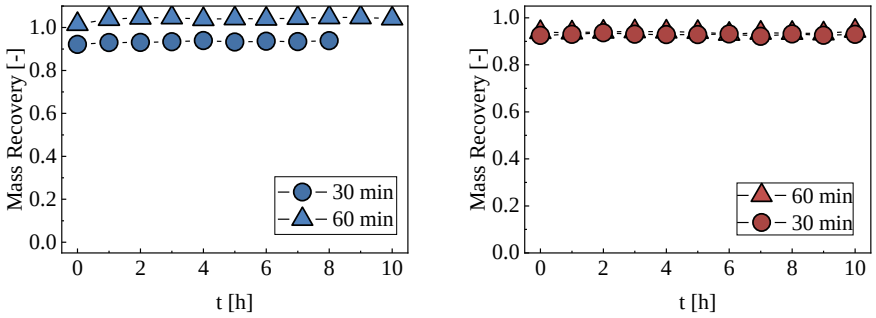

Figure 41 - Mass recovery in SEC analysis as a function of time after neutralization of mAb-1 (left) and mAb-2 (right) solution to pH 5.0 and 100 mM ionic strength for slow neutralization studies.

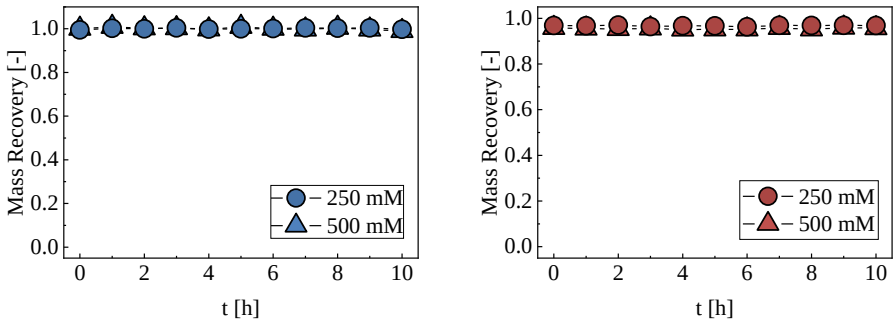

Figure 42 - Mass recovery in SEC analysis as a function of time after neutralization of mAb-1 (left) and mAb-2 (right) solution to pH 5.0 and 100 mM ionic strength for studies in the presence of D-sorbitol.

Appendix 4: Water spectrum

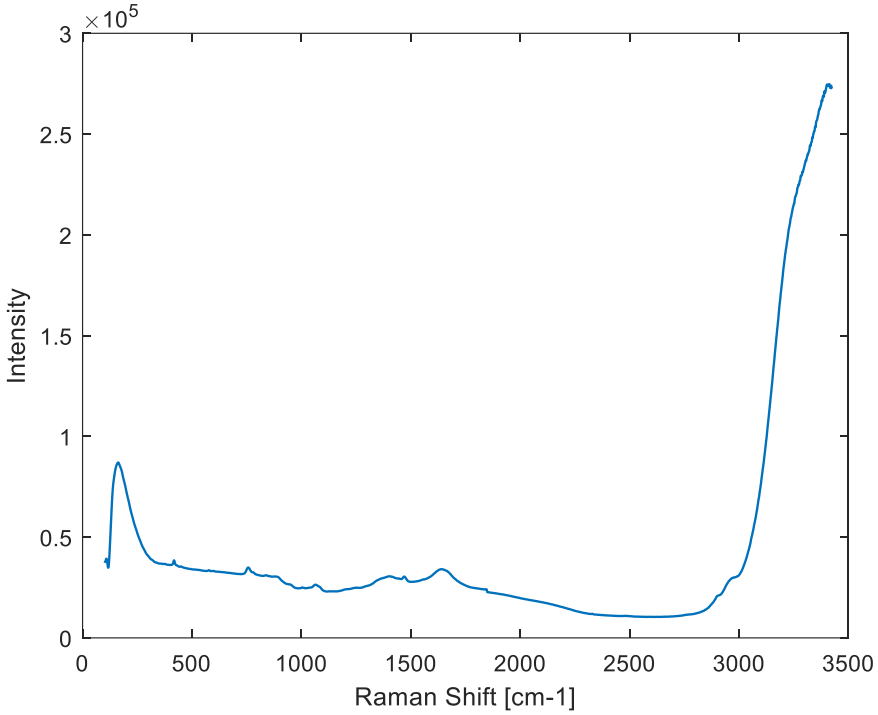

Figure 43 - Raman water Spectrum. The acquisition time was of 30 seconds.

Appendix 5: Score plots for identification of outliers for all four data sets.

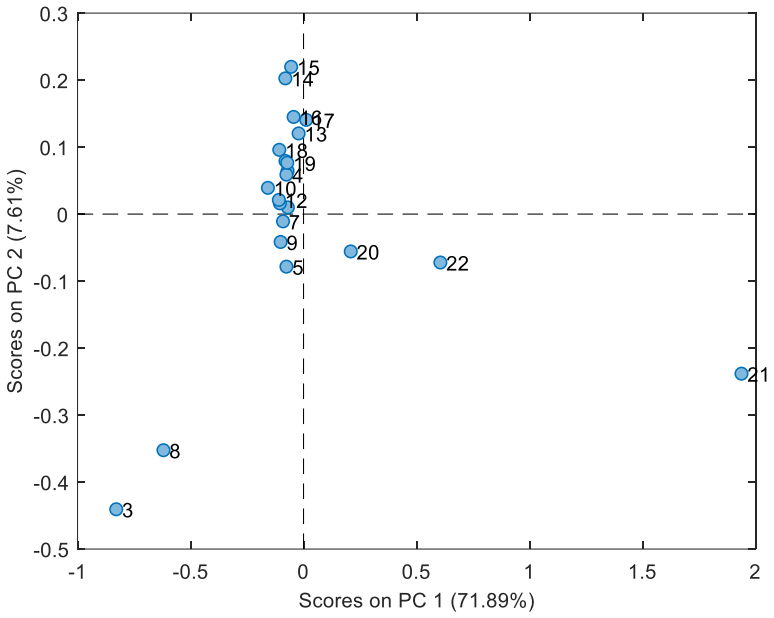

Figure 44 - Score plot from the PCA performed to mAb-1 spectral data set. The scores are identified by the sample number. Identified as outliers were samples 3, 8, 21 and 22.

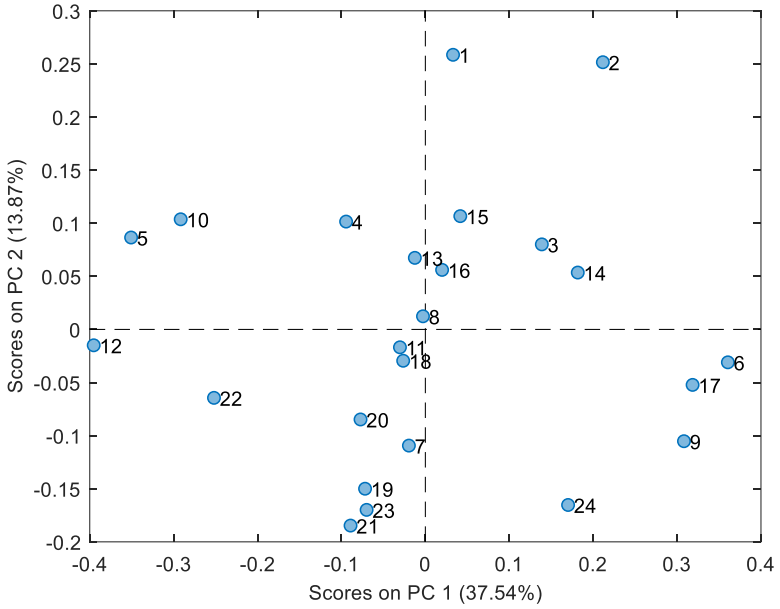

Figure 45 - Score plot from the PCA performed to mAb-2 broad-range aggregate content experiment. The scores are identified by the respective sample number. No samples were identified as outliers.

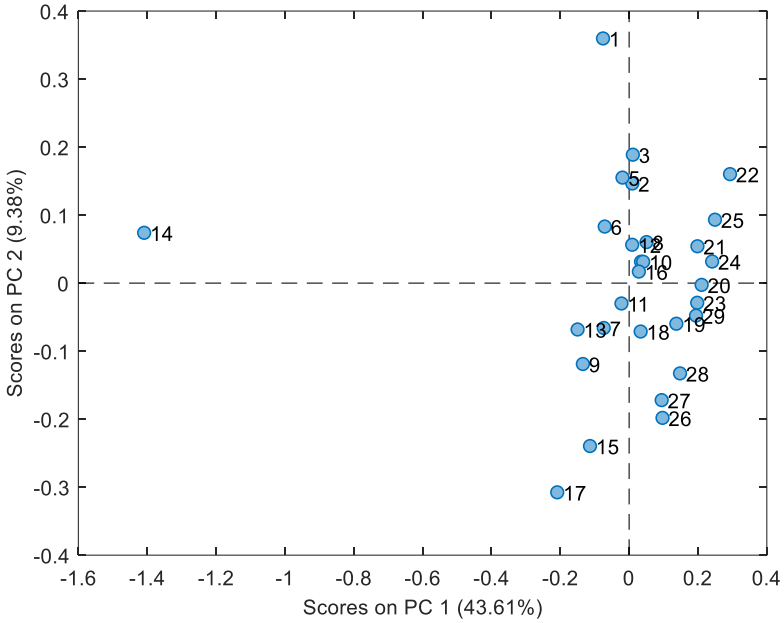

Figure 46 - Score plot from the PCA performed to mAb-2 narrow-range aggregate content experiment without the Notch filter. The scores are identified by the respective sample number. Sample 14 was identified as outlier.

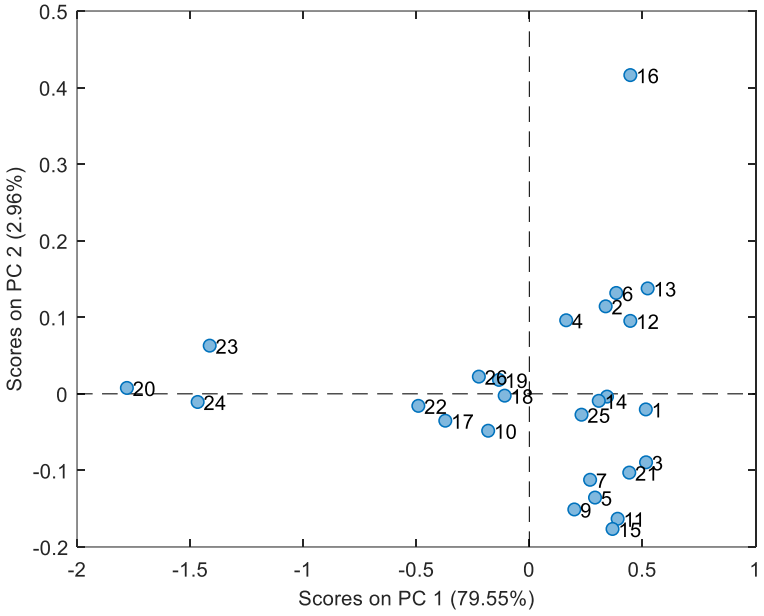

Figure 47 - Score plot from the PCA performed to mAb-2 narrow-range aggregate content experiment in the presence of the Notch filter. The scores are identified by the respective sample number. Sample 16, 20, 23 and 24 were identified as outliers.

Appendix 6: Hyper parameter tuning for monomeric mAb

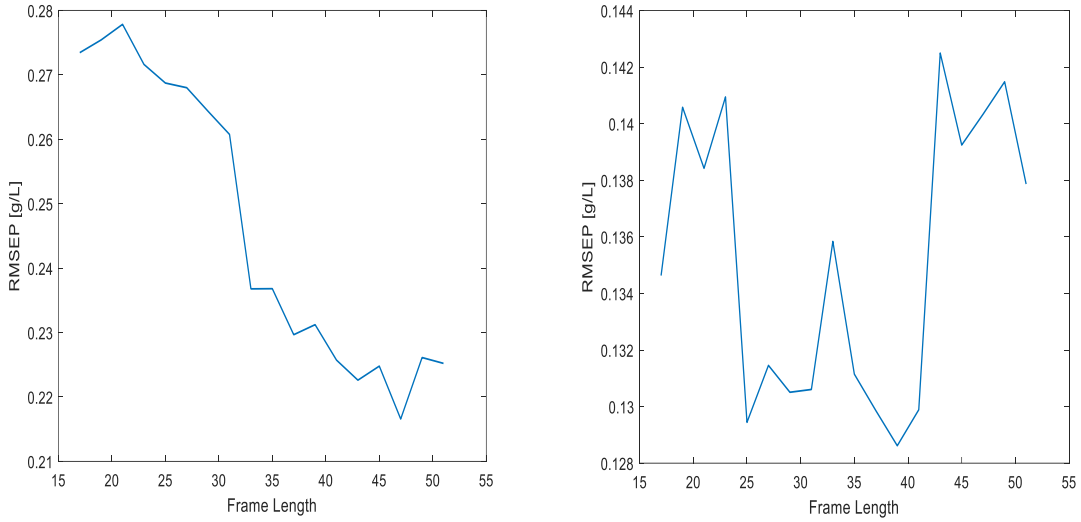

Figure 48 - RMSEP (g/L) result from PLS for the prediction of mAb-2 monomer concentration as a function of the frame length in the SG filter. On the left, using the derivative function of the filter (ord=1); on the right, using only the smoothing option of the function (ord=0).

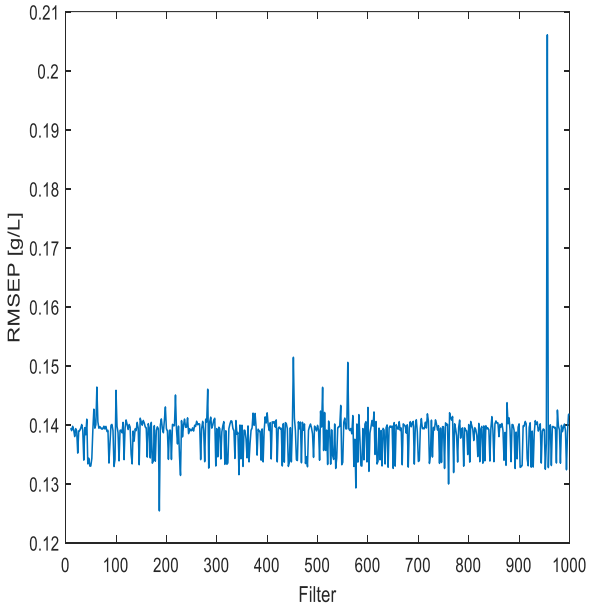

Figure 49 - RMSEP (g/L) result from PLS for the prediction of mAb-2 monomer concentration as a function of the filter after smoothing the data with FT.

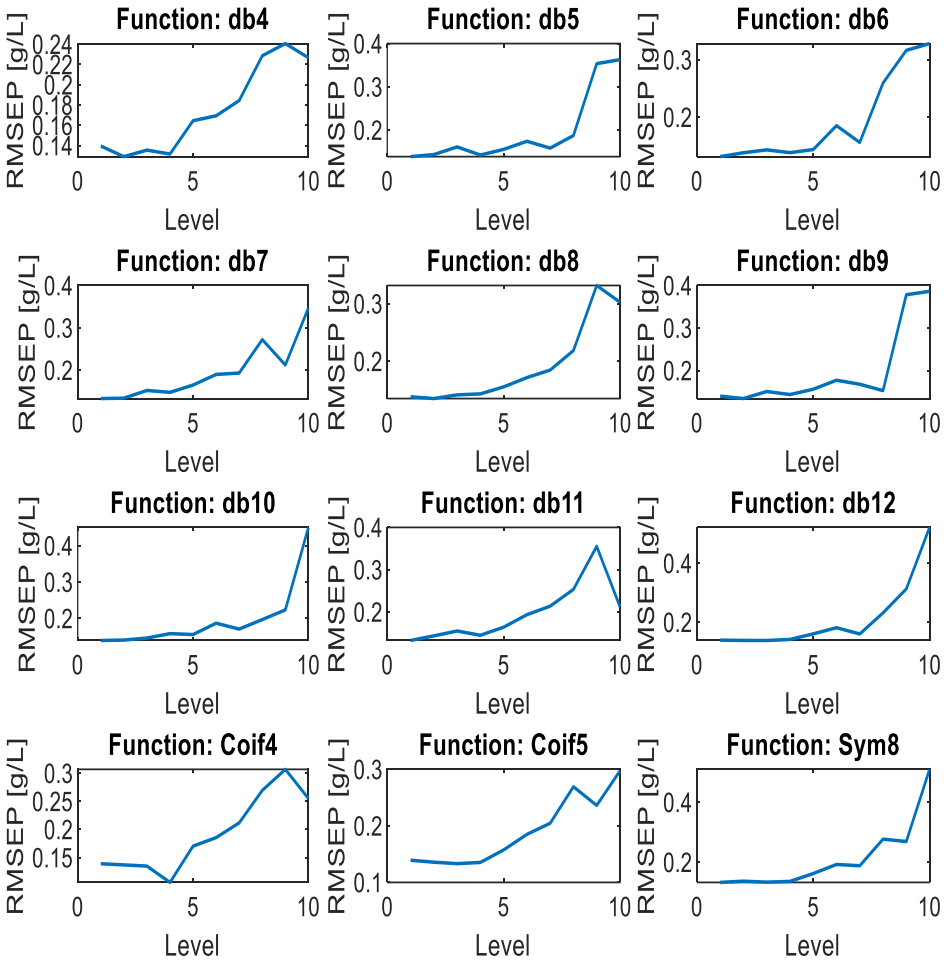

Figure 50 - RMSEP (g/L) result from PLS modelling for the prediction of mAb-2 monomer concentration as a function of the level of decomposition after smoothing the data with WT for each base functions studied.

Appendix 7: PLS results of each experiment discriminated by rotations

Table 11 - PLS prediction results of aggregate and monomer concentration for both mAb-1 (upper table) and mAb-2 (lower table) in the broad-range aggregate content experiments.

Predicted Variable	Rotation	Calibration Set	Calibration Range [g/L]	Prediction Set	Prediction Range [g/L]	Latent Variables	RMSECV [g/L]	Q^2	RMSEP [g/L]	relRMSEP
Monomer Concentration	1	14	0.536-3.81	4	0.751-2.63	1	0.698	-0.976	1.19	0.362
	2	14	0.751-3.81	4	0.536-2.84	8	0.754	0.404	0.697	0.228
	3	14	0.536-2.84	4	1.23-3.81	3	0.493	-2.55	2.10	0.912
	4	15	0.536-3.81	3	0.913-2.04	5	0.819	0.389	0.414	0.126
	5	15	0.536-3.81	3	1.16-1.72	5	1.01	-0.127	0.317	0.097
Aggregate Concentration	1	14	0.0585-2.63	4	0.0145-1.31	6	0.600	0.390	0.413	0.160
	2	14	0.0145-2.63	4	0.0585-1.53	3	0.424	0.358	0.437	0.167
	3	14	0.0145-1.53	4	0.0721-2.63	10	0.295	0.438	0.762	0.503
	4	15	0.0145-2.63	3	0.122-0.976	3	0.542	0.329	0.319	0.122
	5	15	0.0145-2.63	3	0.170-1.31	6	0.422	0.838	0.187	0.0714

Predicted Variable	Rotation	Calibration Set	Calibration Range [g/L]	Prediction Set	Prediction Range [g/L]	Latent Variables	RMSECV [g/L]	Q^2	RMSEP [g/L]	relRMSEP
Monomer Concentration	1	19	0.601-3.76	5	0.437-1.70	19	0.136	0.933	0.162	0.0511
	2	19	0.437-2.61	5	1.07-3.76	6	0.167	0.993	0.084	0.0386
	3	19	0.437-3.76	5	0.601-2.61	19	0.180	0.989	0.075	0.0226
	4	19	0.437-3.76	5	0.925-2.17	18	0.162	0.934	0.112	0.0338
	5	20	0.437-3.76	4	0.874-2.31	6	0.196	0.974	0.0961	0.0289
Aggregate Concentration	1	19	0.0189-1.21	5	0.168-2.13	15	0.081	0.888	0.240	0.202
	2	19	0.0189-2.13	5	0.0768-1.21	19	0.174	0.964	0.072	0.0339
	3	19	0.0189-2.13	5	0.0478-1.07	17	0.174	0.953	0.078	0.0369
	4	19	0.0478-2.13	5	0.0189-0.954	14	0.148	0.847	0.137	0.0659
	5	20	0.0189-2.13	4	0.176-0.897	5	0.213	0.914	0.0772	0.0366

Table 12 - PLS prediction results of aggregate and monomer concentration for mAb-2 (lower table) in the narrow-range aggregate content experiments without the Notch filter (upper table) and with it (lower table).

Predicted Variable	Rotation	Calibration Set	Calibration Range [g/L]	Prediction Set	Prediction Range [g/L]	Latent Variables	RMSECV [g/L]	Q^2	RMSEP [g/L]	relRMSEP
Monomer Concentration	1	22	1.69-4.44	6	2.56-3.66	19	0.158	0.874	0.139	0.0505
	2	22	1.69-4.44	6	2.73-3.24	22	0.163	0.703	0.170	0.0616
	3	22	2.56-4.44	6	1.69-3.92	6	0.125	0.955	0.153	0.0812
	4	23	1.69-4.44	5	2.97-3.85	9	0.139	0.886	0.139	0.0505
	5	23	1.69-3.92	5	2.94-4.44	6	0.132	0.957	0.131	0.0587
Aggregate Concentration	1	22	0.0527-0.401	6	0.0866-0.581	3	0.0786	-0.439	0.195	0.561
	2	22	0.0527-0.581	6	0.0755-0.401	1	0.115	-0.396	0.127	0.240
	3	22	0.0755-0.581	6	0.0527-0.334	6	0.124	-0.105	0.106	0.209
	4	23	0.0527-0.581	5	0.104-0.308	1	0.125	0.321	0.0613	0.116
	5	23	0.0527-0.581	5	0.0900-0.306	1	0.134	0.0589	0.0758	0.143

Predicted Variable	Rotation	Calibration Set	Calibration Range [g/L]	Prediction Set	Prediction Range [g/L]	Latent Variables	RMSECV [g/L]	Q^2	RMSEP [g/L]	relRMSEP
Monomer Concentration	1	20	2.73-4.44	6	1.69-3.66	6	0.153	0.938	0.177	0.104
	2	21	1.69-3.92	5	2.94-4.44	6	0.165	0.863	0.226	0.101
	3	21	1.69-4.44	5	2.97-3.85	6	0.179	0.918	0.113	0.0410
	4	21	1.69-4.44	5	2.86-3.92	6	0.185	0.953	0.0881	0.0320
	5	21	1.69-4.44	5	2.73-3.24	6	0.167	0.742	0.163	0.0591
Aggregate Concentration	1	20	0.090-0.401	6	0.0527-0.656	1	0.0838	-0.113	0.203	0.653
	2	21	0.0527-0.656	5	0.090-0.306	1	0.142	0.277	0.0692	0.115
	3	21	0.0527-0.656	5	0.104-0.308	1	0.146	0.181	0.0690	0.114
	4	21	0.0527-0.656	5	0.107-0.334	1	0.147	0.171	0.0729	0.121
	5	21	0.0527-0.656	5	0.128-0.401	3	0.123	-2.01	0.163	0.270